

UC Berkeley

Dissertations, Department of Linguistics

Title

Basque Word Order and Disorder Principles, Variation, and Prospects

Permalink

<https://escholarship.org/uc/item/82x5f1m3>

Author

Aske, Jon

Publication Date

1997

**Basque Word Order and Disorder
Principles, Variation, and Prospects**

by

Jon Aske

B.A. (University of California, Santa Cruz) 1984

M.A. (University of California, Berkeley) 1990

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Linguistics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:

Professor Charles J. Fillmore, Chair

Professor Karl Zimmer

Professor William H. Jacobsen, Jr.

Professor Dan I. Slobin

Spring 1997

Basque Word Order and Disorder: Principles, Variation, and Prospects

© 1997

by

Jon Aske

The dissertation of Jon Aske is approved:

Charles J. Fillmore 4/17/97
Chair Date

Paul Ziemann 4/17/97
Date

William H. Jacobson, Jr. April 14, 1997
Date

Don L. Rubin 4/18/97
Date

University of California, Berkeley

Spring 1997

Amari, eta Euskal Herriari:
hain urrun triste naiz eta.
Emazte eta seme-alabei:
hain hurbil izanik
arintzen baita nire tristura.

“Zer gertatzen zen?
Gaztelak menperatuta zegoela;
indarrez baino gehiago, halere, liluraz.”

*What was happening? Just that it was subjugated by
Castile; by its fascination, however, more than by force.*

Joxe Azurmendi, *Espainolak eta euskaldunak*, p. 517

ADINA	Sconsigliata, possedea si nobil cor!
DULCAMARA	Vuoi vederti mille amanti spasimar, languire al piede?
ADINA	Non saprei che far di tanti: il mio core un sol ne chiede.
DULCAMARA	Render vuoi gelose, pazze, donne, vedove, ragazze?
ADINA	Non mi alletta, non pi piace di turbar altrui la pace.
DULCAMARA	Conquistar vorresti un ricco?
ADINA	Di ricchezze non mi picco.
DULCAMARA	Un contino, un marchesino?
ADINA	No, non vo' che Nemorino.
ADINA	Il mio rigor dimentica; ti giuro eterno amor. Farti felice or bramo, io bramo.
CONTADINI, SOLDATI	Oh, il gran licore!

Gaetano Donizzeti, *L'Elisir D'Amore*

Table of Contents

Table of Contents	iv
List of Symbols	xxi
Introduction	xxiv
Acknowledgments	xxix
Chapter 1: Theoretical and methodological preliminaries	1
1.1 The functional-typological approach to linguistic analysis	1
1.1.1 The functionalist perspective	1
1.1.2 The functional-typological approach	4
1.1.3 Functional motivations	6
1.2 Some basic underlying assumptions	7
1.2.1 Introduction.....	7
1.2.2 Rules and representations: The list-rule fallacy.....	8
1.2.3 Form and function.....	11
1.2.4 The nature of explanation in linguistics.....	13
1.2.5 Description vs. explanation in linguistic analysis.....	16
1.2.6 Explanation vs. motivation	18
1.2.7 Conclusions.....	20
1.3 Pragmatic principles and the grammar	21
1.3.1 Introduction.....	21

1.3.2 An example: The Setting-Assertion communicative strategy	22
1.3.3 Discourse pragmatic principles and grammar	25
1.4 Form and function revisited.....	30
1.4.1 The separation of form and function	30
1.4.2 The relationship between form and function	31
1.4.3 Functional structure vs. formal structure (constituency)	39
1.4.4 Constructions	43
1.4.5 Structures (constructions) vs. processes	48
1.4.6 Grammaticalization.....	50
1.5 Variation: genres, style, media	57
1.5.1 Spoken language and contextualization.....	57
1.5.2 Genres, styles, and media	59
1.5.3 Genre and word order in Basque	60
1.5.4 More on the differences between speech and writing	62
1.5.5 Concepts for speech analysis	65
1.6 Word order typology	69
1.6.1 Introduction.....	69
1.6.2 Basic word order.....	70
1.6.3 Frequency and basicness.....	72
1.6.4 Markedness and basicness	74
1.6.5 Clause type and basic word order	77
1.6.6 Syntactic or intonational complexity	80
1.6.7 Grammatical vs. pragmatic word order	82

1.6.8 Subject-prominent vs. topic-prominent languages	86
1.6.9 Word order freedom and verbal inflections	88
1.6.10 Subject-object asymmetries in flexibility	89
1.6.11 Verb-initial languages.....	91
1.6.12 Word order flexibility and pragmatic-role changing constructions.....	93
1.6.12.1 Introduction	93
1.6.12.2 Non-default topic assignment.....	95
1.6.12.3 Non-default focus assignment	98
1.6.12.4 Other pragmatic constructions.....	99
1.6.13 Verb-second and the verbal brace.....	100
1.6.14 Other types of variation	103
1.6.15 Basic word order correlations.....	106
1.6.16 Conclusions.....	108
1.7 Typology and explanation	109
1.7.1 Introduction.....	109
1.7.2 Pragmatic principles and the order of clausal constituents.....	110
1.7.2.1 Explaining the subject-first preference.....	110
1.7.2.2 Explaining the OV ~ VO alternation.....	111
1.7.3 Iconicity related explanations	113
1.7.3.1 Iconicity, order, and constituency.....	113
1.7.3.2 Degree of closeness, relevance, and scope	114
1.7.3.3 Negative morpheme positioning.....	117
1.7.3.4 Ordering of complement pairs.....	120

1.7.4 Connector in the middle: the Relator Principle	121
1.7.4.1 Introduction	121
1.7.4.2 The Relator Principle and Basque dependent clauses and phrases	124
1.7.4.3 The cause-and-effect issue and the instantiation issue	126
1.7.5 Diachronic typology and drift.....	128
1.7.6 Analogy: Systemic explanations for the correlations	133
1.7.7 Processing constraints and word order	137
1.8 Language contact and language change	143
1.8.1 Grammatical change	143
1.8.2 Word order change	146
1.8.3 The Basque sociolinguistic situation	148
1.8.4 Language contact and convergence	151
1.8.5 Language contact and simplification	154
1.8.6 Language contact and language decay, attrition, obsolescence, and shift.....	158
1.8.7 Imperfect speakers or 'semi-speakers'	159
1.8.8 Language changes in an (individual or community) attrition situation	161
1.8.9 Further thoughts on the shift from SOV to SVO	166
1.8.10 A seeming contradiction: Pragmatic word order and rigid SVO order	168
1.8.11 The mechanism of change	171
Notes to Chapter 1	174

Chapter 2: Typological overview and overview of the corpus.....	210
2.1 Introduction.....	210
2.2 A typological overview of Basque.....	211
2.2.1 Dryer's correlation pairs.....	211
2.2.2 Declension and postpositions	211
2.2.3 Nominal modifiers: The genitive suffixes	213
2.2.4 Modifiers in apposition.....	214
2.2.5 Demonstratives	215
2.2.6 Numerals.....	215
2.2.7 Relative clauses	216
2.2.8 Adjectives	217
2.2.9 Comparisons	218
2.2.10 Other noun related characteristics	218
2.2.11 Verb-related characteristics	219
2.2.11.1 The Basque conjugation	219
2.2.11.2 The position of the verb.....	221
2.2.11.3 Clausal constituency	222
2.2.11.4 Evidential and aspectual particles.....	224
2.2.11.5 Other 'auxiliary' verbs.....	224
2.2.12 Subordinating relators and dependent clauses.....	229
2.2.13 Constituent order in dependent clauses	231
2.2.13.1 Finite dependent clauses.....	232
2.2.13.2 Non-finite dependent clauses.....	234

2.2.14 Ergativity	235
2.2.14.1 Introduction	235
2.2.14.2 A semantic motivation for ergative coding	236
2.2.14.3 Ergative-absolutive syncretism	239
2.2.15 Conclusions.....	240
2.3 Previous studies of Basque constituent order and focus	241
2.3.1 Introduction.....	241
2.3.2 Azkue 1891, 1923-5.....	242
2.3.3 Altube 1920, 1929	243
2.3.4 Narrow focus and Romance influence.....	245
2.3.5 Negative and emphatic clauses.....	247
2.3.6 Verb and polarity focus	249
2.3.7 The focus rule and subordinate clauses	252
2.3.8 Basic word order and Basque	254
2.3.9 The preferred order of elements	257
2.3.10 Basque as a left-branching language	259
2.3.11 Some unanswered questions	260
2.4 Characteristics of the corpus	261
2.4.1 Introduction.....	261
2.4.2 The units of the Spoken Basque Corpus.....	263
2.4.3 The units of the Written Basque Corpus.....	265
2.4.4 Constituent orders in main clauses in the Spoken Basque Corpus.....	266
2.4.5 Differences among the subjects of the Spoken Basque Corpus	272

2.4.6 Further notes on word order in the Written Basque Corpus	274
2.4.7 Additional comparisons between the spoken and written corpora	276
Notes to Chapter 2	280
Chapter 3: Discourse pragmatics and linguistic form	289
3.1 Discourse pragmatics and the domain of word order	289
3.1.1 Introduction.....	289
3.1.2 The structure of simple assertions	292
3.2 The intonational and informational structure of the sentence	294
3.2.1 Intonation structure and information structure of the sentence	294
3.2.2 The intonation units of the Spoken Basque Corpus.....	296
3.2.3 The ordering of constituents in the sentence	298
3.2.4 Sentence-head intonation units.....	300
3.2.5 Sentence-body (main) intonation units.....	302
3.2.6 Sentence-tail intonation units	303
3.2.7 Conclusion: The domains of word order	306
3.3 Discourse pragmatics and word order	307
3.3.1 Introduction.....	307
3.3.2 The Prague School approach	308
3.3.3 Givenness: Activation, accessibility, identifiability	312
3.3.4 Some complicating factors for the notion of givenness.....	316
3.3.5 The domain of givenness.....	320
3.3.6 Thematic importance and inherent topicality	321

3.3.7 Thematic salience and the content of linguistic expressions	323
3.3.8 Thematic salience and the order of linguistic expressions	325
3.3.9 The coding of nominals	331
3.4 Topics: Thematic grounds and predication base	333
3.4.1 Topics and topicality	333
3.4.2 Topic-comment and all-comment information structure	335
3.4.3 Other explanations for thetic structure	337
3.4.4 What are topics?	339
3.4.5 Topics and subjects	343
3.4.6 Covert vs. overt topics	344
3.4.7 The position of overt topics	345
3.4.8 Split in the topic's sub-roles	347
3.4.9 Other types of assertion grounding (non-thematic grounding)	348
3.4.10 Thematically grounded and ungrounded sentences	349
3.4.11 Obligatory vs. optional thematic grounding	353
3.4.12 Conclusions	356
3.5 Focality and focus	358
3.5.1 Introduction	358
3.5.2 Assertion focus	358
3.5.3 New (regular) assertions and new focus	364
3.5.4 Contrastive assertions and contrastive focus	366
3.5.5 Emphatic assertion and focus	370
3.5.6 Default focus	377

3.5.7 A typology of focus positions.....	381
3.5.8 Assertion final focus (focus extraposition).....	388
3.5.9 Focus typology in two VO languages: English vs. Spanish	392
3.6 Focus structure in Basque.....	397
3.6.1 Main focus position in Basque	397
3.6.2 Unmarked focus in all-new assertions	399
3.6.3 Verb focus.....	400
3.6.4 Rheme/polarity focus.....	401
3.6.5 Delayed or extraposed focus in Basque.....	404
3.6.6 Summary of focus positions in Basque	405
3.6.7 Further examples of focus structure types in Basque	407
3.6.7.1 Copula constructions with two overt complements.....	407
3.6.7.2 One-argument copula constructions.....	412
3.6.7.3 Impersonal assertions	414
3.6.8 Conclusion.....	415
3.7 Focus structure and word order in Basque.....	416
3.7.1 Introduction.....	416
3.7.2 Osa 1990: Basic order and markedness revisited	419
3.7.3 Marked and unmarked orders: A classical example	423
3.7.4 Some speculations on markedness and word order change.....	427
Notes to Chapter 3	429

Chapter 4: Settings, topics, and antitopics	439
4.1 The nature and variation of the topic notion	439
4.1.1 Introduction.....	439
4.1.2 Topic vs. subject prominence	441
4.1.3 Inherent topicality and topics.....	443
4.1.4 Thematic ground or topic as a special case of ground	446
4.1.5 The basicness of the topic-comment configuration	446
4.1.6 Multiple topics	450
4.1.7 Non-topic topical elements	452
4.1.8 Topic-comment structure with non-verbal comments	454
4.1.9 The formal expression and position of grounds	455
4.1.10 Clause-internal vs. clause-external topics.....	457
4.1.11 Pragmatic characteristics of topics and settings	460
4.1.12 Pragmatic characteristics of topicalized elements	462
4.1.13 Information structure vs. syntactic structure: topic-comment vs. NP-VP	466
4.2 Covert and initial overt topics in the Spoken Basque Corpus	469
4.2.1 Introduction.....	469
4.2.2 Non-'subject' topics.....	472
4.2.3 Covert topics in the Basque corpus	475
4.2.4 Overt topics in the Spoken Basque Corpus	480
4.2.5 Dislocated topics in the Basque corpus	484
4.2.6 Split external and internal topics	490
4.2.7 Verb topics.....	493

4.3 Inverted topics, antitopics, right-dislocation and afterthoughts	496
4.3.1 Inverted topics or 'antitopics'	496
4.3.2 Antitopics and right-dislocation	500
4.3.3 Antitopics and afterthoughts	501
4.3.4 Uses of the antitopic construction.....	504
4.3.5 The antitopic construction in the Spoken Basque Corpus	508
4.3.6 Antitopics in narrative written style: the Written Basque Corpus.....	513
4.3.7 Antitopics in journalistic written style.....	521
4.4 Topicalization in Basque	525
4.4.1 Introduction.....	525
4.4.2 Discourse connectors and topicalizers.....	526
4.4.3 The topicalizing, additive particle ERE 'also'	529
4.4.3.1 Introduction	529
4.4.3.2 ERE and unusual topics.....	532
4.4.4 Low topicality topics in existentials	537
Notes to Chapter 4	544
Chapter 5: Topicless clauses, object topics, and antitopics.....	551
5.1 Thematically grounded and ungrounded sentences	551
5.1.1 Introduction.....	551
5.1.2 Theticity: coding strategies and explanations.....	552
5.1.3 Constraints on thetic subject inversion	557
5.1.4 Thetic functions: Typologies of thetic sentences.....	561

5.2 Announcements	563
5.2.1 Sasse's definition and examples.....	563
5.2.2 All-news statements in newspaper headlines	568
5.2.3 All-news sentences in Spanish newspaper headlines	569
5.2.4 All-news sentences in Basque newspaper headlines	573
5.2.5 Interruptions and sudden events	578
5.3 Explanations	580
5.3.1 Explanations vs. announcements	580
5.3.2 Ungrounded explanations	581
5.3.3 Object topics in happenstance sentences	582
5.3.4 Some further examples ofthetic explanations.....	585
5.3.5 A special type explanation construction	587
5.4 Introductions	591
5.4.1 Existentials.....	591
5.4.2 Presentationals and focus.....	592
5.4.3 Circumstantially grounded presentatives.....	595
5.4.4 The form of Basque existentials and otherthetic sentences	598
5.4.5 The delayed focus strategy in presentations	603
5.4.6 Subject inversion in elicited sentences	605
5.4.7 Post-presentational predicates	612
5.4.8 Basque -NA finite predicate clauses.....	615
5.4.9 Basque -LA finite predicate clauses	616
5.4.10 Finite predicate clauses without subordinators.....	618

5.4.11 More on the position of a presentational focus phrase	620
5.4.12 Post-presentative predications and 'extraposition'	622
5.4.13 Presentations in a picture descriptions.....	624
5.5 Descriptions	627
5.5.1 Background descriptions	627
5.5.2 Thetic sentences and the setting function	631
5.5.3 Inversion in background Basque clauses.....	634
5.6 Episode openings.....	638
5.6.1 Episode opening sentences	638
5.6.2 Spanish examples and conventionalized formulae	640
5.6.3 Surprising episode openings in Basque	642
5.7 Conclusions.....	647
Notes to Chapter 5	652
Chapter 6: Focality, focus, and word order	657
6.1 Focality and focus	657
6.1.1 Introduction.....	657
6.1.2 Focality differences and default focus.....	661
6.1.3 Default focus and basic word order	663
6.2 Types of focality and focus.....	664
6.2.1 Contrastiveness and focus.....	664
6.2.2 Intrinsic vs. extrinsic focality.....	667
6.2.3 Assertion types and focus types.....	668

6.2.4	Optional elements and focalizing particles.....	670
6.2.5	Summary of focality and focus types	672
6.3	The formal expression of focus and focality	675
6.3.1	Introduction.....	675
6.3.2	Languages with preverbal focus as the main focus strategy.....	677
6.3.3	Preverbal focus position in Basque.....	680
6.3.4	Emphatic assertions and emphatic focus	683
6.4	Focus extraposition and Basque word order.....	687
6.4.1	Focus extraposition or delay	687
6.4.2	Focus-extraposition in previous studies of Basque syntax	690
6.4.3	The complexity of focus phrases and the markedness of extraposition	693
6.4.4	Extraposed focus with focus placeholders.....	697
6.5	Verb focus and word order	698
6.5.1	Deviations from default focus	698
6.5.2	Basque 'basic word order' revisited	703
6.5.3	Unmarked VO order: Polarity focus and emphatic assertions.....	705
6.6	The position of absolutive complements in the Basque corpus	714
6.6.1	Rhemes with no complements or one complement	714
6.6.2	(A)OV vs. (A)VO order in the Written Basque Corpus	715
6.6.3	(A)OV vs. (A)VO order in the Spoken Basque Corpus	719
6.6.4	OV vs. VO in the Lur subcorpus	721
6.6.5	Some generalizations about the position of the object	725

6.7 The position of non-absolute complements in the corpus	730
6.7.1 SXV vs. SVX clauses	730
6.7.2 SXV vs. SVX sentences in the written corpus	732
6.7.3 SXV and SVX sentences in the spoken corpus	735
6.7.4 The subcorpora compared.....	740
6.8 Conclusions.....	743
Notes to Chapter 6	746
Chapter 7: Polarity, emphasis, and word order.....	751
7.1 Polarity, focality, and focus.....	751
7.2 Negation in Basque	756
7.2.1 Introduction.....	756
7.2.2 The placement of operators in Basque.....	759
7.2.3 Negative assertions in which the assertion is not the focus.....	763
7.2.4 Can an assertion have multiple foci?	771
7.2.5 Focus position in negative assertions	774
7.2.6 Postverbal elements in negative clauses in the spoken corpus	777
7.2.7 Some conclusions	781
7.3 Topicalization in negative clauses.....	785
7.3.1 Fronting in Basque negative clauses.....	785
7.3.2 The expression of restrictive only operator in Basque	789
7.4 Emphatic affirmation ('assertion focus')	795
7.4.1 Introduction.....	795

7.4.2 Polarity focus with Basque synthetic verbs	798
7.4.3 Polarity focus with analytic (periphrastic) verbs in Basque	802
7.4.4 Polarity focus in all-new assertions	805
7.4.5 Polarity focus and word order change	806
7.4.6 BA in subordinate clauses	807
7.5 Intensity adverbials.....	809
7.6 Polarity questions.....	812
7.7 Imperative and desiderative sentences.....	814
7.8 Emphasis, exclamations, and subject inversion	820
7.8.1 The sources of language inversion crosslinguistically	820
7.8.2 Spanish exclamations.....	823
7.8.3 Inverted order in Basque exclamations.....	825
7.9 Conclusions.....	829
Notes to Chapter 7	831
 Bibliography	 836
 Appendices.....	 897
2.1 Finite dependent clauses in the Written Basque Corpus	897
2.2 Finite dependent clauses in the Spoken Basque Corpus	898
2.3 Dependent non-finite clauses in the Spoken Basque Corpus	899
2.4 Dependent non-finite clauses in the Written Basque Corpus.....	900
2.5 Background of speakers in the Spoken Basque Corpus	901

4.1	Clauses with inverted topic in the Spoken Basque Corpus	904
4.2	Clauses with a postverbal overt subject in the Written Basque Corpus	907
4.3	Non-copula antitopic construction clauses in the Written Basque Corpus	909
4.4	Article published in <i>Diario Vasco</i> , 9/27/96.....	910
5.1	Headlines from <i>La Vanguardia</i> , May 3, 1996.....	911
5.2	Headlines from <i>El Periódico</i> , May 3, 1996.....	913
5.3	Headlines from <i>La Jornada</i> , May 3, 1996	914
5.4	Headlines from <i>Euskaldunon Egunkaria</i> , May 3, 1996	915
5.5	Sentences used for eliciting subject inversion.....	916
5.6	Affirmative statements with <i>etorri</i> “come”	923
6.1	AVO clauses in the Written Basque Corpus	926
6.2	AOV clauses in the Written Basque Corpus	927
6.3	VO clauses in the Written Basque Corpus	928
6.4	AOV clauses in the Spoken Basque Corpus.....	929
6.5	AVO clauses in the Spoken Basque Corpus.....	930
6.6	VO clauses in the <i>Lur</i> subcorpus of the Spoken Basque Corpus	932
6.7	OV clauses in the <i>Lur</i> subcorpus of the Spoken Basque Corpus	934
6.8	Examples of the delayed focus construction in Spoken Basque Corpus.....	936
6.9	Written Basque Corpus, SXV sentences	937
6.10	Written Basque Corpus, SVX sentences	610
6.11	Spoken Basque Corpus, SXV sentences	940
6.12	Spoken Basque Corpus, SVX sentences	943

List of Symbols

Symbol	Explanation
<i>text</i> - (<i>text</i>)	Short pause (< ½ sec.)
<i>text</i> .. (<i>text</i>)	Medium length pause (~ ½-1 sec.)
<i>text</i> ... (<i>text</i>)	Long pause (> 1-1½ sec.)
@	Laughter
,	At the end of an intonation unit it indicates continuing (non-final) rising intonation contour; inside an intonation unit it indicates a lesser rising intonation contour
.	At the end of an intonation unit it indicates final, falling intonational contour
^	Sharp rise-fall intonation contour associated with the following word
/	Sharp rising intonational contour (similar to ‘,’)
\	Sharp falling intonational contour (similar to ‘.’)
<i>text</i> =	Lengthened syllable (in hesitations)
= <i>text</i>	There is no pause between this and the preceding intonation unit
< _X ... >	Period in which a certain feature X holds
<i>text</i> -	Sudden interruption of a word
<i>text-text</i>	Dash indicates that the two morphemes form a single phonological word; not used in Basque orthography (see below)
<u><i>text</i></u>	Underlining on an accented syllable denotes that the constituent in which this word is found receives the main accent associated with an intonation unit it is in
ABL	Ablative case (<i>from, about</i>)
ADJ	Morpheme used to form denominal adjectives (e.g. <i>-dun</i>)
ADV	Various adverbial markers (<i>-rik, -ta, -la(rik)</i> , etc.)
ALL	Allative case (<i>to</i>)
COM	Comitative case (<i>with</i>)
COMPL	Complementizer (typically <i>-la</i>). Used in completive complements and finite postpresentational predicates.
DAT	Dative case (<i>-(r)i</i> , etc.)
DEF	‘Definite’ nominal phrase; pragmatically, it may be simply referential (<i>-a</i> , etc.)
DIMIN	Diminutive suffix (<i>-txo, -tto</i> , etc.)
DIRECT	Directional case, a variant of the allative case (<i>-rantz</i> “towards”)
EMPH	So-called emphatic focus marker (e.g. <i>ba-</i> with synthetic verbs; <i>-xe</i> with some pronominals)
ERG	Ergative case (transitive ‘subject’)
FIN-ALL	Final allative (<i>-raino</i> “up to”)

FUT	Future: actually an erstwhile genitive attached to the perfective participle (-ko, -en)
GEN	Genitive case (-ko, -en)
IMPFV	Imperfective aspect for non-finite verbs (-t(z)en).
INSTR	Instrumental case (it can also be used as an adverbializer) (-z, etc.)
LOC	Locative case; it serves locative and temporal functions; equivalent to <i>at, on, in, etc.</i> (-n)
NOMIN	Nominalizing affix for verbs (-t(z)e-)
PART	Partitive case (-(r)ik)
PFV	Perfective aspect for non-finite verbs (-tu, -i, etc.).
PL	Plural (used only when the English gloss cannot be pluralized)
PROX	Proximate: a variant of the definite marker which indicates proximity to the speaker in time or space (-o-)
QN	Question particle, used in some dialects of Basque (e.g. <i>al</i>)
QUOTATIVE	Verbal evidential particle (<i>omen, ei</i>)
REFLX	Used for the particle <i>se</i> in Spanish examples; this particle has a variety of 'anti-causative' senses
REL	Relative clause marker
ROOT	Verbal root, used with certain 'tenses' (<i>zero</i>)
SUBJUNCT	Subjunctive mood; extremely rare in speech.
SUBORD	Nondescript subordination marker, e.g. -n in embedded questions

Notational Conventions

In a departure from Basque orthography, I have in most cases used a hyphen (-) to connect the two parts of periphrastic verbs. I have done this in order to emphasize the fact that the two parts act as a single word in affirmative clauses for purposes of word order. Occasionally and for simplicity's sake I have also given the gloss for the whole verbal complex instead of for each of its parts. Similarly, I typically use a hyphen to connect the negative word *ez* to the finite verb that follows to indicate that it's cliticized to it. This also is not part of standard practice in Basque orthography. When I do not use the hyphen it is not because of any differences between these and the other sentences, but

for the most part to respect the orthography of sources other than my own texts, or because it is not relevant in that context.

Dotted underline: denotes variable accent of topics and other pre-rhematic setting constituents which varies depending on the constituent's degree of intonational integration with the rest of the clause. If fully accented, they would be dislocated. If unaccented, they would be 'cliticized' to the following rheme. A minor, secondary accent indicates an intermediate status, in which the topic (or setting) is part of the clause's overall intonation unit.

For simplicity, absolute noun phrases have been left uncoded for grammatical role. Thus all nominals which do not have a grammatical case role mark in the gloss are absolutes. Plurality has also been coded inside the English gloss itself whenever possible for simplicity (e.g. *apples* instead of *apple:PL*).

Occasionally I have glossed some grammatical morphemes with a 'meaningful' grammatical label, such as MORE (for *-ago-*), IF (for conditional *ba-*), WHETHER, BECAUSE, etc. I have also done this with occasional examples from other sources.

Introduction

Basque has been classified as a verb-final language. It is also said to be a language with very flexible or 'free' constituent order. In other words, constituent order is said to freely change according to so-called 'stylistic' considerations. In this dissertation I have set out to exemplify, quantify, and begin to understand the logic behind this freedom in the ordering of constituents in Basque asserted clauses. In addition, I have attempted to come to grips with the variation that one finds, for instance, among individual speakers and among different discourse genres. I have also attempted to ascertain certain possible ongoing trends towards change which may be connected to that variation.

Basque is the only so-called verb-final (SOV) language in Western Europe, and it has remained verb-final despite the fact that it has been surrounded and influenced in other ways by neighboring so-called verb-medial (SVO) languages, even though word order is a feature said to spread readily. The language contact situation is particularly interesting because at present all the remaining Basque speakers are bilingual in their language and a neighboring Romance language and they are surrounded by monolingual Romance speakers. Other word order characteristics of Basque include: postpositions, genitive-noun and noun-adjective orders, prenominal relative clauses, preverbal focus, and mostly postverbal subordinators, which are (typically) clause-final in non-finite clauses and in some finite subordinate clauses (such as conditional clauses, though not in

asserted complement clauses). Most of these are characteristics which are said to correlated with SOV order.

As far as we know, this flexibility has been a characteristic of the Basque word order for a long time. However, there are indications that modern day spoken Basque may be being influenced by the surrounding languages in its word order characteristics. This study attempts to ascertain whether this is indeed the case and, if so, the nature of that influence.

This study deals with different varieties of Basque, including standard written Basque and, most importantly, Basque as it is actually spoken by different speakers. Standard Basque is a creation of the last thirty years and in the area of word order it follows certain idealized principles of constituent ordering which are not always adhered to faithfully in practice. The ways in which these rules are broken or, rather, bent, provides an interesting glimpse into the nature of Basque constituent order, the differences among varieties of Basque, and the possible directions of change.

A review of the literature on the subject of word order and its application to the Basque data reveals that the main explanatory principles or parameters have already been identified at one time or another to describe Basque and other languages. These are principles of pragmatic information structure, such as the pragmatic roles topic and focus, and not grammatical principles or categories as it is often assumed. These principles are common to all human languages. Different languages apply them somewhat differently in their grammaticalized constructions for a variety of reasons which I explore in this dissertation. In addition to attempting to ascertain the universal explanatory principles of

word order, I try to deal with the parameters of variation which account for the differences found among languages, and within any one language.

In the course of this investigation I have found that previous analysts of Basque word order have often made extremely insightful observations. Still, much work remains to be done, especially in the area of describing the variation found in Basque word order, according to dialectal and contextual differences, such as genre, and according to characteristics of the speakers, such as age, degree of bilingualism, and so on. The present study attempts to add to the growing literature on this subject and to encourage others to follow it with more detailed studies along the lines presented here.

This study concentrates on actual spoken and written Basque as exemplified in a corpus which will be presented in Chapter 2. Until now, most studies have used primarily intuition and introspection to uncover the principles which determine or influence constituent order, a fact which may have limited the types of generalizations which have been reached and which may have masked much of the variation that is found. This is because some of the factors which affect constituent order are not readily amenable to introspection and a detailed analysis of actual texts provides valuable additional data to consider.

This dissertation is divided into seven chapters. Chapter 1 lays down many of the theoretical assumptions which guide this study, which are in great part common to the functionalist and typological perspective(s) in linguistics. It was felt that presenting the detailed background for this study was important, given the lack of a homogenous theoretical approach in the field of linguistics, even among those who share a similar general orientation. In this chapter I also discuss some notions from the study of

language contact and language change, especially as they pertain to grammatical and word order change.

Chapter 2 serves two main purposes. The first is to familiarize the reader with typological facts about Basque. The second is to introduce the corpus used for this study and to present some of the most obvious preliminary generalizations which can be made from it.

Chapter 3 is the major theoretical chapter and in it I discuss the basic notions of discourse pragmatics and information structure, and the ways in which they to a large extent determine the order of constituents in asserted clauses and sentences in different languages, as well as how they interact with other principles of word order, such as grammatical ones.

The remaining chapters, chapters 4-7, deal with different aspects of the general theory presented in Chapter 3 and with the variation that is found in Basque in these different areas. Chapter 4 deals primarily with the pragmatic notion of topicality and the pragmatic role topic, and the formal realization of topics in Basque. Chapter 5 concentrates on sentences which do not have topics, or which have 'unusual' topics, either non-subject topics or postverbal topics (antitopics). Chapter 6 is about the pragmatic notion of focality, which is closely related to the notion of topicality, and with the pragmatic role focus, as well as the formal characteristics of focus constituents in Basque. Finally, Chapter 7 deals in greater depth with assertions other than average affirmative declarative assertions, such as negative assertions, emphatic assertions, and imperatives.

Like other investigators before me, I have detected a trend towards greater use of postverbal complements in Basque. Through the use of actual spoken data from a variety of speakers, as well as from written sources, I have attempted to understand the actual parameters of variation and to uncover the mechanisms by which such variation comes about and to ascertain the reasons, or at least some correlations, for the variation. Much of this variation seems to be due to the different use that different speakers make of existing constructions and focusing strategies in the language and not directly to borrowing of foreign structures. This is in line with current theories of language change and language contact.

Acknowledgments

This long dissertation has been many years in the making and has somewhat of a long history as well. I would like to thank a number of people whose love, friendship, inspiration, help, and support have made it possible.

While I was growing up in the Basque Country, in the city of Donostia, in the 1960's and 1970's, two worlds coexisted side by side, a Basque world and a Spanish one. I lived a rather sheltered life on the Spanish side of this double reality, being quite ignorant about the other side, while those who lived on the other side were not allowed that 'luxury', as they saw their rights stepped on and their culture and language languish under Spanish hegemony and a regime which had so recently crushed an incipient revival of Basque language and culture. Their resentment at the lack of respect lingers to this day and is still the source of great suffering among the Basque people.

The Basque side of my family had long given up their ancestral language for which they found little use in what was, at least on the surface, a Spanish city, even though for many of them it had been their mother tongue, which they had come to forget. I, along with many people of my generation, discovered in the 1970's that we weren't fully Basque, but realized that we weren't Spanish either. I have some wonderful friends to thank for helping me recover from this discovery and for making me feel part of their Basque world. Without their friendship I would have long ago become an un-hyphenated American. I would like to mention my dear friends Karmen Rodriguez Ranz and Joxan Lizarribar in this respect. Also important for me have been my friend Iñaki Heras

Saizarbitoria, who has also helped me so much with this dissertation, and his wonderful and kind family. I wish I could have gotten to know them better.

In my linguistics ‘career’ I have had many wonderful and inspirational teachers. My first memories of discovering the wonders of language and of linguistic analysis are of those of the heated parsing arguments I used to have with Brother Pinillos in high school, one of the few decent teachers that I remember having throughout those difficult years. From my college experience, I remember especially Jorge Hankamer, who introduced me in the early 1980’s to the rigors and the fun of linguistic analysis, even though he never managed to convince me about the viability of autonomous syntax. My most inspiring early memories from graduate school are from a seminar I took in 1985 at UCSD with Margaret Langdon, who introduced me to linguistic typology and language universals. At Berkeley I also had wonderful teachers, such as Chuck Fillmore, Dan Slobin, Jim Matisoff, George Lakoff, Len Talmy, and Robert Van Valin, among others, from whom I learned a great deal. I also want to thank all the linguists whom I have not personally worked with but who have provided great inspiration to me through the years with their work, such as Sandy Thompson, Wally Chafe, Dwight Bolinger, Tom Givón, Joan Bybee, John Haiman, Koldo Mitxelena, and Eusebio Osa Unamuno, in no particular order, among many others.

In the course of researching and writing this dissertation many people have been extremely helpful and to them I am greatly indebted. I would like to thank first all those who helped me while I was conducting my fieldwork in the Basque Country in the fall of 1993: Iñaki Heras for his invaluable help at many of the recording sessions; Iñaki’s friends Koro Nabarro and Juan Mari Arzallus at *Lur Argitaletxea* for the contacts at the

schools; Edurne Yañez, of *Gure Ikastola*, Hendaia, and Edurne Odriozola Sustaeta, of *Ikasbide Ikastola*, for access to their students and very helpful information about their linguistic background; Xabier Falcon of *Deustuko Unibertsitatea* (Donostia Campus), for access to his students; Asun Urzainki Mikeleiz and Jose Angel Ascunce Arrieta, also of *Deustuko Unibertsitatea*, for making our trip that fall possible and for allowing me access to their library; *Bates College* of Lewiston, Maine, also for making that trip possible, as well as for the teaching opportunities they have given me during these last few years and for the invaluable access to library resources; Aitziber Arnaiz Garmendia and Izaskun Mitxelena Santos, then students at *Deustuko Unibertsitatea*, for the first rough draft of the transcripts of the recordings; without their help, I would probably still be transcribing tapes. Thanks are also due to Karmen Rodriguez Ranz for the Aduna connection, which technical difficulties prevented us from taking advantage of, and for so many other things. And, of course, I would like to thank all the patient subjects who shared with me their versions of the *Pear* and *Chaplin* stories.

I also want to thank Bill Jacobsen for his watchful eye for the details in the Basque and Greek examples. During the writing process, my friend Iñaki, my favorite ‘naïve informant’, was especially helpful, answering most of my questions by return (electronic) mail. I couldn’t have done it without him.

Last, but not least, I would like to thank my wonderful family for their continuous support and great patience during the last few years. Avi, first of all, whom I love so dearly, and who supported and put up with me during this stressful period, and also read some of the chapters at the last minute, managing to catch many of my inconsistencies and grammatical deficiencies without really understanding much of the content, thus

proving that form is at least partially independent from meaning. And my children, Sandi and Alaitz, the most wonderful kids in the world, for all the time that I stole from them while I was “working my head off” at the computer. And my parents, of course, who have not forgotten their prodigal son and who made it all possible.

Chapter 1

Theoretical and methodological preliminaries

In its broadest interpretation, the goal of linguistics is to discover how human languages are alike and how they differ, and to propose and test theories that explain the similarities and differences. There are many alternative paths to this goal, for there is much to explain. (Bybee, Perkins and Pagliuca 1994:1)

1.1 THE FUNCTIONAL-TYPOLOGICAL APPROACH TO LINGUISTIC ANALYSIS

1.1.1 The functionalist perspective

Functionalism in linguistics is a perspective on language and grammar which views linguistic function, i.e. semantic or pragmatic meaning, as being intimately tied to linguistic form. Within this general perspective there are many different possibilities. Some functionalists, for instance, emphasize a direct connection between function and utterances themselves, and thus have been said to deny the 'existence' of grammar. However, most functionalists do not deny the existence of partly arbitrary pairings of meaning and form.

At the other extreme there are those who believe that syntax (form) may be studied and analyzed on its own (to a greater or lesser extent), but who also study 'the uses to which syntactic structures are put'. These two have sometimes been called functionalists, as compared to perhaps most 'formalists', who typically do not find this

issue to be central, or even important, to the study of grammar, since for them grammar is totally independent from function.

Defining functionalism and categorizing the differences among functionalists is not an easy task and the several attempts that have been made remain partially unsatisfactory, though they all provide important insights (cf., e.g., Bates 1978; Nichols 1984; Van Valin 1990a, 1993a; Prideaux 1994). Although lines can be drawn in different places, I would make a main division among those who believe that at some level form can be described and, especially, understood without recourse to function and those who disagree (more or less vehemently) with that position. I count myself among the latter.

Functionalists of this latter type differ in many ways among themselves, and there is no single, unified approach. One way in which functionalists differ is the types of functions which they study, be they semantic functions or discourse functions (informational meanings). The two approaches are, of course, not incompatible with each other, but they do lead to different emphases and biases. Even among those who study the pragmatics of discourse many different approaches are found.

One major way in which functionalist approaches differ has to do with the type of relationship which they propose exists between form and function. For some the link is quite direct and iconic and it expresses itself rather directly in every utterance. For others the link is primarily mediated through the constructions and functions of the language. This approach emphasizes the arbitrariness of the form-function relationship. In the end, I think we will find that both approaches are necessary and that both are needed to understand different aspects of the form-function relationship. In the area of discourse functions and constituent order we often find that there is a mixture of direct iconicity,

construction mediated semi-iconicity, and constructional arbitrariness, which isn't always easy to disentangle.

Functionalist analyses have sometimes been criticized for not displaying enough concern for formalization and explicitness. Although I believe it is true that some functionalist analyses are guilty of vagueness, oversimplification, and even sloppiness (though perhaps no more than many non-function bases ones), I think that this is due in part to the complexities associated with the functionalist approach. When one studies meaning, language use, and especially non-categorical functions, with all the inter-relatedness, variation and nuance involved, it is much harder to find convincing formalisms, given the complexities and the poverty of our knowledge. In this approach one is not free to posit formal invariant categories (whether abstract or concrete) and to attribute the variation and counterexamples to 'other factors' outside the scope of the study. Complexity is exacerbated when one seeks to incorporate insights from cross-linguistic comparison (typology) into one's analyses and generalizations, as we shall see.

Some functionalists, such as Dik and Van Valin, have indeed emphasized the importance of formalization, but to some extent with the result of oversimplifying the picture of the form-function relationship. In some ways it is useful to have an idealized picture of 'the forest' before examining 'the trees', and explicitness and thoroughness allow us to see the deficiencies of our analyses. On the other hand, excessive a priori assumptions and idealizations may be more blinding than they are illuminating. To the extent that our knowledge of the functions and forms of language is still quite preliminary and incomplete, explicit formalization may not always be possible or desirable. Functionalists, because they concentrate primarily on the functions of language, which

have different strengths, compete with each other for coding resources, and display much variation, must be careful about hypothesizing structures and units which may be questionable.

1.1.2 The functional-typological approach

The present study follows what is often known as the functional-typological approach, or set of approaches, associated with a number of investigators, such as Greenberg, Givón, Haiman, Thompson, Bybee, Myhill, and many others.

A cross-linguistic or typological perspective is central to functionalism since it allows us to see the different functions or meanings which are important in language, to compare the different ways in which they are formalized, and to ascertain the core aspects from those which are ancillary. The fact that different functions may be formalized in different ways and that the forms that express those functions may have unpredictable (though motivated) extended uses has been seen by some as proof that form-function pairings are fully arbitrary. This, however, is not the only possible conclusion, as I will attempt to show.

Typology, in the sense adopted here, goes beyond mere taxonomy or classification of linguistic structures, and extends to the making of cross-linguistic generalizations (also known as ‘universals’), and, crucially, to seeking explanations for those generalizations (cf. the three senses of the term ‘typology’ in Croft 1995:86ff). This approach closely integrates the functions of language with the structures, keeping sight of the combination of diachronic and cognitive forces which shape the form-function units of each language. In other words, from this perspective, typology, the study of how languages differ, is

inseparable from the study of ‘universals’, the study of how languages are similar and the patterns of similarities which are found.¹

As Croft reminds us, besides the sampling problem, the major methodological problem of typology and “cross-linguistic comparison is essentially how to identify two grammatical phenomena in two different languages as the ‘same’ thing” (Croft 1995:88), i.e. the identification of the relevant **categories** of typology.² The problems that we find here are that the “variation across languages is too great” and that “formal definitions are internal to the structural system of a single language, so they cannot be the basis of a language-independent definition” (Croft 1995:88).³ Some approaches to typology bypass this problem by assuming the existence of universal abstract, formal categories. According to such quasi-Platonic approaches to universals, “if languages resemble each other in certain ways, posit a [abstract] concept in one’s theory that ‘captures’ these resemblances, so that the theory now ‘explains’ the resemblances” (Dryer 1994). This is very different from the functional approach envisioned here, where it is assumed that “underlying the resemblances among grammatical [categories] in different languages are various functional principles and forces that both cause the resemblances and allow the variations” (*ibid.*).⁴ That is why functional-typologists use “semantic, pragmatic, or discourse-based definitions for morphosyntactic phenomena and phonetic definitions for phonological phenomena” (Croft 1995:88).

The functional categories of language which get coded lexically and in constructions do not in any way exist in pristine form before they become ‘formalized’. If they did, variation among languages would be non-existent. Rather these ‘functions’ reflect communicative pressures, needs, and ‘habits’, as well as cognitive predispositions

and constraints, which become institutionalized into linguistic units through a process of lexicalization and grammaticalization (cf., e.g., Hopper and Thompson 1993).⁵

In other words, the similarities among languages are due to the fact that cognitive preferences and constraints, and communicative needs are identical for all humans. Languages differ, however, because those constraints and needs are not categorical and static concepts and do not come preordained but are rather high probability outcomes given the *communicative and cognitive constraints and needs*, working on already established and constantly changing language 'systems' (cf., e.g., Aitchison 1989).⁶

1.1.3 Functional motivations

Functionalists and typologists seek to 'understand' form primarily in terms of external functional 'motivations' which give rise to that form diachronically, and which may motivate them synchronically in the speakers' representations as well. The major functional motivation is perhaps *iconicity*, which postulates a isomorphism between the forms of language and some external aspect of the realm of experience.⁷

In the realm of constituent order, iconic principles are of utmost importance, as we will see. The 'topic-comment' order inside statement clauses (declarative assertions), for example, as well as the more general tendency to place 'setting' elements before the assertion, are intrinsically iconic in nature, as is the positioning of 'elaborations' on such assertions after the assertion proper. Also the 'inverted order' found in some types of emphatic sentences can be seen as iconic. The related tendency to place very 'urgent', or 'salient', rhematic elements in clause and even sentence initial position, with topics and other 'setting' elements after the verb. In some languages, primarily verb-final languages,

any salient rhematic (focus) element comes in rheme-initial, preverbal. In other languages only the most 'urgent' ones, such as contrastive and emphatic ones do.

The functional motivations which leave their imprint on the constructions of language can work in unison to produce certain convergent outcomes. This is often referred to a 'conspiracy' to produce a certain form. Other times, those motivations seem to work at cross-purposes since they attempt to produce different, partly or wholly mutually incompatible outcomes. In such cases we say that the functions are in 'competition' with each other (cf., e.g., Du Bois 1985).⁸

1.2 SOME BASIC UNDERLYING ASSUMPTIONS

1.2.1 Introduction

All approaches to linguistic analysis start off with a series of more or less explicit assumptions about the subject matter: What is language? What deserves to be studied? What is interesting, or most interesting, about language? What is central to language and what is peripheral? Where do regularities come from? Is description the main goal of linguistic analysis, or is explanation also a viable goal? Under what circumstances is explanation a possible goal? What is the nature of explanation? And so on and so forth.

Since I do not have a single point of reference for answers to these questions, and since my assumptions may not be shared by many, this section will present some of the basic assumptions which underlie this study. I will discuss the issue of psychological reality and cognitive representation, the issue of the relationship between form and function, and the issue of explanation in linguistics.

1.2.2 Rules and representations: The list-rule fallacy

One of the major goals of all approaches to linguistic analysis is the extraction of patterns, or regularities, from the data. Patterns and regularities are said to be the predictable, or 'rule governed' aspects of language, as opposed to that which must merely be listed as arbitrary and unmotivated. Thus many approaches have sought to keep the two parts, the predictable and the arbitrary, separate using rules to formalize the former and lists of representations to formalize the latter.

Approaches to language such as the generative approach, which are concerned with representing speakers' knowledge, in addition to capturing linguistic generalizations, have transferred this distinction between rules and representations to the cognitive domain, giving cognitive status to the dichotomy. Even within this general approach there are different possibilities as to how regular, predictable, or redundant a phenomenon has to be in order to be represented by rules (either as linguists' rules or speaker's mental rules) and different approaches have been presented as to which types of regularity, or patterns, are presumed to be 'psychologically real' and which aren't. At one time even the most minor of regularities was assumed to call for a grammatical rule. This gave way to a perspective which dealt only with certain regular patterns and ignored others. Only recently, and under the influence of non-symbolic approaches to cognition, have alternatives surfaced in generative studies to handle minor regular patterns (cf., e.g., Pinker 1991).

This classical approach to the cognitive representation of linguistic generalizations is not without problems, as has often been noted. First of all, it is not at all clear that the linguist's pattern extraction methods correspond to the speakers'

methods of achieving generalizations. Another major problem has to do with the productivity of linguistic patterns. Although some patterns in any one language are quite general, productive, and predictable, and thus perhaps amenable to a rule characterization, most such patterns are not fully productive, but are rather 'semi-productive'. Thus it becomes necessary to make a decision as to how productive a pattern has to be to merit the linguist's, or the speaker's, attention. Some approaches draw the line according to certain assumptions about the nature of cognition and the rules (cf., e.g., Bolinger 1979).⁹

The major problem with the 'list-rule metaphor' in a cognitive-responsible approach to linguistic analysis which cares about representing linguistic knowledge, is that it presupposes a particular view of how the mind is organized and how it works which is quite questionable. This is a view of the mind in which regularity (and thus regularity extraction onto rules and rule application) is maximized and memory (memorization) is minimized. The idea that human beings, and in particular children learning the language, are like little descriptive linguists extracting regularity from the 'data' to place it in rules and to make lists of exceptions out of the remainder has been challenged by many linguists and psycholinguists (cf., e.g., Bolinger 1961a).¹⁰

The model of grammar which I envision is one which incorporates these insights. It allows patterns to be meshed together with the representations themselves, at all levels, types, and degrees of regularity, rather than keeping them separate. Given our current scant knowledge of the nature of cognitive linguistic representations, I believe we should adopt a conservative model in which the basic unit is the construction, a minimal pairing of form and function which is not predictable but must be learned and memorized by the speakers.

Such an approach allows us to incorporate the possibility that certain common expressions are also memorized by speakers. It also allows us to postulate relations between constructions in the form of more general patterns which may, or may not, have psychological reality. In other words, a construction-based approach doesn't consist of a mere listing of constructions. Rather, the constructions of a language may form a highly structured system in ways yet to be determined. Thus, a construction-based approach to grammar is compatible with approaches which reject the symbol-manipulation or rules-and-representations paradigm (such as connectionist and analogy based models).¹¹ I will return to this issue below.

The truth is that we do not really know how much speakers generalize from the linguistic input and how much they store. And we still don't know enough about the nature and characteristics of the cognitive storage medium and the retrieval capabilities and how constructions are related. Because of the precariousness of our knowledge about knowledge representation, I believe we should proceed with caution. We should continue to look for regularities, but without making undue assumptions as to how they are represented, while at the same time searching for the nature of that representation. This general sentiment also applies to the nature of innate knowledge and capabilities, for there are many indications that they are not of the nature assumed in the generative paradigm (for a review of some such approaches which present alternatives to rules see Lima, Corrigan and Iverson 1994). In any case we should always heed Van Valin's (1992) call to avoid all types of reductionism.¹²

1.2.3 Form and function

As I argued above, according to the functionalist view espoused here, form is irrevocably tied to function. This is a basic assumption about how knowledge is represented in the mind, as well as to how it should be represented in linguistic analysis. This would be not unlike Saussure's (1916) view of the sign as a unit composed of two poles: a formal one and a semantic (functional) one. Despite of the obviousness of this approach, at least when applied to lexical units, the challenges made to it for larger structural units (constructions) requires that this position be justified.

The difficulties involved in motivating properties of and constraints on language form, such as constraints on 'extraction' ('long distance dependencies'), from functional considerations have led some to argue that form at the grammatical level, much like form at the lexical level, is unmotivated, arbitrary, and, furthermore, autonomous from function. Many others believe, however, that such failure does not reflect that this is an impossible task, but rather that we have not yet found the relevant functional parameters or the relevant modes of representing linguistic knowledge and generalizations.

I believe that the concepts and categories which get lexicalized, i.e. to which form gets attached, in different languages are not preordained or given by universal, preexisting categories. Rather they depend to a great extent on how we cognize the world and our communicative needs and thus involve a certain amount of variation as well as a great amount of similarities. In the grammar, too, cognitive preferences and constraints, as well as informational needs, to a large extent motivate, or are reflected in, the form of constructions. The patterns of grammaticalization that we find, including those which involve the order of constituents inside constructions, reflect those cognitive and

informational parameters, just like patterns of lexicalization do. Furthermore, I believe that those forces are involved not only in the creation of those syntagmatic patterns in the process of grammaticalization, but to a large extent also in their cognitive representation. The fact that it is not always easy to ascertain how those forces are involved in the creation and the representation of grammatical units should not be enough of a reason for us to deny their existence or to give up on the possibility of finding valid generalizations.¹³

A major obstacle to getting a clear picture of the relationship between form and function in grammar is that true (historical and/or psychologically real) cognitive and informational motivations are not easily disentangled from other aspects of constructions which are more arbitrary (less motivated). Thus a construction may contain some properties which are 'motivated', that is, which can be explained or understood to some extent by 'external' factors, and others which are less motivated, or even completely arbitrary.

Ascertaining the presence of an external motivation, even if it mixed with less motivated factors, is not enough, however. Another major question still remains, namely, to what extent are motivated aspects of constructions and other syntagmatic phenomena motivated synchronically in the speaker's representations, and to what extent is the motivation simply fossilized into the construction at the time in which the construction was first grammaticalized?

These questions regarding degree of motivation of a formal property and the psychological reality of a motivation are not simple questions to answer and it may very well be that we will never find totally satisfactory answers. It seems to me, however, that

we do not need to give up the search before starting it. I believe that great progress can be made, at least in some areas, and that constituent order at the sentence and clausal level in asserted sentences is one such area.

1.2.4 The nature of explanation in linguistics

It is not universally popular at the present time in the field of linguistics to seek (functional) explanations to formal phenomena and indeed there are good reasons for having a healthy dose of skepticism, given past performance and inflated claims.¹⁴ I do think, however, that it is too early to give up on achieving a high degree of explanatory potential and understanding of why form (constructions) is the way it is, although it is also too early to claim victory, given how much there is that we still don't know.

If language is indeed like an ecosystem, and not like a mathematical, formal model, then we have some idea of the sorts of explanations which we will be able to find and the limitations which such explanations will be subject to. We know that the explanations we will be able to find will not be of the kind found in the physical sciences, where full predictability of phenomena is the norm, but, rather, given the complexity and interdependence of the factors involved, that it will be more of the type that we are accustomed to in the biological sciences. Thus, at some level we know that present form can be explained by means of past and present function. We also know that often certain inputs have certain effects on individuals, but that often the predictability of those effects is nowhere near certainty and that often our understanding of the phenomena is limited because we do not know, and perhaps cannot even hope to know, all of the variables involved.

In this study I sometimes use the term “explanation” in a ‘lax’ way, which has been rightly viewed with suspicion by many investigators. Explanation is an ambiguous term, which can refer either to an initial cause or propensity for a certain outcome, as well as for the actual mechanism by which an outcome obtains. To take an example from biology, the most famous case of adaptation in biology textbooks is perhaps the replacement of England’s white-and-black peppered moth by black-winged ones in the 19th century. This replacement is commonly attributed to the fact that industrial soot blackened the trees where these moths rested and killed the lichens on the trees, which made the light moths more susceptible to being spotted by predators. In a lax way it can be said sometimes that the darkening of the trees *explains* the change in the color of the moths. Of course, that is not the explanation of how the replacement actually took place, which has to do with natural selection and adaptation to the (new) environment, but rather the motivation, or in other words, what made it possible for natural selection to produce the results it did.

In linguistics we do not always have explanatory tools analogous to natural selection to account for the facts, especially since linguistic adaptation seems to be subject to a number of variables much more complex than the one in this example. But, of course, in evolutionary biology the ‘motivations’ for change (adaptations) aren’t always as obvious as in the previous example. Changes in an ecosystem which lead to an adaptation (and thus, loosely speaking, explain it), are typically harder to determine. Still, it is not uncommon for scientists to be able to trace a change in a population to a single main source, even if they cannot explain each and every step that followed in the ecosystem from that original event and which resulted in a certain outcome.

In the field of pathology we also find these two extremes. On the one hand there are cases which are fairly straightforward. Thus we know that arsenic ingestion can 'cause' (and 'explain') a person's death. Of course, the ingestion of arsenic doesn't cause or explain the death; rather the explanation lies in the actual effect of the arsenic on organs vital to the sustainment of life. However, given the certain cause and effect involved, we feel justified in saying that arsenic ingestion *explains* someone's death. Many cases are much more complex than this one. Thus, for instance, frequent meat consumption doesn't guarantee a heart attack, it merely correlates with (increases?) one's chances of suffering one. So, can we say that an avid meat eater's heart attack has been caused, or is explained, by eating meat? Or that a smoker's lung cancer was caused by (and only by) his or her smoking? We must be much more careful here. However, speaking in general terms, we can probably be somewhat lax in the use of the term, as long as we are aware that that is what we are doing. In the case of smoking, the connection between this activity and lung cancer seems to be more direct than in the case of meat and heart attack, but still only 1/3 of smokers develop lung cancer and some people who are not exposed to smoke develop cancer as well (cf. Newmeyer 1994:72 for a similar example, attributed to Dryer p.c.).

In the study of language too, by looking at recurring changes and patterns we can often identify the motivations which lead, in a significant number of cases (at least given other accompanying factors) to a certain outcome. That is, we find that different languages follow recurrent paths (solutions) when presented with certain situations, and in that sense we can say that those situations *explain* the outcomes, even if all they do is lead the language in a certain direction (cf. Aitchison 1987, 1989). Of course, the actual

mechanisms for moving along that path are also part of the explanation. We must be concerned with both aspects of a situation, though not always with both of them at the same time.

For all these reasons, instead of the terms explanation and explain, which some insist on associating with explanations in the physical sciences, we may rather use the terms motivation and motivate, as is commonly done in functionalist studies. Thus we say that the name *apple juice* is 'motivated' by the nature of the concept and that it 'makes sense' (to a learner for instance), after the fact. This term is also motivated system internally by the fact that the name for other drinks made from fruits follow the pattern <fruit> juice.¹⁵ The term, however, is still arbitrarily assigned to this drink and the fact that the two parts of this compound have that particular ordering is in some sense arbitrary.¹⁶

1.2.5 Description vs. explanation in linguistic analysis

Description is the primary goal of linguistic analysis.¹⁷ But explanation, in the sense explained above, must go hand in hand with description in a functional-typological approach. The similarities among languages, as well as the differences, must have explanations, and linguists may, or indeed must, try to find the source of those similarities and dissimilarities to the extent that this is possible. It goes without saying that linguists must be extremely cautious in proposing possible explanations for linguistic facts, but that doesn't mean that they shouldn't try to find them after the descriptive job is done, and to the extent that this is possible, while it is being done (cf., e.g., Comrie 1993).¹⁸

Linguistics will rarely approach the levels of (deterministic) explanation associated with the physical sciences, where the number of factors involved is typically small and manageable. However, I believe that sometimes we can achieve high levels of understanding for the form of constructions in terms of the functions the constructions perform, the cognitive constraints on verbalization and processing, and the processes of grammaticalization based on 'institutionalization' of common functions (cf. Haiman 1991a, 1993, 1994, 1995). Only as a last resort do we want to entertain the possibility that form correlations and cross-linguistic commonality (or similarity) of form have a cause which cannot be related, directly or indirectly to functional factors, that is, the possibility that they are arbitrary or perhaps innate (cf. Aitchison 1989:152).¹⁹

The failure to find direct, deterministic mappings between external functions ('motivations') and the actual sentences of a language should not be a reason to abandon all hope of understanding the non-arbitrary types of connections that exist between them. 'External' language functions and constraints do not necessarily express themselves directly on linguistic form, i.e. on the sentences and utterances of a language, but rather, are typically mediated by the linguistic constructions of languages (grammar). Constructions are 'distilled' or 'sedimented' through the process of 'grammaticalization', in which routinary verbal events representing a more direct mapping of (external) function to form (given the preexisting lexical and grammatical resources of the language) become established (cf. Section 1.4.6).

1.2.6 Explanation vs. motivation

I agree to a great extent with Lambrecht, who while agreeing that “grammatical structures arise diachronically under pressure from information structure constraints” (Lambrecht 1994:29), nonetheless argues that he “would find it misleading to say that the communicative requirements of discourse *directly determine* not only the content but also the form of utterances and that information structure can in some sense *EXPLAIN* the structure of sentences” (Lambrecht 1994:26, my italics, J.A.). However, in discussing linguistic phenomena we rarely talk about functions ‘directly determining’ form. Even the most iconic of form-function relationships must perhaps have a minimal of arbitrariness, namely the fact that they are established patterns in the language. What we must establish is the degree and the type of motivation of a particular pattern (Lambrecht himself uses the concept of motivation in the same way it is used here, cf. Lambrecht 1994:29).

The real question is whether a particular ‘motivation’ is purely diachronic and lost in time, as in the case of unmotivated apparent compounds, such as *butterfly*, or whether the motivation has a synchronic reality for speakers, as supposedly in the case of *apple juice*. Linguists tend to take polar positions with respect to this issue: the relation is either diachronic and arbitrary, or it is fully transparent and directly productive synchronically. The former position is typical of formalist approaches (cf. Hyman 1984, Newmeyer 1994). The latter position is perhaps close to the attitude of a minority of functionalists. The truth, it seems to me, and I believe that this attitude is common with most functionalists, must be much more complex and somewhere in between these two extremes.

Another major, and related, obstacle to obtaining clear form-function correspondences in language has to do with the fact that the number of possible expressible functions competing for coding is much greater than the number of functions which actually get formally coded. Different functions thus must compete with each other for the limited coding resources (cf., e.g., Du Bois 1985; Croft 1990:7.4; Gibson 1992; Newmeyer 1994). We do find, however, that some functions are more crucial than others and tend to be coded more pervasively. Also we find differences, for instance, in the number of possible ways in which different functions are coded.

Another point to keep in mind, as I said, is that the fact that something has an explanation doesn't mean that it is meaningful and thus, also, synchronically motivated, or motivated for all speakers. It is fairly clear now that some typological correlations found across languages are merely due to the fact that through the process of grammaticalization some outcomes are more likely or more common than others (cf. Section 1.6.15). So, for instance, languages with OV order and pre-nominal genitives also typically have pre-nominal relative clauses. It seems that the 'immediate' reason for the prenominal order of relative clauses in these cases is often due to the fact that relative clauses are derived from genitive phrases with a clausal genitive phrase. So the fact that relatives are pre-nominal may be 'explained' (or 'motivated') in a way by their diachronic source. On the other hand, it may also be the case that prenominal relative clauses (as well as prenominal genitives) are preferable for some reason in these verb-final languages, perhaps for reasons having to do with processing advantages. Thus, although such ordering for relative and genitive clauses doesn't synchronically code any function

for speakers it is still 'motivated' and not totally arbitrary at some level, though of course, not at all 'directly determined' synchronically by anything.

On the other hand, other types of 'motivations', such as discourse-pragmatic functions of sentence organization in speech act constructions, which are the topic of this study, seem to me to be much more 'alive' ('synchronically real') and constantly relevant in the grammar of a language, though more so in some languages than in others perhaps (cf. Section 1.6.7 below). Thus, for instance, the 'strategy' of verbalizing a referent first and then asserting some information about that referent is perhaps grammaticalized in the statement constructions of most ('subject initial') languages. But even if this strategy is 'grammaticalized' into constructions in these languages in some way or another, the 'strategy' seems to me to be much more general and psychologically real since it is based on a pre-linguistic and iconic communicative principle which is immediately transparent to all speakers, including speakers of those languages which are said to be non-subject initial. (I will return to this example in Section 1.3 below.)

1.2.7 Conclusions

To summarize, the search for explanation and the connection between form and function exists at many levels and cannot be dismissed as an idealistic, misguided, or impossible endeavor. Possible explanations are of many kinds, from purely diachronic reasons which are only cognitively relevant at the time they act on the system by favoring the creation certain structures, to those which, whether they are grammaticalized or not, are also made use of synchronically, either to make sense of and organize those stored linguistic units, or to deploy them more or less freely in the production of discourse.

The types of explanation that we seek are not necessarily deterministic ones for why every element of an utterance is the way it is, for that is an unrealistic goal in such a complex system as language. The type of explanation that we are after seeks at least some understanding of why certain solutions to coding problems are the way they are, why some are so prevalent and some so rare, which ones are more 'functional' and which ones more 'dysfunctional', under what circumstances and why.

The functional-typological approach to linguistic analysis and explanation is still in its infancy, which is why analyses from this perspective often seem like groping in the dark. There are so many interrelated factors, so many motivations, so many possible reasons for things to go one way or another. Still, there are many aspects of language which we are making great progress in beginning to understand, such as, for instance, the general principles involved in the ordering of linguistic units at the discourse, sentence, and clausal levels. To deny that explanation is possible because we cannot explain everything, all at once, or even because not everything is explainable is a defeatist attitude.

1.3 PRAGMATIC PRINCIPLES AND THE GRAMMAR

1.3.1 Introduction

As I said earlier, some functional pressures or preferences manifest themselves at certain points in the history of a language when certain constructions are being created, or recreated, to help determine (motivate) the form of those constructions in the process of grammaticalization. Principles of this kind may or may not continue to express

themselves once their job is done. I gave as a possible example of this the development of prenominal relative clauses in verb-final languages.

On the other hand, there exist discourse principles of what we could call 'information flow', pertaining to the ordering of elements in sentences (including elements within the main clause, as well as elements which precede and follow the main clause), which are often discernible in major grammaticalized syntactic constructions crosslinguistically, which do not seem to be 'forgotten' by speakers. Strategies of organization such as the topic-comment order in statements and comment-topic order ('subject inversion') in emphatic statements, for instance, are quite iconic and are used at other levels beyond the simple clause.

In other words, principles of discourse structure are reflected at two different levels: the grammatical and the pragmatic. At the grammatical level, they may be embedded in the structured collection of constructions that are used for communication (i.e. the grammar). But they also exist at a pragmatic level, accessible to all humans, given the basic character and the natural iconicity of these principles.

1.3.2 An example: The Setting-Assertion communicative strategy

One of the most basic patterns of discourse organization is perhaps the one in which the speaker presents some idea (such as a referent or a proposition), and immediately afterwards says something about that idea or which applies in some way to that idea. Following the terminology used for one of this pattern's subtypes, we may call the first part the 'protasis' and the second part the 'apodosis'. I will often call them 'setting' and 'assertion'.

This extremely iconic principle expresses itself in a variety of ways in communication, including a variety of major constructions in most languages. The preferred position for conditional clauses, temporal clauses, connectors, vocatives, and other such phrases is sentence initial position. The preferred initial position for overt subjects, elements which I will argue in Chapter 3 are a special type of setting, also reflects this pattern, as do many other constructions. We can see a number of examples in Table 1-1 below.

Setting (Protasis)	Assertion (Apodosis)
My mother	is a great dancer.
My mother	she's a great dancer.
All those over 55	out of the dance-floor!
That woman	I really like going dancing with her
There are some shoes	that don't hurt when you dance
Those are the shoes	which I wore dancing last night
Those shoes	I don't like very much for dancing
John	how good a dancer is he?
When I was young	I used to go dancing all the time
I met a man the other day	who could dance very well
On the dance-floor	there is magnificent dancer
If you go dancing	don't forget to bring comfortable shoes
Show me you can dance	and I'll believe you
I was happily dancing	when the lights suddenly went off
Mother	dance with me!
Mother	will you dance with me?

Table 1-1: Some examples of constructions displaying a 'setting-assertion' pattern.

As we can see, the first part of these 'binomials' may contain elements which can also be used as independent assertions. Other protases, however, are quite specialized. All of the instantiations of this pattern are highly iconic, even those which are highly grammaticalized, such as the subject-predicate construction.

This pattern however, is not the only one we find in human language. In fact its very opposite, namely ‘assertion-setting’ order, is also sometimes found under some special, or marked, circumstances, typically having to do with a certain urgency involved in the expression of the assertion, such as in cases when the assertion is emphatic. Thus, whereas the setting-assertion pattern is by far the most common one under most circumstances in communication, if the assertion is ‘salient’ in some way, it may come first, displaying what has been called an ‘inverted order’.

In Table 1-2 we can see some instantiations of this inverted order pattern in English. This pattern, as we can see, is more constrained than the previous one, and thus not all setting-assertion binomials have corresponding assertion-setting ones, though languages may differ as to how unmarked and prevalent this latter pattern is.

Assertion	Setting
She’s a great dancer	my mother
Out of the dance-floor	all those over 55!
I don’t like them very much for dancing	those shoes
I used to go dancing all the time	when I was young
There is magnificent dancer	on the dance-floor
Don’t forget to bring comfortable shoes	if you go dancing
Dance with me	Mother
Won’t you dance with me?	Mother

Table 1-2: Some examples of constructions displaying a assertion-setting ordering pattern.

In the following chapters I will discuss some aspects of these and other general order patterns, primarily as they apply to the Basque language.

1.3.3 Discourse pragmatic principles and grammar

The basic iconic ordering patterns which I have shown in the previous section have been recognized for a long time (cf. Chapter 3). The main challenge is, however, determining how they are manifested in each language and how they interact with the (rest of the) grammar in those languages.

The claim that these iconic patterns are strong ‘synchronic motivators’ and not merely grammaticalized into constructions and then ‘forgotten’ receives support from the proposal that there exist two different modes of communication in language, a ‘pragmatic mode’, one which follows iconic principles more closely, and a ‘grammatical mode’ which relies more on the grammatical(ized) constructions of the language. This dichotomy has been argued for strongly by Givón for instance (cf. Givón 1979:Chapter 5; 1989:Chapter 7; 1990; 1995:359-60).

Givón himself traces this idea to prior studies about different realms of language: diachrony, ontogeny, pidgins and creoles, and more vs. less planned language (‘formal’ and ‘informal’ registers), in which we can discern a contrast between a more ‘pragmatic mode’ and a more ‘grammatical mode’ of communication (cf. Sankoff and Brown 1976, (Ochs) Keenan 1977, 1979; Slobin 1977; Givón 1979:208). Thus in diachrony, “loose parataxis”, as in topic-comment structures, becomes grammaticalized into subject-predicate constructions. Likewise, in child acquisition, pragmatic structures are learned first, only to give way to ‘tighter’ grammatical constructions later on. But also synchronically, for every individual speaker, the two modes coexist, with informal conversation displaying more characteristics of the ‘pragmatic mode’ and more planned, careful (and written) speech displaying fewer such characteristics and more

grammaticalized constructions. More fluent, or 'competent' speakers also produce a greater variety of grammatical constructions, whereas less fluent ones display lesser variety and a greater proportion of more pragmatic structures.²⁰

According to Ochs, in adult language "[a] major condition affecting adult reliance on early communicative patterns is the extent to which the communication has been PLANNED prior to its delivery" (Ochs 1979:53). This would account for many of the differences between spoken language and (most types of) written language, perhaps the most extreme form of planned language. Also, "[s]ome of the most extreme examples of planned language, such as the use of more formal discourse devices (e.g., use of textual cohesion and transitional terms)" seem to be acquired later in life, that is they "draw upon knowledge transmitted through formal education" (Ochs 1979:54). There is nothing inherently 'better' about more grammatical and 'richer' modes of communication, but they do seem to be more efficient, as has often been noted.

One of the most interesting features associated with children's discourse and with unplanned discourse, according to Ochs, is the use of "relatively simple morphosyntactic structures" (Ochs 1979:61). One of the most interesting aspects of this phenomenon is the use of loose 'Referent + Proposition' constructions (in separate intonation units), i.e. setting up a referent and then saying something about it, following the setting-assertion pattern I mentioned above (Ochs 1979:66). These structures are used instead of more 'grammatic' subject-predicate ones. An intermediate construction, also frequent in speech, but much less common in writing and other forms of planned speech is 'left-dislocation' with a 'resumptive pronoun' inside the clause, also a hallmark of spoken language, but relatively rare in writing.

A second manifestation of this phenomenon is the use of adjacent (paratactic) and coordinate clauses instead of subordinated ones, the relation between the two being 'recoverable' from the context, e.g. "I don't like that house. It looks strange" (Ochs 1979:66). This is one of the ways in which "[i]n relatively unplanned discourse more than in planned discourse, speakers rely on the immediate context to express propositions" (Ochs 1979:62).

The grammatical and the pragmatic modes of communication, are of course, idealizations and not actual separate and independent systems. In reality, speakers typically use different blends of more pragmatic and more grammatical structures at different times depending on ability (fluency), degree of planning, and so on. In addition, there are also very strong indications that the grammatical system is built upon the pragmatic system, and it is grafted on to it, without ever totally replacing it. In Table 1-3 we can see the major characteristics of these two major modes of communication as summarized by Givón (1979) drawing primarily from (Ochs) Keenan 1977, 1979.

Pragmatic mode	Syntactic mode
Topic-comment structure; more topicalized, left-dislocated structures	Subject-predicate structure
Loose coordination; less embedding.	Tight subordination
Slow rate of delivery (under several intonation contours)	Fast rate of delivery (under a single intonational contour)
Word-order is governed mostly by one <i>pragmatic</i> principle: old information goes first, new information follows.	Word-order is used to signal <i>semantic</i> case-functions (though it may also be used to indicate pragmatic-topicality relations)
Roughly one-to-one ratio of verbs-to-nouns in discourse, with the verbs being semantically simple	A larger ration of nouns-over-verbs in discourse, with the verbs being semantically complex
Reduced morphology	Elaborate use of grammatical morphology
Prominent intonation—stress marks the focus of new information; topic intonation is less prominent	Very much the same, but perhaps not exhibiting as high a functional load, and at least in some languages totally absent.
Shorter clauses	Longer clauses

Table 1-3: Major characteristics of the pragmatic and the grammatical modes of communication (from Givón 1979:223, ex. 41, and 229).

Givón argues that “[t]he extreme instance of the formal-planned pole is educated, book-written language” (Givón 1979:230), and that linguistic analysis is often based, whether consciously or not, on this more ‘grammatical’ end of the spectrum. It is only when one begins to look in earnest at how people actually speak, by making use of actual recordings and detailed transcripts, that the contrasts between the two modes of communication become quite apparent.

The conclusion that Givón draws is that grammar, the collection of ‘sedimented’ grammatical categories and constructions of the language, is a way of facilitating communication by automating choices of the material to be communicated to allow the speaker to concentrate on the content of communication. In other words, grammar is a way of facilitating the “automatic processing of speech”.²¹

In Table 1-4 we see a more recent list from Givón's work of the differences between the two modes of communication, according to the different domains in which the differences are relevant.

Area		Grammatical mode	Pre-grammatical mode
STRUCTURAL	<i>Morphology</i> <i>Syntactic constructions</i> <i>Word order</i> <i>Pauses</i>	More abundant Complex/Embedded Grammatical (subject/object) Fluent	Less Simple/Conjoined Pragmatic (topic/comment) Halting
FUNCTIONAL	<i>Processing speed</i> <i>Mental effort</i> <i>Error rate</i> <i>Context dependence</i>	Fast Effortless lower lower	Slow Laborious higher higher
COGNITIVE	<i>Processing mode</i> <i>Acquisition</i> <i>Evolution</i>	Automated Late Late	Attended Early Early

Table 1-4: Differences between the pre-grammatical and the grammatical modes of communication; adapted from Givón 1995:360.

As we can see the idea is that the grammatical mode is more complex, more fluent, and acquired later than the pre-grammatical or 'pragmatic' mode.

To summarize, I believe that the area of discourse organization is one in which highly motivated iconic patterns and principles operate. As I will argue in Chapter 3, these principles also have a great influence on the word order inside what is usually known as sentences and even asserted clauses. In other words, constituent order is not only the province of arbitrary grammatical constructions, but also of iconic functional principles or organization. This is true of all languages to some extent, though some languages have a greater tendency to use word order to reflect pragmatic properties, as we will see is the case in Basque. Let us look now in further detail at the relationship between form and function and the nature of grammar.

1.4 FORM AND FUNCTION REVISITED

1.4.1 The separation of form and function

The idea of studying form independently from function in modern linguistics in the United States goes back to some behaviorist structuralists' aversion to the study of meaning. In the generative approach which succeeded this school, the reluctance to incorporate meaning and function probably had more to do with the linguists' view of the grammatical system through what we could call a mathematical system metaphor. The impossibility of fully accounting for form by means of function (meaning) led to the extreme of fully divorcing the two, and the attempt to "do the syntax autonomously" (Ray Jackendoff in Huck and Goldsmith 1995:98).²²

During a certain period in the history of the MIT School some serious attempts were made to relate structure and meaning, but it was done in a way which treated all meaning-form correspondences as being fully systematic, with function (meaning) fully determining form.²³ The failure of this approach was responsible for the return to a formalist approach which concentrated on phenomena of the more categorical (regular) variety, or close enough so that problematic cases could be dismissed in one way or another. In this way structural patterns which were not fully productive, or regular, and thus were subject to semantic or arbitrary considerations, were relegated to 'lesser domains', such as the lexicon or the 'periphery' of the grammar. Not only was semantics separated from syntax, but all context connected syntactic phenomena were relegated to "performance" and "optional rules" and "stylistic rules", for which there was no room in the grammar.²⁴

This formalist, or autonomist, approach has been extremely influential in linguistics. The functional approach, however, with just as long a history in modern linguistics, has remained strong all the while. At present many linguists recognize the need for including functional considerations in grammatical analyses and that grammatical constructions and categories may reflect functional pressures, as well as express functional (semantic and pragmatic) meanings. Typology and grammaticalization are two areas in which very fruitful work has been done in recent years from a functionalist perspective.

Thus we can say that there exists a dichotomy, or perhaps a cline, in the field of linguistics between those who eschew functional notions, at least those which cannot be safely incorporated into the grammatical system and those who are willing to explore messier relationships even at the risk of getting dirty.

In this section I will attempt to further motivate the inclusion of functions and meaning into our linguistic analyses. I will first take a closer look at the nature of the relationship between meaning and form. Next I will analyze the nature of syntactic structure or constituency. Finally, I will discuss the notion of constructions as the basic units of grammar, the dichotomy between structure and process, and finally the process of grammaticalization by which meanings become institutionalized into grammatical units.

1.4.2 The relationship between form and function

The 'forms' of language include grammatical categories, such as nominative or ergative case marking, constructions, such as the different major and minor speech act constructions of a language, or even some aspect of those constructions, such as the

relative position of an element in such constructions, when a choice is available. The functions corresponding to those forms are, of course, their meanings and the uses that they are put to. These meanings, of course, may very well have complex structure, exhibit prototype effects, be polysemous, and have different uses in different contexts, which is what makes matters so complex.

But what does it mean when we say that a particular form codes a function, or alternatively that a function is coded by a particular form? In the most literal sense it would seem to mean that whenever a form is found (in a particular construction or other context perhaps), it expresses that function, and that whenever speakers want to express that function they use that particular form. In other words, we might expect a biunique relation: Form \leftrightarrow Function.

In practice, however, what we find is rarely this simple, which has led some linguists to question the importance, or the centrality, of form-function relationships in language, and thus led them to the position that form can be studied independently of function. Understanding the nature of the relation between form and function is thus crucial to a function based approach to grammar.²⁵

The first thing that we must keep in mind is the complexity of the relationship between form and function. As Bates (1994) notes, the relationship between form and function is not direct and transparent

Because grammars represent complex solutions to an enormous constraint satisfaction problem, their one-to-one relationship to particular meanings or classes of meanings will be rendered necessarily opaque at many points. ... if our approach is correct, it also means that the effort to motivate each and every surface reflex in semantic/pragmatic terms may be self-defeating. (Bates 1994).

On the other hand, the idea that form codes meaning is to a certain extent misleading, since a great part of the meaning that we communicate is not directly coded by the forms of language, but, rather, is implicit. As Slobin argues,

Language *evokes* ideas; it does not represent them. Linguistic expression is thus *not* a natural map of consciousness or thought. It is a highly selective and conventionally schematic map. At the heart of language use is the tacit assumption that most of the message can be left unsaid, because of mutual understanding (and probably also mutual impatience). (Slobin 1982:132)

Many 'meanings' are never coded directly, and the number of possible meanings inherent in events and situations that we use language to express and which could conceivably be coded is always much smaller than the number of actual meanings which get coded more or less directly and schematically.

Another complication is presented by the fact particular forms often code more than one function. In other words, much like words in the lexicon, grammatical forms are often polysemous and rely on the context for the proper interpretation (cf. Haiman 1985a&b; Givón 1992a, 1995). This is because there are many more possible functions (or meanings) than there are forms to express them and very often functions are extended to express related meanings beyond those that they were originally intended. The context is then relied on for the proper interpretation, be it the grammatical context or other types of context. Because a form may code more than one function, as Givón has argued, typically "the predictability is much stronger from function to form than from form to function" (Givón 1992a:316).²⁶

The fact that forms, whether grammatical categories or constructions of different types, get extended to express meanings other than those originally meant is well

recognized. Thus for instance it has been noticed that in English cleft constructions, which are at heart a focalizing device, a way of extracting (and fronting) the most rhematically salient aspect of a proposition which is accessible to the hearer, can be used in a new context, such as in introductions to formal lectures, e.g. *It was Churchill who said ...* (Prince 1978). This does not mean that the original function of the construction has been lost. Rather, the construction has acquired a new use in a very specific context, without losing in any way the original function which so iconically marks the 'clefted' element as salient and the rest (a 'relative clause' like clause) as accessible or backgrounded.

Another example of constructional polysemy might involve the often discussed passive and dative-shift constructions, which have been 'explained' in a variety of ways (cf., e.g., Hopper and Thompson 1993:365). Thus the passive construction, for instance, may in some contexts serve primarily to make a patient the topic of the clause. But the very same construction, for instance, may also serve to code the patient as the focus in clause initial position (cf. Chapter 3).

The importance and frequency of a function also has a lot to do with its being coded formally and in a more explicit and unambiguous manner (cf. Du Bois 1985:363; Hopper and Thompson 1993:357-58; Givón 1995). Very common functions tend to receive stable, unequivocal coding, whereas the same may not be the case for other less frequent or less crucial functions. Likewise, very strong cognitive or processing constraints are much more likely to express themselves in the language than weaker ones. Less important or less frequent functions, on the other hand, may never become

grammaticalized. And weaker or less frequent cognitive constraints are also less likely to have an effect on the form than stronger or more frequently occurring ones.

The fact that a constraint is very strong does not necessarily mean that it will be reflected (coded?) in the grammar. Center-embedding avoidance, for instance, is a very strong constraint in language. Kuno (1974) claimed that this processing difficulty has a clear effect on form, namely that it is responsible for the fact that OV languages have prenominal relative clauses. On the other hand, we find that sentences with multiple embeddings of this type are quite rare in language (as well as easy to avoid or circumvent) and thus it is not clear that center-embedding avoidance could have such a strong repercussion on the grammar of a language.²⁷

We also find that many of the correlations between form and function that have been proposed are imperfect, but still statistically significant. This can be due to a variety of causes. On one hand, a non-categorical or imperfect correlation may derive from the fact that there are additional factors involved which have not yet been identified or properly fit into the picture, such as additional aspects of the context or the particular construction in which the correlation holds.

On the other hand, we should always keep in mind that correlation doesn't always involve causation.²⁸ So just because there is a statistically significant correlation between a form and a function, it doesn't mean that the form was *intended* to code that meaning or that the correlation forms part of the cognitive representation of the construction. The correlation could be indirect. The correlation between imperfective aspect and ('background' or 'setting') non-asserted clauses might seem to be of this kind, for instance (cf. Hopper 1979a&b; Givón 1987). However, the fact that there is a correlation

between imperfective aspect and this type of clauses doesn't mean that imperfective aspect is involved in the *coding* of such sentences. The correlation between perfect and perfective aspects on the one hand, and passive constructions may also be of this type. And the correlations that are found between ergativity and tense/aspect or topicality hierarchies and agentivity do not necessarily suggest that ergative case marking or ergative constructions are used to code those meanings (cf. Garrett 1990).

This doesn't mean that we should be interested in 'text correlations' if they are not also psychologically real correlations in the minds of speakers. Such correlations may provide us with very interesting information about the nature of language. But we must be careful to keep the two matters separate. I do believe, however, that our final goal should be to ascertain which correlations are psychologically real. As Givón argues, "text distributions by themselves are not the **end** of the investigation, they are only its **empirical means**. Our hypotheses, what we endeavor to construct as explanations, must not be **about the text** ... [but] **about the mind** that produced or perceived the text" (Givón 1992a:317).

Alternatively, imperfect form-function correlations in constructions (or even among constructions) may involve the influence of certain principles or constraints at the diachronic level, that is, at the time when such constructions were first created and established in the language (grammaticalized), and it may be that they play no role in the actual coding of anything by the speaker at the synchronic level. The original construction may have even changed significantly since it was originally grammaticalized (responding to other needs or pressures). Such correlations are still of interest to the

linguist for what they tell us about language change, but may not be terribly illuminating about the synchronic system and how it is cognitively organized or used by speakers.

When we look at a particular form and a particular function in a construction, we must also keep in mind how the form and function interact with other forms and functions in the construction, and how they interact with the whole system of constructions. Within a construction meaning is not necessarily compositional, and we may find that not every single formal aspect of a construction is paired with a specific single meaning. That is because the construction may add, subtract, or modify the meaning of the whole in different ways.

Also, as I said above, grammars are not unstructured lists of constructions and categories, but are rather structured, forming a more or less tight system of constructions. In other words, to understand a certain form-function pairing, and its limitations, we must often look at how the whole system is organized, and even 'self-organized'.²⁹ This doesn't imply that the linguistic system is a "formal deductive system" where "everything is explained and everything fits" (Bolinger 1979:110). As has been long recognized, grammatical systems "leak" (cf. Sapir 1921, Bolinger 1979).³⁰ However, we must not underestimate the importance of the whole system when looking at the individual parts.

Another source of the imperfect correlations between form and function is the nature of the categories themselves. As we now know, the categories that humans use are not necessarily like classical essentialist categories and they may have rather complex structure and exhibit 'prototype effects'. As Lakoff (1987) argues, the source of prototype effects is quite varied. Polysemy is one major source, both at the lexical and the grammatical levels. Polysemy comes about by the application of a category to new

situations through motivated extensions of the original meaning. The new sense may be added by making the original sense more general, but the relationship between the senses may also be much more complex and involve contextual parameters for instance (cf., e.g., Lakoff 1987; Fox 1995).

Finally, it seems that even when a form really does code a function (in a particular context or construction), the faithfulness of the coding is rarely categorical in language. It seems that, as Givón has argued, speakers can live with a minor percentage of cases in which the default is overridden for one reason or another (cf. Givón 1992a). Thus for instance, the correlation between the formal category subject in nominative languages and the category topic (see Chapter 3) is very close to categorical in actual language use, i.e. most subjects are topics, and it may very well be that such a category responds to the pressure of coding the default topic of the clause.

On the other hand, it is well known that in actual language use not all subjects are topics. Sometimes they are incorporated into the rheme, either because there is no topic, or because some other argument is the topic (cf. Chapter 3). But languages always have some way of formally indicating that this is the case, of overriding the default, such as by placing the subject argument in focus position. It seems to me that to conclude from the fact that some nominative arguments are not topics that there is a formal category which is totally divorced from the functional category topic is a fundamentally misguided conclusion.

The same thing may be true of the hypothesized formal unit called the verb-phrase. As I will argue below, this unit seems to be a reflex of the topic-comment informational organization found in the vast majority of assertions in language. On the

other hand, not all clauses seem to have such an organization and it is not true that the comment, or 'rheme', is always composed of the clause minus the subject (some other argument could be the topic). I will return to this particular unit in more detail in the next section.

1.4.3 Functional structure vs. formal structure (constituency)

I have argued that there is an inherent unity and interdependence between form and function in language, such as between formal categories and semantic and pragmatic functions, even if the relationship is not always fully transparent and biunique. A similar argument can be made at the level of grammatical structure, or constituency. I believe that structure (constituents) is the result of functional forces which hold elements together and that it doesn't make sense to speak of structure and constituency being a separate realm independent of functional considerations.

The main 'forces' which hold elements together in discourse are of two types: semantic and informational. Semantic 'forces' are responsible for the fact that all the elements of a clause, a verb and its complements, act as a unit. It is also responsible for the fact that noun phrases, a noun and its complements and modifiers, act as units. These 'forces' are not enough to predict all the properties of a construction, such as their relative ordering, but there is little doubt that it is these semantic forces which keep the elements together, both conceptually and formally.

Semantic forces or 'bonding' also seems to be responsible for the fact that different verbal elements in a clause, which may derive from different structures, such as clausal complements, very often fuse into a single verb in many languages. We will see

later on in this study that in Basque, auxiliaries and non-finite verbs, which used to belong to different semantic (and thus grammatical) units, have fused into a single word in some contexts, with interesting repercussions for word order. This type of semantic bonding, which no doubt exists at some level even before lexical fusion takes place, is impossible to represent in a framework for which structure consists of brackets or trees and not of semantic bonds. What's more, semantic bonds may be overlapping, something which cannot be easily represented in current formal theories.

Thus, for example, syntactic tests such as pronominalization and movement tests and theory internal considerations have been used to argue that a sentence such as *John should have given that book to Mary*, has the structure in (1.1) below.³¹

(1.1) [s John [v_P should [have [given [that book] [to Mary]]]]]]

But even if these are (or have been) the units of this clause at some level, it seems obvious that the 'constituency' of the whole clause or of the noun phrase *that book* is of a different nature from the 'constituency' of the (supposed) unit *given that book to Mary*, for example. Also, many would argue that somehow the 'verbal complex' *should have given* also forms a unit of some kind (cf., e.g., Dixon 1991:22).³² Semantically there is a bond among the verbal elements, which is what leads to their 'restructuring' into single words in some languages. I believe that this fact must have a synchronic validity, at some level, something which formal theories do not have a ready way to represent.

Just as problematic here is the so-called VP constituent, exemplified maximally by the string: *should have given that book to Mary*. While noun phrases and clauses are generally considered as grammatical units or constituents, many have questioned its

existence or, at the very least, that it is a unit of the same type as the other two.³³ I agree with this latter assessment. However, the common assumption in many approaches (and not only formal ones) that all types of constituents are of the same type (i.e. 'syntactic units'), forces them to seek an affirmative or a negative answer to the question of whether there is a VP unit in language.³⁴

Because of the difficulties involved, some most investigators have taken the all-or-none route, including arguing that some languages have a VP constituent while others don't, according to some 'configurationality parameter'.³⁵ That, I would suggest, is not a satisfactory answer. It is the type of answer that we are led to by faulty assumptions about the nature of structure. The answer to this conundrum should be sought in the fact that the forces which hold together the elements of the 'predicate' (the clause minus the 'subject'), in those constructions in which those forces exist, are not of the same type as those which hold the elements of the clause or a noun phrase together, for example.

This solution, however, is not readily available in frameworks in which, synchronically speaking, constituency is all or none, as represented by pairs of brackets, boxes, or (sub)trees. In the following chapters I will attempt to show that what makes that part of the sentence which excludes the subject (when one is overtly expressed, which is not very often in actual clauses in many languages), corresponds primarily to a semantic-informational unit, the comment, rheme or predicate.

Thus, an initial attempt to represent the bondings among the elements in a sentence such as the one in (1.1) would include different types of units, perhaps by using different types of brackets for different types of bonds, as in (1.2).³⁶

(1.2) < [John < {should have given} [that book] [to Mary] >] >

Although formalists may view with distaste the introduction of such unconstrained formal devices, it seems to me much more important to attempt to represent the facts about linguistic structure than to build a highly constrained theoretical edifice on questionable assumptions.

Another major problem that I see with constituency in formal theories is that, along with abstract elements, it is often used diacritically in order to make up for the deficiencies caused by the lack of functional information in the representation of constructions. Thus, some investigators have claimed, for instance, that all structures are binary (head + complement), in order to account for example for the fact that the 'syntactic bond' between a verb and a direct object, for example, seems to be greater than the bond between a verb and an indirect object or other complements. This, it seems to me, is an *ad hoc* solution. A solution such as that fails to capture, for instance, the fact that the bond between a verb and a non-referential object (e.g. *eat apples*) is stronger than the bond between a verb and a referential one (e.g. *eat an apple*), which is why in many languages the former becomes a lexical unit through noun incorporation but the latter does not. The bond between a verb and an non-specific, indefinite ('unidentifiable') noun then in turn is greater than the bond between a verb and a specific, definite ('identifiable') nominal (e.g. *eat the apple*).

The 'diacritic' use of structure and constituency has reached new heights with the claim that abstract units, so-called 'functional categories', are present in formal 'slots' in the representation of sentences, which take constituents of various types as their complements. This approach has been pushed to an extreme in the work of Pollock

(1989), for whom ‘tense’, for example, is one such category. This general idea has even been extended to include the functional category ‘negation’ in Basque (Laka 1994/1990).³⁷

1.4.4 Constructions

As I mentioned earlier, I believe that the grammatical unit of analysis is the construction, the grammatical equivalent of the lexical item, a unit consisting of more than one (bound) morpheme (morphology) or more than one word (syntax). Constructions, like morphemes and words, may be polysemous. Any structural unit which has formal or semantic characteristics which are not predictable and which do not automatically follow from some already established construction, no matter how small the difference, constitutes a new construction (or new sense of a construction) which must be learned (‘stored’) by speakers.

Constructions were considered to be basic units at one time in linguistics, only to be ‘outgrown’ by some recent approaches to linguistic analysis, then to be ‘rediscovered’ by other approaches (much like other once dismissed units in phonology, such as the syllable).³⁸ Constructions are not only basic units of analysis, but are ‘psychologically real’ as well, i.e. they correspond to units stored by speakers and actually used in production and comprehension.

This is, for instance, the approach to linguistic analysis proposed by Fillmore and other linguists of the Construction Grammar school (cf., e.g., Fillmore 1985, 1988, 1989, Fillmore, Kay and O’Connor 1988; Kay and Fillmore 1994; Lakoff 1987, 1992, 1993; Lakoff and Brugman 1986; Goldberg 1995).³⁹

This approach, or set of approaches, is based in part on the idea that it is impossible to make a clear-cut distinction on principled grounds between ‘idiomatic’ and ‘regular’ types of constructions, or between highly productive and minor constructions, all of which, it is felt, should be accounted for by the grammar (cf., e.g., Fillmore, Kay and O’Connor 1988). A grammatical construction is thus “any syntactic pattern which is assigned one or more conventional functions in a language, together with whatever is linguistically conventionalized about its contribution to the meaning or the use of structures containing it” (Fillmore 1988:36). More concretely,

Construction grammar is monostratal (no transformations), naive (no empty categories), uniformitarian (not requiring strict separation between lexicon and grammar, or between the idiomatic and the general), and “thick” (not founded on an autonomous syntax). Its practitioners give a great deal of attention to the details of complex idiomatic structures, in the belief that the effort to discover what is truly special or irregular about some body of linguistic facts is equivalent to the effort to discover what is general and regular in those same facts.

A construction is a conventional pattern of linguistic structure which determines a set of linguistic expressions (or constructs) in a language along with, in the typical case, “instructions” on their interpretations. The repertory of constructions available in a language can range from highly abstract principles of constituent order to extremely detailed constraints on the combinatory requirements of specific words. A sentence in a language is seen as having its form and the conventional part of its interpretation determined by the simultaneous satisfaction (through “unification”) of a well-formed assembly of constructions. (Fillmore 1994)

The goal of this framework is to describe the form and functions associated with particular constructions in great detail, as well as determining the ways in which they fit into the are system of constructions and how they are related to each other.

‘Construction grammarians’ differ as to how willing or ready they are to go beyond individual constructions and emphasize the system of commonalities and shared

properties of constructions and the degree that they attribute psychological reality to such linguistic generalizations. Thus, Goldberg, for instance, much like Lakoff, emphasizes the relatedness of constructions when she says that “[t]he collection of constructions is not assumed to consist of an unstructured set of independent entities, but instead it is taken to constitute a highly structured lattice of interrelated information” which is cognitively represented (Goldberg 1995:5).⁴⁰

Others in this framework are more cautious about making cognitive claims for their constructions or about introducing external motivations into the framework. For these linguists, the linguist’s grammar is seen as a “repertory of constructions, plus a set of principles which govern the nesting and superimposition of constructions into or upon one another” (Fillmore 1988:37). In other words the emphasis is on accurate description and capturing linguistic generalizations, rather than on representing cognitive structures or making guesses about the generalizations which speakers might make. Thus they argue that “the construction grammarian is required to develop an explicit system of representation which will *encode economically and without loss of generalization* all the constructions (or patterns) of the language, from the most idiomatic to the most general” (Kay and Fillmore 1994:2; my italics, J.A.).⁴¹ Whether speakers also encode the constructions ‘economically’ and ‘without loss of generalization’ is, for the most part, a matter left open.

In addition to the question of the relatedness of constructions among themselves, which has been a major concern of construction grammarians, there is the question of the relatedness between constructions and ‘extra-grammatical’ functional principles, such as the iconic principles which are more common in the pragmatic mode of communication,

as hypothesized above. Such principles, at least those of a ‘combinatorial’ or syntagmatic nature, such as the setting-assertion binomial, are amenable to a constructional interpretation. This would reduce the relationship between iconic (pragmatic) syntagmatic patterns and grammatical ones to a relationship between constructions, as schematized in Figure 1-1.

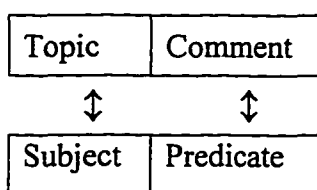


Figure 1-1

To the extent that speakers also make the ‘connection’ between the two it would be a cognitive relation as well.

Constructions constitute a rather heterogeneous set of elements, from the most basic and general statement sentence construction to idiomatic constructions such as the *Let Alone* construction (cf. Fillmore, Kay and O’Connor 1988), or the *Mad Magazine* (speech act) construction (cf. Lambrecht 1990). Different constructions are associated with different speech acts (e.g. imperatives), or variations of speech acts, such as those reflecting different types of information structure in statements (cf. Lambrecht 1994).

There is little doubt that, cognitively speaking, constructions are not merely listed in speakers’ minds. There is no doubt that there exist numerous relationship among constructions and among subsets of the constructions of a language, and thus that they form a ‘system’, albeit an imperfect and ‘leaky’ one. It is up to the analyst to hypothesize

what the relationships between constructions, and thus capture linguistically, and presumably cognitively, relevant generalizations.

On the other hand, the constructions of a language are also at least partially independent from each other, since they all must contain some element which is not predictable from other constructions in the system and thus must be learned independently. This has interesting repercussions for language change since it predicts that constructions may change by acquiring or changing some of their properties without other related constructions being necessarily affected (at least immediately).

New constructions, including variations on old constructions, often arise in a language through analogical and grammaticalization processes. Whether a construction simply changes when it is used in a novel context (and thus becomes more vague or ambiguous) or whether it develops a new and semi-independent sense (polysemy) is a matter of debate, the same debate that is found in lexical semantics studies (cf., e.g., Sweetser 1986, Geeraerts 1993, Tuggy 1993). Occasionally we know that new constructions are indeed created, and that they acquire a 'life of their own' and become (semi-)independent linguistic and cognitive units (constructional spin-offs). It is not clear, however, how new constructions are integrated cognitively and how the relationships among constructions should be represented.

The main reason for adopting a constructional approach to linguistic analysis is that it is a descriptively and cognitively sound strategy. It is a descriptively sound strategy first of all because it emphasizes describing individual constructions on their own terms, thus taking an agnostic attitude towards the organization of the whole system. And it is a descriptively sound strategy because it allows to go beyond the individual

construction and to make hypotheses about what larger organizational units and relations are like. Finally, the constructional framework is a cognitively sound approach because I believe that that is how humans acquire language, by learning new constructions and variations of constructions acquired earlier and incorporating them into a general system of constructions and into a general cognitive system.

1.4.5 Structures (constructions) vs. processes

The constructional approach to linguistic analysis shares with formal approaches a primarily static outlook on the language system, one which concentrates on grammatical structures and conventionalized units of language in general. While I believe that such an approach is valid to a point, it obviously has its limitations, though perhaps not insurmountable ones, especially when dealing with discourse phenomena, i.e. with sequences of speech acts expressed by speech act constructions.

In recent years other approaches to linguistic phenomena, which have emphasized the building of discourses, in particular spoken discourses, have reached conclusions which have repercussions for the study of the units of grammar, as conceived by different formal and functional grammatical theories. These approaches coincide in viewing speech production primarily as a *process* and not merely as a sequencing of relatively static tokens of grammatical constructions (cf., e.g., Enkvist 1982).⁴²

This process perspective on language stems, no doubt, from the shift from the study of individual speech acts out of context to the study of actual language in use, by means of recordings and their transcripts.⁴³ The reasons for the shift in perspective from

static structures to processes has been thrust upon investigators for reasons summarized by Enkvist (1982):

some of the reasons why linguists are increasingly preoccupied with processes [as opposed to structures] are clear enough. There are a number of tasks in which we are compelled to look at processes even at the expense of traditional types of structure. The study of impromptu dialogue ... is a good case in point. Unplanned speech can be highly deviant in terms of syntactic sentence structure, but can still be understood without difficulty in its proper context. (Enkvist 1982:630)

This doesn't deny the possibility of looking at more stable structures which are used in discourse, however:

To emphasize processes does not, however, necessarily mean that we have turned about and are marching in the direction opposite of that of the syntacticians. Indeed the terms 'structure' and 'process' are complementary rather than mutually exclusive. (Enkvist 1982:631)

On the other hand, it is clear that it is not possible for these two approaches, or perspectives, to ignore each other, and in particular for the static approach to ignore the importance of discourse processes. This is because it doesn't seem possible to analyze and understand certain 'constructions', or syntagmatic patterns, without recourse to discourse phenomena of a process-like nature. The example which I gave earlier about the ordering of elements in a discourse, such as two clauses, one of which performs some kind of function with respect to the other, and in general any type of 'construction' which spans more than one intonation unit, follows patterns which are more fluid and process-like than those envisioned by static structures.

Other pragmatic and discourse-related principles and functions, such as those which are internal to assertion constructions, may be easier to integrate into a static

system of constructions while ignoring, though not denying, the importance of wider, and perhaps less precise, processes and principles of discourse organization. This, for instance, is the approach followed by some investigators within the Construction Grammar and related frameworks (cf., e.g., Lambrecht 1994). It seems to me, however, that ideally our final goal should be also to incorporate into our model of language principles and patterns of discourse organization which go beyond the clause and the sentence.

1.4.6 Grammaticalization

Grammaticalization, a concept which I have alluded to above, is the mechanism by which grammatical categories and constructions are created, or 'institutionalized', out of lexical categories and syntagmatic semantic and (discourse related) pragmatic patterns in a process of "ritualization, reification, and emancipation", for the purpose of automating communication and reduce time-consuming choices in communication (cf., e.g., Haiman 1991a, 1993, 1994). Involved in grammaticalization patterns are also cognitive constraints and preferences of linguistic organization and production.

Grammaticalization is a concept that is still imperfectly understood, first postulated early in this century by Meillet (1912), but anticipated by earlier investigators, such as Bréal (1897).⁴⁴ There is at present much work going on to uncover the principles guiding this process (cf., e.g., Lehmann 1982, 1985a, 1985b; Heine and Reh 1984; Heine, Claudi and Hünemeyer 1991; Traugott and Heine, eds., 1991a&b; Hopper and Traugott 1993; Pagliuca, ed., 1994).

The basic idea behind the concept of grammaticalization is that, much as phonological processes are based on phonetic principles and constraints and preferences of phonological organization, so grammatical constructions and categories are grounded on semantic and pragmatic categories and principles, from which they may become emancipated to a certain extent. Gender categories are a classical example for how semantic or cognitive categories with fairly obvious correlates in the real world become to some extent arbitrary, but without completely losing their grounding on the semantic categories on which they are based. The other analogue of grammaticalization is the process of lexicalization and relexicalization for lexical (non-grammatical) items.

Grammatical constructions and categories can become increasingly emancipated from their original 'motivated' sources, such as pragmatic principles of organization in the case of speech act constructions, resulting in less motivated structures or structures embodying a variety of motivations acquired through their diachronic development. It is true that, as others have noted, grammar takes 'a life of its own', but this doesn't mean that its motivational 'inheritance' is wiped away (synchronically speaking) as soon as new uses and arbitrariness are introduced into an iconically motivated construction, such as, for example, the already mentioned new uses for the cleft construction in English. The introduction of less, or differently, motivated aspects into a construction does not mean that the construction necessarily loses touch with the originally motivating factors once emancipation begins as some have argued (cf. Newmeyer 1992, 1994). In actuality there is a continuum as to how much an original function underlies a particular form. The two extremes (total transparency, and total opacity) are probably quite rare. As Haiman has put it in a very insightful paper,

The more iconic a structure, the more likely it is to be up at the 'expressive end' of what I would like to consider the functional spectrum. The more arbitrary a structure, the more likely it is to be down on the phatic end of the same functional spectrum. (Haiman 1993:313)

In other words, the 'action' is actually in between full predictability and full unpredictability, a fact that may be hard to accept and especially hard to formalize. The greatest challenge to linguistic analysis is ascertaining the nature of the relation between grammar and ('external') function, which in its final instance means ascertaining the nature of the relationship in the minds of speakers. Emancipation of form from function is always partial, and dysfunction is a matter of degree.⁴⁵

However, and this is a crucial point, the degree of and toleration for emancipation of grammatical categories and constructions differs greatly from one case to another. Thus, for example, the subordinator *that* in English is probably fully emancipated (in the minds of English speakers) from its original meaning as a demonstrative pronoun.⁴⁶ This causes no dysfunction or contradictions in the organization of the grammar. English at one point 'needed' a subordinator for asserted complement (completive) clauses and used the demonstrative pronoun *that* to iconically point to them, eventually turning it into a 'complementizer'. (In modern English manner adverbial clauses are sometimes used to express such completive meanings, thus causing the adverbial pronoun *how* to act as a new complementizer, e.g. *He told me how his mother was going to come*). Many examples of grammaticalization, especially those involving grammatical morphemes, show that the original motivation for a grammatical fact may be completely, or near completely, obscured, and we can say that the grammatical element is quite arbitrary.

On the other hand, some grammatical categories and constructions would seem to be more motivated synchronically too, by which I mean cognitively motivated. The category subject, for instance, correlates quite strongly with the category 'clausal topic'. And the subject-predicate, or topic-comment, construction is quite transparently a reflex of the setting-assertion binomial, which as I argued above, is a prelinguistic notion or, at least, a highly iconic pattern. In other words, it can be argued that the relatively emancipated (grammaticalized), subject initial, basic (clause-level) constituent order is to some extent synchronically (cognitively) motivated, even in a language with quite rigid word order such as English. And the constituent orderings found in assertion constructions in other languages with 'freer' word order are probably motivated to an even greater extent.

Because of the existing strong tendency to interpret ordering relations from an iconic point of view, the emancipation or grammaticalization of word order from its underlying pragmatic discourse principles would seem to be more 'costly' or more likely to be 'resisted' than the emancipation of, say, a pronoun turned complementizer. Thus, for instance, it seems to me that the fact that in the vast majority of languages with grammatical word order an overt subject comes first in statement constructions (at least when it is a topic) is very likely to be 'interpreted' iconically by language learners at some level yet to be specified.

Most studies of grammaticalization have concentrated on cases which are at a fairly advanced stages of the grammaticalization process. It is much harder, however, to appreciate incipient examples of the phenomenon, especially since probably very few examples of incipient emancipation of form from function (grammaticalization) make it

very far in this process. The ones we hear about are the ‘successful’ cases, but as Hopper and Haiman have often emphasized, we must look at incipient cases to understand the process of grammaticalization. Speaking of Lehmann’s (1985a) proposed “concomitants of grammaticalization”, such as paradigmaticization, obligatorification, condensation, coalescence, and fixation, Hopper notes that

Such principles ... are, however, characteristic of grammaticization which has already attained a fairly advanced stage and is unambiguously recognizable as such. They work best, in fact, when the stage of morphologization has been reached ... The problem of identifying grammaticization when it is not already obvious is precisely that the form or construction in question has not yet reached a stage of being ‘obligatory’, ‘fixed’, and so on. (Hopper 1991:21).⁴⁷

The example given above of *how* being used as a complementizer in English is a case in point of incipient ‘emancipation’, or ‘emergent grammar’ in Hopper’s term (cf. Hopper 1987a, 1988). Of course, not many of these ‘innovations’, novel ways of interpreting form in emancipatory ways prosper to the point of becoming ‘grammatical’ (or grammaticalized) categories and constructions. Other students of the phenomenon, however, have de-emphasized such flirtations with grammaticalization and emphasized the actual stage, which may be quite brief, in which a category becomes ‘grammatical’ or in some way changes its status to a different grammatical category. Thus Givón, for instance, argues that

Several earlier studies of verb serialization (Hyman, 1971; Lord, 1973; Givón, [1975a]) suggested that grammaticalization was a *gradual* process. ... But there are reasons to suggesting [*sic*] the exact opposite perspective: That *cognitively*, grammaticalization is not a gradual process, but rather an *instantaneous* one. It involves the mental act of the mind *recognizing a similarity relation* and thereby exploiting it, putting an erstwhile lexical item into grammatical use in a novel context. The minute a lexical item is

used in a frame that *intends it as grammatical marker*, it is thereby grammaticalized. (Givón 1991a:122)

Clearly, both perspectives may be important and they are not necessarily mutually exclusive. One may choose to emphasize the ‘reification’ or ‘emancipation’ process at its origins, as does Hopper, or the more ‘abrupt’ act of reanalysis in which there is a significant reinterpretation, an item becomes entrenched in the ‘grammar’ from which there is no going back.⁴⁸

On the other hand, there may not (always) be such a cognitive moment in which a radical reinterpretation take place. It may be that change is at least sometimes (relatively) gradual, though there may be periods in which the change is accelerated, and that the obvious change is only evident to a linguist looking at the initial and final states, the two extremes of the continuum (cf. Bybee, Pagliuca and Perkins 1991). Also, let us not forget that reanalysis doesn’t necessarily do away with earlier interpretations of a category or construction, but rather merely adds a new way of interpreting form on top of the old one.

Grammatical constructions themselves may change their form and meaning imperceptibly, resulting, however, in major “reanalyses” of the units involved. Langacker (1977a) defines such reanalysis, which he refers to as “syntactic/semantic reformulation”, as “change in the structure of an expression or class of expressions that does not involve any immediate or intrinsic modification of its surface manifestation” (Langacker 1977a:59). As Heine, Claudi and Hünemeyer (1991) note, in some works, “‘reanalysis’ has been used as a near-synonym for grammaticalization” (Heine, Claudi and Hünemeyer 1991:167), but the two would seem to be quite different phenomena.

The most common examples of grammaticalization are those in which a lexical item takes on a grammatical function (e.g. the complementizer *that*). Meillet himself defined grammaticalization as “the attribution of a grammatical character to a previously autonomous word” (Meillet 1912:131; cited in Hopper 1991:17). But Meillet himself foresaw the application of the same principle to constructions as well. This is in line with much modern functionalist work of the last 30 years, involving cases from morphology (the rise of affixes), syntax (the rise of tighter grammatical constructions from paratactic combinations), and grammatical categories (e.g. the grammaticalization of topics into subjects or adpositions from adverbs and verbs).⁴⁹

In particular, Meillet foresaw the application of the concept of grammaticalization to word order, where more grammatical order in some languages contrasts with a more iconic and pragmatic order in others. Hopper and Traugott 1993 mention that

At the end of the article [Meillet] opens up the possibility that the domain of grammaticalization might be extended to the word order of sentences (pp. 147-8). In Latin, he notes, the role of word order was “expressive,” not grammatical. (By “expressive,” Meillet means something like “semantic” or “pragmatic.”) ... In modern French and English, which lack case morphemes, word order has primarily a grammatical value. The change has two of the hallmarks of grammaticalization: (i) it involves change from “expressive” to grammatical meaning; (ii) it creates new grammatical tools for the language, rather than merely modifying already existent ones. The grammatical fixing of word order, then, is a phenomenon “of the same order” as the grammaticalization of individual words: “The expressive value of word order which we see in Latin was replaced by a grammatical value. The phenomenon is of the same order as the ‘grammaticalization’ of this or that word; instead of a single word, used with others in a group and talking on the character of a ‘morpheme’ by the effect of usage, we have rather a way of grouping words” (p. 148). We see, then, that in this initial study of grammaticalization, Meillet already points to applications of the term that go far beyond the simple change from lexical to grammatical meaning of single words. (Hopper and Traugott 1993:22-23)

I believe that just like phonology is grounded on phonetics, speech act constructions, and the ordering of the elements therein, are grounded on discourse pragmatic principles. And much like phonology can only stray so far from that which is 'natural' from a phonetic viewpoint, so the grammaticalization of word order can only stray so far from what makes sense from a pragmatic-iconic point of view without 'triggering' mechanisms for coping with the opacity produced by changes, be they motivated changes or more blind ones, such as erosion of grammatical morphemes through phonological change. I will discuss this topic further in the following chapters.

1.5 VARIATION: GENRES, STYLE, MEDIA

1.5.1 Spoken language and contextualization

Crystal 1992, referring to an earlier study of conversational English (Crystal and Davy 1975), mentions that for someone trained to view language in terms of sentences and constructions, the spoken language can seem to be quite a different type of entity, one where sentences may be quite difficult to find.

In grammar, the most noticeable feature was the absence of clearly-definable units at sentence level. It was not at all easy to identify "complete sentences". Now this observation alone can cause considerable disquiet. The sentence is probably the most established and cherished fact of grammatical life, and its central role in linguistic theory, school grammar, and personal intuition is undeniable. We write in sentences, and we are supposed to speak in sentences and while occasionally we may lose our way in our expression, and leave an utterance unfinished, for the most part people accept the proposition that educated speakers speak in sentences most of the time. (Crystal 1992:124-125)

If there is any grammatical unit which seems to be organising this kind of speech, it is the clause, not the sentence. The concept of sentence does not seem to be so useful, with this material. (Crystal 1992:126)

That is, in conversation, full sentences, i.e. main clauses with preceding and following material can be quite rare. In narrative, on the other hand, I believe that sentences are easier to identify, although the parts of the sentence must be defined in discourse functional and intonational terms ('settings', 'main clause', 'elaborations'; cf. Chapter 3). But Crystal's point is well taken that spoken language is in many ways different from written language. Unfortunately, linguists typically analyze language in an idealized form which has much more in common with written sentences than language as it is actually spoken, in particular conversational language.

Discourse, and narrative discourse in particular, consists of a sequencing of units, primarily sequences of clausal units, which, as Givón has often emphasized, share a great degree of thematic and other types of coherence with each other for stretches at a time.⁵⁰ The sentences that linguists typically study, are typically decontextualized and thus miss many of the characteristics of actual sentences in discourse, especially those which cannot appear easily out of context.

In many ways it is quite hard to understand the workings of sentences, including their form and the elements which compose them, without taking into account the contexts in which they can be used. To some extent it is feasible for an investigator working through introspection to conjure up the right context for use of sentences. This, however, is not always the case, and it works best when the sentence in question can be used 'out of the blue'. Some clauses and sentences, however, *especially those displaying marked word orders* in some languages and which are used in marked pragmatic contexts, are harder to contextualize, and thus must be studied in their full, rich context. This has been noted by many investigators before, such as, e.g., by Bates *et al.* 1982.⁵¹

1.5.2 Genres, styles, and media

As we have already seen, there seem to be major differences between speech and writing. It is clear that written language is not merely speech committed to paper. This is perhaps not obvious to the untrained eye, since the myth persists that we typically speak in 'grammatical sentences', but modern discourse research leaves very little doubt that the units of speech are somewhat different from the sentences which linguists analyze. It is also well known that writing tends to have its own peculiar characteristics which differ from speech, differences which stem in great part from the differences in the medium and from the cognitive differences involved in their production and comprehension, as well as differences in the types of linguistic resources involved.

It has also been shown that within any one language there may exist significant differences between so-called 'varieties', either spoken or written varieties. Many different factors affect the variation found here, as discussed by sociolinguists and discourse analysts, such as the purpose of the discourse (genres) as well as individual and community preferences (styles). The best known and most common types of genres are conversation and narrative. Analysts have differed as to which one is more 'basic', but they are both important. We must keep in mind that narrative is often found embedded in conversation just like narrative may contain dialogue in it.⁵² Other minor genres include expository and procedural discourse (cf., e.g., Longacre 1982).

It has sometimes been argued that word order preferences may differ noticeably from one genre to another (e.g. Longacre 1982).⁵³ If there are any differences, and it seems that they may exist in languages where word order is highly responsive to pragmatic considerations, the correlation between word order preferences and genres

must certainly be an indirect one. That is, it certainly does not seem to be the case that a particular word order codes a particular genre or that a particular genre calls for a particular order, as is sometimes assumed, or at least suggested. What we are more likely to find is that constructions with different orders associated with them are found in different proportions in different language varieties.

Style is another category often used to classify varieties of language. Style refers to personal, or even culturally shared, ways in which language is actually produced. Although the possibilities for variation here are rather large, we may be able to identify common properties characterizing certain varieties of speech. Some of these may have repercussions for the study of word order differences. Here too the differences may often be related to different proportions in which different commonly available constructions are used in different contexts or by different speakers. Thus for example, I believe I have identified in my research speakers who make greater use than others of more 'vivid' styles of communication, in which statements are often presented with added 'intensity' or 'emphasis', a fact which, as we will see, has definite consequences for the word order in Basque. Such vivid (emphatic) statements often differ in their word order characteristics, as well as in their intonation and accentual patterns, from less marked statements.

1.5.3 Genre and word order in Basque

Basque linguists have noticed that there exist differences in word order according to these three different contexts in which language is found (medium, genre, and style). The Basque linguist Mitxelena, for example, has warned us that the traditional word order

rules that Altube (1920, 1929) proposed for Basque and which are based on the primacy of preverbal ‘focus position’, are very often broken by perfectly good speakers, and writers, especially in *narrative* (Michelena 1953:460, 1978; Mitxelena 1977; see Chapter 2).⁵⁴

Altuna argues that this is because narratives are action centered and the verb itself is the focal element in these cases (Altuna 1979:31-2). Osa agrees that the verb is central to narrative sentences, though he wouldn’t go as far as to say that ‘focal’ verbs are more common in narrative (Osa 1990:214). According to Osa the common situation in narrative is that either there is only a verb in the clause, or if the verb has company, all the elements are equally ‘rhematic’ (Osa 1990:217). I believe that Osa is basically correct that in narrative, where the majority of clauses are continuous, there are fewer clauses with highly focal (‘rhematically salient’) elements. In the Chapters 3-7 we will see the actual proportion of languages in which rhemes are verb-initial, which varies a great deal between speech and writing, and also among speakers.

Even within narrative, Basque grammarians warn us that there may be different styles, or ways of telling stories. Mitxelena, for instance, makes some very interesting observations about the different ways of translating Bible stories by two different authors: Agirre and Lardizabal, both speakers of the same dialect and in the same period (Gipuzkoa 1750-1850; Michelena 1978:224). He finds that Agirre, who has a greater propensity to place complements after the verb, producing a higher proportion for instance of (S)VO clauses, displays a way of telling stories which is “more popular, more like oral narrative, and less bookish” (Michelena 1978:223).⁵⁵

I believe that the difference between these two narrative 'styles' has to do in part with the greater use of emphasis, which, as we will see in later chapters, is often correlated with the placement of a salient element, or in its default, the verb, in rheme-initial position. (Only one element may precede the verb in asserted rhemes in Basque.) Such 'marked' statements seem to be more common in conversation than in narrative, and in certain narrative styles than others. And written narratives tend to be more of the latter type than of the former.⁵⁶

1.5.4 More on the differences between speech and writing

There is a rapidly growing bibliography on the differences between speech and writing.⁵⁷ The major differences between the two derive from two basic facts. The first one is that speech takes place 'on line', i.e. it must be produced and understood 'on the go', whereas with writing the writer is allowed the luxury of editing, and the reader the luxury of following the text at his or her own pace, as well as of backtracking (though this, of course, may be a sign of poor writing in the first place).

The second difference is that writing lacks a major aspect of language production, namely intonation: contours, prominences, tempo differences, pauses, and the like can be only very poorly reflected in writing. Thus some 'meaning', discourse pragmatic meaning, such as emphasis, may be lost, and must be filled in by the reader. This may have repercussions on word order in languages where word order reflects discourse pragmatic considerations. Some intonational information can indeed be committed to marks on paper and some may be recoverable from other cues, since intonational patterns

are often associated with particular constructions, other aspects of which can indeed be written and thus provide cues as to the 'suprasegmental' patterns.⁵⁸

As I already argued, until fairly recently, with the availability of recording equipment, linguistic analysis had a clear "written language bias" (Chafe 1994a:46). When one set out to analyze language, the major permanent forms of language available were written language and the decontextualized sentences that the analyst could conjure up on demand, which had quite a strong similarity to actual written language. It wasn't until actual oral language was recorded and then committed to paper with as much detail as possible that linguists began to notice that actual speech has a logic and structural patterning all of its own, which differs to some extent from the written language. Even after the availability of such possibilities for research there has continued to be in linguistics a strong bias against speech, which is sometimes seen as corrupt and full of 'performance errors'.⁵⁹

In one of the most comprehensive recent studies about the differences between speech and writing, Biber (1988) has shown that one finds many different parameters (such as the use of certain devices and constructions) along which speech and writing differ, such as the use of passives. He also finds, however, that different styles of either writing or speech themselves vary along those same parameters, and thus some styles of speech (e.g. lectures) are quite similar to stereotypical written speech, whereas some styles of writing (e.g. personal correspondence) may be more like stereotypical forms of speech.⁶⁰ In other words, he argues, one often cannot pinpoint absolute characteristics which differentiate speech and writing, the differences having to do with frequency of use.

Thus for instance, the English-like passive construction in Spanish is used almost exclusively in writing, but it is occasionally found in speech as well, and it can be quite rare in some styles of writing. Constructions such as ‘left dislocation’ are quite rare in writing but common in speech. Right dislocation is very rare in writing in English, though it is more common (and conventionalized) in some written genres in languages like Basque, as we will see.⁶¹

The nature of the spoken medium reveals other interesting facts about word order. As we will in the chapters 5-7, in spoken Basque we sometimes find a pause or hesitation after the verb and before a complement which seems to reveal that the clause has not been fully formulated at the time the verb was verbalized, which results in clausal orderings in which the verb comes before the complement (e.g. SV...O). (This situation is far more common than the opposite where there is a pause before the verb.) Such pauses, I will argue, may account for some of the cases of SVO-type order in speech, since they seem to be examples of a type of focus extraposition. Such a situation is unthinkable in writing, where writers can take their time to formulate a sentence and do not have to compose their discourses on the fly.

More generally, in standard written Basque, verbs in statements are much less likely to be rheme initial (i.e. (S)/VX) than in speech, as we will see in Chapter 6. In other words, in writing the verb typically follows a (single) rhematic complement, unless there is a ‘good reason’ not to (such as when the verb or the polarity are the focus). In speech the likelihood of the verb coming first in the rheme is much greater. Several factors may be involved here, besides the one just mentioned. Vividness of the spoken style is one such factor. It may also be that written Basque is more conservative in this

respect, or that ‘variable rules’ are being applied in speech to a greater extent than they are in writing.

Prescriptivism could play a role in the higher degree of verb-final (non-rheme-initial verbs) sentences found in writing. During past centuries written activity in Basque was minimal, limited in great part to translations of religious works. This seems to have favored the imitation of foreign models of writing, including sentences with a high degree of complexity and perhaps even a tendency towards calquing word order patterns in simple sentences.⁶² This led to a backlash in the present century, led by the grammarian Altube and others, towards supposedly more Basque-like patterns, such as shorter sentences with higher faithfulness to a supposedly more authentic verb-final (SOV-like) order pattern (cf. Chapter 2).

1.5.5 Concepts for speech analysis

The analysis of actual discourse requires attention to concepts which are not commonly found in structural approaches to linguistic analysis but which are crucial to the study of discourse. These are factors which become apparent as soon as one begins to transcribe speech, but which are not usually committed to writing, such as intonation units, contours, pauses, and the like.⁶³

The most striking fact about speech is perhaps that people do not speak in strings of sentences, but rather in “spurts of verbalization”, typically lasting about 2 seconds, and “characterized above all by having a single coherent intonation contour,” and typically being separated from other units by pauses. These units go by a variety of names in different linguistic traditions, but, following Chafe, I will call them “intonation units”

(e.g. Chafe 1994a).⁶⁴ For Chafe, an intonation unit is also a cognitive unit since it reflects what he calls a “single ‘focus of consciousness’”, i.e.

a brief perching of the speaker’s consciousness, attention, or short-term memory on a particular small chunk of information. The amount of information that can be included in a focus of consciousness appears to be limited by a wired-in constraint on how much a person can attend to at one time. When people try to focus on more information than short-term memory can handle, they are likely to get into trouble, both conceptually and syntactically. (Chafe 1984:437).

That is, these units have cognitive analogues and aren’t just speech that takes place between breathing interruptions (cf. Chafe 1994a:57).⁶⁵

Pauses are not necessary or sufficient conditions for the delimitation of intonation units. Sometimes the pauses are minimal, i.e. intonation units run into each other, and pauses may occur within units as well. Actually, it has been said that these linguistic units cannot be “identified unambiguously from phonetic properties,” since “the physical manifestations of psychologically relevant units are always going to be messy and inconsistent.” (Chafe 1994a:58). However, there are cues which in the vast majority of cases allow us to determine the boundaries of these units. These cues include: (a) pauses, (b) anacrusis: a pattern of acceleration at the beginning of the intonation unit and deceleration at the end (cf. Cruttenden 1986); (c) pitch contours, such as (i) ‘declination units’, i.e. when the high pitches in a sequence of high pitches are successively lower (cf. Schuetze-Coburn, Shapley, and Weber 1991); and (ii) ‘terminal pitch contours’, such as terminal (transcribed with a period), rising (associated with yes no questions, and various non-terminal contours, usually transcribed with a comma; and (d) voice change, such as creaky voice, laryngealization or “fry” (cf. Chafe 1994a:60).

Within intonation units we find some elements which are intonationally more prominent than others. Following Chafe I will use “the term *accent* for prominences that are realized as pitch deviations from a mid or neutral baseline, usually a higher pitch but occasionally a lower one” (Chafe 1994a:60) Chafe distinguishes between ‘primary accent’: “When one of these accented elements is also either loud or lengthened or both, I say that it has a *primary* accent and show it with an acute accent mark” (*ibid.*); and ‘secondary accent’: “A pitch deviation alone, without accompanying loudness or lengthening” (grave accent mark) (*ibid.*).

Chafe distinguishes between different types of intonation units according to their function, or lack thereof. Thus there are (a) truncated or fragmentary intonation units, which are unsuccessful ones; (b) substantive intonation units, i.e. “those that convey *substantive* ideas of events, states, or referents” (*ibid.*, p. 63); and (c) regulatory intonation units: “those that have *regulatory* functions in the sense of regulating interaction or information flow.” e.g. *well, mhm* (*ibid.*). Substantive intonation units very often coincide with the boundaries of clauses. Chafe estimates that “the mean proportion of single-clause substantive intonation units in the measured sample is about 60%” although “they are often forced to spread the clause across several intonation units” (*ibid.*, pp. 65-66).

A clause represents “the idea of an event or state” (*ibid.*, p. 66). A state “involves a situation or property that exists for a certain period without significant change, whereas an event typically involves a change during a perceptible interval of time” (*ibid.*, p. 66). Within these ideas there are other ideas, the “*participants* in the events or states”, which “are typically the ideas of people, objects, or abstractions, for which the term *referents* is

appropriate” (*ibid.*, p. 67). Whereas events and states “are activated transiently, many *referents* remain active for longer periods than any of the events or states in which they participate,” as we will see in Chapter 3 (*ibid.*, p. 68).

Chafe has argued that in writing too we can distinguish the analogues of intonation units, even though punctuation is not always a useful guide (cf. Chafe 1988, 1992a, 1992b, 1994a). Chafe argues that “many writers and readers experience a kind of auditory imagery that includes intonation contours, accents, and pauses” (Chafe 1992b:273). By reading a written text aloud, written language “take[s] on a prosody that is in some respects closer to that of spoken language” (*ibid.*). Thus for instance, “the intonation units of oral reading consistently show a mean length of about five or six words, regardless of how the passage was punctuated (Chafe 1988). This is the typical mean length of spoken intonation units. It evidently reflects a strong constraint on oral language production” (Chafe 1992b:273).

This is the process which I have followed to break up the texts of the narratives in the Written Basque Corpus into ‘intonation units’. I have not encountered any great difficulties in performing such a break up of sentences, and only in rare cases have I have made arbitrary choices in the coding. We will return to these issues in Chapter 3.

1.6 WORD ORDER TYPOLOGY

1.6.1 Introduction

As I mentioned earlier in this chapter, the study of typology goes beyond mere taxonomy, or classification, to attempting to finding ‘universals’ (cross-linguistic generalizations) and even explanations for these generalizations. These explanations or motivations may be either internal (systemic), or external, such as cognitive limitations or preferences, iconic principles of information distribution, and communicative needs, all equally valid for speakers of all languages.

Modern typology grew in reaction to the prevailing attitude that languages may vary in a myriad ways, without limit. That this is not the case was shown by Greenberg in the field of word order, when he found that some possibilities are much more common than others, thus opening the way for the study of cross-linguistic patterns (‘universals’).

Diachronically, ‘decisions’ made at different times may take a language in different directions, but at each junction there are preferred pathways to be followed (cf., e.g., Aitchison 1989). In other words, there are pressures and constraints (‘motivations’), which have different strengths or priorities, and which act on an existing linguistic state to produce different, but highly constrained, possible results (cf., e.g., Campbell *et al.* 1988).⁶⁶ In this section I will take a brief look at a variety of typological generalizations which may be of help in understanding word order universals, and in particular Basque word order and the possible ‘choices’ that it may be faced with at present.

On the other hand, we should keep in mind that formal categories across languages are only partially equivalent or comparable, due to the fact that they constitute

only partially similar solutions to similar problems (cf., e.g., Downing 1995).⁶⁷ Linguistic categories are not formally, universally, and abstractly defined at any level. Thus for example a subject in language X is never the exact same thing as a subject in language Y, although they may coincide enough for most purposes of comparison (cf., e.g., Croft *et al.* 1990).⁶⁸ If there is some, or even great, resemblance, it's because similar functional forces have shaped those categories. The situation is even more complex when we are dealing with 'word order types' rather than grammatical categories such as subject. In the past few decades we have witnessed a great interest in typing languages according to word order characteristics, even in instances when the choice of word order type was somewhat arbitrary, or at least not straightforward. We will be dealing with this and related issues in this section.⁶⁹

1.6.2 Basic word order

Greenberg (1966), assuming that it was possible to ascertain the 'normal', unmarked, or 'basic' constituent order for (statements in) a language, concluded, on the basis of a small sample of languages, that of the six possible orderings of subject, verb, and object, only 3 are actually found: VSO, SVO and SOV (or types I, II, and III in his terminology), and of these only two are found in large numbers, namely SOV and SVO (over 90% of all languages are said to be either SVO or SOV (in approximately equal proportions). In addition, Greenberg found that other properties of languages, such as whether they have prepositions or postpositions correlate with these 'basic' orderings.

In later years there have been claims that some languages do display the other three basic word orders: VOS (Malagasy, Keenan 1976b, 1978), OVS (Derbyshire 1977,

Derbyshire and Pullum 1981) and OSV (Derbyshire and Pullum 1986). Still, the fact remains that these three are quite rare and not 'basic' in the more clear-cut way that the other three can be basic, such as in being relatively rigid and having high frequency in texts.

Although intuitively appealing, the notion of basic word order is one that we must handle with a great deal of care.. To say that a language has a certain 'basic' word order is not always uncontroversial, for the degree to which a language 'respects' its 'basic order' varies widely. Thus, it is not the same thing to say that English, which has a rather rigid word order, is an SVO language, as to say that Spanish, which allows greater freedom to subjects to postpose and objects to prepose, is an SVO language.⁷⁰ Even more questionable would seem to be saying that Yagua is a VSO language (cf. Payne 1990) or that Basque SOV order (or SVO order for that matter). This is a very different situation from the determination of most other typological facts about a language. This is particularly true of many languages which have been classified as verb-initial (cf., e.g., Campbell, Bubenik, and Saxon 1988), but also of many other languages, such as Ancient Greek or Classical Nahuatl (cf. Lehmann 1990:165).

In a model of language which allows synchronic derivations (or 'transformations') it is easy to say that the 'basic order' is the 'underlying order' and that the deviations are derived by rules which alter that order.⁷¹ On the other hand, if one believes that different constructions and orderings are not synchronically derived from each other, positing a basic order is not such a simple proposition. As Basque linguist Mitxelena very perceptively put it: "We will not get anywhere if we postulate an abstract normal order and consider everything else to be a deviation" (in Villasante 1980:234, *fn2*).⁷²

In general, the significance of determining *the* basic word order of languages which do not have a grammatically fixed order is not clear. In these languages, the order of elements, and even the overt coding of arguments, is to a large extent pragmatically (rhetorically) determined. In such languages such a ‘basic’ order is likely to appear only infrequently (cf. Section 1.6.7 below on grammatical vs. pragmatic word order).

1.6.3 Frequency and basicness

Relative frequency is a test which is often used to determine which order is ‘most basic’. This is so even though in many languages (transitive) clauses with an overt subject and object may themselves be quite rare (cf., e.g., Du Bois 1987, Chafe 1987).⁷³ Also, the determination of what the basic word order of a language is by absolute majority may lead to quite different results depending on the dialect, the genre or the style chosen, as we will see is the case for Basque. The use of frequency to determine ‘basicness’ has been criticized, for instance, by Lehmann (1976), and most recently by Dryer (1995). According to Lehmann, “[i]ngenuous use of statistics, as in charts listing the percentage of VO versus OV patterns, or NG versus GN and the like in a given text or period, is accordingly very misleading” (Lehmann 1976:520).⁷⁴

According to Dryer, relative frequency, which in many cases means rather low absolute frequencies, is “epiphenomenal relative to the grammars of individual languages,” although he also argues that frequency does “play a much larger role than pragmatic markedness in explaining why languages are the way they are” (Dryer 1995:106). Frequency is not necessarily meaningful in a particular language since “[b]ecause the frequency of different linguistic constructions depends entirely on the

frequency with which the different factors conditioning the choice arise, frequency will actually vary from discourse type to discourse type, from text to text, and from subtext to subtext” (Dryer 1995:119). This is directly related to the fact that I discussed earlier, where relative word order frequencies may vary according to dialect, genre, or style.

Basque was said to be (underlyingly) an SOV language by de Rijk (1969), for instance, based in great part on the fact that in some written text counts SOV was almost twice as frequent as SVO. In written dialogues, however, he admits that “S.O.V. is only slightly more frequent than S.V.O” (de Rijk 1969:324). Thus, he argues, assuming “the most frequent order to be the unmarked one, and the unmarked order to be that order that preserves best the order in Deep Structure, we may take this predominance as an argument for an S.O.V. order in Deep Structure” (de Rijk 1969:325).

De Rijk does admit, however, the weakness of this argument when he immediately adds that “the existence of obligatory syntactic transformations makes any such argument extremely weak” and that “[i]t is quite conceivable that the preference for the S.O.V. order is merely a matter of Surface Structure and has nothing to do with Deep Structure at all” (*ibid.*). He thinks he finds confirming evidence for his decision that Basque is an “SOV language”, however, from other typological traits of Basque which Greenberg argued go hand in hand with SOV order, such as postpositions and Aux-Verb order. In addition he finds that internal evidence also suggests, or at least does not contradict, the SOV hypothesis.

1.6.4 Markedness and basicness

Frequency is not the only test for basicness, nor the most important one. It has also often been argued that ‘basic word order’ is the order found in certain pragmatic contexts, typically categorized as ‘unmarked’, such as those which display a “lack of specialized semantic or pragmatic connotations” (Croft 1995:95), or the least presuppositional information structure (cf., e.g., Siewierska 1988, Payne 1990, Croft 1995). For some this means that ‘out of the blue’ sentences with minimal or no context (i.e. some linguists’ favorite sentences), where all the elements are equally ‘new’ and no element is contrastive or emphatic, constitute pragmatically unmarked contexts, even though such sentences are quite rare in context, especially in narratives, as we shall see.

Hawkins has argued that “[b]asicness is established by showing that one order is free from various grammatical restrictions which affect the other” (Hawkins 1986:44).⁷⁵ Dryer too proposes that, for a particular language, “a particular word order can be described as pragmatically unmarked if it is the *default* word order”, the “elsewhere” case (Dryer 1995:105).⁷⁶ Connolly (1991), Lambrecht (1994) and many others propose similar definitions.⁷⁷

Still, we must remember that situations in which a sentence has an overt subject and object are already quite marked in many languages to begin with. It may also be that relative pragmatic markedness depends on a variety of factors, such as genre and style. Also, although some orders may be obviously marked in some languages, there may be situations in which the choice of an order in unmarked situations is not obvious, since more than one possibility is available. Thus we will see, for instance, that in the case of

Basque both SOV and SVO order are available in out-of-context statements for many speakers, particularly in narrative, even if in standard Basque SOV may be preferred.

Basque has often been said to have unmarked SOV order, because in isolated, single sentences OV is the order found inside the predicate ('rheme') both when the object is focused, i.e. is the new or contrastive element, *and* when the whole predicate is new, as in an answer to the question *What happened?* (cf., e.g., Ortiz de Urbina 1991:18).⁷⁸ The order OSV, on the other hand, for example, occurs only in a narrow focus sentence, and thus its range is more restricted, i.e. it is more marked (*ibid.*). (The information structure for both of these sentences is topic-comment, as I will argue later on.) Ortiz de Urbina doesn't mention the SVO possibility, however, which in some contexts, and for some speakers more than for others, would also be a possible answer to the question *What happened?*, and not necessarily a marked one, as we will see.

The notion of pragmatic unmarkedness is not as uncontroversial as we might have hoped either. Dryer (1995), for instance, proposes two possible types of pragmatic markedness, "pragmatic markedness as additional meaning" and "pragmatic markedness in terms of discourse factors". (He doesn't view frequency as a relevant to the issue of markedness, however.)⁷⁹

According to the first definition, a construction/word order would be marked if it contains an additional presupposition, and thus "a construction is pragmatically marked relative to another if the range of contexts in which it is appropriate is a proper subset of the set of contexts in which the unmarked construction is used" (Dryer 1995:112; cf. also Lambrecht 1994). Dryer adds that "[i]t is not clear, however, how often discourse-governed alternations can be characterized in this way, how often a difference can be

adequately characterized in terms of an added component of meaning in one alternant that is absent in the other” (Dryer 1995:112), that is, “intrinsic properties of the clause” (Dryer 1995:128), such as signaling ‘single focus contrast’ or ‘counter-expectation’ (cf. Payne 1990).

The second possible way of describing pragmatic unmarkedness, to which Dryer is partial and which has also been used by other investigators, is in terms of an ‘elsewhere alternation’ for a particular constructions. Thus, Dryer assumes that the discourse grammars of languages with pragmatic word order characterize the “circumstances in which the different orders are used”, and that sometimes, though not always, one of the cases can be characterized as the “elsewhere” case, that is, the unmarked case.⁸⁰ As Dryer concedes, however, this approach to pragmatic markedness “appears to be dependent on a particular analysis” (Dryer 1995:118). Dryer also admits that sometimes “it is very difficult to provide convincing evidence that a given construction is pragmatically unmarked” (Dryer 1995:127).

A third relevant definition of pragmatic markedness has been proposed by Givón (1990), one based on ‘cognitive complexity’, which “involves the amount of effort or the number of cognitive operations that take place in the mind of the hearer in procession” (Dryer 1995:129-30). Dryer mentions this definition, but questions its validity since “[w]hile we can talk of the cognitive complexity of a given utterance in a given context, it is less clear how we can talk about the cognitive complexity of a given construction”, such as the passive construction, since complexity here depends clearly on the context (Dryer 1995:130-131).

We can conclude from these observations that determining that an order is unmarked is not a straightforward and uncontroversial matter, but rather one loaded with assumptions about the grammar, discourse pragmatics, and even cognition. It seems to me that we cannot characterize in any meaningful way what an unmarked (pragmatic) situation is, independent of the speech act construction and independent of the context in which a construction is deployed. We may define a particular type of construction and a particular type of context as unmarked (by definition), but the choice is not likely to be uncontroversial and will not tell us anything about other speech acts and contexts.

It seems to me that the notion of pragmatic markedness, as used by many investigators, relies on the intuition that some utterances seem odd and 'marked' in out-of-context sentences. But defining such a context as the one to measure other contexts by seems to me to be a serious mistake. What is marked, or odd, in some contexts, is perfectly normal in other contexts, contexts which may even be more common, and, as Mitxelena argued in the continuation of the quote that I gave above, what we really should be concerned about is determining the uses and meanings associated with the different orders, as opposed to finding a magic formula to classify languages with.

1.6.5 Clause type and basic word order

Besides frequency and pragmatic unmarkedness, other characteristics have been used to determine basic order, such as the order found in a particular construction. Most approaches to 'basic word order' assume that the basic order is the one found in declarative affirmative sentences. It is in such clauses, however, that the greatest variation of word order is found in all languages, especially those in which word order is

responsive to pragmatic factors. In other clausal constructions, however, such as subordinate clauses, word order may be much more rigid and it may be easier to ascertain a 'basic' or default order for them.

Even in languages with a relatively fixed word order, what is 'basic' in main affirmative clauses may be different from what is basic in other clause types. Thus in languages, such as Dutch, in which main and subordinate clauses have different typical orders, one would have to choose whether it is the order in main or subordinate clauses that is basic.⁸¹

In Basque, most subordinate (and non-asserted) clauses tend to be rigidly verb final, especially non-finite ones. This phenomenon is associated with the fact that subordinators, which fulfill a 'delimiting' function, are typically suffixed to the verb, but it extends to clauses where the subordinator is preverbal, such as conditional clauses. The main exception to this rule involves asserted finite subordinate clauses, of which the main representative is asserted complement (completive) clauses when they follow the main verb, which, although they have a postverbal subordinator, display as much (pragmatically governed) word order freedom as main clauses. In addition other non-asserted subordinate clauses (constructions) which traditionally are strictly verb-final, such as conditionals and causatives, are beginning to display more freedom to break the verb-final mold, especially in speech and more so for some speakers than for others.⁸²

It has often been claimed that word order change begins in main clauses and only later does it spread to subordinate clauses. In Danchev's words, "the general view [is] that subordinate clauses are usually more conservative, cf., e.g., Schwartz (1975), Vennemann (1984), Hock (1986)" (Danchev 1991:113). This, of course, agrees with the

hypothesis that discourse pragmatic principles are involved in the ordering of elements in main (asserted) clauses, but not in (non-asserted) subordinate ones, which are 'free' to have a more consistent word order. Once a change has taken place in main clauses, so the theory goes, that change might spread to subordinate clauses, perhaps for (system-internal) reasons of consistency, though this is not a necessary requirement.

Stockwell and Minkova 1991, on the other hand, have argued that subordinate clauses, and not main clauses, "are central to the process of reanalysis" from verb-second (V2) to SV which took place in English (cf. Section 1.6.13 below). They doubt "that a direct reanalysis of main clause verb-second syntax into subject-verb syntax ever took place" (Stockwell and Minkova 1991:391). They believe, rather, that "the reanalysis in main clauses was mediated by analogy from the subordinate clauses where the reanalysis was completed much earlier; and there were other factors in the reanalysis." (*ibid.*).

In addition to the often mentioned dichotomy between main and subordinate clauses, speech-act constructions for statements and for other speech acts may have different ordering preferences. The Basque imperative construction, for instance, is one in which VO order, i.e. rheme-initial verbs, predominate, although OV order is sometimes also found. This, however, as we will see in Chapter 7, is motivated by the same type of pragmatic principles which account for word order in statements. The difference is that, unlike in statements, in imperatives, if all the elements are relatively 'new', it is the verb, and not a complement, that is seen as the most 'newsworthy' element and thus occupies 'focus position'. I will return to this issue in Chapter 7.

1.6.6 Syntactic or intonational complexity

Sometimes the basic order is said to be the one exhibiting the least amount of syntactic and morphological complexity, which is yet another measure of markedness, what Dryer calls “formal markedness” (cf., e.g., Dryer 1995:110; see footnote 79);. Thus, for instance, as we will see, in Basque clauses with a synthetic (non-periphrastic) verb, the verb in an SVO clause receives an emphatic polarity marker (the prefix *ba-*, somewhat similar to *do*-support in English), and thus we may want to say that these SVO sentences are ‘formally marked’, or more marked than SOV sentences.

However, in practice, it is not clear to me that these VO type clauses are pragmatically more marked than the corresponding OV ones. I don’t think that we can say that the *ba* particle makes the construction marked, since all this particle does is mark that the polarity (or the verbal idea itself), and not the object, is the focus of the assertion. All we can really say is that the two clauses have different foci or most salient elements, each of which is called for in a different type of context.

However, the formal markedness of verb/polarity focus sentences is limited to assertions with synthetic verbs only. Most verb tokens in Basque are not synthetic verbs but, rather, analytic (‘periphrastic’), and in these cases no overt marker is required when the verb is rheme-initial, since the erstwhile non-finite verb (participle), now fused with the finite verb as a single word, occupies the rheme-initial position.

An interesting case in this respect is presented by languages in which main clauses are formally more complex than subordinate clauses. Thus in Dutch, for example, main clauses with a periphrastic verb split the verb in two, placing the finite verb after the topic and the non-finite verb at the end. Also, in French, it seems that certain main clauses

exhibit greater complexity than subordinate clauses, since they are derived from complex constructions (cf. Lambrecht 1986a). Would we want to say that these main clauses are more 'marked' than subordinate clauses in these languages?

Relative pragmatic markedness can often be correlated to the relative complexity of the intonational contours and accentual patterns of a construction. The simplest intonational contour seems to be one which has a single major accent on the rheme, e.g. the word *book* in *He read a book*. In Basque and OV languages in general the tendency is for the rheme's main accent to go on the rheme-initial element. But focus constituents can also be 'extraposed', as we will see, and placed at the end of the rheme or in a separate intonation unit. The resulting sentence is clearly more complex, or 'marked', intonationally speaking. On the other hand, sentences with subject focus, for example, have also been said to be more 'marked', and this seems harder to justify. I will return to these issues in Chapter 3 and later chapters.

It is true that some constructions are more formally marked than others. It is also true that complex constructions may be used in some languages to express meanings which are expressed by simpler constructions in other languages (cf., e.g.,thetic constructions in spoken French, which were originally cleft constructions). I believe, however, that the attempt to use the notion of 'markedness' in word order typology has been somewhat misguided. The reason for this is that the driving force behind the use of this notion is the attempt to get at an 'essential' property of languages, namely their 'basic word order', which doesn't really exist, which doesn't have any independent systemic or psychological reality.

1.6.7 Grammatical vs. pragmatic word order

As I said above, the notion of basic word order is least controversially applicable to languages which make relatively low use of word order for coding the pragmatic characteristics (or roles) of the clausal elements, that is, it is least controversial in languages in which sentence positions are associated with grammatical roles, such as in English (cf. Thompson 1978, Givón 1984a:Chapter 6; Hawkins 1986). The use of word order to code grammatical relations is very often correlated with the lack of other means, such as nominal inflectional morphology, for signaling grammatical function, and it is found primarily in rigid SVO languages (see below).⁸³

The distinction between grammatical and pragmatic order is closely related to Li and Thompson's (1973) distinction between subject-prominent languages and topic-prominent languages (cf. Section 1.6.8 below). The main idea here is that the grammatical category subject is primarily a grammaticalized version of the pragmatic category topic. Topic-prominent languages tend to be verb-final, but not all verb-final languages are topic-prominent (cf. Chapter 3).

It seems to me that the distinction between grammatical and pragmatic word order, just like Li and Thompson's distinction, is not categorical dichotomy. It seems that only a minority of languages use word order to signal primarily grammatical relations and thus display a relatively 'rigid' word order of the kind we find in English. On the other hand, all languages use constituent order, to some degree or another, to indicate pragmatic functions (and roles), such as topicality (and topic status) and contrast or emphasis ('rhematic salience'; cf. Chapter 3). In addition, there is a fairly strong correlation between grammatical relations and pragmatic relations, as we shall see. The pragmatic

principles which regulate word order are quite general and universal. A language may dispense with using word order to signal pragmatic properties in great part because there are other 'channels' for signaling the same information, such as accent and intonation and 'minor' constructions (e.g. passive, clefts, etc.). Thus, the pragmatic use of word order is to some extent unnecessary.

The 'fixing' of word order to code grammatical relations can be seen in some ways as a form of systemic simplification, as we will see in Section 1.8 below.⁸⁴ Since the job of the grammar is to automate language generation and processing, it makes sense for a language to fix its order as much as possible and thus free speakers from having to face ordering choices for every sentence they utter. On the other hand, the choices involved here are rather small, given the fact that most of the main clauses in actual speech for instance have at most 3 elements, usually two, and sometimes only one (cf. Chapter 2). In addition, the high correlation between pragmatic and grammatical roles results in word order in languages with grammatical order not being significantly different (on the surface and statistically speaking) with that found in languages with more pragmatic order.⁸⁵ Finally, all languages make use of word order to indicate pragmatic functions to some extent, such as to code more 'marked' (salient) topics and foci, for example.

As I already mentioned, pragmatic (flexible) ordering typically applies only to asserted clauses, i.e. clauses in which discourse pragmatic notions, such as topicality, focality, topic and focus, are most relevant.⁸⁶ Thus it shouldn't surprise us to find that the constituent order in non-asserted clauses is typically less flexible (or more 'frozen') than the order in asserted ones, which is why some investigators have resorted to such order to

classify languages (cf. Givón 1984a:212). Thus by this logic Dutch and Basque might both be classified as verb-final languages.⁸⁷ Givón also notes, however, that the opposite phenomenon, namely cases in which main clauses are more conservative have also been reported (*ibid.*).

Payne (1992a) has argued that a distinction can be made within pragmatic word order languages, between those that have 'rigid' pragmatic word order and those that don't. She argues that "in some languages the relationship between surface order and cognitive-pragmatic statuses is just as strong and rule governed, as is the relationship in others between order and grammatical relations" (Payne 1992a:3). In other languages, according to Payne, the coding of pragmatic statuses is sometimes less than 100% predictable due to "the interaction of [other] factors" (besides "performance errors") (Payne 1992b:140).⁸⁸ On the other hand, we should not forget, as we will see in more detail in Chapter 3, that pragmatic 'statuses', unlike grammatical ones, are typically not all-or-none categories. Also, I will argue that we must go beyond the notion of 'pragmatic statuses' (such as identifiable vs. non-identifiable) and recognize that pragmatic *functions*, such as topic and focus, are the relevant factors in pragmatic word order.

Dryer (1989a) argues that languages which display pragmatic word order to one extent or another can indeed be classified as having a 'basic word order', at least from a typological perspective, if not in the sense of 'underlying' order. He argues that "the fact that word order in a language is governed by discourse factors does not constitute a reason not to classify it as OV or VO on the basis of frequency of order, even if the

frequency difference is demonstrably attributable to a side effect of the discourse rules” (Dryer 1989a:83).⁸⁹

Dryer argues that in pragmatic word order languages, languages which typically allow both OV and VO order and in which the alternation is discourse-governed, such as Papago and Ute (and Basque), if the ratio of the two different orders is 2:1, then the language can be shown to display typological characteristics, adpositions, genitive order, relative clause order) which are consistent with the statistically dominant order. The ratio of OV to VO order in languages which have been said to have pragmatic word order, such as Papago and Ute, is approximately of that order of magnitude. In Basque, the OV:VO ratio is of that magnitude in some cases, such as in written narrative, as we will see. In other cases, such as in some speakers’ spoken narratives, the ratio is the opposite.

I believe that the explanation for Dryer’s finding is that although pragmatic order languages display strong correlations between pragmatic function and position, languages differ as to what those positions are. The correlation between overt topics and initial position is very strong in most languages. The correlation between foci, which are always overt, and a single position, on the other hand, is not. For some languages the focus position is preverbal (rheme-initial) position. For others it is postverbal (rheme-second). Yet in other languages both positions are available for foci of different characteristics, with very salient foci (contrastive or emphatic) preferring preverbal position and less salient ones preferring postverbal position.

1.6.8 Subject-prominent vs. topic-prominent languages

As I mentioned in the previous section, Li and Thompson (1976a) have proposed a typological distinction between ‘subject-prominent’ and ‘topic-prominent’ languages. Many Indo-European languages, such as English, are said to be of the former type, whereas languages such as Chinese and Lahu are of the latter type. In topic-prominent languages “the basic constructions manifest a topic-comment relation rather than a subject-predicate relation” (Li and Thompson 1976a:459). It would also seem that topic-prominent languages display pragmatic word order. Subject-prominent languages, on the other hand, are those languages in which “the structure of sentences favors a description in which the grammatical relation **subject-predicate** plays a major role” (*ibid.*). These would be, then, languages with primarily grammatical word order.

All languages have ‘pragmatic’ topic-comment constructions (e.g. left-dislocation in English), and the subject-predicate construction may also be seen as reflecting a topic-comment pattern (with the right intonation). But in topic-prominent languages the topic-comment structure is said to be more ‘basic’, whereas in subject-prominent languages topic-comment constructions are more marked, or less basic, somehow, than the subject-predicate construction (cf. Keenan 1976a). Subject prominent languages are also more likely to have alternative constructions, such as the passive construction, the main purpose of which is to promote non-subjects to subject position when they fill the topic role.

As I said, I do not believe that this dichotomy is an absolute one. The basic claim Li and Thompson make is that “some languages can be more insightfully described by taking the concept of topic to be basic, while others can be more insightfully described by

taking the notion of subject as basic” (Li and Thompson 1976a:460). In addition, they argued, there may be languages which are both subject prominent and topic prominent, such as Japanese and Korean, and languages which are neither, such as Tagalog. Also, as I said above, topic comment constructions are also found in subject-prominent languages. In addition, as I said, the grammatical category subject in subject-prominent languages is to a high extent coextensive, statistically speaking, with the category topic in other languages.

Li and Thompson (1976a) mention a number of characteristics shared by topic-prominent languages, all of which I believe apply to Basque to a very great extent. First of all topic-prominent languages lack grammatical coding for subject; Basque codes ergative and absolutive, but not subject. Also, topics are coded, primarily positionally (clause-initially) and intonationally; Basque topics are indeed coded that way. Topic-prominent languages lack (promotional) passive constructions; Basque lacks a passive. Topic-prominent languages lack empty subjects, as does Basque. Topic-prominent languages have so-called ‘double subject constructions’, which, as we will see in the following chapters, I believe is true of Basque. Finally, topic-prominent languages tend to be verb-final, though perhaps not strictly so; this can be said to be true of Basque as well.

This correlation between topic-prominence (and supposedly pragmatic order) and verb-final (OV) order is a very interesting hypothesis.⁹⁰ On the other hand, there are VO languages, such as Spanish, which also exhibit a high degree of pragmatic world order, as we will see. One way in which verb-final languages with pragmatic word order tend to be more consistent than verb-medial ones is in the positioning of focus constituents. Thus, whereas verb-final languages consistently place focus constituents in preverbal position,

verb-medial languages often use both preverbal and postverbal positions. Notice also that verb-final languages tend to have case marking, something which has been seen as a prerequisite for having pragmatic word order. I turn now to this latter issue, and I will return to some of the other issues from this section again in Section 1.7 below.

1.6.9 Word order freedom and verbal inflections

The grammaticalization of word order is often associated with the lack (or loss) of other means to code the grammatical relations subject and object. Rich case marking morphology has been associated with pragmatic word order and its lack with rigid word order.⁹¹ More exactly, it seems that rich inflections are a requirement for having ‘free’, or pragmatic, word order, but that the opposite is not true, for not all languages with a rich system of inflections have free word order (cf., e.g., Lehmann 1992).⁹² In other words, the correlation is one sided, as shown in (1.3a), and doesn’t go the opposite way, as in (1.3b).

- (1.3) a. Free/pragmatic word order → Case inflections
b. * Case inflections → Free/pragmatic word order

The presence or absence of rich verbal ‘coreferencing’ or ‘agreement’ for at least some of the major arguments of the verb also seems to be a major correlate of the degree of word order ‘freedom’. Thus it seems that the loss of case inflections does not guarantee the establishment of a rigid word order unless it is also accompanied by a loss of verbal ‘coreferencing’ or “subject-verb concord” (cf. Givón 1976:175; Danchev 1991). In other words, the implication in (1.4) obtains.

(1.4) NOT (case inflections & coreferencing) → Fixed/grammatical word order

Also, as I mentioned earlier, there is a strong correlation between having case marking morphology and having verb-final order. Greenberg's (1966) universal 41 notes that verb final languages almost always have a case system, cf. (1.5). The inverse implication, on the other hand, is not true, since many VO languages also have case marking.

(1.5) OV → Case marking

I will return to some of these important issues below. In Section 1.7.5, for example, I will discuss the drift towards grammatical, as well as pragmatic word order. And in Section 1.8 I will discuss the relation between language contact and word order rigidification.⁹³

1.6.10 Subject-object asymmetries in flexibility

As I said, languages differ as to the extent and the ways in which word order is employed for pragmatic or grammatical purposes (cf., e.g., Lehmann 1992).⁹⁴ One thing we immediately notice is that 'rigidity' is typically limited to the core arguments of the clause, namely subjects and objects. Furthermore, the position of direct objects tends to be more fixed or rigid than that of the subject cross-linguistically.

Myhill (1992) has classified languages according to the flexibility allowed to the subject and the object. Some languages allow total freedom for both of these arguments. These truly flexible word order languages, such as a number of Native American languages studied by Givón, Payne, Mithun, and others, follow the principle of putting the most unpredictable, important, or 'newsworthy' element first (either 'focused'

elements or non-continuous unpredictable topics), regardless of grammatical role. Other languages, however, may have adopted a more rigid order for either the object, or for both subject and object, and forced to alter that ideal pragmatic ‘most—important-first’ principle somewhat to account for the default (grammatical) order.⁹⁵

Myhill argues that, whereas languages such as English and French are strongly SVO (and other languages are presumably strongly SOV), many languages are primarily VO, but in these languages the order of the S is more flexible, being either preverbal or postverbal. In some of these languages the order of the S is primarily preverbal (SV), such as in Spanish or Rumanian, whereas in other languages it is primarily postverbal (VS), as in Early Biblical Hebrew and Tzotzil.

I believe that the reason for any asymmetries between subjects and objects must be looked for in the topic-comment structure found in most asserted sentences and the strong correlation between grammatical and pragmatic roles. As I will argue in Chapter 3, I believe that in a VO language the object is often the (postverbal) focus of the assertion, but when it is not, such as when the verb or the polarity are the focus, this doesn’t require any reshuffling of object and verb, since the verb is already in a salient position (rheme-initial). When some other complement other than the object is the focus there is no reason for verb and object to shift places either (unless the object is the topic).

The other major correlation we find is that between subject and topic, which is why overt subjects are in most languages primarily clause initial. However, when an overt subject is not the topic, and overt subjects are often not topics in languages which allow covert (topic) subjects, as we will see, subject-verb inversion is a very well motivated operation. This is because when the subject is not the topic, it is either the

focus (typically) or an additional argument of the verb, both of which types of elements are typically placed after the verb in these languages. This is exactly what we find in Spanish, for example. In English, on the other hand, a language with rigid subject order, when the subject is the focus of the assertion it does not invert with the verb. This, however, is also pragmatically motivated, since what happens is that the focus subject is placed in the secondary focus position in this and other VO languages, namely preverbal position. (I will return to this hypothesis in Chapter 3.)

There is perhaps also a semantic reason why the object may have a greater tendency to be more fixed in its (postverbal) position. As I argued earlier, objects are much more likely than subjects to have a strong semantic bond with the verb. The less 'specific' the object is, the greater its bond is to the verb, a fact that may lead to its close conceptual connection to the verb and, in extreme cases, 'incorporation'. This phenomenon, of course, is one of the reasons why subjects are typically (default) topics.

1.6.11 Verb-initial languages

Elsewhere, Myhill (1992) argues that, unlike languages in which overt subjects are primarily preverbal, languages in which subjects are primarily (more than 60%) postverbal use this order in 'temporally sequenced clauses' (i.e. the same thing as Labov's 1972 'narrative clauses' and Hopper's 1979a&b 'foregrounded clauses'). SV order, on the other hand, is used in these languages to show discontinuity, and for focus subjects. That verb-initial languages show discontinuity by 'NP preposing' has also been argued by Fox (1985), Cooreman (1992), and others.⁹⁶

This type of language in which overt subjects are primarily postverbal seems to be a counterexample to the topic-comment pattern which I have argued is universal (found in all languages), and dominant (the most common) in language. On the other hand, notice that the numerical differences between primarily SV and primarily VS languages can be quite small (40/60 vs. 60/40). Before we can understand predominantly ‘verb-initial languages’, we must ask what the proportion of overt subjects is in these as opposed to other languages, since the higher the percentage of overt subjects is, the higher the percentage of preverbal subjects should be (assuming similar proportions of topic subjects, whether overt or covert). Thus we must also ask what proportions of the overt subjects are topics and which ones are not, and if they are not, what they are.

As we will see later on, a significant number of overt subjects in Basque are not topics, but foci or ‘antitopics’, a special type of inverted topic found in (some) ‘emphatic’ sentences. In languages which allow the inversion of subjects when they are foci or when the assertion is emphatic, such as Spanish for instance, minor factors may then be involved in producing a small increase in the percentage of such inverted subjects, such as the loss of markedness of emphatic assertion constructions with subject inversion (cf. Chapter 3).

I must emphasize, however, that the topic-comment information-structure pattern is not absent even from the most strict verb-initial languages, just like the rheme-initial position for foci isn’t absent either. Aissen (1992) reports, for instance, that Mayan languages, most of which are VOS and a few VSO, do use clause initial topics and preverbal foci.⁹⁷ Preverbal topics and foci in these languages seem to be used for marked or prototypical topics and foci, such as when they are contrastive, unpredictable, or

surprising. Clause-chain-initial topics, that is, those which are destined to be major topics of a paragraph would also seem to qualify in this category. Finally, let us not forget that actual sentences with overt subjects in these languages can be quite rare in discourse if they have verbal coreferencing for the subject, as they typically do.

We should not forget either that the topic-comment information structure is, albeit extremely widespread and available even in verb-initial languages, not equally common in all languages. In later Chapter 4 we will see that topic-comment structure is used sometimes in Basque in cases in which other languages would use a topicless structure. Sometimes topicalization is an optional alternative, as with many existential assertions, but sometimes it is obligatory, as with additive topics, equivalent to additive foci (of the *also* type) in other languages. We can perhaps attribute this fact to Basque being an extremely topic-prominent language.

1.6.12 Word order flexibility and pragmatic-role changing constructions

1.6.12.1 Introduction

As I mentioned earlier, languages which have grammatical order (to different degrees) are not in any way deprived by the lack of word order flexibility, for there are additional mechanisms which may make up for the lack of pragmatic ordering, such as accent and intonation, and complex (bi-clausal) constructions.

Phonological cues go a long way towards coding these functions, but the use of specialized constructions is very important as well, especially so-called voice-alternation constructions, such as the passive construction in English. Although such constructions

may have idiosyncratic features associated with their diachronic source and may have even acquired additional uses, it seems that the primary function of many of these constructions is to code pragmatic relations, which is why they are particularly prominent in rigid word order languages such as English.

Traditional ‘voice constructions’, or ‘voice alternations’, such as passive, middle, and inverse, make changes to the assignment of the grammatical relations for ‘agents’ and/or ‘patients’ that is expected in ‘direct’, ‘active’ transitive clauses. Languages differ a great deal as to the characteristics of these constructions depending on their origins and peculiar evolution.

Traditionally, voice constructions have been thought of as grammatical constructions which alter the grammatical relations of the arguments of the verb. Some, however, have included in this category ‘constructions’ in which the pragmatic relations are altered by means of word order change, since these word order change operations perform functions in language which are quite similar to the functions performed by traditional voice constructions in other languages, and in some languages may be the only means of performing those functions (cf. Givón 1994). This latter approach, of course, might force us to treat all word order changes as voice alternations as well.

As well will see below, traditional voice constructions involve changes in the ‘normal’ assignment of the topic relation. They may involve special forms of the verb, as in the English passive or active verb forms with different grammatical role assignment, as with the English ‘middle’. The choice of voice construction (including the ‘active voice’) is typically correlated with topicality related characteristics of the referents of the agent and patient. Sometimes the correlation is mandated by such pragmatic properties, leaving

no option to the speaker, such as the ergative voice in Australian ergative languages or the inverse voice in Algonquian languages.

Languages also have mechanisms for indicating ‘marked’ (unusual) assignment of the focus role, typically focus assignment to the subject role. Such mechanisms include complex constructions, such as clefts, and word order changes, such as in Basque and Spanish. In addition languages often have other means of reflecting other pragmatic properties of the ‘referents’, or other ‘ideas’, of the elements of clauses.

1.6.12.2 Non-default topic assignment

The primary function of what are traditionally known as voice alternations and other constructions (‘transformations’) is to alter the default alignment of grammatical roles and pragmatic roles associated with direct, active clauses. These voice alternations perform functions which in other languages are expressed by simple reordering of the arguments. Thus voice alternations are found typically, or most prolifically, in languages with less free, and more grammatical, word order.

Many of these constructions involve coding either that the subject is not the topic or that an object (typically the ‘direct object’) is the topic. Sometimes it is said that these construction reflect the topicality (a pragmatic-cognitive *status*) of the different arguments, but I believe that that is an indirect correlation. As I will argue in Chapter 3, I believe that while pragmatic *role* assignment is motivated by the pragmatic *statuses* of ‘ideas’ expressed by the arguments, this correlation is not perfect and some degree of speaker choice is typically involved. Since the purpose of these voice alternations is to code unusual (non-default) pragmatic role assignment, thus it is not surprising that we

find correlations between the use of these constructions and the pragmatic statuses of the referents involved, but, as I said, I believe these are indirect correlations.

Traditionally recognized voice constructions are not the only ones involved in changing subject/topic assignment. Many other constructions, such as *tough*-movement, also result in ‘unusual’ elements becoming the subject-topic of a sentence.⁹⁸ Such constructions are also typically found mostly in rigid word order languages such as English (cf. Hawkins 1986; Comrie and Matthews 1990).

The two possible topic-related changes that constructions (or word order changes) perform are: (1) indicating that the subject is not the topic (cf. ‘subject demotion’, as in ‘demotional passives’); and (2) indicating that the object is the topic (cf. ‘object promotion’, as in ‘promotional passives’). So-called passive and middle constructions in many languages perform either one or both of these changes. The English passive construction typically performs both operations at the same time. Other constructions, such as the Spanish ‘reflexive passive’ construction, typically only demote the subject to non-topic due to low topicality of its referent.

Besides their pragmatic role related characteristics, some passive constructions are also correlated with other characteristics of the clause, such as aspectual characteristics (anterior is twice as common in passive constructions as in active ones) and semantic ones (affectedness of the patient) (cf., e.g., Myhill 1992:4.1). As Myhill points out, “[s]ome of the factors are more important than others in the case of particular constructions”, depending on the origin of the passive (e.g. in English *be* passives and *get* passives differ somewhat in patient affectedness) (Myhill 1992:109).

These semantic and aspectual ‘preferences’ found in passives sentences are obviously correlated with, and not totally independent from, the choice of a patient for the topic role, for example. Passive constructions also seem to be the diachronic source of ergative constructions in ergative languages with an tense-aspect split (cf., e.g., Anderson 1977, Trask 1979).⁹⁹ In other words, in these languages passive constructions became specialized for coding the tenses or aspects which are commonly correlated with all passive constructions.¹⁰⁰

Another major so-called voice alternation is the ergative-antipassive one found in some ergative languages. This construction “intransitivizes the verb by removing the argument in the unmarked (absolutive) case (the P) from its role as a direct argument of the verb and having its case (the absolutive) taken by the A [ergative subject] in the same way that the passive ... intransitivizes the verb...” (Myhill 1992:115).¹⁰¹ These constructions do not seem to be related to changes in the topic status of the ‘subject’, but rather to transitivity related aspects of the clause, such as tense or aspect (cf. Hopper and Thompson 1980).

Basque, an ergative language, does not have a formal anti-passive construction, though the construction that codes continuous imperfective aspect for most predicates has many of the properties of antipassive constructions. This construction is not unlike the English continuous aspect construction (*be -ing*), with the difference that the Basque construction hasn’t been fully reanalyzed and the complement ‘clause’ is not fully integrated with the main one. Thus the subject of the whole complex clause is always coded as the (intransitive, or absolutive) subject of the auxiliary verb (*ari* “be busy, occupied”), regardless of the role it plays in the ‘lower’, or ‘complement’ clause.

There are other constructions which are indeed involved in topic changes, such as the ‘middle’ and the ‘inverse’ constructions found in some languages. There is considerable overlap between these constructions and so-called passive constructions in other languages since they seem to perform similar operations. The actual uses and constraints on such constructions are somewhat different from the more familiar passive construction, due to their different origins (cf., e.g., Myhill 1992:Chapter 4; Fox and Hopper, eds., 1994; Givón, ed., 1994).

Basque lacks an English-like promotional passive, despite claims to the contrary based on the existence of passive-like stative constructions.¹⁰² Basque does have a very common agent suppressing, or detransitivizing, construction which deletes the ergative role from the valence of the verb. In some cases this implies that there is no agent involved in a certain event (e.g. *the water boiled*), or that the agent is ‘indefinite’ (comparable to impersonal subject clauses with *they* or *one* in English). The absolutive argument in these detransitivized or decausativized clauses may or may not be the topic. In neither case does its formal absolutive role change from the role it would have had in a transitive construction.

1.6.12.3 Non-default focus assignment

In addition to default topic-assignment altering ‘strategies’, languages also have ways of indicating that the assignment of the focus role is different from what might have been expected. Again, such changes may involve simply phonological or word order changes in the construction, but languages also often have specialized constructions which are used for this purpose, particularly for indicating that the subject, the main topic

candidate, is the focus. These constructions are often complex, bi-clausal structures, which may originate as means of coding very marked, or salient foci, but whose use becomes generalized for coding less salient ones.

Cleft constructions, found in many languages, place the focus constituent in clause-initial position. So-called pseudo-cleft constructions achieve a similar effect by placing the focus in clause final, and sometimes dislocated, position (cf. chapters 3 and 6). Bi-clausal structures such as these may become grammaticalized to such degree that they are reanalyzed as simple sentences. This is what seems to be happening in modern French (cf. Lambrecht 1986a).

Just as in the case of voice constructions, these iconic, focalizing constructions also develop their own ‘personalities’ and idiosyncrasies through time in different languages, and may become polysemous to some extent. Thus, for instance, the English cleft construction has developed new (motivated) ‘senses’ or ‘uses’ which differ from the main use of the construction (cf. Prince 1978).¹⁰³

The relevant constructions in different languages also differ as to how exactly they are employed, i.e. what degree and type of salience calls for their use. Thus even Basque, which uses word order to code pragmatic relations, occasionally uses cleft and pseudo-cleft constructions. They might seem to be more ‘marked’, however, than the corresponding constructions in English.

1.6.12.4 Other pragmatic constructions

Languages, specially rigid order languages, may have other minor constructions which ‘rearrange’ the elements of the clause to reflect pragmatic characteristics of those

elements. These constructions accomplish goals which are similar to those achieved by simple word order reorderings in languages with more pragmatic word order. As it has often been noted, English, a language with very fixed word order, has a significant number of constructions the purpose of which is to place an element in subject position, more so than many other languages (cf., e.g., Hawkins 1986).¹⁰⁴

English, also has, for instance, a ‘dative-shift’ construction, which is cross-linguistically much less common than the passive (cf., e.g., Wierzbicka 1986, Thompson 1990).¹⁰⁵ This construction allows either an indirect object to be placed in postverbal focus position, or a direct object to be placed in clause-final focus position (cf. Chapter 3). Constructions such as this one are more grammaticalized, and thus more idiosyncratic, than the simple, pragmatically driven rearrangement of elements found in other languages, such as Spanish or Basque.

English also has some minor constructions which exhibit subject inversion, such as ‘emphatic’ assertion constructions (cf., e.g., Sadock 1971, McCawley 1973, Green 1980), and ‘presentative’ inversions (such as the *there*-construction), which can be understood in terms of the same pragmatic principles which cause subject inversion in languages with more flexible word order, such as Spanish or Basque (cf. Chapter 3). The English constructions, however, are typically more idiosyncratic, lexically bound, and exception-ridden, than the similar, ‘reordered clauses’ in other languages.

1.6.13 Verb-second and the verbal brace

There is yet another type of construction in which the elements of the clause are ‘rearranged’, which also seems to have to do with pragmatic properties, and which is

found in the history of Germanic and other Indo-European languages, as well as, in one form or another, in other languages. I am referring to the verb-second (V2) construction. This type of construction is particularly relevant to the study of Basque word order, since it seems to be similar in at least one respect to the Basque negation construction, which I will analyze in Chapter 7, namely the ‘verbal brace’.

In the verb-second construction found in Germanic languages the finite verb of asserted clauses is placed in ‘clause-second position’.¹⁰⁶ In other words, there can only be one (un-dislocated) element before the finite verb. If the initial element is not the subject, e.g. if it is a setting adverbial, then the subject must then ‘move’ after the finite verb (cf., e.g., Danchev 1991). This construction is found in Dutch, German, Frisian and the Scandinavian Germanic languages, as well as in Old English. Another characteristic of this construction is that if the verb is periphrastic, then the non-finite verb is ‘moved’ to the end of the clause. Thus the two verbs, the finite and the non-finite, form what is called a ‘verbal brace’, a phenomenon known as ‘embraciation’.¹⁰⁷

In some Germanic languages, including English, the verb-second construction seems to have been a major intermediate step in the change from being verb-final to being verb-medial (cf., e.g., Stockwell 1977, 1984; Bean 1983; Breivik 1991; Anderson 1993). This verb-second stage is still present, and stable, in German and Dutch finite clauses. In late Old English prose the ‘V2 rule’ was dominant, though seemingly not fully categorical (cf. Stockwell 1984).¹⁰⁸

As I said, English supposedly became a verb-medial (SVO) language through a diachronic process in which the first step was the V2 construction, followed by exbraciation (to the ‘right’), when the verb was periphrastic, of rhematic elements inside

the brace and thus the 're-union' of the two verbs, although now in a different relative order ($V_{\text{FIN-AUX}}V_{\text{MAIN}}$) from what it used to be ($V_{\text{MAIN}}V_{\text{FIN-AUX}}$). (As we will see below, V-Aux order is associated with OV, verb-final, languages and Aux-V order with VO languages.) In (1.6) below we can see the different stages of such a diachronic development (V_f = finite verb; V_m = non-finite verb; X = initial element; Y = braced or exbraciated element).

(1.6) From verb-final, to verb-second, to verb medial:

- a) Verb final stage: $XY(V)V_f$
- b) Verb second stage: $XV_fY(V)$
- c) Verb medial stage: $XV_f(V)Y$

The classical explanation for exbraciation has been that exbraciated elements are 'afterthoughts' or material which doesn't come to the mind of the speaker until after the brace has been closed. (As we will see, this has also been used as an explanation for other situations in which a verb-final language becomes verb-medial, cf., e.g., Hyman 1975.)

The afterthought account would certainly seem to be suspect, since it doesn't account for the fact that the brace may be quite stable in languages in which supposedly speakers also have things come to mind after the brace has been closed. Even if afterthoughts do play a role, it is obvious that other 'forces' must have been at play in languages like English where the brace has been lost.

In the case of the Basque negative construction we will see that exbraciation is common in speech (though not in writing). This, however, doesn't seem to be related to the expression of afterthoughts at all. Afterthoughts exist, but they are always marked intonationally, and do not seem to me to constitute the prime, or even a minor,

mechanism by which languages change ordering preferences. It should be noted that the term 'afterthought' has also been used to cover a wide assortment of phenomena beyond the most obvious sense (e.g. Kuno 1978).¹⁰⁹

Haiman (1974, 1991b) has associated the existence of the verb-second constraint with what Perlmutter (1971) called Type A languages, i.e. languages such as English which share a number of cross-linguistically speaking unusual characteristics, such as dummy subjects and obligatory subject pronouns.¹¹⁰ Type B languages on the other hand, which comprise the vast majority of the languages of the world, lack these characteristics. According to Haiman, the existence of type A languages is very closely tied with the V2 construction. It is even possible, he argues, that "only those languages which have or had the V/2 constraint ever become 'type A'" languages (Haiman 1991b:135-36).

In other words, he believes that "superficial pronoun subjects are motivated by word order requirements" associated with rigid grammatical constituent order (Haiman 1991b:136). It seems to be, however, that although there may be some connection between the two phenomena, the V2 construction doesn't necessarily have to be responsible for the development of Type A language properties.

1.6.14 Other types of variation

When we speak of the dichotomy between (S)OV vs. (S)VO orderings, this may be interpreted as a shorthand for orderings which have all their rhematic, or predicate-internal, complements either before or after the verb. However, it has been argued that in some languages different types of complements and modifiers may have different ordering preferences within any one language.

In OV languages such as Basque only one rhematic complement may be found in preverbal position, thus if a verb has more than one complement in the predicate, only one may go before the verb in what we'll see is the focus position, unless, of course, it is topicalized and thus outside the assertion proper. We should keep in mind, however, that in actual practice, rhemes rarely have more than one or at most two complements in their rhemes and that only one is typically 'new' and truly rhematic (cf. Chapter 3).

We will also see that the verb may also be in rheme-initial position in Basque, in which case all the complements are postverbal. We will see that this is more likely to happen when the complement is an absolutive argument than when it is a different type of complement, such as a non-verbal predicate or an adverbial phrase. This, however, is not due to the grammatical nature of the complement, but rather to their pragmatic properties. These pragmatic properties correlate with the tendency of an element to become the focus of the assertion.

In other languages, however, the ordering preferences for different complement types is said to depend on the grammatical category. Thus, Givón has argued, for instance, that "[o]ften in ex-SOV languages, direct objects display the more conservative OV word-order" (Givón 1984a:214). This has been said to be the case in several Mende and Voltaic (Niger-Congo) languages. Supposedly in such cases the correlation is grammatical and not determined primarily by pragmatic considerations.

The 'structural' characteristics of a complement, such as length and/or complexity, are also known to be a factor in the ordering preferences of a complement with respect to the verb. Thus pronominal and sentential complements often differ in their ordering preferences from other nominal complements. So, for instance, as Spanish, an

SVO language, rigidified its object order, object pronouns settled, or became ‘frozen’, before the verb, whereas all other objects (nominal and sentential) settled after the verb (cf., e.g., Givón 1984a).

On the other hand, in quite rigid SOV languages, where most complements are found in preverbal position, we find that often complex complements, such as sentential ones, are ‘extraposed’ to postverbal position (*ibid.*). In Basque too, as we will see, sentential complements are much more likely to be postverbal than preverbal, almost categorically so in our Spoken Basque Corpus. The ‘mechanism’ by which such sentential complements are placed in postverbal position seems, what I will call focus extraposition, may be generalized to other less complex complements. This, I will argue, is the reason for the large number of postverbal complements in spoken Basque. (I should note, however, that object complexity is not the only possible motivation for extraposing a focus complement, and that not all postverbal complements are foci.)¹¹

The difference in behavior between pronouns, regular nominals and sentential complements has often been attributed to a ‘principle’ according to which ‘light’ elements prefer to come early in the clause, whereas ‘heavy’ ones prefer to come late. Although this may indeed be a factor, having to do with processing considerations (cf. below), we should not underestimate the pragmatic differences which commonly exist between the different types of complements. Thus, whereas sentential complements are almost always ‘new’ (and may even constitute independent assertions), pronouns are always highly topical (though not necessarily topics), and have a well known tendency to gravitate towards the verb, for reasons unknown, and even to become cliticized to them, as is the case with object pronouns in modern Spanish.

Other elements also gravitate towards the verb and sometimes become ‘frozen’ there. Givón reports that in Ute, “[n]on-referential (generic) direct objects are quite often incorporated into the verb, and thus become morphologically frozen” (Givón 1984a:214), much like the Spanish pronouns just mentioned, thus displaying less freedom of movement. This might seem to be a grammatical constraint on movement. On the other hand, we will see that Basque non-referential objects are also much less likely to ‘move’ to postverbal position, even though they are not strictly incorporated into the verb. This fact has been noted for other OV languages as well (cf. Comrie 1984 for Armenian). I will argue in Chapter 6 that this constraint responds to pragmatic factors and not to grammatical ones.

1.6.15 Basic word order correlations

Greenberg (1966/1963/1961) found, or rediscovered, that the different ‘basic orders’ correlate, or ‘harmonize’, with other characteristics of language, such as whether they have postpositions or prepositions, the order of genitive phrases and adjectives with respect to the noun, and so on.¹¹² Lehmann (1973, 1978) and Vennemann (1973, 1974a, 1974b, 1976) have suggested that the correlations, or ‘harmonic patterns’, are one-sided, that is, that object-verb order is primary and that all other correlated ordering relations depend on it. In other words they argue that there is an implicational ‘universal’ from word order to the other, ‘harmonic’ properties.

The correlations found by Greenberg, and refined by others, are not anywhere near exceptionless, however. As many investigators have shown, the majority of languages are not ‘consistent’, and there is no clear way in which those languages can be said to be

'in transition' from one 'perfect state' to another. Also, some of Greenberg's correlations have been found to be spurious, once better samples have been devised. Many have labored to elaborate and refine Greenberg's correlations. The sampling problem, however, has proven to be a hard point of contention among typologists.

Hawkins (1983), in an attempt to eliminate exceptions from Greenberg's correlations, has made the implicational formulae more complex, by including multiple terms. It is not always clear, however, that these more precise formulae which remove exceptions in a particular sample are true (and explanatory) generalizations about language. Hawkins (1983) has also sought to provide 'explanations' for the preferred patterns in terms of head-modifier relations, supplemented by two other principles: heaviness and mobility. More recently he has attempted to find an explanation based on a production model (cf. below).

Dryer 1992a represents one of the most thorough attempts to determine the extent to which the correlations which have been proposed are true correlations. Dryer, like both Lehmann and Vennemann before him, but unlike other investigators, "assume[s] that the word order correlations can be discussed in terms of a contrast between VO languages and OV languages", and that "SVO languages pattern like verb-initial languages" (Dryer 1992a:87; cf. Dryer 1991).

Based on a sample obtained by a novel, and controversial, sampling method (cf. Dryer 1989b),¹¹³ Dryer argues that only about one third of all the proposed correlations are true correlations (though far from being perfect ones), a little under a third are non-correlations, and another third are somewhere in between, i.e. controversial.¹¹⁴ In (1.7) below we can see the 'true' correlation pairs, according to Dryer.

- (1.7) OV: Postp., Gen-N, Std:Adj, PP-V, Adv.-V, Pred-Cop., Verb-*Want*
 VO: Prep., N-Gen, Adj-Std, V-PP, V-Adv., Cop.-Pred., *Want-Verb*

In Table 1-5 we can see the complete set of ‘true’ correlations, non-correlations, and controversial correlations, as well as their relative frequencies.

CORRELATION PAIRS			NON-CORRELATIONS			CONTROVERSIAL PAIRS		
	OV	VO		OV	VO		OV	VO
1. Adposition:NP			1. Adjective:Noun			1. Noun:Relativizer		
# of genera	114	82	# of genera	129	95	# of genera	63	61
%N-Postp.	.94	.15	%Adj-N	.39	.47	%Rel-N	.42	.01
2. Noun:Genitive			2. Demonstrative:Noun			2. T/A-Aux:VP		
# of genera	124	93	# of genera	111	96	# of genera	39	32
%Gen-N	.89	.45	%Dem-N	.70	.68	%V-Aux	.92	.13
3. Adj.:Std. of Comp.			3. Intensifier:Adjective			3. Neg. Aux:VP		
# of genera	36	32	# of genera	64	52	# of genera	11	14
%Adj-Std	.77	.02	%Inst-Adj	.60	.51	%V-NegV	.73	.07
4. Verb-PPhrase			4. Neg. part.:Verb			4. Complementizer:S		
# of genera	72	60	# of genera	42	50	There is a correlation, but there isn't enough data.		
%PP-V	.88	.02	%V-Neg	.24	.10	5. Question part.:S		
5. Verb: Manner adv.			5. T/A part.:Verb			5. Question part.:S		
# of genera	70	58	# of genera	24	29	# of genera	45	47
%Adv-V	.89	.24	V-T/A	.16	.15	%S-Q	.75	.30
6. Copula:Predicate						6. Adv. subordinator:S		
# of genera	72	55				# of genera	55	60
%Pred-Cop.	.85	.36				S-Subord.	.70	.02
7. Want:Subord. verb						7. Article:N'		
# of genera	39	46				# of genera	32	52
%V-want	.72	.17				%N-Art	.63	.21
						8. Plural word:N'		
						# of genera	16	23
						%N-Plural	1.00	.35
						9. Verb:Subject		
						# of genera	135	107
						%S-V	.98	.64

Table 1-5: Correlations, non-correlations, and controversial correlations between OV/VO order and other characteristics. Adapted from Dryer 1992a.

1.6.16 Conclusions

In this section I have summarized some facts about word order typology and have shown some of the problems hidden behind a simplistic classification of languages

according to ‘basic word order’ types. I have discussed, for instance, issues such as word order freedom and the use of minor constructions for expressing pragmatic meanings associated with word order rearrangements in languages with freer word order.

In the next section I will look in greater detail at the issue of explanation in typology, and in particular the use of typology and typological generalizations, that is, universals, for explaining facts about languages, and in particular word order facts. I will attempt to show further that explanation of word order patterns must rely primarily on discourse pragmatic notions, and in particular the pragmatic roles *topic* and *focus*, which will be further motivated in Chapter 3. I will also discuss other likely explanations for word order patterns and preferences, such as iconic motivations and processing constraints. I will also explore the issue of word order change.

1.7 TYPOLOGY AND EXPLANATION

1.7.1 Introduction

The differences among language ‘types’, as well as the similarities and the correlations that we have seen, are not accidental or arbitrary, but, as has often been pointed out, are quite amenable to explanation. This doesn’t mean that we can synchronically explain every pattern and every construction that we find in every language, but, rather, that we can understand the ‘forces’, or the ‘motivations’, which mold these patterns.

The explanatory factors are primarily ‘external’ ones, including cognitive and processing preferences or constraints and iconic discourse pragmatic patterns of

information structure. Crucial to understanding these patterns are also system ‘internal’ factors related to the process of grammaticalization, i.e. the diachronic process by which certain patterns become part of the grammatical inventory of the language and transform themselves in the process, fixing and molding the grammatical system.

The primary principles involved in ordering patterns seem to be those related to discourse pragmatic (information packaging) principles of the sort described in Section 1.3 (which will be further motivated in Chapter 3). That is why the relative order of clausal constituents seems to be the central starting point for the correlations or harmonies. Other external principles are also important, as we will see, as are some more or less arbitrary choices, such as the choice of primary position for the focus constituent.

1.7.2 Pragmatic principles and the order of clausal constituents

The pragmatic principles which I believe motivate, directly or indirectly, the different basic constituent order possibilities will be expanded on and motivated in the following chapters, but I believe that we are ready now for a preview of those motivations. The first principle involved is one we have already seen, namely the iconic setting-assertion pattern.

1.7.2.1 Explaining the subject-first preference

One readily available conclusion from the relative frequencies of the possible word order types is that overwhelmingly subjects tend to be clause-initial, the result of a strong pressure, or motivation, no doubt. I believe that the main motivation for this fact has to do the main information structure type found in language (topic-comment

structure), and with the fact that the pragmatic role topic is largely (though not completely) coextensive, statistically speaking, with the 'grammatical' category 'subject'.¹¹⁵ The strong tendency for subjects to be topics and the strong tendency for topics to be clause-initial would account for the preference for subjects to be initial.

We must keep in mind however, that not all overt subjects are topics, especially in those languages in which topic subjects with active referents are not overtly expressed as nominals. Subjects are not topics when they are rhematic arguments (whether foci or not). Subjects may also be inverted topics or antitopics, as we will see.

Of course, subjects which are foci may also appear in clause-initial position since clause-initial position is the most salient position in which foci are found. This is what we find in OV languages, but also in erstwhile OV, rigid SVO languages, such as English. Of course, when the subject is the focus of the assertion some other covert argument may be filling the topic role and thus precede the focus subject.

1.7.2.2 Explaining the OV ~ VO alternation

The major remaining issue in need of a general explanation is the OV ~ VO alternation, the major typological parameter found in language, according to some typologists. As I said, Lehmann, Vennemann, and others have argued that this dichotomy actually establishes the two major word order types cross-linguistically: OV and VO, existing in approximately equal proportions. For them OV languages are verb-final languages VO languages are verb-initial languages *and* verb-medial (SVO) languages as well. Some have argued against lumping SVO languages with verb-initial languages, e.g. Comrie (1981), Hawkins (1980, 1983) and Mallinson and Blake (1981), arguing that

SVO languages do not pattern like verb initial languages in their corresponding ‘harmonics’.¹¹⁶ Dryer (1991), however, has argued forcefully that Lehmann and Vennemann were right and that “with certain well-defined exceptions, the word order properties of SVO languages differ little from those of VSO and VOS languages” and that “we do find a basic split between VO and OV languages” (Dryer 1991:443).¹¹⁷

Be that as it may, I believe that the OV~VO dichotomy is related to the preferred location for the informatively most salient element of the ‘assertion’, namely the focus. As I argued above, main clauses rarely have more than one complement, and when they do, all but one tend to be very accessible (later I will call them ‘extra-rhematic’). The single complement, or the less-accessible, most informative or ‘newsworthy’ element is the focus of the assertion.

Assuming the reality of the focus pragmatic role is real, the question now is whether this element has a preferred position (like topics do), and, if so, what that preferred position is. The facts show that in fact there are a number of preferred positions cross-linguistically, with some languages making use of only one such position, other languages making use of two of them, and yet others making use of all three of them.

Preverbal position is typically used for the most salient or urgent foci in most languages, including VO languages. This is also the main position for all foci in OV languages. The main position for non-urgent foci in VO languages, however, is postverbal position. Finally, most languages, except perhaps some rigidly verb-final ones, allow foci to be ‘extraposed’ to the end of the clause, or even to a new intonation unit under some circumstances.

Here lies, I believe, the true nature of the difference between OV and VO languages. That is, OV languages are those in which the major or single focus position is preverbal (and rheme-initial); and VO languages are those in which the major (though perhaps not single) focus position is postverbal (rheme-second). I will return to this issue in Chapter 3.

1.7.3 Iconicity related explanations

1.7.3.1 Iconicity, order, and constituency

Iconicity is a very powerful motivation in determining the flow of discourse, including information structure at the sentence and the discourse level. Linguistic iconicity is the notion that the formal characteristics of the language, in this case constituent order, reflect iconically or isomorphically deeper characteristics of the meaning/function of those constituents. I have already discussed the setting-assertion template for communication, as well as its 'inverse' version, assertion-setting template, and its iconic analogues found in the speech-act constructions of many languages. Iconicity also seems to be a very strong motivating factor in language change.

The connection between iconicity and word order goes back at least to Behaghel (1932),¹¹⁸ who stated it as his 'first law' of word order: "That what belongs together mentally [semantically] is placed close together [syntactically]" (Behaghel 1932:4, cited and translated by Vennemann 1974b:339).¹¹⁹ Haiman, who along with others has revived the notion of iconicity in recent years, has also argued for the connection between order and iconicity. Haiman argues that "[s]ince it is impossible to say everything at once, words

must appear in a certain order,” and it seems that “[s]everal types of iconic motivation exploit the resulting linearity of the linguistic sign” (Haiman 1980:528).

Citing the work of Greenberg (1966), Haiman mentions examples of iconicity in constituent order such as: “[n]arrative sequence (e.g. *Vēnī, vīdī, vīcī*), ... the all-but-universal precedence of subjects over objects, ... the fact that conditional protases almost always precede apodoses ... and the way in which initial position in phrasal conjunctions often reflects a priority of rank” (*ibid.*). In other words, Haiman and others have argued for the same iconic setting-assertion pattern I argued for above, as well as other ‘natural’ patterns of discourse organization.

1.7.3.2 Degree of closeness, relevance, and scope

Not only sequential order, but other ordering relationships and even the degree of closeness between two elements can often be understood by means of iconic principles. Thus the attraction and even phonological fusion often found between the verb of a complement clause and the ‘higher’ verb, mentioned in Section 1.4.3 above, might be attributable to the close semantic affinity existing between the two verbs, especially if the higher verb is semantically ‘light’, such as when it is a copula or other auxiliary verb.

The semantic (conceptual) closeness between a verb and a non-referential complement may also explain the fact that they are typically next to each other and sometimes even fused into a single phonological word (noun incorporation). In general it is probably safe to say that objects are much more likely to be low in topicality than subjects, which would account for their lack of referential independence and thus the greater need to be interpreted along with the verb (cf. Chapter 3). This principle could

also be partly responsible for the fact that objects are closer to the verb than subjects in SOV-type languages, for example, and that the word order of objects is typically less flexible than that of subjects (cf., e.g., Haiman 1983; Sheintuch 1981; Keenan 1984, Siewierska 1988).¹²⁰ I believe, however, that the major reason for this fact is that the object is typically part of the rheme (an information unit) and that the subject is typically the topic. The correlation, however, is hardly accidental.

More generally, it seems to me that iconic principles, and not any abstract syntactic ones, are ‘responsible’ for the fact that modifiers, complements and the like form structural units (constituents or constructions) with the elements they modify and complement (barring exceptional, and somehow sanctioned, ‘long distance dependencies’).

In general, we can expect the relative order of modifiers, complements, operators, quantifiers and so on with respect to a phrasal ‘head’ to reflect to some extent the degree of semantic or functional ‘proximity’ to that head. Haiman proposed a ‘distance principle’, based on Behaghel’s First Law, according to which “[t]he linguistic distance between expressions corresponds to the conceptual distance between them” (Haiman 1983:782). Bybee (1985a&b) also proposed a ‘principle of relevance’, according to which “the degree of fusion of grammatical material with lexical depends upon the semantic relevance of the grammatical morpheme to the lexical: the extent to which the grammatical meaning directly affects or modifies the semantic content of the lexical morpheme” (Bybee 1988a:359).¹²¹

When the constituents in question are **operators**, that is, modifiers of a complex unit, these two principles, and in particular Bybee’s principle of relevance, are very close

related to the perhaps more familiar principle of **scope**. Thus we expect modal, aspectual, tense, or negation particles, for instance, to be ordered iconically according to their relative scope, and ideally at the edges of the unit over which they have scope. This is what we often find, and especially in Creole languages, where such markers arise to a large extent all at once, thus showing that whenever possible, in the ideal situation, the relevance or scope ordering principle prevails.¹²²

Bybee stresses that these are ‘motivating’ principles, and not explanatory ones in the nomological sense. Thus she argues that “despite making correct predictions, [these principles] do not qualify as explanations,” although “[t]hey certainly tell us what to expect” (Bybee 1988a:359). In order for them to be truly explanatory “they would need to explain how these structures become conventionalized,” in other words, the mechanisms by which these iconic principles are converted or incorporated into constructions (*ibid.*).

In other words, Bybee argues that iconic principles motivate or encourage the development of certain orders, structures, or constructions diachronically (cf. ‘smoking causes cancer’), but do not actually ‘explain’ how the actual synchronic constructions come about (cf. ‘how person X gets cancer’). Iconic principles do not guarantee or deterministically predict a final outcome (given a particular initial state), rather the probabilistically motivate it.

The other major question here, of course, is the one of psychological reality of these motivations at the cognitive synchronic level. Although iconic motivations do not predict the actual constructions and orderings found in language, it seems to me that we cannot deny that it is very likely that speakers are unconsciously aware of the inherent

iconicity deposited in constructions, especially in the more transparent cases, and that this awareness is used in the process of storing, retrieving, and making sense of those grammaticalized (and thus learned) patterns. I believe that this is certainly the case with iconic word order patterns such as the topic-comment binomial, or with the possible positions for focus elements and their relative salience, or the fact that the negative particle is rheme-peripheral in a language (as it typically is). In the next section I will look at the cross-linguistic preferences for the coding of negative operators, something which will help us understand the Basque negative assertion construction later on (cf. Chapter 7).

1.7.3.3 Negative morpheme positioning

The negation operator morpheme provides an interesting case study of ordering preferences for a scopal operator cross-linguistically. The negative operator has scope over the whole assertion, excluding setting elements, such as the clausal topic and other setting elements (cf., e.g., Givón 1979, 1984a). Dahl (1979) and Dryer (1988b) have studied the position of negation morphemes crosslinguistically and came up with some interesting findings. Dryer looks at the position of the negative morpheme whether it is a free or dependent morpheme. Dahl makes a major distinction between morphological negation (inflectional) negation and syntactic negation, i.e. by means of uninflected particles, auxiliaries, or “dummy auxiliaries” (the rarest method, cf. English).

Dryer’s statistics for the different negative positions in verb-initial, verb-medial, and verb-final languages can be seen in Table 1-6.

SVO	Lgs	Fams	SOV	Lgs	Fms	V...	Lgs	Fams
negSVO	4	3	negSOV	8	5	negVXY	44	all
SnegVO	47	13	SnegOV	6	3	VnegXY	0/0	0/0
SVnegO	3	1	SONegV	39	15	VXnegY	0/0	0/0
SVOneg	13	4	SOVneg	64	18	VXYneg	0/0	0/0
	67	15		117	23		44	

Table 1-6: Order of negative morphemes in languages with different basic word orders, according to Dryer 1988, by number of languages and number of language families.

As we can see, for each positional possibility, Dryer gives the number of languages which displayed that order *and* the number of families represented.

One major generalization is that the negative particle tends strongly to gravitate towards the verb, and especially towards preverbal position. The preverbal preference is categorical for verb-initial languages, a strong preference for verb-medial languages (50/65 languages), and the position for at least one third of verb-final languages (39/117) and probably more if we include some hard to classify, flexible verb-final languages (cf. Dryer 1988b).¹²³

Another generalization, which doesn't necessarily stand out in this chart, is that the negative word has a strong preference to be at the outside edge of the whole rheme over which it has scope, i.e. the clause minus the subject (when it is the topic). The most preferred peripheral position is rheme-initial position, over rheme-final position.

These two cross-linguistic preferences for preverbal and rheme-peripheral placement, conspire to make the negative morpheme be rheme-initial. This tendency is perfectly compatible with verb-initial languages (neg-VOS, neg-VSO), highly compatible with SVO languages (SnegVO), but more problematic in verb-final languages, where placing the negative particle in preverbal position prevents it from being rheme-peripheral

(and also interferes with the closeness of the verb and the object), unless, of course, the object ‘moves’ out of the way, which is what happens in Basque, as we will see.

Thus in OV (verb-final) languages we find a classical case of competing motivations, the same motivations which in VO languages conspire to produce certain outcomes. There is a motivation for the negative particle to be rheme peripheral, and a motivation for it to be close to the verb, and a motivation for the verb and the object to be close to each other. However, all of these iconic motivations cannot be satisfied at the same time, which is why these languages must find some sort of compromise. The compromises which Basque has found, however, have produced a somewhat delicate balance which may end up producing structures which are much more like VO structures than like OV ones.

The order of the negative morpheme can, and should, also be studied from the point of view of its diachronic source, as Givón 1984a has done, at least when this is at all possible. Such sources typically explain their actual position in the clause. Thus Dahl (1979) found, for instance, that in OV languages the negative particles which follow the verb tend to be derived from (higher) negative verbs, whereas other particle types tend to be preverbal (as in Basque). On the other hand one must assume that these morphemes were chosen at one time because they were acceptable, i.e. motivated, choices.

Two major sources of negative morphemes are negative verbs, and nouns originally used as intensifiers (e.g. French *pas* “step”). Negative adverbials are another possible source. In many cases, however, such as in Spanish and Basque, the source of the negative morpheme is not known, but the negative particle (cf. *not*) is identical to the negation word (cf. *no*).

1.7.3.4 Ordering of complement pairs

Haiman argued that the relative order of subject and object in transitive clauses might iconically reflect the order agent-patient, since the agent initiates the action and the patient is at the 'receiving end' (cf. also Bates and MacWhinney 1979:190).¹²⁴

Hajičová (1991) argues that iconicity is responsible for a variety of default orderings among other pairs of elements inside clauses indicating cause and effect, source and goal, and the like. Thus, for instance, in English, and perhaps in most languages, "the order Directional-*from* Directional-*to* can be regarded as basic" (Hajičová 1991:101).¹²⁵ Other orderings are more variable from language to language, e.g. Object < Instrumental is basic for English, but the opposite seems to be more basic, or unmarked, for Czech.

In earlier studies Hajičová and Sgall have discussed the preferred orderings of semantic complements of verbs in different languages, what they call "the systemic ordering of verb complementations" for each language (cf. Sgall, Hajičová and Benešová 1973; Hajičová and Sgall 1987). In English they argue that the "systemic ordering" is the following:

(1.8) Actor < Addressee (Dative) < Objective < Origin (Source) < Effect (Result) < Manner < Directional-*from* < Instrument (Means) < Directional-*to* < Locative

The 'systemic orderings' of different languages may differ from each other somewhat, they argue, but some relative orderings may be highly preferred crosslinguistically. These relative order preferences may have differing strengths and may interact with other iconic principles of ordering in different ways, but they may be important nonetheless, all other things being equal.

Notice also that some of the pair relationships involved here seem to go beyond ordering preferences and involve also adjacency and even semantic bonding of the type which perhaps also constituency. Thus directional-*from* and directional-*to* phrases typically occur not only in that order, but also adjacent to each other, and sharing a single intonational contour. Such pairs of phrases may also change positions together, as when they are used as setting adverbials.

1.7.4 Connector in the middle: the Relator Principle

1.7.4.1 Introduction

There is yet another very interesting iconic principle which may be involved in motivating certain word order preferences. This principle is mentioned for instance by Dik (1989) who refers to it as “the Relator Principle”. By relators, Dik means coordinators, adpositions, case markers and subordinators.¹²⁶ According to this principle, “[r]elators have their preferred position: (i) in between their two relata; (ii) at the periphery of the relatum with which they form one constituent (if they do so)” (Dik 1989:346).¹²⁷ In (1.9) we can see the two preferred possibilities.

- (1.9) a. Complement/Modifier - **Relator** - Head
b. Head - **Relator** - Complement/Modifier

This principle would motivate for instance the fact that postpositions and other phrase final markers correlate quite strongly crosslinguistically with the order modifier/complement-head, (e.g. [*my mother's*] *house*) and that prepositions and other phrase initial markers correlate with the order head-modifier/complement (*the house* [*of the Prime Minister*]).

This principle would also motivate the cross-linguistic correlation between **postpositions** and other phrase-final **case markers** and **subordinators** with verb-final order, and the correlation between **prepositions** and other phrase-initial markers and subordinators and verb initial and verb-medial order.

This principle would also seem to motivate the correlation between prenominal **relative clauses** and post-clausal relativization markers, as well as postnominal relative clauses and pre-clausal relativization markers. Indirectly, it might seem to also account for the correlation between prenominal relative clauses and verb-final clausal order, and, inversely, for the correlation between postnominal relative clauses and verb-initial (or verb-medial) clausal order. In other words, it would seem to motivate the **branching consistency** (left- vs. right-branching) which has been so often observed, a preference which has been assumed to be basic or underlying for all language in some theories of grammar. In Table 1-7 we can see some of the predictions made by the Relator Principle.

OV languages	NP-POSTP> [[Rel. Cl.]-REL> NP-CASE> Verb _{head}
VO languages	Verb _{head} <CASE-NP <REL-[Rel. Cl.] <PREP-NP

Table 1-7: Some preferences for relator position as predicted by the Relator Principle.

As we can see this principle has the potential for helping us understand a great number of the correlations that have been found to exist with different basic word orders. The Relator Principle would motivate the development of phrase-final relators (or dependency markers) in OV languages and of phrase-initial ones in VO languages, or, alternatively, the development of OV order in languages with phrase-final relators and of VO order in languages with phrase-initial ones.

On the other hand, that the existence of a great number of exceptions to this principle makes it clear that it is probably not a very strong motivator, or at least that sometimes its motivation can be easily overridden if there are reasons to do so. It is also a preference which may take a great deal of time to fulfill itself in any one language, diachronically speaking.

It is not clear how dysfunctional structures which violate this principle actually are, since so many languages allow exceptions. The presence of exceptions, however, could be correlated with other characteristics of the units in question, such as whether the related element is clause internal or clause external. Thus, for instance, this principle, in and of itself, would seem to motivate (in all languages) clause-final subordinating conjunctions for setting clauses, which typically come before the main clause, such as for temporal or conditional clauses, both of which typically function as settings. On the other hand, the potential dysfunction caused by clause initial clauses with initial subordinators would seem to be very small given that such clauses are always clearly delimited intonationally. In addition, a clause initial subordinator in setting clauses is also motivated since it allows the hearer to immediately classify a clause as a setting clause. With intonationally less-clearly delimited, clause-internal phrases and clauses, on the other hand, the delimiting and orienting function of relators might be more important. In the next section I will look at the relevance of these observations for Basque.

The actual reason behind the Relator principle, it would seem, is not to produce systemic symmetry among constructions, i.e. to produce head-initial vs. head-final structures, as has often been thought by typologists trying to understand Greenberg's correlations, but rather to facilitate processing through iconic ordering of elements. This

principle, like the head-ordering symmetries in general, is also consistent with the preference for avoiding center-embedding, which has also been shown to be related to processing difficulties (cf. Kuno 1974).

1.7.4.2 The Relator Principle and Basque dependent clauses and phrases

Basque presents an interesting testing ground for this principle when it comes to dependent clauses. To begin with, Basque **setting clauses** are quite strongly verb and subordinator final, e.g. the absolute-like adverbial *-larik* clause and temporal and conditional clauses. This fact thus agrees with the relator principle. I should note, however, that there is a new, budding adverbial, setting construction in Basque, modeled on a similar Romance construction, which is subordinator-initial, cf. the adverbial *nola* “how; since” construction. This still limited addition to the Basque inventory of dependent constructions is taking root despite the fact that it goes against the relator principle and that Basque already has equivalent subordinator-final constructions.

Another interesting case is presented by **asserted dependent clauses**, i.e. finite complement clauses which function as dependent assertions following the main asserted clause in their own intonation units, such as *for/cause* causative clauses in English. These clauses, like main asserted clauses, display a great degree of word order freedom, and are not strictly verb-final. One of the most common of such clauses is, in Basque as in other languages, dependent causative assertions, e.g. *He told me he was going to come, cause/for/since he had to pick up some books*. Such asserted clauses are not strictly verb-final in any version of Basque, but in standard Basque they have a preverbal ‘subordinator’ (*bait-*). This, of course, violates the Relator Principle. But in spoken

Basque, very often an assertion initial subordinator/relator, the particle *ze(ren)* “(of) what”, is added to such clauses. It isn’t clear whether the development of this subordination marker is directly, or indirectly, attributable to foreign influence, but it may very well be in part motivated by the Relator Principle’s influence.

Basque **completive clauses** are almost always postverbal in speech, and very often they are in separate intonation units from the main clause. The completive subordinator (*-la* with affirmative main verbs) is verb-final, and thus can never be ‘in the middle’ if the complement clause is postverbal, thus violating the Relator Principle. On the other hand, these clauses are always well-delimited intonationally and ‘pre-announced’ by a verb which takes such a complement clause, thus the dysfunctionality of this violation might be lessened. I should mention, however, some speakers, also very likely under the influence of Romance, sometimes use *nola* ‘adverbial’ clauses as completive clauses, with assertion-initial adverbial pronoun *nola* used as the de-facto subordinator, much like in English, e.g. *He told me how he was going to come tomorrow*. Unlike the previously mentioned adverbial *nola* construction, this new addition is in full agreement with the Relator Principle.

Interesting observations can also be made about the position of relators and the order of (non-clausal) non-clausal **complements** and **modifiers** inside clauses. We may note, for instance, that, as we will see in Chapter 6, absolutive (direct) objects, which have zero marking, i.e. no overt relators, are more likely to postpose than other types of complements, all of which have case endings or postpositions. In other words, the proportion of XV to VX orders (where X = a non-absolutive complement or modifier) is higher than the proportion of OV to VO orders. Although I will argue that the reason for

this difference is related to the typical pragmatic differences between O's and X's, it is interesting to note that the X's are typically coded by a more complex relator than are O's and, thus, that this seems to be consistent with the Relator Principle.

1.7.4.3 The cause-and-effect issue and the instantiation issue

As I mentioned earlier, the mechanism which would 'instantiate' the Relator Principle in the grammar of a language would have to be a diachronic one. That is, this principle could encourage, whenever new relators are created or elements are 'rearranged', to follow certain paths and not others.

It is not clear, however, whether the Relator Principle acts by 'encouraging' heads to be placed on the side of the relator, and/or by 'encouraging' complements and modifiers to place themselves so that the relator is next to the head. It is of course possible for the principle to work both ways, depending on the circumstances. Thus, for instance, if a complement changes position with respect to the verb for some reason, the Relator Principle might eventually encourage the development of new relators on the opposite side. On the other hand, change could proceed in the opposite direction as well. Thus the development (or borrowing) of a construction with a relator which violates the Relator Principle might eventually encourage the shift in position of that construction with respect to its head.

The diachronic mechanism of instantiation of iconic principles is, of course, the mechanism of grammaticalization and restructuring I mentioned earlier. We know, for instance, that adpositions (a type of relator) arise from nouns and adverbials which express relations and take nominal complements. This type of development always

produces structures which respect the Relator Principle. On the other hand, the development of relators from verbs in serial verb constructions is known to produce relators which violate the relator principle, in addition to producing complements on the opposite side of the verb from which they are expected (cf. Li and Thompson 1973, 1974; Lord 1973; Givón 1975a, 1984a).¹²⁸

One may argue that the Relator Principle is not a cognitive principle at all but merely a descriptive generalization of , and that the types of structures that it predicts, just like the word order correlations, are the result of 'blind' diachronic changes of grammaticalization, such as the development of postpositions from clause-final verbs. In other words, one might argue that the correlations (or harmonies) are indeed the result of diachronic changes, but deny that there is any teleology involved in such changes. Thus, if a language is verb final and an adposition develops out of a verb, of course it will be a postposition and not a preposition. Or if a tense or aspect marker develops from a main verb, the position of the morpheme will reflect its original position as a main verb (cf., e.g., Givón 1975a, 1984a). Also, the longer a language has been either OV or VO, the more consistently it displays these correlations, simply because there has been more time for the new constructions to become grammaticalized. (Givón also argues, however, that the correlations are stronger for OV languages than for VO languages.)

Some, however, have argued that although this mechanistic (non-teleological) explanation may be valid for some correlations, it may not be valid for all of them. Thus Dryer (1992a) agrees that 'blind' grammaticalization changes may be able to explain the correlations between the order of verb and object, and the order of verbal auxiliaries and some affixes, but not correlations such as such as the order of verb and adpositional

phrase (Dryer 1992a:127). As we will see below, Dryer argues for a consistent branching generalization which is fully compatible with the Relator Principle.

It is clear that if some change is truly adaptive, i.e. beneficial, given enough time a language could find a way of realizing it. I have mentioned the use of adverbial embedded 'questions' in Basque as complement clauses which yield clause initial 'subordinators'. To take another farfetched example we can see how if phrase-final relators were indeed 'beneficial', English could easily 'create' them by means of its genitive construction.¹²⁹

As we can see from these examples, the relationship between typological correlations ('universals') and diachrony is very tight indeed. The question of the teleology and psychological reality of these universals, on the other hand, is far from resolved. In the next section we will look more closely at this deep relationship between typology and diachrony.

1.7.5 Diachronic typology and drift

Typology studies the patterns (absolute patterns and statistical correlations) that are preferred and those that are less likely cross-linguistically. But patterns change through time and the process of how preferred and 'dispreferred' patterns come about and disappear in diachrony is crucial to the study of typology. Greenberg, the leading figure in modern typology, is also a strong advocate of the close connection between the study of typology and diachrony (cf., e.g., Greenberg 1968, 1979; Croft, Denning and Kemmer, eds., 1990). He has argued that "[s]ynchronic regularities are merely the consequence of [diachronic] forces. It is not so much ... that 'exceptions' are explained historically, but

that the true regularity is contained in the dynamic principles themselves” (Greenberg 1969:186).¹³⁰ This means that

Diachronic typology “dynamicizes” the synchronic model by reinterpreting synchronic language states as stages in a diachronic process. One then analyzes the transitions between states or stages and the stability and frequency of occurrence of language states. Synchronic typological generalizations become diachronic sequences of stages. So-called “exceptions” to synchronic universals, such as a combination SOV and Noun-Genitive word orders, usually turn out to be unstable transitory stages in a more or less gradual diachronic process. (Croft, Denning, and Kemmer 1990:xiii).

Greenberg and other typologists are well aware that diachrony is the key to typology, for languages come to be of a certain type, and to change types, by means of diachronic processes. The assumption is that “languages do not stand still, and the ways that they change are constrained whether the change occurs externally through contact or internally via ‘drift’ (Sapir, 1921)” (Croft 1991:2). The mechanisms of change are “a large class of diachronic processes ... that go under the name of grammaticalization” (Croft 1991:3), and, as has been shown, there is directionality to these processes.

One should not think, however, of mixed states as being somehow impure or imperfect. As Croft tells us,

The terms ‘intermediate’ and ‘transitional’ have the unfortunate and unwarranted connotation that such language states/stages are diachronically unstable, i.e. are inherently dysfunctional and prone to change. Variation does not in itself imply instability; Weinreich, Labov and Herzog make it quite clear that they do not have this in mind (1968, 101). Stability of a language state is an independent factor that must be argued for on the basis of functional considerations. (Croft 1991:3).

The Sapirian concept of ‘drift’ has played an important role in recent studies of word order. By ‘drift’ Sapir (1921) meant an “unconscious selection on the part of [a

language's] speakers of those individual variations that are cumulative in some special direction" (Sapir 1921:155).¹³¹ Of course, there need not be anything mysterious about this 'unconscious selection' at all, if we view it as being 'primed' or 'motivated' by 'external' cognitive principles, such as a preference for iconic structures or structures which are easier to produce and to process, as well as perhaps system internal principles of grammatical organization.¹³²

The main example which Sapir used to exemplify the phenomenon of drift involved the loss of inflections and word order freedom and the concomitant word order change from verb-final to verb-medial in English. Sapir saw these two latter changes as being the result of the former, the loss of case distinctions, which resulted in the use of word order to code grammatical relations (cf. also Lakoff 1972, Vennemann 1975, Malkiel 1981).¹³³

The hypothesized connection between loss of case marking and the shift in word order from OV to VO is extremely suggestive in light of the existing strong correlation between case marking and OV order mentioned earlier. How exactly the change ('drift') would take place across time, through generations of speakers, is not completely clear, however. Also, it is not clear that the fixing of the order is caused by the loss of case distinctions and not that, say, the two went hand in hand. As Hawkins (1986) and others have argued, it would seem that the changing from OV to VO should precede, or at the very least be simultaneous with, case loss, or else we should be able to find languages which are OV and caseless.

Besides the actual sequence of events, we should consider whether either one of these two changes is the cause of the other one, or whether they are independent. Sapir

(1921) and others, for instance, have viewed case syncretism, which according to him came first, as being the direct cause of the fixing of word order, which they have seen it as a compensatory strategy. Others, such as Jespersen (1922) have argued that word order change and fixing came first, and that case syncretism followed, presumably since case marking was no longer 'needed'.¹³⁴

Regarding the possible causal relationships between these two facts we should not forget a third factor in the equation, namely verbal 'coreferencing' of arguments, as we saw in Section 1.6.9 above. I will return to this issue in Section 1.8 below where we will see one possible way in which the two changes, loss of inflections and shift to VO order, may come about through language contact and both be caused by other factors which conspire to do away with inflections and to fix the word order.

The issue of the connection between word order and case making, and case syncretism in particular, is quite relevant for the study of modern Basque. Basque has a fairly strong case system for distinguishing subjects and objects, but in practice it seems that this system is losing some of its 'strength'. Even in the standard dialect we find case syncretism between the definite ergative singular and the definite absolutive plural markings for nominals (both *-ak*). In some of the major dialects, however, there is further syncretism between the singular and plural ergative (again, both, *-ak*, as opposed to the eastern varieties and the standard, which have the ergative plural in *-ek*). The system is complicated by accentual differences which in some varieties allow speakers to differentiate nominals with identical case markings. All these issues have been discussed in great detail by Jacobsen (1974).

The problem is aggravated by the fact that at least some speakers of western dialects are not always ‘careful’ about adding ergative case marking to singular nominals, i.e. about differentiating singular absolutes and ergatives (just like they don’t differentiate between plural absolutes and ergatives). As we will see in Chapter 2, 15.5% of all clauses containing an overt ergative argument in the Spoken Basque Corpus display case syncretism, i.e. the ergative argument is unmarked, i.e. it is marked just as an absolute. On the other hand, because of the rich verbal ‘coreferencing’ system, only about 20% of all ergative arguments are overt in spoken narrative (80% are ‘covert’ or ‘elided’), which means that ergative syncretism is present only in about 3% of all clauses containing ergative arguments.

In other words, we see here that a ‘free’ word order language such as Basque may live with some case syncretism without this causing grave dysfunction in the system.¹³⁵ On the other hand, it is possible that such incipient loss may be a preview of graver things to come, i.e. that it may reflect an ongoing ‘drift’ towards the loss (or more severe curtailment) of case marking and the fixing of word order. More serious would be in this case the loss of verbal coreferencing, but that system seems to be very strong and stable, allowing the coding of all major argument types.¹³⁶

The concept of drift, and long term change in general, is one of the major challenges to the study of diachrony. How can teleological change be possible across generations, without any one generation seeing the whole picture? Children cannot possibly learn the history of the language, but only the synchronic state to which they are exposed. This is in fact one major reason behind the separation of synchrony from diachrony which has characterized 20th century linguistics since Saussure (cf., e.g., Lass

1984; for the opposite perspective, see Ferguson 1990).¹³⁷ On the other hand, there must be an explanation, since the facts are not spurious. Recent sociolinguistic work also points to the fact that children accurately recreate conditioning for variables which seem to have no linguistic significance, and thus “[i]n many ways, the child is a perfect historian of the language” (Labov 1989:85).¹³⁸

The answer to the question of long term change must lie, I believe, at least in part, in the existence of internal and/or external ‘motivating’ forces which lead languages in different directions. Different generations of speakers may ‘conspire’ to take a language in a certain direction because, given the current state of the language and the forces that act on it (and on the language’s speakers), that direction is the ‘only way to go’, or at least a “high probability outcome” (cf. Aitchison 1989).

1.7.6 Analogy: Systemic explanations for the correlations

Although the mechanism for achieving harmonic constructions is generally admitted to be diachronic, most investigators have sought a teleological explanation for these diachronic changes, or at least for some of them. We saw one likely candidate for such a teleological motivator of change in the Relator Principle. The Relator Principle is primarily an iconic and thus external motivation presumably with a cognitive base.

There is another possibility, however, one which has been adopted by several investigators, and this is a system internal explanation which reformulates the correlations in terms of abstract properties that all the correlations share. There have been several similar versions of this principle which Dryer calls generically ‘Head-Dependent Theory’ (HDT). According to this type of theory, the word order correlations or “harmonics”

“reflect a tendency to order grammatical heads consistently with respect to their dependents” (Dryer 1992a:87). In other words, languages tend to be either head-initial or head-final. This is basically what was proposed by Lehmann and by Vennemann, as well as by Hawkins (1983), and others.¹³⁹ This is at heart a system internal motivation based on systemic consistency among constructions.

This principle of system internal coherence or harmony is readily amenable to a purely formal interpretation, which is probably why it has proven to be so appealing to formalist approaches to language, for which full harmony is seen as the default situation in language. The inherent underlying systemic harmony is represented by the MIT School, for instance, by means of ‘X-bar theory’, directionality of Case and/or Theta-role assignment, or by the ‘word order parameter’.¹⁴⁰ This formal (systemic) approach does not explain, however, the fact that languages only rarely prove to be totally harmonic and seems to imply that non-harmonic constructions are somehow deficient and flawed.

Theories in the mould of the ‘Head-Dependent Theory’, however, need not be purely formalistic (over-)generalizations of Greenberg’s harmonics. This type of explanation for the tendency to place certain elements in certain positions across constructions may also be seen from a functional perspective, that is, it can be seen a system internal motivation, namely analogy. Thus, presumably, the ‘desire’ for a more perfect or symmetrical system would motivate speakers to prefer or come up with certain types of structures and not others.

Dryer has argued against previous versions of what he calls the ‘head-dependent theory’ on the basis that it is inaccurate. He argues that “the predictions of the HDT depend ... on what assumptions one makes as to which element—if either—is head” and

often there is disagreement about this, such as in the case of nouns and determiners in noun phrases (Dryer 1992a:89).¹⁴¹ In addition, the HDT often makes incorrect predictions. Thus he finds, for instance, that in just as many OV languages do adjectives follow the noun as precede it (according to the HDT they should always precede it).¹⁴²

Dryer proposes a similar, but seemingly more accurate generalization. Noticing that the dependents that pattern like objects are all phrasal (e.g. genitive phrases), whereas the dependents which do not are all non-phrasal (e.g. adjectives), he argues that we should refine the generalization to say that “languages tend towards one of two ideals: right-branching languages, in which phrasal categories FOLLOW non-phrasal categories, and left-branching, languages in which phrasal categories PRECEDE non-phrasal categories” (Dryer 1992a:89). He calls this the Branching Direction Theory (BDT).¹⁴³

According to the BDT, elements which pattern like the verb (‘verb patterners’) are “non-phrasal (non-branching, lexical) categories,” whereas those that pattern like objects (‘object patterners’) are “phrasal (branching) categories” (Dryer 1992a:109). Thus, although many languages are still not consistent, the exceptions are much easier to account for, though some problems, such as articles and adjectives, remain (Dryer 1992a:118-128).¹⁴⁴

Notice that this restatement of the ‘head-dependent theory’, besides being more accurate, is consistent with the Relator Principle, since phrasal categories are typically connected to their ‘heads’ by means of relators. The Relator Principle, however, has nothing to say about those cases in which the phrases don’t have any relators. This includes elements which do not display any case marking, such as subject and object in English and many other VO languages (OV languages typically have case marking, as we

have seen). The subject position in SVO languages, however, would seem to be a problem for either theory. Obviously there must be other motivations involved which can account for these problematic cases.

Principles such as the ‘head-dependent’ principle or the ‘consistent branching’ principle, or even the Relator Principle, are not necessarily motivated by system internal preferences for harmony and symmetry. For Dryer, as for many others, the generalization is not the final explanation for the patterns. If any of these generalizations obtains is because it helps with processing, either in speech production or in speech perception. That is, the correlations arise in part at least because they help with processing, that is “the word order correlations ultimately reflect the nature of the human parser” (Dryer 1992a:128). How the consistencies help, and how crucial their help is in the different situations, is still open to debate, for it is obvious that languages can live quite happily with a great deal of inconsistency. I will return to this issue in the next section.

There are still others, who maintain that such overarching principles are uncalled for and that the correlations are better explained in terms of two-way-interactions, as would be predicted by a model which accounts for the correlations in terms of non-teleological diachronic processes (cf. Croft 1995). Thus, for instance, if the genitive is postnominal in a language then it will typically have postpositions, since these often develop from constructions containing genitives (e.g. *top of > on*), and it will have prenominal relative clauses, since the relative morpheme often develops from such genitive morphemes (cf. Aristar 1991). Or, as we saw earlier, the order of negative markers depends on the order of the source morphemes (cf. Croft 1995:103-4 for other examples).

The binary relationship model might seem to account for the fact that some relationships (harmonies) are much stronger than others, since some diachronic developments may be more likely than others, for a variety of reasons. On the other hand, it fails to explain why languages tend to have certain preferences in the choice of sources for derived relators, such as the preference of OV languages for genitive markers as the source of relative clause markers, or the preference of VO languages for clause initial elements as the source of theirs, such as the *wh*-pronouns that are found in English for example.

1.7.7 Processing constraints and word order

As I just mentioned, processing constraints have often been pointed to as a factor motivating the form of constructions as well as changes in constructions. Hawkins explains the nature of these constraints thus:

Processing explanations for language universals invoke principles of comprehension and production that have been derived from controlled experiments on real-time language use. These explanations appeal ultimately to two major considerations: human beings are limited capacity machines—there are, for example, real limitations on short-term memory; and different linguistic structures may be associated with different degrees of processing ease or difficulty. (Hawkins 1988:15)

This last sentence suggests, as Hawkins adds later on, that “processing difficulty is a gradient notion, with empirical consequences for language frequencies and implicationally defined co-occurrences of properties” (Hawkins 1988:21). Some processing difficulties may be insurmountable, while others may be a mere annoyance.

The notion of processing difficulties and constraints goes back to Yngve (1960, 1961), who dealt with “sentence depth” and short term memory, and to Chomsky 1964/1961 and Miller and Chomsky 1963, which are a reaction to the former and deal with “center embedding” (cf., also, Bever 1970, Slobin 1973). Chomsky noticed that whereas there are no limits on left- and right-embedding, or branching, speakers have serious problems with center embedding, as exemplified sentence in which the subject contains a relative clause, which in turn contains a subject with a relative clause, and so on.¹⁴⁵

Kuno (1973, 1974) attempted to explain some of Greenberg’s correlations, such as the preferred order of relative clauses in different word order languages, by means of the tendency to avoid center embedding. It is not clear, however, that these rare structures could have such a great influence. To begin with, multiple embeddings are quite rare in language. Secondly, intonation can go a long ways towards disambiguating ambiguous structures. Furthermore, languages may develop strategies, such as extraposition, for dealing with such complex structures. Thus Dryer (1980) shows how verb-final languages may postpose sentential complements to avoid center-embedding, some obligatorily, as Persian and Turkish, and some optionally. Another, less favored, strategy consists in preposing the embedded clause.

One preference which may be related to processing constraints is the one against having long and complex units come early in the sentence, i.e. before the verb. This is especially true when the long constituent is not clause initial and if it contains right-branching structures. Such a preference may be behind some of the rules of extraposition

that have been proposed.¹⁴⁶ It might also be a factor involved in the fact that complement clauses are much more likely to be postverbal than nominal complements in Basque.

Hawkins has proposed in recent years that processing principles (i.e. external motivations), and not “principles of a purely grammatical nature” (i.e. internal motivations) and other *ad hoc* principles which he earlier supported, are “capable of explaining why languages arrange words in the orders they do” (Hawkins 1993:239). In earlier work Hawkins had proposed that analogy- or harmony based principles of the ‘head-dependent’ type could account for the different cross-linguistic regularities of word order, as long as these were supplemented by two sub-principles accounting for the fact that, all things being equal, “‘heavy’ categories”, i.e. complex constituents, prefer to come late in the construction, and ‘light’ ones prefer to come early (cf. Hawkins 1993:234).¹⁴⁷

In his new approach, Hawkins views ‘ease of processing’ as the primary determinant of the word order patterns or preferences found cross-linguistically.¹⁴⁸ He begins by noticing, and applauding, Dryer’s (1988b) already mentioned principle according to which “[l]anguages tend toward consistent left-branching or right-branching” (Dryer 1988b:191). This principle, according to Hawkins, accounts better for the early order of non-branching categories, such as adjectives. However, it does not account for the observed *overall cross-linguistic preference for right branching*. For explaining this latter preference, he argues, what we need is *a theory of sentence processing*, by which he actually means structural parsing, since meaning and intonation seem to play no role in his theory of processing (cf. Hawkins 1990, 1993, 1994).

The first parsing principle which Hawkins proposes, the “Principle of Early Immediate Constituents” or EIC, says, basically, that it is easier to process strings which

allow the early construction of the tree than those which don't. In other words, the listener wants to get the whole structural picture out of as few phrases as possible. What allows the processor to construct the tree is finding "mother node constructing categories" (cf. Fodor 1983), such as prepositions, a notion that is largely, but not fully, coextensive with the notion 'head' (cf., e.g., Hawkins 1993:247).

The EIC predicts, according to Hawkins, that right and left-branching languages can be equally optimal with "short simple sentences", given the existence of a 'buffer' which allows some leeway in having to wait for the 'mother node constructing category'.¹⁴⁹ In addition, he argues, this principle also "explains the existence across languages of re-arrangement rules such as Heavy NP Shift and Extraposition" (Hawkins 1993:238), predicting right dislocation of 'head initial' constructions and 'left dislocation' for 'head final' ones.

The second parsing principle that Hawkins proposes is the Minimal Attachment Linear Order Principle. This principle is very similar to the principle first formulated by Antinucci *et al.* 1979, according to which "[t]he human parser prefers linear orderings ... that don't invite garden-path interpretations," i.e. erroneous parsings (Hawkins 1993). Here there is an asymmetry between right- and left-branching languages, since complement clauses and relative clauses in the latter invite misinterpretation, since their heads or relators (their "Mother Node Constructing Categories") come late in the clause.¹⁵⁰ As Hall 1992 has noted, this type of reasoning has led many psycholinguists to argue for the existence of a "Head First Preference" in language.¹⁵¹

Different experiments have shown that that left-branching dependent sentences are more prone to misinterpretation. On the other hand, it must be noted that the types of

dependent sentences used in these experiments are unlike the vast majority of sentences that one actually encounters in discourse and that they are more like the sentences of complex writing. It thus seems unlikely to me that this 'minor' problem with left-branching structures would actually be a strong motivator for grammatical constructions. If it were, we wouldn't find, as Hall reminds us, that half of the languages of the world are OV.

One of the major problems that I find with the studies which propose processing explanations is that they assume that humans parse meaningless strings and construct meaningless structures rather than meaningful ones complete with very rich intonational cues as to the constituency and the function of the parts. Thus, as I will argue in Chapter 3, it is clear that from an informational and intonational perspective we must distinguish different parts of a sentence: the settings, the core clause, and the post-clausal elaborations (of the three only the core clause is needed).

Basque grammarian Villasante (1956) made some very perceptive comments supporting the idea that cognitive difficulties are caused by excessive left-branching. Villasante complained about some authors who, following prescriptivist pressures, had started to follow the verb final rule too strictly. He says that this

is a kind of syntax [ordering] which forces one to keep one's attention in suspense until the end of the sentence, when the main clause [verb] arrives, and this is certainly disorienting for the inexperienced reader. ... It seems like a paradox that the syntax which the people follow when they speak Basque can be so difficult to them when they see it in writing. But nonetheless, that is how it is. And this is because the spoken language, because of the intonation and other such resources, easily avoids such drawbacks (in addition to the fact that in speech we do not find such long and rather complicated units). (Villasante 1956:16; my translation, J.A.)¹⁵²

As Hawkins predicts, Villasante noted that having to wait for the main verb makes such sentences very hard to follow. He noticed that the verb-final rule, which is so easy to follow in speech, is really not so appropriate when we're dealing with long complex written sentences with a great deal of subordination. He also notices that sentences in actual speech are never as long and complex as the ones we find in (some) writing, and that in speech, unlike in writing, the speaker and listener can rely on intonational and other such resources which make constituency more obvious.¹⁵³

These are crucial points which proponents of the influence of parsing on the structure of language, such as Hawkins, seem to miss sometimes. Speakers do not parse mere syntactic strings, but rather meaningful, intonationally rich ones, which in the vast majority of cases are nothing like the long and complex sentences of some styles of writing. This is why the difficulties which are indeed associated with 'excessive' left branching seem to have so little effect on the actual structure of languages.

Compensation for the difficulties associated with left-branching is more likely to take the form of 'remedial strategies'. Thus we find in Basque that sentential complements are much more likely to be extraposed than nominal ones, following a focus extraposition strategy. The extraposition is to clause final position or to a new intonation unit following the main clause. Interestingly, however, the extraposing of complement clauses is much greater in speech than in writing. On the other hand, postposing of any type of complement is always greater in speech than in writing, as we will see. Also, I should note that there may well be other reasons that motivate the postposing of complement clauses, besides those associated with processing difficulties, such as the fact

that they, and not the main clause, are typically the main assertion of the sentence. I will look at these constructions again in Chapter 6.

1.8 LANGUAGE CONTACT AND LANGUAGE CHANGE

1.8.1 Grammatical change

Languages change. They do so for a variety of reasons (motivations), and they change at different rates, depending on a variety of circumstances. The motivations may be language internal (systemic), language external (e.g. processing, iconic principles), or they may lie in the influence of other languages, or a combination of these factors.¹⁵⁴ In this section I will present some basic concepts of grammatical change, especially as they relate constituent to order change and to the changes which may be taking place in the grammar of Basque.

Changes in the grammar consist of changes in constructions and other grammatical units. Unlike in phonology, semantic and/or pragmatic meaning (i.e. function) is at the heart of many grammatical changes. Grammatical change may involve changes in existing constructions or grammatical categories or the creation of new constructions out of paratactic structures. It may also involve borrowing grammatical elements, though it has sometimes been argued that lexical elements are much more easily borrowed than grammatical structures, and some have even denied that grammar may be borrowed at all. Others, however, disagree and believe that some aspects of grammar can indeed be borrowed, at least directly, unless perhaps when the grammar is already moving in the direction of the changes in the contact language.¹⁵⁵

Thus, for example, one type of grammatical change involves the reinterpretation of grammatical units (categories and constructions) so that they begin to be used in new contexts, i.e. their functional 'niche' alters, and this may take place under the influence of another language. Change of this type, of course, has repercussions on the other parts of the grammar as well. Such reinterpretations have been claimed to be at the root of diachronic shifts in preferred word order patterns. Changes in the constituent order in main clauses may precipitate, or encourage, changes in other constructions or categories in the grammatical system, as well as changes in the constructions derived from those constructions through grammaticalization. In other words, the influence of another language may go beyond mere borrowing of alien constructions (cf. Weinreich 1953; Lehiste 1988; Thomason and Kaufman 1988; Jahr, ed., 1992).¹⁵⁶

Extensive language contact through bilingualism of the type we will see exists in the Basque context, is certainly a potential motivator for changes to occur in a language. Recent studies have presented us with new perspectives on the types of changes that are likely to result from language contact situations, especially those that a non-prestige and/or receding language undergoes.

As we will see in this section, the main types of changes experienced by languages in contact situations seem to be primarily of two kinds: 'convergence' and 'simplification'. Both of these types of changes seem to have a cognitive motivation. As Silva-Corvalán 1994 has put it, "in language contact situations, bilinguals develop strategies aimed at lightening the cognitive load of having to remember and use two different linguistic systems" (Silva-Corvalán 1994:6). This is a concept that goes back at least as far as Gumperz and Wilson's (1971) well known study of grammatical

convergence.¹⁵⁷ This cognitive claim is based on the current belief that stable and balanced bilingualism, in which two or more languages are kept completely separate and equally accessible, is not the norm for a majority of speakers in most bilingual situations (cf. Fishman 1972; Seliger and Vago 1991).¹⁵⁸

Occasionally we find that a language does borrow a construction (or a grammatical category), seemingly ‘wholesale’, from another language. This is facilitated if it is relatively ‘transparent’ (iconic) and if it fits easily into the grammatical system of the borrowing language. It seems also that this type of borrowing typically involves minor or ‘peripheral’ constructions, even idiomatic calques, which complement the existing constructions of the language, and not, say, major speech-act constructions. I will identify some possible borrowings of minor constructions in later chapters. I will also try to show that Basque has not borrowed major constructions directly.

Sometimes the borrowing (transfer) is more subtle and it involves changes in native grammatical elements under the influence of another language. In other words, the meaning and uses (functions) of a grammatical element may change to become more like the meaning and uses of a superficially (formally) similar element in the source language. Thus, for instance, Kachru (1975) claims that a certain passive construction in Hindi which used to have a very restricted use has become much more prevalent through the influence of the English passive construction (cited in Cowan 1995:44). This type of borrowing may be rather common, especially in translation, as has often been noted. I will return to this issue in Section 1.8.4. This type of change is known as ‘convergence’, a type of transfer (see below).

We should not think of the influence of a language on another as ‘reaching out’ to another (‘foreign’) language to borrow from it. It is clear that borrowing takes place primarily in the mind of the bilingual speaker, who must handle those languages at the same time within a single cognitive realm. Much hinges here on the type of bilingualism involved. We are just beginning to realize what consequences such close contact may have for the shared languages, and in particular for a non-dominant one in an unbalanced or diglossic language contact situation.

1.8.2 Word order change

Despite claims that grammar is not typically borrowed, some investigators have argued that word order can indeed be borrowed, cf. Comrie 1981; Moravcsik 1978. Of course it may well be that borrowing doesn’t mean wholesale transfer but rather convergence of the type we have mentioned or even another type of change common in language contact situation that we will see below, namely simplification.

It seems clear that language contact may result in the word order of the different languages converging. Thus it has been noted by Dryer (1989b) and others that areal effects in word order are found throughout the world. Hock too has spoken of a “common continental European shift” towards SVO order caused by language contact, that is, as a Sprachbund feature (Hock 1986:510). In Western Europe Basque seems to be the only language which has resisted such a change, although we may wonder whether it will continue to do so for long.

If language contact resulted in borrowing, however, we would expect to find attested changes from any one word order type to any other type. In fact it turns out that

change to SVO seems to be much more common. This suggests the possibility that at least sometimes convergence is not the result of borrowing, but rather of other changes which result from the language contact situation, such as the loss of inflections (cf. Givón 1979). As Silva-Corvalán, for instance, mentions, following Givón, “there seems to be widespread agreement that grammaticalized distinctions marked by bound morphology are dispreferred in situations of linguistic stress (e.g. language acquisition, pidginization, cf. Givón 1979) because of their low semantic transparency, as compared with more or less corresponding lexical and periphrastic constructions” (Silva-Corvalán 1991:166).

Loss of case morphology doesn't guarantee that a language will become SVO, and languages have changed to SVO without losing their case morphology (e.g. Finnish, cf. Danchev 1991).¹⁵⁹ And word order doesn't necessarily become fixed when case inflections are lost, unless verbal coreferencing is lost as well, as we saw earlier. In fact, the loss of inflections, and in particular the loss of case inflections, may *follow*, rather than precede, a change to SVO order, since in such languages inflections are presumably less needed.

The history of the Basque language, or I should say the prehistory, is interesting with respect to the claimed naturalness of or inevitability of shift to SVO in a language contact situation. There is evidence that Basque was at one time was in rather close contact with Latin and was influenced a great deal by it. Despite myths to the contrary it seems that Basques did achieve a comfortable *modus vivendi* with the Roman Empire and that there was much contact between Basque speakers and speakers of Latin. (Latin was a flexible verb final language, but Late Latin, and Vulgar Latin in particular, seem to have been much less so.)

The proof of this contact is found throughout the lexicon of Basque. Most of the Latin borrowings are undoubtedly very old, as evidenced by phonological characteristics (e.g. Basque *pakea* < Latin *pacem*) and by the fact that they are typically not recognizable as cognates to the untrained eye. In addition, a not insignificant part of Basque derivational morphology seems to have been extracted from such borrowings. The rich inflectional morphology of Basque, on the other hand, including the rich verbal morphology and coreferencing system, doesn't seem to have been affected by this contact. This might perhaps have been the reason why the constituent order of Basque did not change at that time of intense contact. The fall of the Roman Empire, however, interrupted the contact between Basque and its Latinate neighbors for a long time, interrupting that period of close language contact, and perhaps giving Basque time to 'recover' from this contact.¹⁶⁰

Before analyzing language change issues resulting from language contact I turn now to the sociolinguistic situation of Basque, which will give us a sense of the psycholinguistic issues facing its speakers and of the reasons why language contact in this situation might result in changes in the grammar of Basque.

1.8.3 The Basque sociolinguistic situation

Basque has been a receding language for quite some time now, with the greatest loss during the last few hundred years (cf., e.g., Hualde *et al.* 1995 and papers therein). In a large part of the historical Basque speaking area, Basque has already fully disappeared. In these areas, loss of Basque has always been preceded, relatively briefly, by widespread bilingualism. For a number of reasons, bilingualism in the Basque Country has not been

a stable phenomenon. Bilingualism has always been the precursor of Romance monolingualism. This pattern is common to many languages of low prestige elsewhere. The situation for Basque has always had a strong class component, so that elites became bilingual first and subsequently became Romance monolinguals when there were enough bilinguals around so that speaking Basque was not necessary for communicating effectively with people in their communication 'networks' (cf., e.g., Milroy and Milroy 1985).

Basque elites have for a long time been dependent on the Spanish and French elites and the centers of power in Madrid and Paris, especially since the beginning of the Spanish Empire and the unification of the Kingdom of Spain and the French Revolution.¹⁶¹ Thus these elites, with very few exceptions, never did anything to unify and develop Basque into a language of culture and education, thus resulting in a diglossic situation for bilinguals which in the end favored the spread of Romance, first as a second language, then as the replacement language, a change which can happen within a few generations, particularly in urban areas where Romance speakers were mostly concentrated.

The fact is that because of both natural sociolinguistic tendencies and because of a forceful effort on the part of State governments *and* local elites to extend the state ('national') languages (Spanish and French), only a minority of Basques are Basque speakers at present, the percentages varying greatly from region to region. Furthermore all Basque speakers are currently bilingual and may actually be dominant in the other language (either French or Spanish).

To complicate the sociolinguistic situation, we find that because of recent political changes in the Spanish State which have favored a humble linguistic renaissance, a not insignificant number of children with monolingual Romance parents are going to all-Basque schools and learning Basque in an educational setting. An even larger number perhaps are attending schools in which at least part of the curriculum is carried out in Basque, thus creating a number of speakers for whom Basque may be more like a foreign language since they use it in very limited situations.

As a final complicating issue, the Basque Language Academy, starting in 1968, has been involved in a process of standardization of the sometimes quite widely diverging Basque dialects, which was no doubt a factor in the adoption of Spanish and French as a language of wider communication within the Basque Country in earlier days. The standardization has affected primarily the written language and the mass media, but it is extending itself in part to speech as well, primarily through Basque-language television and radio and the schools. This means that speakers of Basque are exposed to their own local dialect, plus a still unstable, standard variety, as well as a number of other dialects which speakers of earlier generations were generally not exposed to when communication between members of different communities was more limited and there weren't any Basque media. As we will see below, this increased linguistic variation in the Basque speech community may have consequences for the structure in the long run.

However, despite its recent resurgence at some levels, Basque is a language in dire straits and in great danger of disappearing as a 'normal' language and of becoming a living relic, much as has happened with Irish. Although the Basque-speaking population

is said to comprise at least a quarter of the inhabitants of the Basque Country, many Basque speakers use this language only in very few, limited contexts.

These are some of the sociolinguistic variables that a study of language change in Basque has to contend with. I give these facts because they may help us understand certain possible ongoing changes in the grammar of this language. The language contact situation between Basque and Romance is extremely interesting from the point of view of word order because these languages are (at least in theory) so radically different typologically, the former being primarily an OV language and the latter VO languages. As we will see later on, however, some of the VO-like structures which characterize Romance languages already exist in Basque, even if they are not always used in the same contexts, a situation which is perhaps ideal for the tendency toward convergence mentioned above.

1.8.4 Language contact and convergence

As I said, some linguists have claimed that linguistic transfer of major grammatical elements cannot actually take place unless they are in accord with the general patterns of change already present in the language. In other words language contact may speed along changes which are already in process. Even then some grammatical elements may be more resistant to change than others. This view has been challenged recently by a number of investigators, such as Thomason and Kaufman (1988). Silva-Corvalán too finds examples in the Spanish of Spanish-English bilinguals in Los Angeles of what might seem to be clear transfers from English, such as the elision of the complementizer and the relative pronoun (*que* “that”), obligatory SVX order (no

subject inversion), and obligatory overt subjects (Silva-Corvalán 1993:21). These changes, however, are clearly not ongoing tendencies in any varieties of Spanish.

Transfer, as defined by Weinreich, involves either (1) the replacement of one form by another, (2) the transfer of meaning from a form in the source language to a form in the target language, or (3) the change in the use of a particular form in the target language (Weinreich 1974:30-31, cf. Silva-Corvalán 1991:153). Only the first of these changes has been said to constitute true borrowing, although if the form in question is very iconic and conforms to the patterns of the borrowing language, its transfer may be highly motivated. The other two types of transfer can be seen as cases of convergence, a milder type of transfer (Silva-Corvalán 1991, Prince 1992a).

Silva-Corvalán, for instance, has identified in several recent publications (1991, 1993, 1994) examples of grammatical convergence in the Spanish of English-Spanish bilinguals. These cases of convergence crucially depend on existing structures in the target language. Transfer, she argues, is much more likely to take place in the form of convergence of structures in the target language with those of the donor language. In other words, transfer seems to be most likely to occur when 'similar structures' exist in the borrowing language which can be extended to cover new contexts and thus become more similar to the structures of the source language.¹⁶²

Silva-Corvalán shows that this type of convergence is found in the realm of word order for example. Thus she finds a decreasing tendency to postpone subjects which are not topics in English-dominant bilinguals.¹⁶³ But preverbal focus subjects are indeed possible in Spanish, as we will see in Chapter 3. They are just more marked than postverbal ones, being typically interpreted as contrastive or otherwise emphatic.

Prince (1989, 1992a) too has discussed two interesting cases of convergence, one from Slavic into Yiddish, and one from Yiddish into the English of some Yiddish-English speakers. The latter example, 'Yiddish Movement', involves object focus fronting, a rather marked, but possible strategy in English, as we will also see in Chapter 3. She finds that bilingual Yiddish-English speakers, as well as those influenced by their Yiddish-influenced dialect of English, use this construction in contexts in which it would not be used by most speakers of English, presumably under the influence of Yiddish (cf. also Prince 1981a).

According to Prince, for a given form in one language to become associated with a given form in a second language, first of all "the two syntactic forms must be 'matched up' by the speakers, following the string order of the constituents" (Prince 1992a:107). In addition "a discourse function associated with a form in the borrowing language [must be] consistent with a discourse function associated with a form in the source language" (Prince 1992a:108). In the case of focus fronting, it is clear that what we find is a natural extension of an already existing English construction, a construction which at any rate is based on a universal cross-linguistic pattern of information structuring inside asserted clauses.

It would seem that convergence would be most likely if the construction in the borrowing language is an optional one, which may be used to convey, for example, optional rhetorical effects. As I will show later on, I believe that several optional Basque constructions may have been extended to new contexts or uses under the influence of Romance. The first construction is the 'focus extraposition construction', which, as we saw above, is commonly used in all dialects of Basque when the focus constituent is

complex, but which is now increasingly being used with 'simpler' complement types. This construction results in SV,O type orders in Basque. A second such construction is the *emphatic assertion* construction, in which the finite verb (the bearer of the polarity) is the focus and thus precedes all (non-topicalized) rhematic complements. This construction typically requires that the complements' 'referents' (or 'ideas') be 'given' or easy to 'accommodate', but this requirement gives the speaker great leeway for interpretation. A third construction which seems to be used increasingly in Basque is formally indistinguishable from the former. In this construction the focus is the 'verbal idea', also instantiated by the verb. This construction is quasi-obligatory when the complement is very accessible, or even 'active', but 'optional' when it is less so, thus also leaving great leeway to the speaker to extend its use to new contexts.

As a final comment, we should keep in mind that convergence and (outright) transfer, as rightly emphasized by Weinreich and others, may sometimes indeed simply accelerate latent changes already underway in the target language or take advantage of 'weaknesses' or inherent tendencies in the system. At the very least we can say that a language is unlikely to borrow structures which are alien to its overall system, unless perhaps if they are highly motivated (iconic) and transparent.

1.8.5 Language contact and simplification

Recent studies of language change have proposed that one common type of change in contact situations, especially when the contact is 'unbalanced', involves the simplification of structures in the target language (cf., e.g., Dorian 1981, Gal 1984, Silva-Corvalán 1986, Trudgill 1992). It seems that language contact brings complexity into a

language by introducing extraneous elements into the language and upsetting the balance of the system. This complexity may cause interference and norm confusion and may induce reactive changes which aim to simplify the language system. These changes are also of the type that make the language easier to use for speakers who can dedicate limited cognitive resources to it.

Trudgill compares the situation of languages in contact to the situation in central dialects of a language, dialects which are known to be more innovative than the more conservative peripheral dialects. Not only is the rate of change faster in the central dialects but the *type* of change is also of a particular kind. The reason why central dialects are more innovative would seem to be related to the fact that they are in contact with the greatest number of other dialects. As to the type of changes, Trudgill argues that

In dialect contact generally, it seems that we most often encounter a process of koinéization, in which levelling and simplification both play a role. By levelling we mean the loss of minority, marked, or complex variants present in the dialect-mixture in favour of majority, unmarked, or simpler forms also present. By simplification is meant in this case especially the growth of new of interdialect forms (see Trudgill, 1986) that were not actually present in the initial mixture but developed out of interaction between forms that were present, where these interdialect forms are more regular than their predecessors. (Trudgill 1989:228-29)

In peripheral dialect and low-contact language situations, on the other hand, change is slower, and the type of change is different.¹⁶⁴ In these cases, change seems to go in the opposite direction and result in increased grammatical 'complexity' of the type that causes no problem for fluent native speakers (cf. Andersen 1988, Trudgill 1989).

Simplification is found at all levels of the language. In phonology it results in reduction of allomorphy and other (delicately balanced) phonological complexity. In the

realm of grammar it involves for instance the leveling of paradigms and, crucially for us, diminished reliance on inflections and synthetic structures in general, and greater reliance on analytic ones.¹⁶⁵ It may also include greater reliance on structures associated with the ‘pragmatic mode’ of communication to the detriment of tighter, more ‘grammatical’ (and less iconic) structures associated with the ‘grammatical’ mode (cf., e.g., Trudgill 1992:197).

Silva-Corvalán (1993) concludes from her experience with Spanish speakers in the Los Angeles area that “under conditions of intense contact and strong cultural pressure speakers of the secondary language *simplify* or *overgeneralize* grammatical elements”, although she argues that they “do not introduce elements which cause radical changes in the system of this language” (Silva-Corvalán 1993:20). According to Silva-Corvalán, simplification “involves the higher frequency of use of a form X in context Y (i.e. generalization) at the expense of a form Z, usually in competition with and semantically closely related to X, where both X and Z existed in the language prior to the initiation of simplification. Thus, X is an **expanding form**, while Z is a **shrinking/contracting form**” (Silva-Corvalán 1991:152). In other words, the variation and interference introduced into the system by language contact is resolved by getting rid of those forms which are hardest for non-fluent speakers to deal with.

Although simplificatory changes and a faster pace of change are typical of all contact situations, as we have seen, the existence of significant numbers of non-native speakers and imperfect speakers, or ‘semi-speakers’, may speed up and exacerbate these tendencies. Semi-speakers are those who either never learned the language fluently, or have lost some of their fluency through disuse and thus may have imperfect control of it.

Imperfect learning of a language as a child, or unlearning it later on, may have the same effect as the necessarily imperfect learning of adults (cf. Dorian 1977, Silva-Corvalán 1991, Watson 1989, Kuter 1989).¹⁶⁶ We will expand on this issue in the next two sections in which I will deal with language attrition, both at the individual level and at the speech community level.

Bilinguals in languages A and B who are imperfect speakers of B are more likely to introduce elements of A into B, thus adding ‘bad complexity’ into the language for all speakers, the kind of complexity which introduces norm confusion and random noise, not the kind of ‘good complexity’ which organizes the language system. The result may be an added impetus to simplify the language in directions which make it easier for imperfect speakers to use.

This process is very similar to the one found in pidgin and creole situations. In the case of pidgins, imperfect learners develop a communication system which in some ways is simple and relies on paratactic and analytic structures, but in other ways is complex, cumbersome, and limited in resources. Native speakers of this code then change it by introducing ‘good’ complexity (through grammaticalization) and by reducing ‘bad’ complexity and variation.

Stylistic shrinkage is also associated with non-fluent speakers. As Trudgill notes, “[n]on-native speakers have great difficulty with native speakers when the latter speak in restricted code (Bernstein 1971)” (Trudgill 1992:201). Such observation can be extended to non-fluent speakers and unbalanced bilinguals. Such situations seem to promote the preference for more regular, simple, analytic, iconic structures, as opposed to more subtle, complex, synthetic ones.

1.8.6 Language contact and language decay, attrition, obsolescence, and shift

What is known as ‘language decay’, ‘language attrition’, and (finally) ‘language shift’, which may also lead to ‘language death’, is found in ‘intense’ language contact situations. Language attrition seems to be an extreme case of normal changes which take place in regular language contact situations. Just like multilingual individuals may lose their fluency in a particular language through disuse, this phenomenon may be generalized to a whole community in which all speakers use other languages besides the ‘decaying’ one.

Unlike in normal contact situations, in some situations of intense, unbalanced contact the language degenerates without being regenerated, or repaired, by its speakers, and it degenerates to such an extent that it may cease to be a full-fledged language (cf. Campbell and Muntzel 1989).¹⁶⁷ The changes involve “[m]assive loans from the dominant language ..., radical simplification in inflectional morphology ... and generalization of periphrastic forms in the verb system ... are among the phenomena repeatedly observed in different areas. (Giacalone Ramat 1992:321; cf. Gal 1979; Dorian 1981; Dressler, Wodak, and Leodolter 1977; Dressler 1988, etc.).

Language shift involves “the gradual displacement of one language by another in the lives of the community members,” with or without an initial period of ‘attrition’ or decay, since shift may be too swift for decay to be noticeable other than in the last remaining speakers (Dorian 1982:44). Language shift “occurs most typically where there is a sharp difference in prestige and in levels of official support for the two (or more) languages concerned” (Dorian 1982:44). This, of course, is the phenomenon which has taken place in region after region in the Basque Country during the past few hundred

years, and language shift is probably the fate that awaits the remaining regions unless this trend is somehow reversed.

1.8.7 Imperfect speakers or ‘semi-speakers’

As I said earlier, Dorian has used the notion of “semi-speaker” (Dorian 1977), as well as the notion of “imperfect speaker” (Dorian 1982), to refer to speakers in language shift situations.¹⁶⁸ Referring to semi-speakers of East Sutherland Gaelic, Dorian says that “[m]ost of the imperfect speakers have always been exactly that: they lacked sufficient exposure or motivation to become fully proficient speakers ... and have never at any time spoken a grammatically normal form of the language” (Dorian 1982:51). And, she adds, “[l]ike the second language learner, the imperfect speaker of a dying language controls some other language much better than the home language he has either partially forgotten or never fully mastered (or both)” (Dorian 1982:56).

Andersen too argues that “there are many gradations between a fully-competent” individual and “a person whose competence in a language has eroded as a result of language attrition” and “who has lost all ability to use the language under study”, or one “who never became fully competent” in the first place (Andersen 1982:84).

This type of situation in which many speakers are less than fully competent has important consequences for the language. As Dorian argues, restricted or “reduced use of a language [is] accompanied by a reduction in form” (Dorian 1977:30). Restricted use may involve reduced overall use, or increasingly reduced number of contexts in which the language is spoken, and thus ‘stylistic shrinkage’. And of course, reduction in the forms available to a speaker itself leads to reduced use of the language since a speaker will use

the language she or he is more comfortable with, and which has more 'expressive power' for her or him, as long as the interlocutor knows the speaker's dominant language. In a situation in which everyone speaks the dominant language, as is the case in the Basque Country, the minority language has far fewer opportunities for being used.

It is very hard to know how many of the speakers of a declining language are actually less than proficient speakers. Dorian has argued that native speakers know who is a semi-speaker (Dorian 1977:30). This may not always be the case, however. Semi-speakers can often get by with what they do know and may seem to know more than they do.¹⁶⁶ As Mithun (1990) has remarked, "[s]peakers generally use the linguistic resources they have, ... and semi-speakers are notoriously adept at exploiting the structures they control; structures they do not control simply fail to appear in their speech" (Mithun 1990:3). Mithun also argues however that lexical deficiencies are harder to mask than grammatical ones: "Lexical choice accounts for many of the seemingly idiosyncratic grammatical patterns among speakers less fluent in the language" (Mithun 1990:19). And, of course, lexical deficiencies account for the large number of lexical borrowings that take place in language contact situations, which are much more unimpeded than changes affecting the grammatical system.

Andersen too argues that lack or loss of fluency can often be compensated for or 'covered up' by speakers:

A person who has lost a certain amount of his original competence in a language may be able to hold his own for a long time under many circumstances. Initially, the gaps in his linguistic performance may pass unnoticed and not interfere with communication or reflect unfavorably on him. If his competence in the language continues to fade, however, he will have to resort to a variety of compensatory strategies to keep his half of the conversation going. At some point he will begin to get negative feedback

from some of his conversational partners. This may be indirect if he fails to make himself understood or has to circumlocute, paraphrase, resort to borrowing, innovations, or transfer frequently to get his point across. He will begin to realize that his listener is being subjected to an additional burden by his compensatory strategies. In addition, some listeners may overtly comment on his inadequacies or question his use of certain words, forms, or expressions (probably to derive from them the intended meaning, rather than as a metalinguistic comment). The LA ['language attriter' or 'semi-speaker'] will eventually feel insecure in his use of the language. A natural consequence of this insecurity may be avoidance of situations where he would have to use the language or, at least, of topics he does not feel he can adequately deal with. Such a reaction to the negative feedback he receives can produce more of the very problem he is reacting to. (Andersen 1982:112)

The end result of such a progressively deteriorating communicative situation, and the increasing insecurity and even "feelings of inferiority" that comes along with it, is the abandonment of the language, i.e. language attrition, for an individual or for a community of individuals. When this process finally reaches the last speakers of a language, we speak of language death.

1.8.8 Language changes in an (individual or community) attrition situation

The types of 'interference' that affect the language of speakers in attrition situations and, more generally, declining languages, in which a majority of speakers have been 'affected' by these changes, have been studied in detail Dorian, Andersen, and others in recent years. Dorian notices that the source language's most obvious type of interference is lexical, and not grammatical. As for the grammatical changes, the most noticeable ones are those which come through simplification (cf. Dorian 1982).¹⁶⁹

Andersen (1982) has some very specific suggestions as to the exact forms which simplification, or reduction, takes, some of which are phonological, some morphological,

and some syntactic. With respect to syntactic reduction, Andersen argues that (LA = “language attriter” or “semi-speaker”, LC = fluent speaker):¹⁷⁰

- “An LA will use a smaller number of syntactic devices (transformations, constructions)” than a more ‘fluent’ speaker.
- “An LA will preserve and overuse syntactic constructions that more transparently reflect the underlying semantic and syntactic relations.” In other words, more iconic structures are preferred over less iconic ones.
- “Where there is more than one possible surface structure for a given underlying relation (ex. negation), the LA will tend to collapse the different surface structures into one. (This will be evidenced by a smaller variety of surface structures for the LA as compared to an LC.)”
- “Where elimination of a transformation (or a syntactic construction of a given type) would result in informational loss, an LA will (1) tend to preserve that transformation in his use of the language or (2) eliminate the transformation but compensate for its elimination in some other way (see “Compensatory Strategies” below)” (Andersen 1982:99).

As to the simplificatory ‘compensatory strategies’, Andersen suggests two general mechanisms are found:

- “GENERAL STRATEGY I: Whenever possible use free morphemes (or at least syllabic morphemes) strung together linearly in the most transparent fashion to express your meaning” (Andersen 1982:100)

In other words, transparent analytic structures are to favored over less self-evident synthetic ones. This strategy goes back to Slobin's (1973, 1977) "charge" to "be clear" in language acquisition, which favors a one-to-one correspondence between form and meaning over complex ('portmanteau') morphemes with complex morphophonemic alternation. Elsewhere, Andersen refers to this as the "One-to-One Principle" (cf. Andersen 1984:79, 1989:388),¹⁷¹ also known as the principle of "semantic transparency" (cf., e.g., Silva-Corvalán 1991:154). This principle is also quite "evident in the formation of pidgin and creole languages" (Andersen 1982:101-102).

We can interpret this strategy also in terms of the two different modes of communication proposed by Ochs and Givón: the pragmatic mode vs. the grammatical mode. It seems that in a situation in which the grammatical system falters, the pragmatic mode will become more prevalent since it is more iconic and 'basic'.

The second general compensatory, and simplificatory, mechanism that Andersen says is found in attrition situations is the following:

- "GENERAL STRATEGY II: Whenever there are different devices to express the same basic meaning, use only one of these devices" (Andersen 1982:102).

Andersen argues that this principle also applies across the different languages in the bilingual mind, as first described by Gumperz and Wilson (1971).

One further simplificatory compensatory strategy that Andersen mentions involves the use of "paraphrasing and circumlocution", although this would seem to be part of the same general strategy towards analyticity and away from syntheticity.

In addition to these simplificatory patterns, the language attrition situation also results in lexical and grammatical transfer and innovations. As I mentioned earlier, grammatical (morpho-syntactic) transfer, it is argued, occurs primarily, or only, when the pattern already exists in some way in the target language, as argued by Weinreich (1953) (cf. Andersen 1982:108). Thus Anderson finds that:

- “An LA will rely on transfer as a compensatory strategy more often in those cases where forces internal to the language being transferred to (i.e., his weaker language) would also tend to produce the same construction through overgeneralization (General Strategy II)” (Andersen 1982:109).

Andersen refers to this as the “Transfer to Somewhere” hypothesis (cf. Andersen 1983).

Transfer and innovation are also subject to, or guided by, the General Strategy I:

- “An LA will innovate morphological and/or syntactic constructions in his weaker language by using as a model similar constructions in his stronger language that utilize free, regular, and invariant morphemes in that construction” (Andersen 1982:109)

It has been noted that the processes which take place in language attrition, both for the individual and for a ‘language’, mirror developments which take place in both ontogeny, i.e. in child language acquisition, and phylogeny, as in pidginization and “creolization in reverse” (Trudgill 1976-77; cf. Romaine 1989).¹⁷² In the case of ontogeny, it has been argued that children first follow assumptions such as the one expressed by the “semantic transparency” principle until the grammatical constructions of the language, with all their idiosyncrasies, are gradually assimilated (cf. Slobin 1973,

1977, 1985). The child starts with iconic pragmatic structures and eventually assimilates the grammatical structures of the language. And if those do not exist, as is the case with pidgins, the child creates them from the pragmatic structures and lexical items themselves.

Pidginization is equivalent to adult learning of a language and produces much simplification along iconic lines and introduces great variation and uncertainty into the system. Creolization is equivalent to language change in a contact situation in that native learners try to make sense of the newly introduced variation by cleaning up the mess and do so by simplifying and regularizing the system. And children are able to regularize the language in these situations unimpeded by the brake usually provided by adults, given the prevailing high degree of variation and uncertainty (cf. Romaine 1989:372).¹⁷³

In all the contexts which I have mentioned (central dialects, languages in contact, pidgins-creoles) the introduction of variation and iconic pre-grammatical analytic structures into the system provides an impetus to settle on more 'basic' or 'transparent' *grammatical* structures.

These findings are potentially quite relevant to the study of present day Basque, since its speakers are now exposed to other dialects and varieties of the language and to increased variation (including that introduced by non-native speakers). This variation and uncertainty may have an influence on children's language, especially for children who do not have consistent models in the family situation, such as those who learn Basque at school. The fact that Basque speakers also know a second language and that perhaps most of them use that second language more often than they do Basque are additional

relevant factors. In the next section we will look at the effects of such simplification on word order.

1.8.9 Further thoughts on the shift from SOV to SVO

Some linguists and psychologists have discussed the supposed ‘basicness’ and ‘naturalness’ of SVO order. Danchev (1991) for instance believes that “SVO order should be regarded as simpler, less marked and more iconic than SOV (and other word orders)” (Danchev 1991:116). Givón (1979) argues that although SOV order “may have been the most suitable word-order at the stage of mono-propositional discourse ... somehow it seems that either SVO or V-first are more compatible with topic-oriented, multipropositional discourse” (Givón 1979:309).

It is very interesting in this regard that change from SOV to SVO is commonly attested, whereas the opposite is much more rare. Vennemann (1975), Givón (1977), Stockwell (1977) and others have noticed this ‘natural’ drift. This is what Givón calls “the SOV mystery” (Givón 1979:Chapter 7). It seems that the SOV > SVO shift can happen relatively quickly under a certain set of circumstances, such as those brought about by language contact, whereas a SVO > SOV shift takes much longer and perhaps can only take place in a situation of relative isolation.

The claim that SVO order is simpler, more iconic, and more natural, of course, is not new. It goes back to some eighteenth century grammarians who viewed SVO order as the language of thought (e.g. Gabriel Girard, quoted in Ramat 1995).¹⁷⁴ Even late 20th century psychologists have been known to make very similar claims (e.g. Osgood and Tanz 1976).¹⁷⁵ Nonetheless, this claim seems to me to be totally unsubstantiated. The

fact that about half of the languages of the world are SOV is enough to call this claim into question. It may be more 'natural' or 'better motivated' under some circumstances, such as when the language lacks inflections and agreement, but I cannot see how the claim could be generalized beyond that.

SVO order has been said to be the order found in obsolescent languages and in pidgin and creole languages, presumably a result of the simplificatory processes mentioned above. Romaine (1989), however, argues that it is not true that SVO order is an automatic consequence of obsolescent languages at all, and that it is not even a characteristic of all pidgins and creoles.¹⁷⁶ Unfortunately we lack enough information and examples about word order change in this type of situation. It does seem to be the case that at least sometimes language contact and simplification bring about word order change to SVO order as well as word order rigidification.

It is not at all obvious that SVO order is the preferred order in acquisition either, as has sometimes been claimed, at least not in all circumstances. Slobin (1982/1978) and Slobin and Bever (1982) for instance have shown that when inflections are consistent, regular and general in an SOV language, as is the case in Turkish for instance, children learn very early to attend to them and do not rely on word order for coding grammatical relations (or semantic roles). It is only when inflections are inconsistent and irregular, as for instance in Serbo-Croatian, that children are more likely to rely on word order to code grammatical (or semantic or pragmatic) relations. Later they learn to rely on inflections when they are available for that purpose, and revert to word order cues when inflections are not available.

This discussion, however, contains what would seem to be a contradiction. On one hand, we have seen that languages undergoing stressful contact, particularly erstwhile SOV languages, may become SVO languages, and order may even be rigidified to some extent. I have also argued, however, that language contact results in increased use of iconic pragmatic structures and, as we saw earlier, SOV languages are said to correlate with pragmatic word-order and topic-prominence. How can this be? I will attempt to answer this question in the next section.

1.8.10 A seeming contradiction: Pragmatic word order and rigid SVO order

I believe that increased importance of pragmatic word order, and in particular the importance of the pragmatic topic-comment configuration, is indeed part of the attrition process, as it is also part of the first language learning process and of the pidginization process. On the other hand, attrition may result in certain grammatical changes, such as the loss (or loss of transparency) of case marking and/or verbal coreferencing morphology, which may result in the need for the use of word order to code grammatical relations. These changes would motivate word order rigidification (grammaticization) and SVO order, especially if the language is not rigidly verb-final and already allows postverbal elements.

In addition, language contact may introduce variation in an area of word order which is more complex, less clearly iconic, and more delicately balanced, than the one relating to the coding of overt topics by placing them clause initial or dislocated position. I am referring to variation in the correlation between word order and the pragmatic role focus (something which I will discuss at length in Chapter 3). As I said earlier, unlike in

the case of topics, there isn't a single clausal position associated with focus constituents cross-linguistically. Rheme-initial (preverbal), rheme-second (postverbal), and rheme final or extraposed, are all possibilities which are found cross-linguistically and most languages use more than one of these positions, or even all three. (Clause final position, as we will see, seems to result from reanalysis of focus extraposition.)

As we will see later on, so-called OV languages are most consistent in this respect, using a single position, rheme-initial and preverbal, for focus constituents of any kind: subjects, objects, modifiers, or even the verb. These languages may also 'extrapose' focus constituents under some circumstances. So-called VO languages, on the other hand typically use two focus positions: postverbal (under 'normal' circumstances), and preverbal (for verbs, very salient objects, and even regularly for subjects, as is the case in English). VO languages, like OV ones, also use the strategy of focus extraposition, though seemingly to a greater extent.

The variation that language contact may introduce in the area of focus coding is of several types. To begin with, the preferences for focus positioning may change under the influence of the source language, as in Prince's Yiddish English example above. Focus extraposition also may become more common in an OV language which uses primarily preverbal position for focus constituents (and never postverbal position), especially if the source language is VO, since this strategy is shared by both language types (although it is used differently). The introduction of variation, and thus inconsistency, into the system may very well result into a simplificatory backlash, which could entail the elimination, or simplification, of the available choices.

Language contact may also introduce variation into the choice of focus constituent, in particular situations. As I suggested above, focus choice in some contexts is somewhat arbitrary and subject to defaults. In Chapter 3 I will argue that in Basque, in cases in which the referent of a rhematic (non-topicalized) complement is 'new', it is typically chosen by default as the focus of an assertion, even if the verb is also 'new'. When the complement's referent is very 'given' ('active'), and if there is no better candidate for focus, then the verb is typically the focus, also by default (unless the complement is contrastive or emphatic). But languages differ as to how 'accessible' a referent has to be for it to 'give up' its default focus role (or risk being interpreted as contrastive). Here too, contact with a language which makes somewhat different choices may introduce variation which may result in simplificatory measures of the sort mentioned above.

To summarize, I believe that word order rigidification, and in particular rigidification as SVO, may arise from the loss of inflections or from the interference caused by another language with a different focus positioning system in a protracted language contact situations. The reason for this is that a somewhat rigid SVO order is able to code grammatical relations most efficiently. It also allows for three positions for focus constituents, which may result in something close to *in situ* focusing, as in English, or different positions for foci with different degrees of salience. Notice also that this order has the added value of respecting the pragmatic topic-comment order.

1.8.11 The mechanism of change

As I mentioned earlier, easing of the cognitive load in language contact situations has often cited as the explanation, or final motivation, for both transfer/convergence and simplification. Even if this is so, we still need to explain the actual mechanics of how the changes occur, the ‘immediate explanation’ for those changes. I believe that the actual mechanism for such changes, just like for any other diachronic changes, must be understood within a social theory of linguistic change, as proposed by Weinreich *et al.* 1968. As Henning Andersen (1989) explains, there is a

usual course of events in diachronic developments: an innovation arises, the new entity (of usage or of grammar) cooccurs or coexists for some time with the corresponding traditional one and is then eventually established as traditional itself—if it does not go out of use, yielding to the traditional one or to a new innovation. To understand any such particular development it is necessary to understand the reason for the initial innovation, why it was accepted, adopted or acquired, or duplicated by others, and finally, why it was generalized or given up in competition with alternative linguistic entities. (Andersen 1989:13)

Language contact explains how an innovation may be introduced into a language, namely (1) by exposure to, and convergence with, certain elements in the other language, or (2) by reaching for ‘pragmatic mode’ and other iconic and transparent structures in a context of norm-confusion or diminished competence. Innovations such as these may actually be introduced into any language, but in a ‘contact’ situation these innovations will be more common, and they will be less likely to be rejected by other speakers.

Furthermore, these are types of changes which, because of their iconicity and simplicity, do not in any serious way impede understanding, especially if all speakers of the weaker language also speak the stronger language which it is the source of, or model,

for some of the changes. In addition, innovations which bring cognitive benefits to a significant number of speakers will be more likely to be repeated and eventually generalized.

The special circumstances associated with intensive and unbalanced language contact don't guarantee that innovations will catch on and spread, but they increase the chances that they will. All these changes indeed do not need to take place in any large scale. In fact the language be abandoned (or even die off in a particular location) before they take place, as the language becomes replaced by the dominant language within a relatively short period of time, say a few generations.

As this picture shows, we are not able to predict when and where a certain change will take place. In that sense we cannot say that a certain change is *explained* by language contact. There are many factors that still remain to be clarified so that we may increase the predictive power of this theory. How much contact and of what kind does it take for certain innovations to surface with significant frequency? What kinds of contact situations make it more likely that an innovation will spread? How long does it take for an innovation to spread? How exactly do 'transfer' and 'simplification' ease the cognitive load for bilingual speakers? These and others are questions which we still do not have clear answers for. More case studies, such as the present one, are needed.

In the following chapters I will show that the constituent order in Basque seems to be reacting to some of the pressures introduced by the unbalanced language contact situation that it finds itself in. It seems that these pressures may be causing Basque to change in ways which make it more like its neighboring languages, at least for a significant number of speakers. The changes, however, have not become generalized to

all speakers. They also seem to be primarily of the convergence variety, and not direct transfer of constructions, as the theory predicts.

NOTES TO CHAPTER 1

1. In Comrie's words: "I consider the area of 'language universals and linguistic typology' to be a single area, and I will therefore use the terms 'universals' and 'typology' more or less interchangeably to refer to this general area" (Comrie 1993:3). And as Shibatani and Bynon 1995 put it, "[i]nitial reactions that while typology seeks differences among languages, universals research seeks what is universal, i.e. common to all human languages, should give way to a more informed perspective once it is recalled that typology in the first place seeks exhaustive classification of human languages ... in terms of variant features. Secondly, the current universals research seeks not only what is common to all human languages but also the range, and the permitted variation within the range of possible natural languages. In other words, the two fields have the same goal of characterizing human language and understanding it" (Shibatani and Bynon 1995:19).

2. For issues related to sampling see, e.g., Comrie 1989/1981, 1993; Dryer 1989b; Croft 1990, 1995; Perkins 1989; Bybee, Perkins and Pagliuca 1994.

3. Comrie and Matthews 1990 are more optimistic when they say that "[t]he vast literature stimulated by Greenberg's seminal paper on constituent order typology (Greenberg 1966) has served, in our opinion, in general to validate Greenberg's intuition that such notions as subject, object, verb (predicate) are sufficiently comparable crosslinguistically, both formally and functionally, to enable a valid typologization of clause order to be undertaken; indeed, later work has served to refine our understanding of the notions used by Greenberg rather than to undermine his original work." (Comrie and Matthews 1990:43). Still, they admit there are problems: "Sometimes, however, it will turn out that apparently comparable phenomena are, on closer analysis, rather different crosslinguistically", e.g. "the crosslinguistic equivalents of Tough Movement", e.g. This book is easy to read (Comrie and Matthews 1990:43), which they argue in their paper have very different structures in German, English in Serbo-Croatian, despite surface semantic similarities.

4. Dryer's quote refers actually more specifically to 'grammatical relations' such as 'subject'. He adds that "It may be convenient to use labels like 'subject' for grammatical relations in different languages that are similar. But if one then looks at languages where things work somewhat differently, as in Philippine languages, it is a mistake, on my view, to try to ask whether these languages 'have subjects' and if so what are subjects in these languages. It is perfectly clear what the grammatical relations in Philippine languages are. It may not be clear what, if anything, we want to call subjects in these languages, but on the view I have presented here, that question is an empty terminological question. The Philippine languages are interesting in that the functional principles grammaticized differently in these languages so that there isn't the same degree of isomorphism with other languages that we more often find. And I believe that these languages can contribute to our understanding of the functional principles that underlie the more common pattern" (Dryer 1994).

5. In the words of Hopper and Thompson (1993): “grammars are shaped by patterns in the way people talk ... Regularities of all kinds arise because of certain strategies people habitually use in negotiating what they have to say with their hearers ... Many of these regularities, specifically those that have become temporarily stabilized in the language, are of the kind that linguists recognize and identify as ‘grammatical’ ones (Hopper [1987a]). ... As Du Bois (1985:363) has put it, ‘grammars code best what speakers do most’” (Hopper and Thompson 1993:357-358).

Grammaticalization is often used solely to refer to the process by which lexical items become grammatical items. The name has also been used for the process by which particular constructions are created, since the process of creating grammatical morphemes goes hand in hand with the creation of the constructions in which they are used. Conceivably, however, the two processes are separable and one could speak of the process of ‘constructionalization’ to refer to the creation of syntactic patterns or constructions.

6. In Aitchison’s words: “Given certain recurring communicative requirements, and some fairly general assumptions about language, one can sometimes see why particular options are preferred, and others passed over. In this way, one can not only map out a number of preferred pathways for language, but might also find out that some apparent ‘constraints’ are simply low probability outcomes. ‘No way out’ signs on a spaghetti junction may be rare, but only a small proportion of possible exit routes might be selected” (Aitchison 1989:152). That is, why a language takes a particular path in its development is due to a variety of reasons: “A number of principles combined to account for the pathways taken, principles based jointly on general linguistic capabilities, cognitive abilities, and communicative needs. The route taken is therefore the result of the rediscovery of viable options, rather than the effect of an inevitable bioprogram. At the spaghetti junctions of language, few exits are truly closed. However, a number of converging factors lead speakers to take certain recurrent routes. An overall aim in future research, then, must be to predict and explain the preferred pathways of language evolution” (Aitchison 1989:170).

7. In Croft’s terms: “Perhaps the most important external motivation that has been invoked to explain the typology of grammatical structure is iconic motivation (iconicity), the principle that the structure of language should, as closely as possible, reflect the structure of experience, that is, the structure of what is being expressed by language. By this means, the theory of grammatical structure can be based on a theory of semantic structure. ... Iconic motivation has been explored in depth in two simple cases, isomorphism and conceptual distance, particularly by Haiman ([1974a], 1976, 1978a, 1978b, 1980, 1983, 1985[b])” (Croft 1995:129).

8. The idea of competing motivations was perhaps first considered at length by Du Bois 1985, though it goes back a long way in functional studies. As reported by Osgood and Tanz, “[w]hen one considers *statistical* universals, it must be kept in mind that ‘...we

always have a number of tendencies at work at the same time' [Greenberg, 6/21/74 letter to C.E.O[sgood].] and these competing factors are likely to produce counterexamples..." (Osgood and Tanz 1976:539). For a related perspective, cf. the psycholinguistic 'competition model' of Bates *et al.* 1982; Bates and MacWhinney 1987, 1989; MacWhinney 1987, 1989a, 1989b; Gibson 1992. For a formalist critique of 'competing motivations', cf. Newmeyer 1994. Cf. also Croft 1995:134-5.

9. Generative grammar, for instance, has reacted to this challenge by reducing the scope of their study to the more predictable phenomena which one can more-or-less safely pretend is categorical and fully regular, and by relegating all other linguistic phenomena, perhaps the vast majority of linguistic phenomena to an ill-defined periphery.

10. Bolinger, for instance, made the following observation 35 years ago: "At present we have no way of telling the extent to which a sentence like *I went home* is the result of invention, and the extent to which it is a result of repetition, countless speakers before us having already said it and transmitted it to us in toto. Is grammar something where speakers 'produce' (i.e. originate) constructions, or where they 'reach for' them, from a preestablished inventory, when the occasion presents itself? If the latter, then the MATCHING technique of traditional grammar is the better picture—from this point of view, constructions are not produced one from another or from a stock of abstract components, but filed side by side, and their interrelationships are not derivative but mnemonic—it is easier to reach for something that has been stored in an appropriate place. Probably grammar is both of these things, but meanwhile the transformationist cannot afford to slight the spectrum of utterances which are first of all the raw material of his generalizations and last of all the test of their accuracy" (Bolinger 1961a:381).

Antila said something very similar: "Memory or brain storage is on a much more extravagant scale than we would like to think; even the most 'obvious' cases can be stored separately" (Antila 1972:349, cited in Bolinger 1979:97). Ladefoged too: "The indications from neurophysiology and psychology are that, instead of storing a small number of primitives and organizing them in terms of a [relatively] large number of rules, we store a large number of complex items which we manipulate with comparatively simple operations" (Ladefoged 1972:282, cited in Bolinger 1979:96). And, as Corrigan and Lima argue, summarizing much recent work on connectionist patterning, "[r]ule-like behavior can emerge without the mental representation of symbolic rules, as in the connectionist models described..." (Corrigan & Lima 1994:xiv). For 'connectionist' approaches to linguistics which reject the list-rule dichotomy, cf. Lakoff 1987, 1991; MacWhinney 1989a; Goldsmith 1994; Lima, Corrigan and Iverson, eds., 1994. For 'analogical' approaches, cf. Derwing and Skousen 1989, 1994; Skousen 1990, 1992. Cf. also Derwing 1977, Aitchison 1992.

11. In recent years there have been attempts in the generativist school, notably Pinker 1991, to incorporate the criticisms of the rule-&-rote approach and to supplement the rule component of the grammar with an associative "component", not unlike the one proposed

by connectionist models, which would handle semi-regularity of the type displayed by irregular past tense items in English. Pinker and Prince call the earlier approach the “traditional rote-&-rule theory”, cf. Pinker and Prince 1988, 1994. Pinker has proposed a modified version of the traditional theory, cf. Pinker 1991; Daugherty and Seidenberg 1994. This theory does for syntax what is by now commonplace in most phonological theories, namely to deal with regularities in the lexicon with redundancy-rule-like objects: “The conclusion we draw is that generative theories are fundamentally correct in their characterization of productive rules and structures, but deficient in the way they characterize memory of less predictable material, which must be associative and dynamic, somewhat as connectionism portrays it. ... Regulars are computed by an implementation of a classic symbolic rule of grammar, which concatenates an affix with a variable that stands for the stem. Irregulars are memorized pairs of words, but the linkages between the pair members are stored in an associative memory structure with certain connectionist-like properties (cf. Bybee 1988[b])” (Pinker and Prince 1994:326). “The new theory, like associationist theories, proposes that irregular generalizations are based on memory-driven analogies, but it is not clear what this means in mechanistic terms” (Pinker and Prince 1994:333).

12. Van Valin has in mind a number of reductionisms often incurred in linguistic analysis, such as “the reduction of the scope of linguistics from language to grammar”, reductionism of rule types, as well as the reduction of syntax to semantics (Van Valin 1992:199). “Over the past three decades there has been a pervasive tendency in linguistic theorizing, especially in syntactic theory, towards theoretical reductionism; that is, theorists have attempted either to reduce complex phenomena to a single unitary phenomenon or to explain phenomena in terms of a small number of unifying principles of some kind. Reduction of complex phenomena or principles to simpler, better understood ones is a natural part of scientific explanation, but the danger in this course lies in ‘over-reduction’ to the point where the phenomena or the principles become distorted; determining the correct reduction is an important consideration in explanatory scientific theories” (*ibid.*).

13. Although solutions to many ‘formal’ issues from a functional perspective, including for example ‘subjacency’ or ‘binding conditions’, are still only partial and preliminary, enough success has been achieved to warrant treating these phenomena as purely formal issues. For functional work on ‘subjacency’ and ‘binding conditions’, for example, cf. Erteschik-Shir and Lappin 1979; Van Valin 1986, 1994; Goldsmith 1985; Lakoff 1986; Deane 1988, 1991a, 1991b, 1992; Levinson 1987, 1991.

14. As Frazier has humorously put it, “one of the most intriguing questions in psycholinguistics is whether any of the properties of specific natural languages, or any principles of universal grammar, can be explained by appeal to the basic structure of mode of operation of the human sentence processor that determine what linguistic structures could not be processed by humans in real time. Jim McCawley has remarked that when a linguist says the word *explanation* you had better put your hand on your wallet. May people think that when a psycholinguist says the word *explanation* you had better keep an eye on your watch, too” (Frazier 1985:129-130).

15. Notice that there are exceptions to this pattern. Thus the term *apple cider* is used in English for a type of ‘apple juice’ made with a particular method of extraction the juice. The same thing is true with the vehicle example, where, for instance, the term *wrecker* can be used for ‘tow trucks under certain circumstances. (*Wrecker*, “Also called < tow car, tow truck.> a vehicle equipped with a mechanical apparatus for hoisting and pulling, used to tow wrecked, disabled, or stalled automobiles” (*Random House Webster’s Electronic Dictionary and Thesaurus, College Edition, v. 1.0.*)

16. This assumes that the speaker knows that orange juice comes from oranges, something which may not be the case for some speakers, such as children, who may be exposed to the juice before they are exposed to oranges. The equivalent of ‘external motivations’ in phonology would be phonetics, and the equivalent of ‘internal motivations’ would be the influence of the sound system. Ohala 1989, however, is highly skeptical of internal explanations: “Symmetry undoubtedly pleases the aesthetic sense of the linguist but it is not clear what its value is for the native speaker for whom the sole function of the sound inventory is its capacity to carry signals, not to serve as an object of beauty. All human languages manifest asymmetry or ‘disequilibrium’ in some part of their phonology but seem, nevertheless, to function adequately for communication” (Ohala 1989:192). But he does admit that they may play a role in language change: “Still, ‘hole filling’ sound changes and the fact that some sound changes seem to affect whole classes of sounds, e.g. Grimm’s Law, suggest that somehow the phonological system of a language does play a role (*ibid.*). Still, “There is not need to appeal to ill-defined notions such as ‘system pressure’, ‘pattern symmetry’, ‘equilibrium’, and the like, nor to maintain that the language is any better or any more ‘fit’ as a result of these sound changes” (*ibid.*).

17. As Bever has argued, for instance, description is always primary to formalization and explanation: “The description of what language is must proceed independent of particular functional explanations of why it is that way. At a practical level, we cannot know what to explain until we have isolated and described it” (Bever 1975:585).

18. Just because the goal is not clear, it doesn’t mean that we can give up without trying. In Comrie’s words, “it is important to explain language universals, or at least to try to do so. Of course, as in other domains of human inquiry, there are many instances where the level of our empirical investigation of a topic is very different from the level of our understanding of the motivation for the generalizations we have uncovered. ...

Conversely, there are instances where we have plausible 'explanations' but where the phenomena to be explained are not so clear! ... Thus empirical investigation and explanation will often be out of synchronization, but it nonetheless behooves us to try to advance on both fronts" (Comrie 1993:10-11).

19. In the end some phenomena may very well be 'arbitrary': "It should be noted that this search for links between linguistic and linguistic generalizations by no means excludes the possibility that certain explanations may amount essentially to saying that language and cognition are structured in such and such a way simply because human beings are structured in such and such a way... Thus, this approach is not necessarily incompatible with the generativists' belief in the innateness of at least a substantial subset of language universals; although equally it allows for universals that do not fall under innateness" (Comrie 1993:11).

20. In Keenan's words: "Becoming more competent in one's language involves increasing one's knowledge of the potential range of structures (e.g. morphosyntactic, discourse) available for use and increasing one's ability to use them. In this view, communicative strategies characteristic of any one state are not replaced. Rather, they are retained, to be relied upon under certain communicative conditions" (Keenan Ochs 1979:52). This is a "view of language development as a broadening of knowledge of the language's potential expressive power" (*ibid.*). "The extent to which earlier patterns continue to remain prominent (i.e., are used) depend [*sic*] upon the linguistic structures under consideration and the developmental period observed" (*ibid.*).

21. According to Givón, "[o]ne may view the rise of rigid word-order, condensed—tight, subordinate—syntactic constructions, rich grammatical morphology, and tight syntactic constraints-rules, as the rise of *automatic processing* of speech. The pragmatic mode is analytic and slow, and it has a high degree—for all practical purposes 100%—of communicative fidelity, and one-to-one correlation between code and message [?]. But it is also a cumbersome mode, requiring a high degree of face-to-face interaction and often limited to topics, tasks, and context that are not too far removed from here, now and us. In the syntactic mode one may view the more evolved coding devices as the *tools* of automatic processing. But while one gains speed, one loses resolution..." (Givón 1979:233). This is the same conclusion reached by, e.g., Bybee *et al.* 1994, when they say that "the grammatical framework, consisting of both word order regularities and grammatical morphemes, provides a means of facilitating production through automation. The obligatory grams and the rule-like nature of word order patterns are automatically supplied so that the speaker's attention may be directed to the propositional content of the utterance. The advantages of automation can explain why categories or structures become obligatory despite the fact that they are not always strictly necessary for communication. Automation thus leads to the semantic reduction that propels the inexorable cycling of grammatical material" (Bybee *et al.* 1994:299-300).

22. Jackendoff says that “if you look back at *Syntactic Structures*, Chomsky said, ‘Semantics is semi-systematically connected with syntax - systematically enough that we want to account for it, but not systematically enough that we can use it as a key to determine how the syntax works. We have to do the syntax autonomously.’ His program was to show we have to (and can) do the syntax autonomously. And he never really worried about a systematic connection to semantics. I think it was really Katz and Postal who forced him to it” (Ray Jackendoff, in conversation with John Goldsmith; in Huck and Goldsmith 1995:98-99).

23. For a history of modern linguistics in the U.S., cf., e.g., Harris 1993, Croft 1993, Matthews 1993. For a critique of Generative Semantics, cf., e.g., Lakoff 1987. Croft 1993 argues that Chomsky’s approach to universals was “a reaction against the behavioristic psychology adopted by American structuralism” (Croft 1993:17). Modern typology, on the other hand, which started with the work of Greenberg, “represents a reaction against anthropological relativism (American linguistics grew out of anthropological research” (*ibid.*). In Europe the functionalist and typological outlook goes back, uninterrupted, to the Prague School and German linguists of the turn of the century (cf. Croft 1993:15-16), even through the structuralist period, where *system internal* considerations (also called ‘functional’) and explanation were paramount.

24. So-called optional and stylistic rules are central to the study of word order. The whole idea of optional or stylistic rules is rather odd, however. As Thompson and Mulac argue, “[t]he work of Dwight Bolinger challenged the idea of ‘optional’ ‘rules’ almost as soon as this concept appeared in the writings of grammarians oriented towards transformational descriptions. Thanks in large part to his efforts, it is now widely accepted that what had been thought of as ‘stylistic options’ are better seen as genuine choices of the speaker based on such factors as attitude, emotional stance, information flow, and discourse structure” (Thompson and Mulac 1991b:238). See also Lavandera 1978; Romaine 1981; Traugott and Romaine 1985:10-12. For a more up to date version of the status of ‘stylistic rules’ in formal grammar, cf., e.g., Rochemont and Culicover 1990.

25. Many, if not all, of the issues involved were discussed in a very interesting exchange in the COGLING discussion list in April 1994, involving Elizabeth Bates, George Lakoff, Dan Slobin, and others.

26. This indeterminacy leaves some leeway for misinterpretation, albeit often of a minor kind which doesn’t impair communication, but which may lead to language change through minor cumulative changes.

27. Antinucci, Duranti and Gebert 1979 is another study which concentrate on the processing difficulties associated with long relative clauses. Frequency is crucial for determining what is basic in language and what becomes grammaticalized. In Givón’s words: “frequency of behavior is of paramount importance, since grammatical

competence is believed to be both reflected in and shaped by the frequency of experience. Functionally oriented typologists would thus tend to suspect that the typological difference between VSO, VOS and SVO, therefore also between OVS and SOV, is relatively shallow at the level that really counts—actual speech production and comprehension” (Givón 1995:194-95). This is the case since we find that in actual discourse, close to 90% of transitive clauses, or “fully 86% of Sacapultec transitive clauses [cf. Du Bois 1987] and 93% of Hebrew transitive clauses ... appear without the subject NP, thus with the verb and the object adjacent to each other” (Givón 1995:194).

28. More generally, causation is never an easy thing to prove. As Ohala aptly put it, “As we know from centuries of philosophical discussion, causation is a complicated matter” (Ohala 1989:174). However, not being able to prove causation doesn’t mean that explanations cannot be proposed: “there are those who believe that unless a complete, nomological account can be given of the causes of sound change that is, including why a particular sound change happened in a specific language at a specific time, the whole enterprise is bankrupt (Dinnsen 1980; Lass 1980). I have answered these critics (Ohala 1987) by pointing out that completely nomological (exceptionless, lawful) accounts of natural phenomena do not exist in any scientific domain; all sciences resort ultimately to probabilistic explanations. That is, those causes they have under their control they deal with; those they do not, they handle in a probabilistic way” (Ohala 1989:174-175).

29. The term ‘self-organized’ is taken from Lindblom, MacNeilage and Studdert-Kennedy 1984, who claim we find that type of self-organization in phonology. As we mentioned above, other phonologists, such as Ohala (1989) are more skeptical about attributing too much importance to systemic forces.

30. In Bolinger’s words: “For a long time now linguists have been reveling in Theory with a capital T. If you assume that language is a system *où tout se tient*—where everything hangs together—then it follows that a connecting principle is at work, and the linguist’s job is to construct a one-piece model to account for everything. It can be a piece with many parts and subparts, but everything has to mesh. That has been the overriding aim for the past fifteen years. But more and more evidence is turning up that this view of language cannot be maintained without excluding altogether too much of what language is supposed to be about. In place of a monolithic homogeneity, we are finding homogeneity within heterogeneity. Language may be an edifice where everything hangs together, but it has more patching and gluing about it than architectonics. Not every monad carries a microcosm of the universe inside; a brick can crumble here and a termite can nibble there without setting off tremors from cellar to attic. I want to suggest that language is a structure, but in some ways a jerrybuilt structure. That it can be described not just as homogeneous *and* tightly organized, but in certain of its aspects as heterogeneous *but* tightly organized.

Specifically what I want to challenge is the prevailing reductionism—the analysis of syntax and phonology into determinate rules, of words into determinate morphemes, and

of meanings into determinate features. I want to take an idiomatic rather than an analytic view and argue that analyzability always goes along with it opposite at whatever level, and that our language does not expect us to build everything starting with lumber, nails, and blueprint. Instead it provides us with an incredibly large number of prefabs, which have the magical property of persisting even when we knock some of them apart and put them together in unpredictable ways” (Bolinger 1979:95-96).

31. For a review of some of these tests see, e.g., Givón 1995:§5.4. The sentence that I have in mind here has a topic-comment, or subject-predicate, structure. I do not have in mind a sentence in which the subject (*John*) is contrastive and rhematic (as in the answer to the question: *Who should have given that book to Mary?*). We will discuss such cases in later chapters.

32. Dixon doesn't follow the modern VP convention for instance: “I use ‘verb phrase’ (VP) to describe a string of verbs (an alternative use, which is not followed here, is to say that the VP also includes an object NP). A VP must include a main verb as head. This may optionally be preceded by auxiliary verbs...” (Dixon 1991:22).

33. The division of the sentence into subject and predicate goes back all the way to Aristotle (cf., e.g., Givón 1995, Chapter 5). In general *all* “intermediate nodes” have been problematic, not just the VP node, as Givón rightly argues (Givón 1995:180).

34. Approaches which incorporate semantic and functional information into their descriptions have often been guilty of the same oversimplification. In some theories, brackets (or boxes) are used to represent such structures, assuming their existence, and in others they are not, presuming their non-existence. The idea that there might be *different types of brackets* has not been put forth, as far as I know.

35. The configurationality parameter was introduced by Hale, cf. Hale 1982, 1983, 1989. For other sources on this ‘parameter’, cf. Abraham and de Meij 1986; Farkas 1984; Kiss 1987; Marác and Muysken 1989; McCloskey 1983; Moravcsik 1988. For the value for this parameter in Basque, cf. Rebuschi 1985a, 1985b, 1986, 1989; Oyharçabal 1989; Levin 1989. Not surprisingly, Basque has been claimed at different times to be configurational and non-configurational.

36. Here the angle brackets $\langle \rangle$ represent information units, square brackets represent semantic units such as complementation and modification, and curly brackets $\{ \}$ represent the type of semantic bond existing among elements of the verb, which sometimes leads to lexical restructuring and fusion.

37. In a “generative description of a language ... syntactic representations are highly abstract objects considerably removed from anything observable in the linguistic data” (Comrie 1989:2). Abstract structure and invisible ‘functional categories’ and ‘traces’ in such *syntactic* descriptions simply makes up for the missing functional information.

Abstract or “imaginary segments” (cf. Zimmer 1975), act as “place holders” in the syntax for the missing semantics and pragmatics, much like in classical generative phonology.

38. In Lakoff’s words: “In traditional grammar, a grammatical construction was a syntactic configuration (often a complex configuration), paired with conditions on meaning and use. ... Within recent formal theories of grammar, the concept of the construction has almost completely disappeared. ... the complex construction—including constraints on pragmatics—has virtually disappeared from the literature on formal syntax and semantics. Grammatical constructions are at best considered an epiphenomenon, to be accounted for by other, more general principles; at worst, they are ignored” (Lakoff 1984:472).

39. For work on similar approaches, cf. Langacker 1987, 1988, 1991; Dixon 1991.

40. Goldberg later on in Chapter 3 specifies what she means by this, e.g. “The repertoire of constructions is not an unstructured set. There are systematic generalizations across constructions ... It is argued that constructions form a network and are linked by inheritance relations which motivate many of the properties of particular constructions. The inheritance network lets us capture generalizations across constructions while at the same time allowing for subregularities and exceptions” (Goldberg 1995:67). Thus the relationship between constructions is one of ‘motivation’, and motivation “lies between predictability and arbitrariness. In an intuitive sense, it often constitutes explanation” (Goldberg 1995:69).

41. Construction Grammar has been in practice a bottom up approach to linguistic analysis. Some might even argue that it has gotten lost in the bottom, looking at idiomatic constructions, and that not enough emphasis is placed on the upper layers of language. This of course contrasts with the approach taken by some formal theories which are in fact top-down approaches, which sometimes may ‘get lost’ in the ‘upper levels’ of grammar and fail to see the connection to those aspects of language which are closer to the lexicon. In Construction Grammar, on the other hand, there aren’t any theoretical obstacles to taking either a top-bottom or a bottom-down approach, or both.

42. Enkvist, for instance, in his report at a symposium on text analysis and generation, reports that his “most powerful, single, overall impression of our work in the past several days is that we are taking part in a shift in basic approaches to language. There seems to be a strong movement towards the study of language in terms of processes and not only in terms of structures” (Enkvist 1982:629).

43. This shift can perhaps be compared to the difference between studying the sounds of words in citation form in typical elicitation contexts, with the study of sounds in actual unguarded rapid and connected speech (‘rapid speech phenomena’).

44. Matisoff mentions how “[t]he pioneer semanticist (and reputedly the inventor of the word *sémantique*, Michel Bréal, conceived of grammatization [*sic*] in terms of

specialization rather than evacuation” (Matisoff 1991:383). According to Matisoff, “[g]rammatization may also be viewed as a subtype of *metaphor* (etymologically ‘carrying beyond’), our most general term for a meaning shift, or *glissement sémantique*. Grammatization is a metaphorical shift toward the abstract” (Matisoff 1991:384). For him, like for Givón, the motivation for grammaticalization must be sought in the automatization of language: “Human beings are preeminently creatures of habit, partly because of laziness, but partly because of the rational need to save effort. ... If a new collocation of morphemes proves useful (if it satisfies a *Sprachnot*), it can be stored in the brain as a unit, much like creating a ‘macro’ key on a computer” (Matisoff 1991:384-385).

45. The study of grammaticalization must deal with semi-motivated and semi-grammatical categories and constructions more often than fully motivated or fully grammatical one. As Traugott and König argue, “[t]he study of grammaticalization challenges the concept of a sharp divide between *langue* and *parole*, and focuses on the interaction of the two. It also challenges the concept of categoriality, and takes as central the concept of a continuum of bondedness from independent units occurring in syntactically relatively free constructions at one end of the continuum to less dependent units such as clitics, connectives, particles, or auxiliaries, to fused agglutinative constructions, inflections and finally to zero ...” (Traugott and König 1991:189).

Haiman argues that arbitrariness, i.e. emancipation from the original motivation, is not a sign of ‘autonomy’ of form from function: “It is an open question whether there are any limits in principle to the degree of emancipation which may occur. Hale’s important study on ‘deep-surface canonical disparities’ (Hale 1971) suggested that language acquisition data set bounds (at least in phonology) on the degree to which phonological ‘deep structures’, however plausible, could be emancipated from brute phonetic facts” (Haiman 1993:316). “Matters are similar in the kingdom of life. While much of evolution seems to proceed on the basis of a random walk rather than onwards sloping motions infinite, making for one sure goal, natural selection limits how much downright dysfunction the real world will tolerate in any of its creatures” (Haiman 1993:317).

He goes on to argue that “[b]iologically innate features are often as arbitrary, but no more arbitrary, than cultural institutions. Proponents of the ‘radical autonomy’ of syntax emphasize the sheer dysfunctionality of linguistic rules in order to validate their claims that language is not a functionally motivated structure” (Haiman 1993:317). And “dysfunctionality, far from pointing unambiguously at innateness, can have cultural as well as natural causes. Dysfunction is rife in culture, and it occurs in culture through emancipation. ... dysfunction even in nature has a similar origin” (Haiman 1993:318).

46. But see Bolinger 1972a about synchronic motivation of this morpheme. Perhaps a better example would be the Romance *que* complementizer.

47. Hopper goes on to “suggest some further principles ... [which] will supplement those suggested by Lehmann in being characteristic of grammaticization not only at the later, more easily identifiable stages, but also at the incipient stages where *variable* phenomena occur, and where the question more cogently arises as to whether we might speak of grammaticization. ... they are to a large extent part of the general lore about language change...” (Hopper 1991:21). These principles “speak only to the question of ‘more’ or ‘less’ grammaticized, not to the question of ‘in’ or ‘out’ of grammar. Worse still, they do not discriminate between processes of change which result in grammaticization and processes of change which do not result in grammaticization” (Hopper 1991:32).

48. Some investigators concentrate on the types of notions which tend to become grammaticalized cross-linguistically, i.e. the ones which, presumably because they are more central to human conceptualization, are more likely to become grammaticalized, e.g. Talmy 1985, 1988.

49. Adpositions, which are grammaticalized to different degrees, may derive from either nouns, verbs or adverbs, with different languages displaying different preferences. As Matisoff notices, “[a]lthough many grammatizational tendencies are doubtless universal, there are certainly areal differences of nuance. Bréal (1897:16 and passim) points out that Indo-European prepositions usually develop from *adverbs* (an independent and well-defined form-class). In the SVO languages of SE Asia, as we have seen, prepositions derive typically from full verbs” (Matisoff 1991:444). The study of the grammaticalization of the category topic and the interaction with other correlated pragmatic categories, such as the semantic notion *agent*, has been the focus of many studies in recent years, e.g. Vennemann 1974b; Schachter 1976, 1977; Keenan 1976a&b; Li and Thompson 1976a; Chafe 1976; Shibatani 1991.

50. As Givón argues, referring primarily to narrative, “[h]uman communication—message, discourse—is most typically NOT monopropositional, but rather *multipropositional*.” Thus “[t]he thematic coherence of discourse is multipropositional. Coherence over multipropositional spans means *continuity* or *recurrence* of some elements. The most easily identifiable recurring elements of thematic coherence, and the ones most conspicuously coded in grammars, are (a) referential coherence; (b) temporal coherence; (c) locational coherence; (d) action-event coherence. Put another way, coherent discourse tends to maintain, over a span of several propositions, respectively, (a) the same referent (“topic”); (b) the same or contiguous time; (c) the same or contiguous location; (d) sequential action” (Givón 1992b:7).

51. Referring to the word order possibilities in Italian, they mention that “all possible orders of subject, verb and object can and do occur in informal discourse, under certain conditions”: OSV, VSO, VOS, OVS, SOV (Bates *et al.* 1982:252). However, “[a]lthough such sequences are quite common, they have so far eluded formal description. For one thing, it is difficult for Italians to judge such orders one at a time.

out of context, since they are apparently conditioned by the pragmatics of connected discourse” (ibid., my italics, J.A.).

52. Longacre (e.g. 1982) argues that narrative is the most basic language type and that that should be the basis of our analysis of ‘standard word order’ for instance. However, as Downing tells us, others disagree, and “an increasing number argue that we should look instead to conversational texts as our primary source of data” e.g. Kyu-hyum Kim (Downing 1995:20). This means, according to Downing and Payne that some languages may be assigned to a different type if conversational data is used.

53. Longacre 1982 argues that VSO order is common “in main line of narrative and predictive discourse” whereas in procedural discourse SV(O) (topic-comment) order is more dominant.

54. This idea will then be echoed by later grammarians, such as Villasante (e.g. 1980) and Goenaga (1980:202ff), among others.

55. According to Mitxelena the two authors use two different styles of narrating: “Even based on such a small sample, I can’t help but conclude that Aguirre and Lardizabal follow two very different types of ordering words and that these are nothing but the manifest expression of two very different ways of narrating things. One could say, by guess, that Aguirre’s is the ost popular, more like in oral narrative, and less book-like” (Michelena 1978:223; my translation, J.A.) [“Aun basándose en una muestra tan reducida, me es difícil escapar a la convicción de que en Aguirre y en Lardizabal hay dos ordenaciones muy diferentes de palabras y que éstas no son más que la cara manifiesta de dos maneras muy distintas de contar. Así, a ojo, uno diría que la de Aguirre es la más popular, más atendida al relato oral, y menos libresca”].

In a writing from the same period Mitxelena (1977/1975:733) also makes some brief and cryptic remarks about the difference between *permanent* and *fleeting language* (*hizkera iraunkorra* and *hizkera iheskorra*, respectively), which Villasante (1980) comments on (Villasante 1980:17-18, 241). Fleeting language would be plain everyday, colloquial language, where (S)OV order reigns. Permanent language, whether spoken or written, is marked language, found in proverbs, poetry, etc. Although Villasante doesn’t say so exactly, (S)VO order is more prevalent in this type of language. It is not very clear to me what Villasante, or Mitxelena for that matter, are trying to say (this is one of the few places in Villasante 1980 where Mitxelena adds nothing to Villasante’s cryptic remarks). Perhaps the intuition they are trying to capture is related to what I will call rhematic salience of the verb (cf. Chapters 3 and 6 below).

Another expression of Lardizabal’s stronger verb-final strategy is the lack of auxiliary inversion in negatives (cf. Michelena 1978:223-4; cf. Chapter 7).

56. Osa too seems to agree with this interpretation when he says that SVO (despite not being the canonical or unmarked order according to him) is the most common order in

popular speech, something which he attributes (in most cases) to the fronting of the salient verb (*aditza aitzinera* “verb fronting”, cf. (Osa 1990:224ff, section 5.6.4).

57. See for instance: Biber 1984, 1985, 1986a, 1986b, 1987, 1988; Tannen, ed., 1982, 1984; Tannen 1982a&b, 1989, 1992; Chafe 1982, 1984a, 1985, 1986a, 1986b, 1988a, 1988b, 1991, 1992a, 1992b, 1994a, Danielewicz and Chafe 1985, Chafe and Danielewicz 1987, Chafe and Tannen 1987; Green and Morgan 1981; Mithun 1985; Halliday 1987; Olson 1994; Cooper and Greenbaum, eds., 1986; Downing, Lima and Noonan, eds., 1992; Horowitz and Samuels, eds., 1987.

58. As Chafe argues, “[s]peaking allows maximum exploitation of prosody—the pitches, prominences, pauses, and changes in tempo and voice quality that greatly enrich spoken expression. Prosody is an absolutely essential part of speaking, and one of the ways observations of language have often fallen short has been in not knowing what to do with it. Writing systems have never developed ways of representing anything that even approaches the range of spoken prosodic phenomena, and in that sense written language is seriously impoverished” (Chafe 1994a:43).

59. Some may think that Chafe exaggerates his argument to make a point. Even with imagined sentences we use imagined contexts to evaluate them, and one may reach valid (albeit incomplete) conclusions, from analyzing written language and ‘imagined language’. His point, however, is well taken, and it is not easily comprehended until one begins to analyze speech systematically. Although recording instruments appeared at the same time as transformational linguistics was taking over the field, the bias continued: “Real spoken language remained in this view [Chomsky 1965] corrupt: ‘A record of natural speech will show numerous false starts, deviations from rules, changes of plan in mid-course, and so on’ (p. 4), an attitude that precluded any interest in exploiting such data to arrive at an understanding of what the mind is doing when people speak. Or even when they write. Linguists were led to devote all their energies to studying the grammaticality (or lack of it) of isolated pseudosentences” (Chafe 1994a:47).

60. As Chafe says, “[i]t is at least possible that the differences [between speech/ conversation and writing] amount to simply a redistribution, though often radical redistribution, of the frequencies with which various linguistic devices are employed (cf. Tannen 1989).” (Chafe 1994a:49).

61. Both within speech and writing we find very specific types of styles, such as newspaper headlines and short articles which have very specific characteristics which derive from the particular characteristics of that speech. Thus Chafe tells us that “[i]f ‘grammars code best what speakers do most’ (Du Bois 1987, p. 851), it is also the case that each mode of language use produces a kind of language that codes best what the consumers of that kind of language find most adaptive. Front-page news reports use participles and prepositional phrases to pack a maximum amount of information into a minimum space ... advertisers attack us with short, snappy bursts,” etc. (Chafe 1994a:48).

62. This, of course, probably could not have been done if the possibilities for those orderings didn't already exist in the language, but it may very well be that those particular constructions were used to a greater extent in writing under the influence of foreign models. As we will see, the potential exists in Basque for 'lawfully' placing verbs before objects (and other complements), producing SVO-type orderings, but how much that potential is made use of varies according to a number of variables.

63. For a thorough background of these tools I refer the reader to, e.g., Du Bois, Cumming and Schuetze-Coburn 1988; Du Bois, Cumming, Schuetze-Coburn and Paolino 1992; Edwards and Lampert, eds., 1993, and the papers there; and Chafe 1994a.

64. In earlier writings Chafe referred to them as "idea units". Cf. also Cruttenden 1986.

65. According to Chafe, there is a high degree of consistency regarding these facts: "[i]n spite of problematic cases, intonation units emerge from the stream of speech with a high degree of satisfying consistency ... That fact suggests that they play an important functional role in the production and comprehension of language." (Chafe 1994a:62).

66. Campbell *et al.* argue that "[p]erhaps the degree to which a universal can have exceptions (i.e., be statistical rather than absolute) is correlated to the strength of its underlying explanations" (Campbell, Bubenik, and Saxon 1988:222). That is, unlike the universal that all languages have vowels, when it comes to word order universals, "externally motivated exceptions would prove less disruptive and therefore more likely to exist" (Campbell, Bubenik, and Saxon 1988:222). And an external factor like borrowing "can sometimes motivate exceptions to otherwise exceptionless universals (cf. Campbell 1980; A. Harris 1985:425; M. Harris 1984; Smith 1981)." (Campbell, Bubenik, and Saxon 1988:222). Thus they claim that "most exceptional languages fail to conform because borrowing or some such external factor has been involved" (Campbell, Bubenik, and Saxon 1988:222). They conclude that there is nothing absolute about universals of word order, and they "relativiz[e] universals, correlating the possibility of exceptions with the degree of disruptive influence to efficient functioning/processing such exceptions would produce in languages not conforming wholeheartedly to the universal" (Campbell, Bubenik, and Saxon 1988:222).

67. As Downing warns us, "[t]he striking profusion of classificatory schemes here is due in part to our discovery, over the years, that not all languages are necessarily describable in the same terms", as shown by the fact that for instance there are "languages in which it is difficult to justify calling any particular noun phrase the subject" (Downing 1995:18).

68. "Synchronic states' are so complex as to be essentially unique for any particular language unless filtered through a considerable amount of idealization. But the variation that makes synchronic analysis so difficult is easily accommodated by the diachronic analysis, which in this case is a sequence of quantitative changes, beginning with VSO > SOV, then Noun-Adjective > Adjective-Noun, then Noun-Genitive > Genitive-Noun, then Noun-Possessive > Possessive-Noun (by extension from Genitive-Noun), and finally

Preposition-Noun > Noun-Postposition (by the grammaticalization of genitive constructions as new adpositional constructions). *Process* is the true regularity here, the true principle of ‘Universal Grammar’” (Croft, Denning, and Kemmer 1990:xiv).

69. See, for example, Downing 1995: “all languages in discourse exhibit some word order variability, and *it is up to the analyst to decide upon one of these orders as “basic” or “unmarked” or “dominant”, if the language is to be typed.* Within the typological tradition, the nod has often been given in cases of competition to the most frequently occurring surface word orders; in cases of close calls, however, a variety of criteria are invoked (Keenan 1976[a]; Hawkins 1983; Givón 1979). Do clauses exhibiting the potentially “unmarked” word order allow the full range of tense, mood, aspect, and voice distinctions permitted within the language? Are clauses with this word order usable in contexts that are relatively free of presupposition, or does use of the construction carry some “special” pragmatic force, such as contrastive emphasis? Is this construction used in a wide range of pragmatic contexts, and within texts of all sorts, and does it appear relatively frequently within all these contexts?” (Downing 1995:19, my italics, J.A.)

70. In Spanish subjects are often elided and sometimes postverbal and objects are preverbal when pronominalized and when topicalized or focalized.

71. In some versions of ‘formal’ theory, word order is an epiphenomenon of the directionality of case and theta-role assignment, and thus of government. A directionality parameter with two possible values: ‘progressive’ or ‘regressive’ supposedly determines whether a language is VX or XV. Unique directionality across constructions is supposedly valued, since it is more general, though it is not required (cf., e.g., Haider 1986). In addition a basic word order is an underlying construct at deep structure level, so it need not be identical to the surface order, especially since there are plenty of empty positions (or ‘functional’ categories) to which an element may move. Thus Goodall (1993) for example argues that Spanish is underlyingly a VOS language. Vallduví reaches a similar conclusion for Catalan from an information structure/packaging perspective (cf. Vallduví 1993).

72. This quote is found in Mitxelena’s running commentary at the foot of the page to Villasante 1980. The original quote says, in Spanish: “No llegaremos a ninguna parte si postulamos un orden (en abstracto) normal y todo lo demás se considera desviación. Hay casi siempre más de una posibilidad de ordenación: lo que hay que determinar es si influyen en el significado, si son o no igualmente corrientes, etc...”

73. It is well known by now that in most languages clauses with both subject and object noun phrases are very rare in discourse (cf., e.g., Du Bois 1987, Chafe 1987), so that exactly what it means to have a certain ‘basic’ ordering of subject verb and object is not necessarily obvious.

74. Lehmann adds that “virtually any syntactic arrangement may be modified to bring about a difference in meaning or to achieve a special effect, whether in poetry, rhetorical

usage, or everyday patterns. ... Such deliberate modifications resulting from the widespread use of marking in language require linguists to interpret texts sensitively, especially texts in earlier stages of languages” (Lehmann 1976:520).

75. Thus for instance, in German “the order *DO + IO is blocked when DO is indefinite” (Hawkins 1986:44).

76. Connolly (1991) says pretty much the same thing: “Marked and unmarked orders can be characterised either statistically or non-statistically. According to the non-statistical view, an unmarked sequence is a default ordering, which is used unless there is some particular reason (such as emphasis) to depart from it, in which case the marked sequence is selected instead... According to the statistical view, on the other hand, a sequence is unmarked if it is the most frequently occurring order among a set of alternatives ...” (Connolly 1991:16).

77. Lambrecht assumes that English, French and Italian have SVO as pragmatically unmarked constituent order; and that the pragmatically unmarked sentence-accent position is clause final (or near-final) (Lambrecht 1994:16). That this is their “pragmatically unmarked (or canonical) constituent order and [the] unmarked focus-accent position” doesn’t mean for Lambrecht “that sentences having these formal properties are “pragmatically neutral” (Lambrecht 1994:16). For him, “given a pair of allosentences, one member is pragmatically unmarked if it serves two discourse functions while the other member serves only one of them. While the marked member is positively specified for some pragmatic feature, the unmarked member is neutral with respect to this feature.” (Lambrecht 1994:17). Thus, “in calling SV(O) constituent order and clause-final focus-accent position “pragmatically unmarked” in our three languages I am referring to the fact that this pattern has greater DISTRIBUTIONAL FREEDOM than alternative patterns and, as a corollary, that it has greater overall frequency of occurrence. In am NOT implying that alternative, i.e. marked, patterns are somehow “stylistically remarkable” or “abnormal.”” (Lambrecht 1994:17-18). It seems to me that we could dispense with the label ‘unmarked’ here and just say that a certain word order has greater or lesser distributional freedom. The labels ‘unmarked’ and ‘basic’ might seem counterproductive, since they don’t help us understand things. Also, when you say that SV(O) is the unmarked order in these three languages, unlike when one speaks of greater or lesser distributional freedom, it makes it sound that it is unmarked in the same way, which is never the case and thus is misleading.

78. For Ortiz de Urbina, working in a formalist framework, SOV order is the underlying order. Other orderings arise from movement of the focused element to Spec of COMP. When this happens, the verb moves to COMP and the other element, the ‘topic’, moves to Spec of S position (optionally) (cf. Ortiz de Urbina 1991:21-22).

79. Frequency may be a valid measure of markedness (to the extent that we may want to use the same name for the two different measures) in a typological context. Thus, Dryer

argues that “[w]hile one might say that SV order is crosslinguistically unmarked, because SV order is more common, such usage is clearly irrelevant to the question whether SV or VS order in a particular language is unmarked” (Dryer 1995:110). Additional types of markedness are, according to Dryer, (1) “formal markedness”, “sometimes referred to as morphological, morphosyntactic, or structural markedness,” and which “refers to the presence vs. Absence of morphemes, or, sometimes, to the relative phonological size of morphemes” (Dryer 1995:110); (2) ‘distributional markedness’, where “one form is said to be distributionally marked if it occurs in a proper subset of the morphosyntactic contexts in which the other occurs. This notion is relevant in the area of word order since there exist instances in which one word order has a more restricted distribution than another”, such as main clauses (Dryer 1995:111). Dryer questions the characterization by Brody (1984) of Tojolabal as VOS; that of Tomlin and Rhodes (1979) of Ojibwa as VOS; that of Payne (1990) of Yagua as VSO; and that of Abbott (1991) of Macushi as OVS.

80. According to Dryer, “[s]ituations in which one word order is the order used when conditioning factors are neutralized is [*sic*] one type of situation in which a word order can be described as pragmatically unmarked. Such situations can be viewed as special instances of a more general type of situation in which one word order is the default order in the sense that the easiest way to characterize the contexts in which that order is used is to specify when other orders are used and then state that the order in question is used *elsewhere*” (Dryer 1995:116).

81. Dutch has VO order in main clauses with a simple tense verb; $V_{\text{finite}}O(V_{\text{main}})$ order in compound tenses, and OV order in subordinate clauses (including relative and infinitive clauses) (cf. Gerritsen 1980, 1984, 1992). Prepositional phrases typically occur preverbally in Dutch, though occasionally postposing is allowed (cf. Gerritsen 1992:356). In other ways the word order is not typically verb final, such as in the fact that subordinators are clause initial and not clause final.

82. In actual spoken language, at least for some speakers, no clause type is exempted from the possibility of allowing elements after the verb, though typically only in a small percentage of cases. Relative clauses are always verb final, since the verbal suffix is the element which connects to the noun, much like in a genitive clause. On the other hand, I have found a small percentage of headless relative clauses in the Spoken Basque Corpus which are not verb final.

83. According to Givón, rigid languages are found with most basic word order types: SVO (English), SOV (Sherpa), VSO (Jacalteco), VOS (Malagasy) (Givón 1984a:Ch. 6.4). Hawkins 1986 argues for the covariation of case marking and rigid word order: “It is plausible to argue that the case system of German is responsible for the greater clause-internal word order freedom of that language. Across languages the existence of rich surface case marking typically correlates with word order freedom of the kind we have seen in German. The reason most commonly advanced for this is that ‘fixed’ word order at the sentence level in a language like English encodes grammatical relations” (Hawkins

1986:40) This principle he attributes to Keenan (1978): “This insight has been expressed as the ‘principle of covariation of functional equivalents’ by Keenan (1978): ‘Syntactic (and morphological) processes which have the same ‘function’ covary in their distribution across languages. (Keenan 1978:120).’ He explains: ‘By ‘have the same function’ we mean something like ‘code the same semantic or syntactic information’ ... The principle predicts then that the more we assign a language overt case marking the freer can be its basic word order and conversely.’ (Keenan 1978:120-21)” (Hawkins 1986:40-41). Thus, argues Hawkins, “Keenan’s principle predicts that the loss of the case system should be accompanied by a relative freezing of the permitted word order permutations, and this is, of course, what happened. Sapir’s (1921) drift theory linking case syncretism to the fixing of word order is completely in accordance with Keenan’s principle” (Hawkins 1986:41).

Japanese is quite rigidly verb-final despite its case marking system (aside from some ‘afterthought’ elements, cf. Kuno 1978). In main clauses, however, we find freedom of ordering of the *preverbal* elements, though in other clauses and in casual conversations, where case marking can be omitted, the order we find is S-IO-DO-V. Clause initial position is reserved for special topics and preverbal position for focus (cf. Kim 1988, 1995).

84. As we will see in Section 1.8, rigid grammatical order is often associated with languages which have undergone a great amount of language contact in their history, including extended bilingualism.

85. It has been claimed that in some languages, one may not be able to make any ordering generalizations in terms of grammatical relations at all. This is what Payne 1990 has suggested for Papago. Meanwhile others have managed to classify it as SOV, and others still as VSO (cf. Campbell, Bubenik, and Saxon 1988).

86. Asserted clauses display what has been called ‘main clause phenomena effects’ (Green 1976, Bolinger 1977), ‘root transformations’ (Emonds 1970, 1976; Hooper and Thompson 1973; Langendoen 1973; Gary 1976; N. McCawley 1977; Newmeyer 1987; Ogle 1981); and the ‘penthouse principle’ (Ross 1973; Andersson and Dahl 1974).

87. Notice, however, that Dutch and Basque subordinate clauses differ in other ways, such as that Dutch has clause initial subordinators, whereas Basque for the most part has clause final ones, cliticized to the finite verb.

88. Payne (1990, 1992b), in her studies of ‘O’odham (Papago) word order refers primarily to pragmatic statuses (identifiable vs. non-identifiable), and not to pragmatic relations or salience relations; cf. Chapter 3. Papago (a.k.a. ‘O’odham) is a notoriously hard language to classify: “There is ... considerable disagreement about the basic word order of Papago resulting from the relatively free word order not only of the basic sentential elements (S,V, and O), but also of the elements within adpositional phrases and noun phrases. Thus Hale (Hale *et al.* 1977) ... describes Papago as SOV, GN and

postpositional (i.e., type 23), though he admits the possibility that it might be non-configurational (without a basic word order). Saxton and Saxton (1969) and Saxton (1982) have suggested VSO, with NG (not the GN of Hawkins), prepositional, and AN (i.e., type 2), a suggestion which Langacker (1977b:24) seconds. Zepeda (1983), a native speaker, gives SOV as basic, but with much variation. Payne (1985:464) suggests that ‘Papago order cannot be characterized in terms of syntactic role; rather, it uses pragmatic word order, sensitive to a feature of indefiniteness’. ... None of these descriptions matches the word-order patterns presented by Hawkins for Papago” (Campbell, Bubenik, and Saxon 1988:211-12).

89. The thesis that he presents in this paper is “that languages with discourse-governed word order often exhibit the word order characteristics associated with the more frequent word order in the language. This is the case apparently despite the fact that there is no reason to believe that there is any rule in the grammar of these languages that states that the more frequent order is in any sense the basic order. Rather, this order is apparently more frequent due to the interaction of the rules associating discourse functions with particular orders and the relative frequency with which the relevant discourse contexts occur in typical discourse” (Dryer 1989a:86).

90. Li and Thompson 1976a attribute this last observation to Hsieh Hsin-I and W.P. Lehmann (personal communication). The only language which seems to contradict this is Chinese, which has been said to be SVO, but they claim that Chinese is actually in the process of becoming SVO (cf. Li and Thompson 1974a and 1974b). This claim, however, has been challenged by Givón for instance.

91. It has sometimes been claimed that the rigidification of English word order was caused by the loss of inflections (cf. Sapir, Vennemann), but others have disagreed with this assessment (cf. Jespersen 1922, Danchev 1991).

92. In Christian Lehmann’s words: “The following can be established as an implicational tendency: if a language has free word order, then it has rich inflectional morphology. This is empirically confirmed in most cases (see Mallinson - Blake 1981; ch. 3.4 for possible exceptions). The converse of this law, however, is invalid: it is not the case that if a language has rich inflectional morphology, it has free word order. This would be falsified by numerous languages with dominantly agglutinative morphology, e.g., by Turkish and Yucatec Maya. It thus becomes evident that there can be no question about a mutual compensation between word order and inflectional morphology. The unilateral implicational relationship between the two phenomena indicates that they are not hierarchically equal, but instead in a dependency relationship” (C. Lehmann 1992:400).

93. I will argue that language contact (1) has the potential of confusing the ‘delicately balanced’ rules of pragmatic ordering, thus potentially leading to its fixing in a ‘search’ for simplicity, and (2) favors grammatical, and in particular, morphological simplicity, which also favors the loss of inflections.

94. In Lehmann's (1992) words, "it should be clear that the distinction between fixed and free word order is not an absolute one. On the one hand, there are possibilities in German of varying the verb position even within one and the same sentence type... On the other hand, even in Latin there is a tendency, which is independent of functional sentence perspective, towards final position of the verb" (C. Lehmann 1992:397).

95. Givón 1988 has argued forcefully that the 'unpredictable-first' or 'important-first' principle is not confined to languages like Ute, but that is a basic underlying principle for all languages. In more rigid word order languages the most-important-first ordering principle may still be followed by complements other than the subject and the object, especially by optional complements.

96. The correlations that Myhill gives, however, are rather weak and must reflect some other aspect of the pragmatics of constituents and the positions yet to be determined. I do not think that we can say in any way that VS order 'codes' a clause as sequential, for example. Hopper (1979a) claimed that SV order is associated with 'background' clauses (i.e. non-foreground clauses) in Old English, cf. also Hopper 1986a. More recently, Hopper (1987b) has argued, based on the findings of Du Bois (1985, 1987) that it is V-NP, and not VS that is associated with 'foregrounding', where NP correlates with the category 'absolutive'.

97. Mayan languages differ as to how grammaticalized the 'topic' is, i.e. whether it is a clausal topic or a sentential (dislocated) topic: "Tzotzil and Jakaltek topics are less integrated into basic clause structure, being essentially prefixed to what is otherwise a fully well-formed clause. The connection to the following clause is pragmatic, not syntactic: the clause must be 'about' the topic. By contrast, Tz'utujil topics appear to be much more tightly connected to the clause that follows. Two specific differences are that Tzotzil and Jakaltek topics do not occur in embedded clauses, while those of Tz'utujil do. Another is that third-person pronouns do not generally function as topics in Tzotzil and Jakaltek, but they do so freely in Tz'utujil. These differences suggest two sorts of entities" (Aissen 1992:44).

98. Comrie and Matthews for instance mention that "It has often been noted that where English uses grammatical-relation changing processes such as Tough Movement, Slavic languages tend not to change grammatical relations but rather to make use of their free word order possibilities to give different linear orders of the same constituents" (Comrie and Matthews 1990:50). They also mention Hawkins' (1986) claim that English tends to be much more free in its use of constructions which 'move' constituents out of their clauses than other languages, such as German.

99. The view that ergative constructions are erstwhile passive constructions goes back to 19th century grammarians, such as Schuchardt, some of whom even argued that ergative constructions were actually passive, synchronically speaking. In recent years the passive theory for the origin of ergativity has been revived, for example, by Hale 1970.

100. One interesting case in which the passive has developed into something quite unlike what I have just described is Irish. The Irish passive-like construction differs a great deal from the ‘prototypical’ passive. This construction is used in contexts where we typically expect non-passive clauses, since it is used as often as active clauses. Also, it is associated with higher topicality of the agent instead of the opposite. In addition, it has an overt agent 80% of the time, unlike in most other languages, where agentive passives are between 10-30%, if they are possible at all (Myhill 1992:111-13).

101. An interesting fact about these two alternations is that in some languages it is not clear which one we are dealing with. This seems to be because in these languages the ergative construction stems, diachronically from a passive construction which became the unmarked form for assertions, cf., e.g., Myhill 1992:161-162.

102. Notice that Basque does not have a true, English-like passive construction, despite certain claims to the contrary based on the existence of a passive-like stative constructions, cf. Bollenbacher 1977, Perlmutter 1978. For a different perspective on these constructions, cf., e.g., Eguzkitza 1981; Trask 1985a. See also Jacobsen 1985; Osa 1990:241; Euskaltzaindia 1993:428.

103. Prince 1978 identifies two different types of *it*-clefts in English, namely the regular or “stressed focus *it*-clefts” and what she calls “informative-presupposition *it*-clefts (cf. also Ball 1991, 1994, Lambrecht 1994). We do not expect to find this particular (motivated) extension of the construction in cleft constructions in other languages.

104. German too has a number of “clause-internal bounded movement rules” (Hawkins 1986:37). Hawkins 1986 presents an interesting comparison of the relative similarities and differences between English and German. As Hawkins argues, “[o]ne of the major determinants [for these ‘rules’] does seem to be Theme-Rheme organisation, ... although a number of other subtle pragmatic functions appear to play a role as well” (Hawkins 1986:44).

105. According to Givón the primary function of dative-shift in English, and dative-accusative order shift more generally in more flexible order languages, is to ‘code’, or rather reflect, topicality differences. He argues, for instance, that “objects are consistently *more topical* than indirect objects, e.g. “A study of written English text [Givón 1984b] ... has found that that the typically human dative-benefactive object, the most topical object type, appears 84% of the time as *direct* object (DO), and only 16% as *indirect* object (IO). In contrast, the non-human locative object appears 100% as IO” (Givón 1990:762-763). Also, “the common denominator of promotion to DO cross-linguistically is word-order: The more topical object is most commonly placed *before* the less topical one.” (Givón 1990:763). This would reflect the tendency to place the most important element first (Givón 1990:764). I am not convinced by this explanation. I feel that rhematic salience, to be defined and exemplified in following chapters, may turn out to be the crucial element in explaining this structural alternation.

106. By second position, one of two things can be meant, the position after the first word in the sentence (e.g. Classical Greek) or the position after the first major (phrasal) constituent (e.g. Germanic languages). Wackernagel 1892 tried to reconcile both phenomena as being related and having to do with the cliticization of unaccented verbs to second position. Anderson 1992, 1993 also attempts a more sophisticated, but similar, solution. Here I will deal solely with the Germanic version.

The Basque negation construction is similar to the verb-second construction in other ways too. In this construction the negated auxiliary is in rheme-initial position, which means that only the clausal topic may precede the rheme, other setting elements being always dislocated, i.e. in intonation units outside the clause proper. In Basque, however, there doesn't have to be an element before the rheme (such as when the topic is 'elided') and thus the finite verb may come in clause-initial position and not just rheme-initial position.

107. In German *Einklammerung* or *Umklammerung*, cf. Gerritsen 1984. The phenomenon is present in Modern German Frisian and Dutch, but not in the Scandinavian Germanic languages. Gerritsen (1984) reports that in "spoken Norwegian, German and Dutch show that about 40% of all declarative main clauses have some other constituent than the subject in front position in all three languages (Faarlund 1981, Jansen 1978, 1981)" (Gerritsen 1984:110), though "fronting of other constituents than subject is [not] subject to the same syntactic and pragmatic conditions for all the Germanic inversion languages" (*ibid.*).

108. Inversion and exbraciation were present in the Germanic languages in variable form at one point. Inversion disappeared in English and became obligatory in German and Dutch. Embraciation disappeared in English and in the Scandinavian languages, but it became obligatory (grammaticalized) in German, Frisian and Dutch (cf. Gerritsen 1984). Gerritsen proposes that this "divergence might be explained by the fact that the English and Scandinavian verb systems were more synthetic when those languages entered the exbraciation stage than the verb systems of German, Frisian and Dutch when they began to exbraciate" (Gerritsen 1984:119). In addition, "The disappearance of embraciation in English ... may also be due partly to its creole character or to influences from French ..., since the brace is one of the first syntactic phenomena that disappear under influence of language contact (cf. Clyne 1980:30)" (Gerritsen 1984:127). English has some minor constructions which are reflexes of the V2 rule, e.g. *In/<Into the room> came the police inspector*. Typically a locative comes before the verb. In the English reflexes, the initial element can be either pre-rhematic (setting) adverbials or rhematic (focal) ones.

109. Herring (1994) mentions that "Kuno (1978) uses the term 'afterthoughts' to characterize what in our terms would be both afterthoughts and backgrounding in Japanese (cf. also Fujii 1991), and Erguvanli [*sic*] (1984) labels as 'backgrounding' a similar diverse set of phenomena in Turkish. The same three functions, more or less, as are found in Tamil are attributed by Kim and Shin (1992) to Korean and by Junghare (1985) to Indo-Aryan: afterthoughts ('corrections'), de-emphasis, and emphasis" (Herring

1994:149). It is obvious that we are dealing with very different types of constituents, some of which must be distinguishable from each other through intonation. There is no doubt, however, that languages differ as to what types of elements they allow to appear postverbally. True afterthoughts, however, are not ‘postverbal elements’ however, other than in the obvious sense that they come temporally after the verb, cf. Chapter 3.

110. The (relatively) rare characteristics shared by these languages are the following (from Breivik 1991): (i) Allow no deletion of unstressed personal pronoun subjects; (ii) Must have subjects for impersonal verbs; (iii) Have special indefinite pronoun subjects like *on* and *man*; (iv) Have dummy subjects to replace extraposed sentences; (v) Have a dummy pronoun *there* (or some equivalent) to stand in the place of logical subjects that have been displaced from sentence initial position; (vi) Have overt subjects for continuous (topical, predictable) subjects, where other languages would omit them.

111. Complement clause postposing may also be the result of the presence of another element in the single (most salient) rhematic position available before the verb. Such complements typically ‘refer’ to propositions which are accessible (and thus non-salient) or easy to accommodate.

112. Greenberg’s findings weren’t as surprising as one might think. As Bean (1983) reminds us, Weil (1844) was already “aware of the fact that some languages are characterized by the use of prepositions, SVO and Noun-Modifier order while others are characterized by postpositions, SXV and Modifier-Noun order. He states, in remarkably modern terminology, that these are the two ‘extreme points between which languages may oscillate. It is, on the one hand, the order ... which places the qualifying word after the word qualified ... on the other, the order which places first the governed word, then the governing’ (p. 56)” (Bean 1983:18). Later on, “McKnight (1897) takes up Weil’s ideas and develops them in a more objective fashion. He employs the terms ‘subjective’ and ‘objective’ order of words to refer to functional (topic-comment) and grammatical (subject-verb-object) ordering relationships respectively” (Bean 1983:19). In addition, “McKnight also seems to have discovered the basic idea underlying Behaghel’s First Law” (*ibid.*).

113. Dryer’s sampling methods, based on counting genera which contain a particular characteristic, are described in Dryer 1989b and constitute an attempt to avoid genetic and areal biases. For a critical view, see, e.g., Croft 1995. One reason for his sampling methods is that word order diffuses areally rather readily (Dryer 1992a:83). Dryer’s data comes mostly from a 543 language sample (out of 625) for which he could determine the relative (‘basic’) order of verb and object. For the remaining 82 languages this was not possible. Also, as can be plainly seen in the table, only for a minority of the 543 languages could the different characteristics be determined (Dryer 1992a:83fn1).

114. The emphasis on the idea that SVO languages patterned like verb-initial languages goes back to Lehmann 1973 and Vennemann 1974a. In the 1980s there were several

attempts to prove Lehmann and Vennemann wrong by trying to show that SVO languages do not really pattern like verb-initial languages, e.g. Hawkins 1980, 1983; Comrie 1981/1989, and Mallinson and Blake 1981.

115. It is also largely coextensive with the semantic category 'subject. The 'semantic subject' is nothing more than the most likely argument to become the topic, the default topic', with multi-argument predicates, i.e. the argument which displays the highest intrinsic 'thematic salience'.

116. C. Lehmann (1992) argues that "[i]n diametrical contrast to traditional historical-comparative linguistics", Lehmann (1973) and Vennemann (1974a) "saw the syntax, especially word order, in the center of the functioning of the language system. They tried to base language types on order types and to reduce grammatical change to order change. Today, these attempts may be considered to have failed. ... One of their methodological mistakes was oversimplification paired with immunization against falsification." (C. Lehmann 1992:396)

117. The characteristics for which Dryer argues SVO languages pattern like verb-initial languages are the following: Prepositions, N-Rel, Adj-Std., Copula-Pred, Plural word-Noun, V-PP, V-Adv, T/A-Aux-V, Neg. Aux- V, and Comp-S. The characteristics for which SVO languages pattern as intermediate between V-initial and V-final languages are: Gen-N/N-Gen, S-Qn.Part./Qn.Part-S, Wh. in-situ. In addition for three characteristics all three types pattern similarly: Adj.-N/N-Adj. (cf. Dryer 1988b), Dem.-N/N-Dem., Intensifier-Adj./Adj.-Intensifier.

118. John Haiman, who has written extensively on iconicity, says the following about it: "The notion that formal distance corresponds to conceptual distance is not a new one. Although Behaghel (1932:4) is credited with the first explicit statement of the principle, Scaglione (1972:243) argues that it was first adumbrated by Condillac's *Essai sur l'origine des connaissances humaines* of 1746. Jespersen (1949:56) speaks of a principle of cohesion, that 'ideas that are closely connected tend to be placed together.' Bolinger (1975:137) speaks of the same principle in his introductory textbook, *Aspects of Language*. The idea has been effectively 'rediscovered' in several recent papers, for example Bybee, forthcoming; Haiman 1983a,b; Lambrecht 1984; where attempts have been made to associate the principle with fairly extensive bodies of data" (Haiman 1985b:106). For recent collections of iconicity studies, cf. Haiman, ed., 1985; Simone, ed., 1995.

119. Later, Lehmann translates it as: "what is closely related in meaning is also placed closely together" (Lehmann 1976:519). The original says: "Das oberste Gesetz ist dieses, daß das geistig eng Zusammengehörige auch eng zusammengestellt wird."

120. VSO languages, such as Irish, constitute apparent counterexamples to this rule. But remember that in actual discourse transitive sentences with overt subjects are rare and, furthermore, as Lehmann notes, VSO languages tend to make abundant use of the passive

construction: “the extent of development of passive constructions among [VSO languages] is remarkably great” (Lehmann 1976:522). For interesting proposals along these lines, see Moravcsik 1978; Sheintuch 1981; Haiman 1983:790; Keenan 1984:200-3; Siewierska 1988:55.

121. Bybee argues that the principle of relevance accounts for “even finer degrees of fusion or ‘distance’ between linguistic units [than Haiman’s distance principle], such as the ordering of affixes, the existence of allomorphy, irregularity, stem change and suppletion” (Bybee 1988a:359).

122. For some interesting studies of relative ordering of particles and adverbials based on their different scopes, see e.g. Greenbaum 1969, Quirk *et al.* 1972, 1985. For a similar perspective in Functional Grammar based on the notion of ‘layers’ (predicate, predication, proposition, illocution, or clause), see Hengeveld 1989, 1990, Dik 1989:50fn4. Van Valin too presents a ‘layered’ perspective on the clause, distinguishing between nuclear operators (aspect, Directionals with respect to the event/action), core operators (directionals with respect to the participants; modality: root modals, e.g. ability, permission, obligation; internal negation); and clausal operators (status: epistemic modals, external negation; tense, evidentials, illocutionary force) (cf. Van Valin 1990b).

123. The percentage of verb-final languages with preverbal negation is even higher “there are 22 additional SOV languages ... in which the negative precedes the verb, but for which the basic order of negative, subject, and object is not clear. This situation arises either because my grammatical source is not specific on this point (it may just say that *the negative precedes the verb*), or *because no single position before the verb is clearly basic*” (Dryer 1988a:96, emphasis added). In other words, flexible verb-final languages.

124. I do not think that the iconic ordering of agent before patient is, however, the primary motivation behind the fact that in the majority of the world’s languages ‘basic order’ the subject precedes the object. It may be a reinforcing motivation, but not the primary one, which, as I said, lies in the iconic ordering of topic before comment (rheme).

125. As Hajičová argues, this phenomenon is independent from discourse pragmatic related ordering preferences: “Systemic ordering should be distinguished apart from communicative dynamism, which is the concept underlying deep word order” (Hajičová 1991:103).

126. Dik categorizes the “relator principle” as a sub-principle (“specific principle”) of the “principle of head proximity”, which says that “[c]onstituent ordering rules conspire to keep the heads of different domains as close together as possible” (Dik 1989:343).

127. It is irrelevant whether the relators are suffixes, clitics, or independent morphemes; or whether they are part of the head or of the ‘dependent’ (cf. Nichols’ 1986 distinction between “head marking” and “dependent-marking” languages). Hawkins and co-investigators (Cutler, Hawkins, and Gilligan 1985; Hawkins and Cutler 1988) have

argued that there exists a suffixing preference cross-linguistically, i.e. that function morphemes, such as relators, which follow a word or phrase, are more likely to become part of the preceding word than those which precede them. Bybee and her co-investigators, on the other hand, have questioned such an assumption (Bybee, Pagliuca, and Perkins 1990) (cf., e.g., Croft 1995:104-5). (For a review of Bybee, Pagliuca and Perkins 1990, see Dryer 1992c.)

128. In the case of nominal morphology (case markers) arise from reanalysis of either serial verbs (verb > preposition > case marker, e.g. *boy take door shut*, “the boy shut the door”, where ‘take’ becomes an object marker, as in Niger-Congo languages for instance), or of genitival compounds (adposition + noun > adverb/adposition, e.g. Spanish *en cima de la mesa* “on top of the table” > *encima (de) la mesa* “on the table”) (Givón 1984a:228). As for verbal morphology, Givón argues the correlations here are stronger, for tense-aspect-mood auxiliaries and morphology almost always arises out of main verbs. Thus, in an OV language we will get suffixes and in a VO language prefixes (Givón 1984a:230). Same thing for ‘verb derivational morphemes’, as in transitivity (‘causativization’) and detransitivization (‘passivization’, ‘reciprocation’, ‘stativization’, and ‘reflexivization’) (Givón 1984a:234). Negation markers follow Greenberg’s correlation only when they arise from negative modality verbs (such as *fail*, *stop*, etc.), but not when they arise from negative intensifiers, as in modern French *pas*, *personne*, or *rien* (Givón 1984a:232-33).

129. The following sentence shows how a directional postposition could conceivably be developed over time: “Time was, Mars was a busy place. Everybody, it seemed—or at least everybody at NASA—wanted to fling something **the Red Planet’s way**. ...” (*Time*, Nov. 11, 1996, p. 73) (cf. *to the Red Planet*).

130. Quoted in Croft, Denning, and Kemmer 1990:xiii. Shibatani and Bynon 1995 have expressed the same idea: “Typology is often characterized as ahistorical classification of languages and is contrasted with genetic classification, which is historical. ... This does not, however, imply that typological and historical comparison belong to two entirely unrelated subfields of linguistics, for ... both evolution and structure are constrained by general principles of language, and typological studies have important implications for historical linguistics” (Shibatani and Bynon 1995:20).

131. There is a tendency to use “‘drift’ as a near-synonym of ‘slowly-occurring change’” (Malkiel 1981:566), and that “‘Drift’ tends to have—perhaps to a higher degree in 1980 than sixty years ago—the connotation not only of a slow, gradual movement, but also, in particular figuratively, of an ill-directed, almost aimless motion...” (Malkiel 1981:540). Malkiel introduces a new term, namely *slant*, for the apocryphal version of Sapir’s *drift*, which is mentioned briefly in Chapter 8 of Sapir (1921) (Malkiel 1981:553). This version of *drift* has to do with ‘convergence’ of change in dialects after they separate. For the notion of ‘convergence’, see Meillet 1918. Sapir also was not totally clear or explicit with his concept of drift, as Malkiel 1981 has loquaciously argued.

132. These are *external* motivating principles. Language *internal* motivating principles may also play a role in slowly occurring change of this type.

133. Lakoff 1972 and Vennemann 1975 brought back the concept of drift to the typological study of word order. Lakoff 1972 is a paper given at a conference in 1969, a conference which was attended by Vennemann, as Malkiel 1981 has reminded us.

134. Jespersen 1922 argued that word order change took place before the loss of case markings, whereas Sapir and Vennemann for instance argued that case loss came first. Sapir also argued that the loss of case markings (i.e. case syncretism) *caused* the fixing of word order (case syncretism → [OV > VO]). Other linguists have doubted the veracity of Sapir's suggested chain of events, arguing that either the two events are independent or causation goes the other way, cf. Li and Thompson 1974a, Hawkins 1983, Hawkins 1986, Danchev 1991. Hawkins notes that, "although OV languages typically have a case system, VO languages occur both with and without one. Diachronically this means that an OV & +Case language ... has two pathways of change to reach VO & Case: it can acquire VO first, before it sheds its case system ...; or it can acquire VO and shed its case system at the same time ... But it cannot shed its case system before acquiring VO ... The existence of OV & +Case ... and VO & +Case ... languages means that the synchronic date are compatible with verb-object reordering *preceding* case syncretism, as well as with verb-object reordering being simultaneous with case syncretism. But case syncretism is therefore not a *necessary* cause of the OV > VO shift (since this latter may occur alone), though it may be argued to be a *sufficient* one" (Hawkins 1986:49-50).

135. Jacobsen concludes that this "moderate amount of syncretism [in Basque] between potentially important case endings is kept tolerable because of evidence provided by the total context" (Jacobsen 1974:109). Furthermore, he argues, "[t]here is no reason to assume that the system is either insufficient or unstable" (*ibid.*).

136. The only exception involves continuous imperfective clauses with the already mentioned *ari* construction. This construction prevents the coreferencing of the lower verb's arguments.

137. The notion of drift doesn't easily fit into a self-contained theory of language. Lass (1984), for instance, argues that "a segment does not know where it came from" (Lass 1984:178). Ferguson (1990) counters that "just the opposite view seems more useful as a working assumption. Every segment provides, by its phonetic characteristics, allophonic variation, phonotactic limitations, morphological alternations, relative frequency, and a host of details of acquisition, dialect variation, etc., a set of clues as to its source and its possible directions of change, and in some instances it offers a remarkably clear picture of its diachrony" (Ferguson 1990:60-61).

138. The issue is the following: "many stable linguistic variables with no synchronic motivation show historical continuity with little change over long periods of time. Children acquire at an early age historically transmitted constraints on variables that

appear to have no communicative significance, such as the grammatical conditioning of (ING) in English” (Labov 1989:85) In other words, the grammatical conditioning of (ING), is that “/in/ was favored most in progressives and participles, less in adjectives, even less in gerunds and least of all in nouns such as *ceiling* and *morning*” (Labov 1989:87).

139. The generalization goes back to the Greenberg’s landmark study, although he didn’t make it a central, and even earlier. As Payne (1990) reminds us, “[t]he operator (modifier) operand (modified) distinction is commonly attributed to Lehmann and Vennemann (cf. Lehmann 1973; Vennemann 1974[b]; Vennemann and Harlow 1977); but Greenberg (1963) and Lepsius before him noted that in most languages there is a tendency to put either the modified element before the modifier, or viceversa. ... Similarly, the seeds of Hawkins’ Cross-Category Harmony principle (cf. Hawkins [1982], 1983) are found in Greenberg’s discussion of harmonic and disharmonic relations among distinct rules of order, presumably associated with psychological generalization” (Payne 1990:16).

Vennemann, for instance, describes this approach in the following way: “From the work of” Greenberg, Lehmann, Bartsch, and Vennemann “a certain concept of consistent basic serialization (CBS) has developed which is understood as follows: a language has CBS if specifiers (modifiers, qualifiers, *déterminants*, *unterscheidende Glieder*, operators) either all precede or all follow the specified (modified, qualified, *déterminé*, *unterschiedene Glied*, operand. A language with CBS is called a consistent pre-specifying language (*langue montante* or *centripète*, OV language, XV language) if specifiers precede the specified; otherwise it is called a consistent post-specifying language (*langue descendante* or *centrifuge*, VO language, VX language)” (Vennemann 1976:615). “The goal of this paper is to give a precise definition of the concept of specification and, consequently, of CBS in a simple categorial grammar” (Vennemann 1976:615).

140. Whether a language is head-initial or head-final is determined by a parameter or, alternatively, by direction of assignment of Case and/or Theta-role (which are formal, not semantic categories). According to Chomsky, “[t]he possible phrase structures of a language are fixed by general principles and are invariant among languages, but there are some switches to be set. One has to do with order of elements. In English, for example, nouns, verbs, adjectives and prepositions precede their objects; in Japanese, the comparable elements follow their objects. English is what is called a “head-first language”, Japanese a “head-last language”. These facts can be determined from very simple sentences, for example, the sentences “John ate an apple” (in English) or “John an apple ate” (in Japanese). To acquire a language, the child’s mind must determine how the switches are set, and simple data must suffice to determine the switch settings, as in this case.” (Chomsky 1992:23).

141. About whether determiners are heads or not, see Hudson 1984, 1990, Hewson 1991. For a summary and criticism see, e.g., van Langendonck 1994 (although he doesn't question the notion of 'head'). See also Haspelmath 1994.

142. Dryer mentions the possibility of modifying the head-dependent theory and talking about complements rather than dependents, to account for some of the non-correlation pairs. However there are pairs in which adjuncts and not complements are involved (such as verb + manner adverb), which ARE correlations.

143. The final, or alternate, version of the BDT that Dryer provides is the following: "Verb patterners are heads and object patterners are fully recursive phrasal dependents ..." (Dryer 1992a:116).

144. There are some facts which are still unexplained, such as the case with manner adverbs, which are object patterners although they are non-branching. Dryer suggests that this may be an effect of analogy since adpositional phrases ARE very strong object patterners and have a similar function (cf. Croft 1995:103).

145. Frazier (1978, 1979, 1985) has also dealt extensively with parsing difficulties and word order (cf. also Frazier and Fodor 1978, 1987; Frazier and Rayner 1988). Frazier's proposal is based on a "complexity metric", not unlike the proposal by Miller and Chomsky 1963.

146. Iranian, however does have SOV+NRel orders, something which causes center embedding. In this language, however, we find "alternative strategies for avoiding center embedding of relative clauses, such as extraposing the relative clause" (Comrie 1989:27).

147. Hawkins analogical principle was called the 'Naturalness Serialization Principle.' The other two principles went by the name of the 'Heaviness Serialization Principle' and the 'Mobility Principle' (Hawkins 1983). Heavy or complex constituents are said to favor clause peripheral position, cf. Jespersen 1949, Quirk *et al.* 1972.

148. Others disagree about the relative importance of parsing mechanisms. Thus, Fodor 1984, who believes in "the explanatory potential of limitations on the sentence-parsing mechanism", is of the opinion "that only rather minor, peripheral constraints on language can be specifically related to parsing problems" (Fodor 1984:10). Among the issues involved is that "[s]ome parsing problems apparently do have an impact on the grammar, but others do not" (Fodor 1984:12).

149. The clock doesn't start ticking until the first node is encountered. Until that node can be constructed, a 'buffer' is used for that material. Once the first node can be constructed, the ideal, according to the EIC, is for the other nodes to be constructed as soon as possible, with minimal delay (i.e. with the least amount of phrases possible). Multiple left branching might seem to complicate things somewhat, however, an issue which goes back to Yngve's notion of 'depth'.

150. Yngve (1960, 1961) also argued that left-branching structures were less preferable than right-branching ones, and that English, for instance, only tolerates a certain number of the former. He attributed this difference to the higher constraints on short term memory posed by left-branching structures. In actual language, especially spoken language, the truth is that multiple embeddings are rare. It remains to be seen whether multiple left-branching embeddings (in OV languages) are as (un)common as right-branching embeddings are (in VO languages).

151. Hall writes that “[i]n language comprehension there is certainly motivation for syntactic heads to serialize initially, as Hawkins suggests for NPs (1983:98). The reality of some sort of Head First Preference (HFP) enjoyed considerable support in early psycholinguistic work (e.g. Clark and Clark [1978]; Fodor, Bever and Garrett 1974; Bever 1970; Moore 1972). It was suggested that the early recognition of the head of a phrase in real time allows for the efficient building of higher level structure in some bottom-up parsing system. ... However, current distributional evidence casts considerable doubt on the HFP as *sole* determiner of word order: the current ration of OV to VO languages is roughly fifty-fifty ... then it behooves us to identify a conflicting principle of equal power which lines up heads at the *end* of their phrases in as many other languages. It is not clear to me how such a counter-principle could be motivated from a parsing point of view” (Hall 1992:78-9).

152. The original text, in Spanish, says: “Es una sintaxis que obliga a tener suspensa la atención hasta el final de la frase, en que viene la oración principal, y esto desorienta ciertamente al lector no avezado. ... Parece una paradoja que la sintaxis que observa el pueblo cuando habla euskera, se le haga a este mismo pueblo difícil cuando la ve aplicada en el escrito. Y sin embargo, así es. Y es que el lenguaje hablado, por la inflexión de la voz y otros recursos, salva fácilmente los inconvenientes (además de que no suele haber en él períodos largos o un tanto complicados).”

153. Villasante noted that “the spoken language is always much simpler and more direct, and in such a language there is no complication involved in following such [verb final] norms; but in writing one must occasionally use rather long segments, combine several subordinated clauses, etc., and then keeping to the [verb final] rules in a strict way produces quite often a heavy and obscure style for writing. And don’t let anybody say that such sentences should not be used in Basque, for there are genres and modes of conceptualizing and expressing oneself that require such a style, whether in Basque or in any other language” (Villasante 1956:14; my translation, J.A.) [“el lenguaje oral es siempre mucho mas sencillo y directo, y en el no ofrece complication mayor el expresarse segun dichas normas; pero en el escrito es forzoso recurrir muchs veces a incisos un tanto largos, combinar varias oraciones subordinadas, etc., y entonces, el guardar a punta de lanza dichas reglas contribuye no pocas veces a dar pesadez al estilo y a oscurecerlo para la lectura. Y no se diga que tales oraciones largas no deben usarse en euskera, pues hay

generos y modos de concebir y de expresarse que exigen tal estilo, tanto en euskera como en cualquier lengua”].

154. Andersen notes that “historical linguists, at least since the nineteenth century, have been concerned to make a distinction between the linguistic and the extra-linguistic (or nonlinguistic), but have disagreed both on where the boundary between these two domains should be drawn and on the very relevance to their inquiry of allegedly extra-linguistic facts” (H. Andersen 1989:9).

155. Some disagree that grammatical borrowing is not possible, e.g. Birnbaum 1984, Danchev 1984 (cited in Fischer 1992). This might seem to be an argument against the view of grammar as an extension of the lexicon (cf. Construction Grammar). Most would probably agree that actual wholesale borrowing is not the primary form of grammatical transfer. This also applies to monolingual situations, where speakers would seem to be more willing to accept new words than new constructions. On the other hand, the kind of lexical items which speakers are readily willing to incorporate to their lexicons might be of the ‘additive’ type, items which do not interfere with the lexical-semantic system. Unlike in the case of lexical items, it is hard to think of constructions which a language would ‘need’ and which wouldn’t interfere with the existing system of constructions. Minor, highly specialized constructions might be of this type, but not constructions to express common declarative sentences, for instance. As Fischer notices about Danchev’s (1984) examples of syntactic borrowing, the “concern the borrowing of idiomatic phrases (p. 51) and of prepositions in prepositional phrases (pp. 51-53)” (Fischer 1992:20).

156. Language contact has in recent years become an active field in linguistics. As Nelde remarks, “the sudden increase in language contact research in the late 1970s” is responsible for “the newly coined expression for yet another so-called ‘hyphenated linguistics’, contact linguistics (*linguistique de contact*, *Kontaktlinguistik*)” (Nelde 1992:381). “After the first world congress on the topic of language contact and conflict in Brussels June 1979, this term has gradually been accepted” (*ibid.*).

157. Gumperz and Wilson note that “[h]istorical linguists frequently point to bilingualism as a major determinant of language convergence” (Gumperz and Wilson 1971:151) There are “striking cases of grammatical borrowing among otherwise unrelated languages” (*ibid.*). And “[a]lthough lexical items are by far the most frequently borrowed, it seems clear that borrowing extends to all aspects of the grammatical systems” (*ibid.*).

158. As Seliger and Vago argue, referring back to Fishman 1972, “[t]he languages spoken by the bilingual may be said, metaphorically, to coexist in a state of competition for a finite amount of memory and processing space in the mind of the speaker. Except for the case of the so-called balanced bilingual, the languages of the bilingual develop patterns of dominance or strength, usually in relation to the domains in which the languages are used (Fishman 1972)” (Seliger and Vago 1991:4).

159. Danchev finds that occasionally word order change from SOV to SVO has taken place without a concomitant loss of nominal inflections, as in the case of the Uralic languages Finnish and Lapp, which had contact with European SVO languages (but not in more remote Ostyak and Samoyed). Danchev hints that perhaps case was not lost in these cases because of the lack of a “common core vocabulary and/or general inter-translatability” between the relevant languages in contact (Danchev 1991:119fn7).

160. Had the Roman Empire not fallen, it is hard to know what the consequences might have been for the Basque language, which could have gone the way of all most of the Western Europe’s other earlier languages. As it was, we know that those who replaced the Romans were constantly at war with the Basques for many centuries and contact with them (by now Romance speakers) was rather limited until the end of the Middle Ages and even later.

161. A different, less dependent situation on the part of the elites could have produced very different results, as shown by the Catalan case, for instance. In Catalonia, Catalan remains to this day a strong language, with real chances of becoming *the* main language of the region. On the other hand, the fact that Basque has survived to this day is quite a miracle itself. However, the sociolinguistic situation which allowed its persistent presence for thousands of years after intense contact with their neighbors started have now changed and, unless quick and drastic measures are taken, with the cooperation of a strong majority of the Basque population, are taken quickly, its chances of becoming a strong, living language are probably somewhat dim.

162. In Silva-Corvalán words: “In opposition to Weinreich ... my research indicates that the permeability of a grammar to foreign influence does not depend on *its structural weaknesses* but rather on the existence of superficially (in terms of string order, cf. Prince [1992a]) parallel structures in the languages in contact [adoption of compatible structures]. Given a primary language A and a secondary language B, the permeability of B will not be evident in the incorporation of new syntactic structures on the model of A, but first and foremost on the following: (a) the extension of the discourse-pragmatic functions of a structure B according to the model of the functions of the parallel structure in A; (b) the preferential use in B of a structure parallel to one in A to the detriment of variants in B (which may or may not convey different meanings); and (c) the loss of semantic-pragmatic constraints governing the use of the variants of a syntactic variable in B when the corresponding structure in A is not sensitive to such constraints” (Silva-Corvalán 1993:20-21).

163. Silva-Corvalán says that she has “conducted a fairly rough quantitative analysis of word order patterns in data from fourteen speakers, marking the pre- versus post-verbal placement of the subject without considering differences in types of verbs (e.g. transitives, intransitives), nor variations in number of expressed constituents in the sentence. The results show higher percentages of preverbal subjects in English dominant bilinguals, which could be interpreted to reflect a trend towards a more fixed SVX order.

However, this possible final outcome is still far from being reached” (Silva-Corvalán 1993:26).

164. Trudgill compares a low-contact language, such as Faroese, which is known to be conservative, to Danish, which is “considerably more innovating” (Trudgill 1989:231).

165. Danchev 1991 for instance mentions that “[t]here is no dearth of cross-language data showing that language contact leads to the loss of noun morphology” for instance (Danchev 1991:114).

166. In a bilingual situation, more so than in a monolingual context, we find a great diversity in the actual fluency of speakers. As Silva-Corvalán (1991) mentions about another, perhaps more drastic, bilingual situation, “[w]e find a bilingual continuum, and people can move along the continuum in their lives” (Silva-Corvalán 1991:151). Even defining who a speaker is can be complicated. Thus, speaking of Irish and Scottish Gaelic, Watson says that “[w]hen a language has been in retreat for a long time and its distribution has been shrinking at the same time that its functions have been dwindling, difficulties are very likely to arise in even such basic matters as determining just who should be considered a ‘speaker’ or a ‘member’ of the speaker community. The ‘native speaker’ population itself may not agree on who falls within that category: some people may claim speaker status when others would not accept them as such; some may say that they are *not* speakers when others would include them as speakers” (Watson 1989:41).

And speaking of Breton, Kuter argues that “[s]tatistics can be deceiving. Those who know Breton are different both from those capable of speaking it and from those who regularly do speak it as their everyday language” (Kuter 1989:75).

167. Campbell and Muntzel describe the typical language-shift situation: “Such [gradual language death] situations have an intermediate stage of bilingualism in which the dominant language comes to be employed by an ever increasing number of individuals in a growing number of contexts where the subordinate language was formerly used. This situation is characterized by a proficiency continuum determined principally by age (but also by attitude and other factors). Younger generations have greater proficiency in the dominant language and learn the obsolescing language imperfectly, if at all. Some terms employed in discussions of such language death situations are: imperfect learning, partial learning, restricted code, semi-speaker, last speaker, healthy speaker/ preterminal speaker/ terminal speaker, better/worse terminal speakers, ‘best’ speakers/fluent speakers of single sentences/inserters of words/understanders, passive bilinguals, hybrid language, intermediate bilingualism, interlanguage, creolization in reverse, deacquisition, language decay, linguistic obsolescence, broken-down or eroded language, linguistic atrophy, language attrition, etc.” (Campbell and Muntzel 1989:185).

168. Another term that is often used, and especially abused, in this context is “semilingualism” and the corresponding “semilingual”. This would supposedly refer to

individuals who are semi-speakers of all the languages they use. For a critique of this term see Romaine and Martin-Jones 1986.

169. Dorian notices how minor the effect of actual direct transfer can be in an attrition situation: “It has always surprised me, in working with ESG, how little *structural* interference there is from English in the speech of imperfect speakers, all of whom are, of course, fully fluent in English. There is a great deal of lexical interference, and some of the weakest imperfect speakers introduce unnatural syntactic patterns from English” (Dorian 1982:56).

“This suggests that we cannot simply assume that the person who is forgetting a language will substitute his dominant language’s structures for whatever he has forgotten in the second language. It is well established now that by no means all errors in the classroom learning of a second language are the of interference from the native language (e.g., Richards, 1973; Stenson, 1974). Perhaps the errors in a half-forgotten language have a logic of their own to (that is, arise from properties of the language being forgotten or from the structure and order of the forgetting process itself), and are not simple interference phenomena” (Dorian 1982:57).

170. Andersen uses the terms LA for a “semi-speaker”, or “* attriter”, “a person whose competence in a language has eroded as a result of language attrition” (Andersen 1982:83-84). He also uses the term LC for a “‘linguistically-competent’ individual, one who is fully fluent and competent in language X, whether as a native or a non-native speaker” (Andersen 1982:84).

171. According the One-to-One principle, “‘an interlanguage system should be constructed in such a way that an intended underlying meaning is expressed with one clear invariant surface form (or construction)’ (Andersen [1984]:79)” (Andersen 1989:388).

172. These parallels have been noticed for quite a long time now. Romaine notes that “Dressler and Wodak-Leodolter (1977) were among the first to suggest an analogy between language death and pidginization, while Trudgill (1978) drew parallels between language death and creolization in referring to the former as *a kind of ‘creolization in reverse’*” (Romaine 1989:375-76; my italics, J.A.).

173. Romaine argues that “[i]n normal communities the expectation is that adults act as brakes on the innovations produced by children so that analogical and other deviant forms like *foots* get corrected and do not persist. In the case of dying and pidgin languages it may be that children have greater scope to act as norm-makers due to the fact that a great deal of variability exists among the adult community” (Romaine 1989:372-73).

174. Girard’s well-known contrast between ‘analogous’ and ‘transpositive’ languages was later adopted by Adam Smith and the compilers of the *Encyclopédie* (Beauzée, under the entry ‘Langue’, vol ix (1765); ... ‘Analogous’ languages include Hebrew, French, Italian,

and Spanish where the word order is basically SVO (Subject + Predicate + Object) and similar (analogous) to the analytical, 'natural', order of the thinking process, since here, it is said, the agent comes first, followed by the action, and then by whatever is involved in the action (objects or adverbial phrases). The grammar of these languages is relatively lacking in inflected forms. Old Slavic and Latin, on the other hand, are 'transpositive' languages, in the sense that the 'ordo naturalis' may be changed by, for example, transposing the subject after the predicate and even after the object. The word order is freer and the grammar rich in inflected forms. A third, 'mixed' (or 'amphilogique') type is represented by Ancient Greek and German" (Ramat 1995:29)

175. Osgood and Tanz, for instance, argue for the 'naturalness' of SVO order: "Given both the nature of the physical world with its space/time coordinates and the nature of human organisms generally, it seemed likely that the most basic simple cognitions would be Actor/Instrument Action Recipient/Object for action relations and Figure State Ground for stative relations. Even the not-so-astute reader will recognize these simple cognitions as characterizing SVO languages" (Osgood and Tanz 1976:537). But, if this is so, they wonder, how do we account for the existence of so many non-verb-final languages?: "But if the "natural" mode of prelinguistic cognizing for all humans is, indeed, SVO, how then does one explain the existence of languages with dominant SOV and VSO orders?" (Osgood and Tanz 1976:539).

176. The correlation between language contact and SVO order is not absolute, however. As Romaine mentions, "[u]nlike many pidgins and creoles, dying languages do not on the whole show a tendency to uniformity of word order. Moreover, there are pidgins and creoles which do not show SVO word order. ... Mithun ([1989]) found that the pragmatic basis of Cayuga word order remained intact..." (Romaine 1989:377).

Chapter 2

Typological overview and overview of the corpus

2.1 INTRODUCTION

This purpose of this chapter is to situate the study of Basque word order by providing some background about the typological characteristics of Basque, including major characteristics of word order typology, and, secondly, characteristics of the corpus which has been used and which will be analyzed in the following chapters.

Section 2.2 presents a basic typological analysis of Basque, exclusive of word order, including primarily typological parameters which have been said to correlate with word order parameters.

Section 2.3 gives an overview of previous studies of Basque word order and some of the conclusions they have reached and the questions they have raised. Issues dealt with here include basic word order, branching direction, and the coding of information structure and pragmatic roles by means of word order.

Finally, section 2.4 will look at the characteristics of both the spoken and written corpora which were used in this study, including characteristics of the speakers in the case of the spoken corpus. I will also discuss some major and obvious findings about word order, and about the differences found among the corpora, stemming from this preliminary analysis.

2.2 A TYPOLOGICAL OVERVIEW OF BASQUE

2.2.1 Dryer's correlation pairs

Before analyzing the word order of major Basque clauses, I would like to go over a number of other typological characteristics which have been associated with the different 'basic word orders' that are found cross-linguistically (the 'correlations' or 'harmonies' of Greenberg 1966). As we can see in Table 2-1, if we look at Dryer's (1992a) correlations, for instance, Basque would seem to qualify as an (S)OV language (cf. Chapter 1).

True OV correlations		Non-correlations		Controversial pairs	
Postpositions	Yes	Adj-N	No	Rel-N	Yes
Genitive:N	Yes	Dem-N	No	Verb-Aux	Mostly
Std.-Adj of Compar.	Yes	Intens-Adj.	Both	S-Complementizer	Yes
PP-Verb	Mostly	V-Neg	No	S-Subordinator	Yes
Manner Adv-Verb	Mostly	Verb-Tense/Aspect	Mostly	Noun-Article	Yes
Predicate-Copula	Mostly			Noun-Plural	Yes
Verb- <i>want</i>	Mostly			Subject-Verb	Yes

Table 2-1: True correlations, controversial correlations, and non-correlations between OV and other formal properties according to Dryer 1992a and their application to Basque.

Basque for the most part displays all of the 'true' correlations, as specified by Dryer, and even all of the 'controversial' ones. In this section I will take a quick look at these and other typological characteristics of Basque.

2.2.2 Declension and postpositions

Basque has 14 phrase-final nominal cases, as well as 'postpositions', which are also strictly phrase final, formed out of erstwhile nouns and adverbials governing a complement noun phrase in one of the core cases (absolute, genitive, dative). A

number of morphophonemic alternations, as well as some complications involving ‘human nouns’, obscure a fairly regular declension schema, which I show in Table 2-2 with the case endings for inanimate nouns ending in a vowel.¹ (Although the traditional labels for the different columns are definite and indefinite, referential and non-referential might be more appropriate labels, as I will argue in Chapter 3.)

Case	Indefinite	Definite Sg.	Definite Pl.
Absolutive	-	A	AK
Ergative	K	AK	EK/AK
Dative	r-I	ARI	EI
Poss. Genitive.	r-EN	AREN	EN
Destinative “for”	r-ENTZAT	ARENTZAT	ENTZAT
Instrumental	Z	AZ	EZ/ETAZ
Inessive	TA-N	AN	ETAN
Locative genitive	TA-KO	KO	ETAKO
Allative	TA-RA	RA	ETARA
Final allative	TA-RAINO	RAINO	ETARAINO
Ablative	TA-TIK	TIK	ETATIK
Partitive	r-IK	-	-
Prolative	TZAT	N/A	N/A

Table 2-2: Basque declension for inanimate nouns ending in a vowel (*r* = linking consonant)

Case markers are always the last element in a noun phrase and they are attached to the last word in it, be it the noun, an adjective, a demonstrative, or, if these three elements are elided, a nominal modifier such as an adjective or even a relative clause.

In addition, Basque has a number of so-called postpositions or adverbial (function) words, which, as I said, are typically nouns or declined erstwhile nouns which govern a noun phrase in one of the major cases. Table 2-3 shows the major Basque postpositions according to the case they govern and their function (cf., e.g., Euskaltzaindia 1993:499-501).

Complement	Postpositions
Genitive	LOCATIVES (add a locative suffix): <i>gain-</i> “top”, <i>arte-</i> “between”, <i>aurre-</i> “front”, <i>azpi-</i> “under”, <i>barru-</i> “inside”, <i>ondo-</i> “next”, <i>inguru-</i> “around”, <i>pare(an)</i> “next”, <i>arte-</i> “interval; between; until”, <i>alde-</i> “side, next” MANNER: <i>kabu(z)</i> “in evaluation of”, <i>parte(z)</i> , <i>bide-z</i> , <i>bi-tarte-z</i> “by means of”, <i>gisa</i> “in manner of”, <i>leku-an</i> , <i>orde-z</i> “instead of”, <i>zain</i> “awaiting”, <i>bila</i> “searching for”, <i>arabera</i> “according to”, <i>esku</i> , <i>menpe(an)</i> , <i>truk(e)</i> “in exchange for”, <i>eske</i> “requesting”, <i>mail-an</i> “in the category of”, <i>alde</i> “in favor of”, <i>aurka</i> “against”
Absolutive or. Partitive	<i>gabelbarik</i> “without”, <i>eze-an/ez-ik</i> “except”, <i>falta-n</i> “lacking”
Dative	<i>buru-z</i> “about”, <i>esker</i> “thanks to”, <i>begira</i> “looking at”
Locative	<i>zehar</i> “across”, <i>gaindi</i> , <i>barrena</i> , <i>gora</i> “upwards”, <i>behera</i> “down(wards)”
Ablative	<i>kanpo(ra)</i> , <i>at</i> “outside of”
Instrumental	<i>gain(era)</i> , <i>bestalde</i> “besides”, <i>lekore</i> , <i>landa</i> “outside of; except for”

Table 2-3: Basque ‘postpositions’. The majority are themselves nouns with some case ending acting adverbially (cf. e.g. Zubiri and Zubiri 1995:2.6.6)

It is in fact hard to decide whether these are actual adpositions or merely complement taking nouns. The distinction would be in fact quite arbitrary. The more the phrasal head performs a linking function the more it looks like an adposition. If we decide to call them postpositions, as many have done, then the fact that Basque has postpositions derives directly from the fact that adpositions are derived from nouns with phrasal complements, which, always come before the noun (i.e. are left-branching), with the linking case acting as a go-between (cf. the Relator Principle).

2.2.3 Nominal modifiers: The genitive suffixes

In addition to nominal compounds in which the modifying noun simply precedes the head noun without any overt marker (as in English), Basque has two genitive suffixes which relate a nominal to another (the head) in a modifying relationship, one which codes

a primarily possessive relation (cf. the *-(r)en* suffix) and one which codes a primarily locative relation (cf. the *-ko* suffix). Both of these markers have extended meanings, however.²

As mentioned, these genitive case markers are suffixes which attach themselves to the rightmost word in the dependent phrase and the dependent (complement) genitive phrase always precedes the head noun. Thus Basque is a consistently Gen-N language. As I said, this also accounts for the fact that postpositions and other function words follow the phrase they ‘govern’, since they are typically formed historically by means of the genitive construction.

2.2.4 Modifiers in apposition

Occasionally we find modifying phrases in apposition to the head noun which follow, rather than precede that noun. Although they are in apposition, such phrases seem to form a semantic, and intonational, unit together with the head noun. We can see an example of such a phrase in (2.1) (from Goenaga 1980:201).

- (2.1) <Altuberen legeak | hitzen ordenari buruz>, ezagunak dira.
Altube:GEN laws:DEF words:GEN order:DEF:DAT head:INSTR know:PFV:DEF:PL
they.are
“<Altube’s laws | about word order> are well known.”

In this complex phrase it is obvious that intonationally there are two parts, each one receiving an equal accent. On the other hand a single unifying intonational contour makes those two parts hang together, thus reflecting the semantic closeness of the two elements and the modifying relation of the second with respect to the first. Such appositional phrases could conceivably be a diachronic mechanism by which phrasal

complements of nouns ‘change position’, supposedly under some kind of pressure, an internal or external motivation. The functional load of these phrases seems to be rather small, but it should not be underestimated.

In (2.2) we see another example of a nominal phrase in apposition to (and to the right of) the head nominal (from a personal letter). The appositional phrase seems to form a unit, at least semantically, with the head noun it modifies.

(2.2) Ondoren datoz <nire erantzunak | zure galderei>.
next:GEN they.are.coming my answers:DEF your questions:DEF:DAT
“Here come <my answers | to your questions>.”

2.2.5 Demonstratives

Demonstrative adjectives, coding three degrees of proximity, when present are in most dialects always the last element in the phrase. Thus they are the element of the noun phrase which receives the case endings (when present). When there is no overt head noun they act as pronouns. There are some minor sub-dialects of the Bizkaian dialect, where they seem to have ‘moved’ to the front of the phrase, while retaining their case markings and agreements (cf. e.g. Villasante 1980:231).

2.2.6 Numerals

Basque numerals are inconsistent as to their position with respect to the noun they typically modify (when not used as pronouns). The numeral *bat* “one” is phrase final, just like demonstrative and other determiners. In fact, this numeral is often used as a determiner, i.e. as an indefinite marker, although with a lesser degree of grammaticalization than, for instance, English *a* (also historically derived from the

numeral ‘one’). Indefinite noun phrases are also often marked in the same way as definite ones, thus neutralizing the distinction.³

The numeral *bi* ‘two’ is also phrase-final in some western dialects. All other Basque numerals, however, are phrase initial. That is, unlike other non-branching modifiers, such as adjectives and demonstratives, numerals in Basque, with only those exceptions I just mentioned, are actually phrase initial.⁴ However, as we have seen, the relative position of non-branching modifiers with respect to the noun in the noun phrase does not seem to be correlated with word order type and there is much variation.

2.2.7 Relative clauses

Basque relative clauses are formed by means of an invariant suffix which attaches itself to the end of a (finite or non-finite) modifying clause. They thus precede the noun they modify.⁵ Finite relative clauses are formed with the suffix *-(e)n* attached to the finite verb. Although historically this suffix is the same as the genitive suffix, it is not clear that it is actually felt by speakers to be the ‘same’ suffix synchronically. We can see an example of a finite relative clause in (2.3) below.

- (2.3) *Ogia erosi du-en gizon txiki horrek istripu bat izan du*
bread:DEF buy:PFV he.has.it:REL man small that:ERG accident one be/have:PFV
he.has.it
“That small man who bought (the) bread had an accident”

Because there is no relative pronoun linking the relative clause to the head noun and thus indicating the role that the referent of the head noun plays in the modifying clause, relativization is, in most dialects, restricted to those arguments of the clause which are coded on the verb.⁶

Non-finite relative clauses in Basque are formed with the other ‘genitive’ suffix, namely *-ko*. These relative clauses are actually ‘adverbialized’ or nominalized clauses to which the genitive marker is added. We can see an example of the former in (2.4).

- (2.4) <Hor eseritako> emakumea gure irakaslea da.
there sit:PFV:ADV:GEN woman:DEF our teacher:DEF she.is
“That woman <sitting there> is our teacher.”

Relative clauses of either type may themselves be declined, just like any nominal modifier can, when the head noun is ‘elided’, i.e. ‘understood’ (cf. English *(the) one(who/-ing)...*).

2.2.8 Adjectives

Although the word class of adjectives performs a similar modifying function as genitives and relative clauses, in Basque they always follow the noun. As I mentioned in Chapter 1, it was thought at one time that adjectives in verb-final languages were primarily pre-nominal, i.e. that the order Adj-N correlates with OV order. This has been challenged by Dryer (1988b), who claims that there is no such correlation.

Thus it seems that adjectives, as a typically non-branching category, do not have the same tendency to precede their heads in OV languages as branching categories (such as adpositional phrases or relative clauses) do. This fact may be related to the iconic preference to keep nominal modifiers near the noun, which is accomplished by spreading them around, with some being prenominal (such as genitives) and others postnominal (adjectives).

It should be noted that many concepts which in other languages might be expressed by means of adjectives, in Basque are expressed by means of modifying nouns,

which precede the head, or denominal derived nouns, by means of the genitive construction, which also precede the head noun, e.g. *lurrezko ontzia* “the earthen pot”. Deverbal adjectives, however, formed by zero derivation from the perfective participle of the verb, always follow the noun, e.g. *gizon nekatua* “the tired man”.

Adjectives are also occasionally found in apposition to the noun, i.e. they follow the closure of the nominal phrase proper (following a determiner, for instance). This seems to happen often with indefinite (and thus ‘new’) nominals (cf., e.g., Agirre Berezibar 1991:633). These adjectives thus seem to act as predicates, not as strict modifiers, of the noun they go with.⁷

2.2.9 Comparisons

Basque comparisons are done in the typical manner of verb-final languages, with the standard of comparison preceding the comparative adjective, and linked to it by the relator *baino*. We can see an example of such as construction in (2.5).

- (2.5) Zu <nire anaia *baino* handi-ago-a> zara
you my brother:DEF than big-MORE-DEF you.are
“You are bigger than my brother”

The fact that the standard of comparison may contain a (left-branching) sentential structure may be the primary motivator for this type of modifier preceding the modified head.

2.2.10 Other noun related characteristics

Basque is eminently a suffixing language. It does have a few long-standing derivational prefixes, however, obtained through Latin borrowings: *des-* “un-”, and *ber-*

“re-”, e.g. *desegin* “undo”, *berregin* “redo”. In addition there are a few prefixes which have been introduced more recently to translate foreign Latinate terms, such as *aurre-* “pre-” and *azpi-* “under-”, e.g. *auresan* “predict” (cf. *esan* “say”), *azpiegitura* “infrastructure; underlying structure” (cf. *egitura* “structure, composition”).

Finite (conjugated) verb forms also show a number of prefixes in addition to suffixes and infixes. The origin of such prefixes is for the most part lost in time, although some of these prefixes are ‘motivated’ by (similar in form to) free-standing pronouns.⁸

2.2.11 Verb-related characteristics

2.2.11.1 *The Basque conjugation*

Most Basque verbs are conjugated only periphrastically, i.e. their verb ‘forms’ consist of a non-finite, or participial, form of the verb and a finite auxiliary, in that order (except for in negative clauses, as we will see).⁹ The auxiliary codes the tense and some aspects and modalities, as well as the person and number of the three main cases of arguments of the verb: ergative, absolutive and dative/benefactive.¹⁰ There are a handful of (very common) verbs which, in addition to the periphrastic or analytic conjugation, also have synthetic past and present continuous imperfective verb forms.

The non-finite and finite verbs in modern Basque form a phonological unit in affirmative sentences. That is, they are phonological words and they may even be on their way to becoming fused as lexical words in some dialects.¹¹ This means that for the purpose of establishing preverbal position the two parts of the word act as a verbal unit. This property doesn’t extend to other verbal units, such as those containing a non-

referential noun and an auxiliary verb, which together form a semantic unit, e.g. *maite izan* “love”. Here the auxiliary *is* the verb, and if some other element must appear in preverbal position, the noun must be placed *after* the verb. (Phonological fusion does extend, however, to a couple of very common N+V verbal complexes in which the noun has taken verbal characteristics, cf. *nahi* “want” and *behar* “need”.)

The periphrastic verb complex is not a phonological unit in all the constructions of Basque, however. In negative clauses, the two parts of the verb are separated *and* inverted, as we will see in more detail in Section 2.3.5 below and in Chapter 7. In asserted negative sentences the finite verb, preceded by the negative particle *ez* cliticized to the front of it, is (typically) in rheme initial position and the non-finite verb is obligatorily postposed, along with any other complements. Focused elements may appear in rheme-initial position in these sentences *if* they are outside the scope of negation, as we will see.

In the periphrastic conjugation, the main verb may be in one of four different forms: root (used, e.g., with some irrealis modalities), imperfective (suffix *-t(z)en*), perfective (suffix *-tu* for the regular cases, or else *-i*, *-n* or zero), and ‘future’ (either *-ko* or *-en*, the ‘genitive’ suffixes, depending on the dialect). The auxiliaries are traditionally categorized by modality as either indicative, potential (with root and epistemic versions), conditional (condition and consequence), subjunctive, or imperative. Most of these divisions can be either present or past (cf. Lafon 1943; Euskaltzaindia 1987).

2.2.11.2 *The position of the verb*

Finite verb-forms have a well-known formal constraint in Basque. In asserted clauses the finite verb cannot be rheme- or assertion-initial (whether it is preceded or not by clause-internal 'setting' phrases, such as topics). Thus, for instance, a finite verb that is the only element of the rheme or which finds itself in rheme initial position must be preceded by the emphatic particle *ba-*, whether the assertion is emphatic or not (true emphasis is coded intonationally). This generalization does not extend to periphrastic verbs, however, since, historically at least, the non-finite verb preceded the finite verb proper.

In other words, in finite, asserted clauses the finite verb proper must be preceded by a (single) element such as a complement, unless the verb itself is 'focal' and marked as such or has is preceded by the non-finite part of the verb.¹² This phenomenon is related to the fact that in Basque asserted clauses the most important, or informative element of the rheme typically comes in rheme-initial position and in the default case the most important element is not the verb. As we shall see, the verb is rheme-initial when either it is 'focal' (e.g. contrastive) or when the whole rheme is emphatic or contrastive (not just plain 'new').

There is an exception to the rule I have just mentioned. In Basque there is a marked construction in which the focal or most important element can be 'extraposed' to rheme-final position. That this construction is marked is evidenced by its intonational properties. Such clauses have to main accents on the rheme, one on the rheme-initial verb and one on the focal element. The 'important' element is thus intonationally separated from the rest of the rheme. There may even be a pause in between the two,

meaning that the rheme could consist of two full intonational units, and not just one unit with two parts.

In VO languages a totally parallel construction is found (e.g. *Thàt wás, márvelous*). The only difference is that in VO languages this construction is typically not interpreted as involving ‘movement’, or ‘dislocation’, of any kind (with respect to the unmarked construction). The effect, however, is identical, namely giving the salient element its own intonational contour and thus, iconically, marking its salience. In this construction a rheme-initial finite verb doesn’t receive the *ba-* marker since it is the extraposed element, and not the verb or the polarity, that is focal (cf. Euskaltzaindia 1987:494-5).

We will see in the following chapters that this construction is becoming relatively less marked in spoken Basque than it is either in written Basque or in languages such as English. The motivation for the ‘de-marking’ of this construction may be found perhaps in a ‘conspiracy’ to produce structures which are on the surface verb-medial. The ‘focus-extraposition construction’ is actually the unmarked construction when the focal element is an asserted clause, i.e. a completive clause, which are predominantly postverbal, as we shall see. But notice that unlike in VO languages, the main verb in such sentences has to be accented, thus breaking the rheme or assertion in two, intonationally speaking, and it is thus intonationally ‘marked’.

2.2.11.3 Clausal constituency

As I said, the relative order of the (non-finite) ‘main’ verb participle and the auxiliary is obligatorily that one: Verb-Auxiliary. Some recent analyses of Basque

grammar assume that the structure of a clause with a periphrastic verb is such that the auxiliary is the main verb, which takes a non-finite clause as its complement, such as in (2.6) cf., e.g., Wilbur 1979. These ‘hypothetical structures’ are in square brackets.

(2.6) [Aitak [amari ogia *ekarri*] *du*]
father:DEF:ERG mother:DAT bread:DEF bring:PFV he.has.it
“Father brought Mother the bread”

This, of course, is the same type of structure as that hypothesized for English clauses. There are problems with this syntactic structure, however. To begin with, we have seen that the two parts of the verb act as a unit phonologically and informationally. The structure in (2.6) may be historically correct, but there is no evidence that in main affirmative constructions the non-finite verb and its complements form a unit dependent on the finite auxiliary.

Secondly, it doesn’t seem to me that the dative argument (*amari* “to mother”) in any way forms a structural unit with the absolutive one and the (main) verb. Since there can only be one rhematic element before the verb, i.e. here *ogia* “the bread”, the dative argument *amari* “the mother” must be a ‘setting’ element, much like the ‘subject’ *Aitak* “father”, a fact that is clearly marked intonationally. Thus, we could say that the units in this sentence should actually be as marked in (2.7): two setting elements and an assertion/predicate *about* those settings.

(2.7) {SETTING Aitak } {SETTING amari } {ASSERTION ogia <*ekarri du*> }

A very similar structure to this one has been proposed by de Rijk 1978, who clearly recognized that the only complex unit below the sentence is the one including the element before the verb and the verb (he doesn’t discuss where postverbal elements

belong), which varies according to which element is placed before the verb. As to the kind of element this is, he has a simple answer: “What else, then, could it be but VP?” (de Rijk 1978:107). For de Rijk, who works in a transformational framework, this VP is a surface unit, since the underlying structure of the clause is flat. Others have also claimed much the same thing, arguing that Basque is a ‘non-configurational’ language (e.g., Mitxelena 1981:79; Rebuschi 1985b).¹³ What many authors do not always recognize is that the ‘forces’ which tie verbs to some complements more than others are not syntactic, but rather primarily informational and also in part semantic (though never strong enough to qualify as constituency the way noun-phrases or clauses are structural constituents).

2.2.11.4 Evidential and aspectual particles

Before briefly looking at other auxiliary verbs, I should mention a special class of evidential and aspectual particles which cliticize themselves to the front of the finite verb: *ote* or *ete*, which is used in rhetorical questions, indicates doubt or suspicion (“perhaps”); *omen* or *ei*, indicates likelihood or expectation (“supposedly”; ‘quotative’); *bide* indicates that something is evident (“obviously”); imperfective aspectual *ohi* indicates repetition (“usually, used to”); and, in some dialects *al*, derived from *ahal* “ability, possibility”, is used in polarity question words. These clausal operators are completely integrated into the verb. They are also the only element that may intervene between the negative word *ez* “not” and the finite verb.

2.2.11.5 Other ‘auxiliary’ verbs

In addition to the periphrastic conjugational possibilities mentioned above, Basque also makes use of other, much less-closely integrated ‘auxiliary’ nouns and verbs,

which act as aspect and modal verbs. We can divide them into two groups: one which has only 3 verbs which take perfective non-finite (*-tu*) clauses as their ‘complements’ and a much larger one which take imperfective non-finite (*-t(z)en*) clauses as complements.

2.2.11.5.1 *Clauses with perfective complements*

Thus, for instance, some dialects use the noun/verb *ahal* ‘ability’ to indicate the (root) potential modality. The noun/verb *nahi* ‘desire’ (in some dialects *gura*) is used to indicate the volitional modality. The noun/verb *behar* ‘need’ is used to indicate necessity modality, as well as the impending (or ‘present-relevance’) future tense. These three words, which are nouns at heart, have some verbal properties in this context. They all take clausal complements in *-tu* (perfective non-finite participle), cf. (2.8). In fact, they are the only ‘verbs’ which take such complements (in addition to *egin* ‘make, do’, as we will see).

(2.8) [... V-*tu*] *ahal/nahil/behar* AUX_{FIN}

The non-finite clauses which, in theory, are complements of these noun/verbs, in practice have gone a long way towards melting into the upper clause. Restructuring is almost a *fait accompli* in these clauses, as evidenced by the fact that the elements of the lower clause do not have to be adjacent anymore, although they can be. We can see in these clauses the influence of two semantic forces intersecting: one which holds the lower verb’s elements together and one which blends the lower verb to the upper aspect-modality verb which tends to dissolve the boundaries of the lower clause. This, of course, is the well known phenomenon sometimes known as ‘clause union’, a type of ‘syntactic restructuring’.

2.2.11.5.2 Restructuring with 'behar'

The breakup of the 'lower clause' into elements which go before and after the verb is rare in standard written Basque, though the whole lower clause is sometimes postposed as a whole. Such break-up is not rare in speech, however. In the Spoken Basque Corpus 10/35 instances of clauses with *behar* "need", for example, have the lower verb come before *behar* and a complement of the lower verb come after the main verb (cf. Table 2-4). Only in 8/35 do lower verb complements precede the lower verb. In an additional 12/35, the lower verb has no complements. An additional 3 are negative sentences, 2 with postverbal *behar*, and 2 with a focused element in which the lower verb, *behar*, and the auxiliary, act as a verbal-unit.¹⁴

	Order	Number
a)	<Vtu> behar Aux	12
b)	<... Vtu> behar Aux	8
c)	<Vtu> behar Aux <...>	10
d)	ez Aux <Vtu> behar <...>	1
e)	ez Aux <Vtu> behar	1
f)	ez Aux behar <Vtu>	1
g)	Focus <Vtu> behar Aux	2
	TOTAL	35

Table 2-4: The structure of *behar* clauses in the Spoken Basque Corpus.

2.2.11.5.3 Restructuring with 'nahi'

With *nahi* "want", the ordering preferences are slightly different. First of all there is an overall preference for the lower verb to follow the main verb, something which is quite rare with *behar*. This suggests that the lower verb is more independent from, or less fused to, the main verb ('want'), even though restructuring is equally strong here. In

27/42 of the affirmative clauses with *nahi* in the Spoken Basque Corpus the lower verb, typically with all of its complements, follows the main verb. Considering that some of the these 42 *nahi* clauses are dependent clauses, and thus more likely to be verb final (including some relative clauses), that is a rather high number.

As we can see in Table 2-5, in 23 of the 27 clauses in which the lower verb follows the main verb, the lower verb has complements, and in all but one of these clauses the lower verb's complements are also postverbal: in 8 clauses the complements precede the lower verb, in 11 it follows them and in 3 there is one on each side. As we can see the picture is rather complex, but there is a strong tendency for this sentential complement to be postverbal, much like the finite completive complements mentioned earlier.¹⁵

Order	Number	%
(a) <Vtu> <nahi Aux>	3	7.1%
(b) <...Vtu> <nahi Aux>	8	19.0%
(c) <nahi Aux> Vtu	4	9.5%
(d) <nahi Aux> <...Vtu>	8	19.0%
(e) <nahi Aux> <Vtu ...>	11	26.2%
(f) <nahi Aux> <...Vtu...>	3	7.1%
(g) ... <nahi Aux> Vtu	1	2.4%
(h) Vtu <nahi Aux> ...	4	9.5%
(i) TOTAL	42	100.0%

Table 2-5: The structure of affirmative *nahi* “want” clauses in the Spoken Basque Corpus.

2.2.11.5.4 Restructuring of imperfective non-finite complements

Aspect-modality verbs which take imperfective non-finite complements seem to be much less advanced in the restructuring process. In other words, the lower clause seems to act much more like a semantically, and thus also syntactically, independent unit

(i.e. as a complement). These are verbs such as *hasi* “begin”, *saiatu* “try”, etc. In Table 2-6 we can see the verbs of this type of which there were 10 or more tokens in the Spoken Basque Corpus.¹⁶

Construction	Translation	Clause type	No.	#V	%V
Ari_Tzen	<i>Be busy/engaged in</i>	ASP	48	17	35.4%
Egon_Tzen	<i>Be -ing (imperf.)</i>	ASP	13	11	84.6%
Hasi_Tzen	<i>Begin -ing</i>	ASP	88	76	86.4%
Lagundu Tzen	<i>Help -ing</i>	COMPL	12	11	91.7%
TOTAL			161	115	

Table 2-6: Constructions with non-finite clausal COMPLs and their types and numbers in the Spoken Basque Corpus.

As we can see, there is a strong tendency to place the complement clause after the main verb. In 155/190 of clauses in the Spoken Basque Corpus containing one of these imperfective complement clauses, the complement clause goes after the verb. The most common of these verbs by far is *hasi* “begin”, with 88 tokens. Of these, 76 or 86.4% have the complement after the verb. The lower clauses which precede the main verb are all verb-final or single-verb, but among the ones that follow we find greater variety. The majority are single-verb (29) or verb-final (32), but 10 are verb-initial and 5 are verb-medial, suggesting an incipient tendency to ‘re-unify’ the two verbs.

In the written corpus there are only 5 examples of *hasi* with a *-tzen* complement. In fact in 3 of these *hasi* is adjectivized (*hasia*). In all five cases the complement precedes the main verb. In a search on a different, three novel corpus, postverbal *-tzen* complements with *hasi* were found in about ¼ of all tokens.

Another common aspectual verb is *ari* “be busy; occupied”, which is the main verb for expressing continuous imperfective aspect for verbs which do not have a

synthetic conjugation. There are 48 tokens of sentences with *ari* in the Spoken Basque Corpus. Unlike for *hasi*, in a majority of these clauses, or 31/48 (64.6%) the non-finite complement clause precedes the main verb. As for the order inside the ‘lower clause’, in those clauses which precede the main verb they are all single verb (17) or verb-final (14). In the ones that follow the main verb, however, 8 are verb-final, 5 single verb, 3 verb initial, and 1 verb-medial.

There are 9 instances of *ari* in the written corpus, only 6 of which have a *tzen*-complement clauses.¹⁷ In 5 of the these 6, the *tzen*-clause precedes the main verb (3 *XtzenV*, 2 *SXtzenV*). In 1 token the dependent clause, a single verb, follows the main verb.

2.2.12 Subordinating relators and dependent clauses

Basque subordinators are for the most part clause-final and suffixed to the verb. These clauses are typically verb-final. The reason why these subordinators are verb-final seems to be found in the diachronic origin of such subordinators. Non-finite subordinate clauses, which are derived from nominal forms of the verb, in *-t(z)e-*, take nominal, and thus phrase-final, endings as subordinators. In (2.9) we can see one such purpose clause, which takes a genitive ending after the nominalization ending, much like the English version uses an erstwhile dative preposition.

- (2.9) Aita etorriko da <ogia ekar-tze-ko >
 father come:FUT he.is bread:DEF bring-NOMIN-GEN
 “Father will come <to bring (the) bread>”

Some finite subordinate clauses have final subordinators for the simple reason that they are formed from finite relative clauses with, e.g., a temporal noun head, or simply a

temporal case ending added to a headless relative clause. In (2.10a) we see a ‘plain’ temporal clause, and in (2.10b) a more specialized one.

- (2.10) a. *Ama etorri de-n-ean*
 mother come:PFV she.is-REL-LOC
 “When mother comes”
- b. *Ama etorri de-n bakoitz-ean*
 mother come:PFV she.is-REL each time-LOC
 “Each time mother comes”

A ‘relative clause’ may also take an adverbial ending *-z* and become an adverbial clause. Such an adverbial clause may itself be the complement of an adverb. We see an example of such an ‘adverbial’ clause in (2.11), which may or may not have the adverb *gero* ‘later’ following it. (Without *gero* this clause can only be a ‘setting causative’; with *gero* it is ambiguous between a temporal sense and a ‘setting causative’ sense.)

- (2.11) *Ogia ekarri zuenez (gero),*
 bread:DEF bring:PFV s/he.had.it:REL:INSTR (after/later)
 “Since s/he brought the bread, ...”

Other dependent clauses are formed on the pattern of comparative constructions and thus are also verb- and subordinator final, e.g. *V-tu baino lehen(ago)* “before V-ing”, lit. “V but/than (more)before”. Following a similar pattern we commonly find what would seem to be erstwhile adpositional phrases following an adverbial clauses, e.g. *V-tu eta gero* “after V”, lit. “V and later”.

There are a couple of subordinators which actually precede the finite verb: *ba-* “if” and *bait-* “because” (in post-clausal ‘dependent assertions’). We can see an example of the former in (2.12). Such clauses seem to be erstwhile emphatic clauses (*ba* for

instance seems to be related to the emphatic particle, perhaps under the influence of Romance, (in Spanish *si* = “yes” and “if”).

- (2.12) Ogia ekartzen ez baduzu,
bread:DEF bring:IMPFV not IF:you.have.it
“If you don’t bring the bread, ...”

These clauses are also traditionally verb-final and do not invert in the negative, for instance, even though the subordinator is not strictly final. In modern Basque, however, postposing and inversion in such clauses is not unheard of (cf. below).

The main subordinator for asserted, affirmative complement (completive) clauses is *-la*, apparently derived from an adverbial ending.¹⁸ These clauses, when they are preverbal, which is hardly ever in the Spoken Basque Corpus and about 40% of the time in the Written Basque Corpus, are strictly verb- and subordinator-final. In these cases it seems that the whole clause is focal. When they are postverbal, however, they seem to act often as main assertions, and they display the same ‘free’ (pragmatic) word order characteristics as asserted main clauses. The suffix *-la* (sometimes *-larik*) has another use as with (non-asserted) adverbial (absolute) clauses in setting position. These setting clauses are quite strictly verb-final.

2.2.13 Constituent order in dependent clauses

As I said above, many Basque dependent clauses are strictly verb-final. However, it is interesting to note that clauses which are often strictly verb-final in writing allow postverbal elements in the spoken language, and that clauses which allow ‘postposing’ in the standard language, make greater use of this ‘right’ in speech. I refer the reader to the figures in Appendices 2.1-2.4, where the statistics for all the finite and non-finite

dependent clauses in the written and spoken corpora can be seen. Although the constructions represented there do not constitute all the subordinate construction types available in Basque, they are by far the most common ones and a very representative sample of those constructions and their relative proportions in speech and writing.

2.2.13.1 Finite dependent clauses

As we can see in the mentioned tables most finite clauses in the written corpus (cf. Appendix 2.1) are strictly verb final. There are some exceptions, however. They involve first of all clauses with pre-verbal subordinators: *bait-* and *ba-*. First, *bait-* ‘tail’ causatives (cf. Chapter 3) are about as often verb-final as not (whether they also have the optional clause initial ‘subordinator’ *zeren* or not). Conditional *ba-* clauses are 16% non-verb final. Completive *-la* clauses are $\frac{3}{4}$ verb-final or single-verb, and non-verb final in 26% of all non-single-verb affirmative clauses.

In the spoken corpus, the number of constructions which allow some degree of postposing (at least for some speakers) is much larger as is the amount of postposing in the clauses in which in writing too postposing is allowed. Complement *-la* clauses have postverbal elements in 37% of all non-single-verb affirmative clauses (as opposed to 26% in the written corpus). The *-lako* (clause-internal *and* clause-external) causative has postverbal elements in 9 out of the 22 non-single-verb affirmative clauses (none in the written corpus). The temporal clause in *-nean* which in writing is strictly verb-final, has postverbal elements in almost 30% of the non-single-verb affirmative clauses. Even headless relative clauses, which should always be verb final (as relative clauses always

are in both speech and writing), display postverbal elements in over 31% of the non-single-verb affirmative clauses.

In addition, in the spoken corpus there is a not insignificant number of complement clauses, 7.7%, which use the modal adverbial construction (*nola* “how”) instead of the post-verbal complementizer *-la* (cf. English: *She told me how she was going to come*). In this construction, what was an manner ‘pronoun’ in focus position has become, in effect, a clause initial subordinator or complementizer (not necessarily in focus position).

The position of finite (declarative) *-la* complement clauses themselves with respect to the main verb is another interesting issue involving dependent clauses. In the written corpus, 59% (26/44) of the *-la* complement clauses follow the main verb and 41% precede it. In the spoken corpus the percentage that follows it is close to 100%: only slightly more than 1% of these clauses, 3/238, are preverbal. These clauses in the spoken corpus are not only postverbal, but typically they are also in their own intonation unit, as befits separate assertions. (Also remember that postposing is typically the result of the application of the focus extraposition ‘rule’, which splits the intonational contour.) A full 56% of the postverbal completive clauses (132/235) seem to be in their own intonation units, separated from the main clause. This is no doubt related to the fact that very often these clauses act like true assertions, with the main verb acting more like a ‘modifier’ of the lower clause, expressing, for instance, the source of the assertion (with speech verbs) or some other evidential or other type of modality.

2.2.13.2 Non-finite dependent clauses

We find analogous results when we look at the orderings in non-finite clauses, although here in general the tendency for verb finality is stronger than with finite dependent clauses for both cases, barring some exceptions. In fact *all* non-finite clauses in the written corpus are verb-final.

Not so in the spoken corpus. Thus for instance, in clauses in *-tu*, which are primarily perfective complement clauses, 14.3% of all non-single-verb affirmative clauses are not verb-final.¹⁹ This is due in great part to the fact that restructuring of *-tu* complements of verbs like *nahi* “want” is taking place in the spoken language (cf. above). The derived adverbial clause in *-tu-z* is another cause in point. Here a full 44% of the non-single-verb affirmative clauses are not verb-final.

Complement clauses in *-tzen* are not verb-final in 25.7% of the non-single-verb affirmative clauses (about 45% of all these complements are single verbs, however). This is not the case only when these clauses are complements of aspect-modal verbs, but also when they are independent secondary predicates. Thus, for instance, the most common type of purpose clause in Basque, the non-finite *-tzeke* clause, is not verb-final in 26.9% of all the non-single-verb affirmative clauses, all of which are in fact verb-initial.

2.2.14 Ergativity

2.2.14.1 Introduction

Basque can be said to be a morphologically ergative, but syntactically accusative, language. This means that the ‘subject’ of intransitive sentences, i.e. the most-likely topic candidate in such clauses, is not coded the same way as the most-likely topic-candidate in transitive clauses (the ‘ergative subject’), but, rather, it is marked like the object of such clauses. The default-topic argument is also the default ‘controlled’ argument in most reduced dependent clauses for instance, and in such clauses, the controlled argument is typically the same one as in any so-called syntactically accusative language, such as English, namely the most intrinsically topical argument according to a number of properties (cf. Chapter 3).

Many investigators have viewed ‘morphological ergativity’ at best as an formal anomaly without any obvious semantic or pragmatic correlates (or syntactic ones, for that matter). There do seem to be some functional correlates of ergativity, but they are often weak and seemingly indirect (cf., e.g., Anderson 1977, Trask 1979). In split ergative languages, in which not all clauses follow an ergative pattern (the majority of ergative languages), correlations have been found, for instance, between ergative coding and either the aspect of the clause (the Indo-Iranian type) or topicality related intrinsic topicality characteristics of the transitive ‘subject’ (the Australian type; cf. Silverstein 1976).

Du Bois (1981, 1987) has identified the ergative as the argument which is typically ‘elided’, or ‘covert’, in narrative, as opposed to the absolutive argument, which is typically overtly expressed. In Basque it is true that ergative arguments are ‘covertly’

expressed (in the verbal inflection) more often than absolutes are, as we can see in Table 2-7.²⁰

	All	Covert S		Covert O		Covert A		Covert I	
Abs	1013	589	58.1%	—	—	—	—	—	—
Abs_Dat	56	11	19.6%	—	—	—	—	49	87.5%
Abs_Erg	969	—	—	240	24.8%	789	81.4%	—	—
Abs_Erg_Dat	312	—	—	27	8.7%	254	81.4%	230	7.4%

Table 2-7: Number and percentage of major arguments (those coded on the verb) which are expressed covertly in affirmative main statements in the Spoken Basque Corpus, according to the number of arguments in the clause.

On the other hand, the correlations are not anywhere near as categorical as we would like them to be to claim that causation, or cognitive connection, is necessarily involved here. It is much more likely that the correlations are indirect, due to the fact that whereas ergative arguments are overwhelmingly topics in (narrative) discourse, absolute ones are not. When they are, they seem to be as likely to be covert as ergative ones are, in particular intransitive subjects. Notice that intransitive subjects are also quite often elided, typically when they are topics. When there is a dative involved in the clause, however, a more likely topic-candidate, the absolute argument is typically overt. I will return to this issue in Chapter 4.

2.2.14.2 A semantic motivation for ergative coding

Basque is unusual in that it is an ergative language without splits. Basque also shows that ergative marking can be a strongly motivated coding pattern in constructions involving a significant sector of the verbal lexicon. It has often been noted that the subjects of many intransitive predicates ('stative' or 'unaccusative' predicates) have semantic roles which resemble more those of the objects of transitive verbs than those of

their subjects. An important subset of these intransitive predicates are those which have transitive equivalents formed, basically, by the addition of an agent/cause argument.

In a language like Basque, which has very productive mechanisms for detransitivizing (decausativizing) causative predicates and for causativizing ‘stative’ ones, the ergative pattern of nominal morphology is to some extent quite strongly motivated semantically speaking (cf. Aske 1987). Some of these constructions can be seen as ‘demoting passives’, i.e. constructions in which the ergative argument is suppressed (much like the Spanish detransitivizing passive.)

Among the detransitivizing and detransitivizing alternations which are highly productive in Basque we find impersonal transitives (sometimes called the ‘middle voice’) and agentive vs. agentless inchoatives (cf., e.g., Talmy 1985; Givón 1994).²¹ Below we can see two pairs of examples which show this semantic motivation for coding the object of transitives and the subject of (some) intransitives the same way.

In (2.13) we see a typical transitive clause, and in (2.14) we see its corresponding detransitivized version, with an ‘understood’ indefinite agent. Unlike in an accusative language, the argument with the ‘theme’ semantic role in both cases is marked in the same way.

(2.13) Jon_{ERG} uhartea ikusten du
Jon:ERG island:DEF see:IMPFV he.has.it
“Jon sees the island”

(2.14) Uhartea ikusten da
island:DEF see:IMPFV it.is
“The island is (can be) seen”

The intransitive sentence is equivalent to a (demoting) passive construction in languages with such constructions, in which the agent is not expressed. It is also equivalent to middle constructions in other languages.²²

In (2.15) we see a sentence with an intransitive, unaccusative predicate. In (2.16) we see the equivalent causativized version of these sentence. Notice that the coding of the causative version leaves the object unchanged. Also the word order of the core (absolute) complement does not depend on whether the clause is transitive or not. These alternations are quite productive in Basque, and they are formed simply by changing the auxiliary from intransitive to transitive, or vice versa, without involving any lexical mechanisms.

(2.15) Ura berotu zen
water:DEF heat:PFV it.was
“The water heated up”

(2.16) Jonek ura berotu zuen
Jon:ERG water:DEF heat:PFV he.had.it
“Jon heated (the) water”

This I believe is probably the main motivation behind ergative coding in Basque. This is quite unlike the main motivation for accusative coding (i.e. for coding the ‘subject’ of transitives and intransitives the same way), which, as has often been proposed, seems to be related to the pragmatic notion topic (cf. Chapters 3, 4). That is, the grammatical category subject is basically the default topic, the argument that is most commonly the topic of the clause. In Basque such elements, however, are coded positionally and intonationally and not morphologically.

2.2.14.3 Ergative-absolutive syncretism

As I mentioned in Chapter 1, Basque displays some forms of ergative-absolutive syncretism, i.e. cases in which an ergative phrase is coded the same way as an absolutive subject. This in practice doesn't seem to be detrimental to communication. To begin with, there is redundancy in the coding system, for the verb codes both ergative and absolutive arguments for person and number, and thus actual possible cases of ambiguity are very few. In addition, as we saw above, ergative arguments are elided in the vast majority of sentences, at least in narrative. In Table 2-7 above we saw the percentage of ergative arguments in main clauses which were not overt. Table 2-8 shows that about 80% of the ergative arguments in *all* finite clauses are not overt nominals.

	All	Fin Cl*	Ergative ellipsis	NO erg. ellipsis	% erg. ellipsis
Abs_Erg	1677	1287	992	251	80.5%
Abs_Dat_Erg	414	361	295	66	81.7%
TOTAL	2091	1604	1287	317	80.2%

Table 2-8: Clauses containing an ergative argument in all finite clauses in the Spoken Basque Corpus (Unit type: *Fin_Cl**).

In the remaining 20% of clauses with an overt ergative argument, I have found ergative syncretism in about 15% of the clauses, as can be seen in Table 2-9.²³

	Overt Erg.	Sg. Erg. syncretism	Pl. Erg. syncretism	Total erg. syncretism	% erg. syncretism
ABS_ERG	251	22 +	15	= 37	14.7%
ABS_DAT_ERG	66	12 +	0	= 12	18.2%
TOTAL	317	34 +	15	= 49	15.5%

Table 2-9: Finite clauses in the Spoken Basque Corpus containing an overt ergative argument and displaying ergative-absolutive case syncretism.

One type of syncretism occurs when a singular ergative argument does not receive the *-k* ending characteristic of ergatives. This is a variable type of syncretism found among many speakers. Often we find this type of syncretism when the speaker has not yet mentally formed the sentence and does not know what grammatical role this nominal, which invariably bears the topic pragmatic role, will have in the clause. Another type of syncretism involves plural definite ergative arguments which are coded the same way as absolutive plurals (with the ending *-ak*) and not as ergative ones (in *-ek*). This is a fully regular, non-variable phenomenon, but only in some dialects (though not in standard Basque, for instance).

There is yet another way in which syncretism comes about in Basque. This happens with certain auxiliary verbs like the ones we saw in Section 2.2.11.5 above. Thus, with verbs derived from nouns, such as *nahi* “want” and *behar* “need”, the ‘subject’ is always an ergative argument, regardless of the valence of the ‘lower’ verb. All other auxiliary verbs, however, which take complements in *-tzen*, such as *ari* “busy; engaged” and *hasi* “begin”, are intransitive and their ‘subjects’ are absolutives, regardless of the valence of the ‘lower’ verb.

2.2.15 Conclusions

In this section I have described some of the major typological characteristics of Basque. We have seen, for instance, that Basque is a fully ergative language, although there are some contexts in which of ergative-absolutive syncretism creeps into the language.

Basque also displays most of the typological properties associated with languages said to have (S)OV order, such as prenominal relative clauses and genitives. Basque also has postverbal auxiliaries, as OV languages typically do, at least when it comes to the main tense-aspect-mood auxiliaries in affirmative clauses. In the case of less grammaticalized auxiliary verbs and in negative clauses, however, auxiliary verbs may, or even must, precede the main, lexical verb.

In the next section I will look at the major word order properties of Basque and some of the exceptions. I will also summarize previous analyses of Basque word order and point out some of the major weaknesses of these prior studies. This will serve as a prelude to my analysis of Basque word order and word order variation in the following chapters. The main source of data for those analyses will be described in Section 2.4.

2.3 PREVIOUS STUDIES OF BASQUE CONSTITUENT ORDER AND FOCUS

2.3.1 Introduction

The basic facts about Basque word order have been known for quite a long time. Long before linguists worried about fitting Basque into a rigid typology of underlying or basic types, quite capable grammarians had discovered many basic organizing principles of Basque word order. Authors in this tradition, and particularly Altube, have provided many important insights into the factors of information structure influencing word order.

The first one of these in the Basque linguistic tradition is Azkue 1891. He already ‘knew’ that Basque constituent order, as different as it was in some respects from the order of neighboring languages, had a logic all its own.

It has been said more than once that the Basque sentence doesn't follow any rules; I will show (to the best of my ability) that there cannot be a more rule directed sentence than the Basque sentence. (Azkue 1891:334; my translation, J.A.)²⁴

It would be too much to say that Azkue's hyperbolic promise was fulfilled, but it is also true that some of the basics of the logic of Basque constituent order were laid out in his work.

2.3.2 Azkue 1891, 1923-5

In the section on clausal constituent order, Azkue for the first time mentions the 'first rule' of Basque word order: "place first the most important element and the less important ones after" (Azkue 1891:341).²⁵ This would be equivalent to saying, in modern day terminology, to place the focus, the 'inquired element', 'the most important element' first in the predicate (after the setting elements).

Azkue had the right idea, but he didn't know how to deal with the seeming exceptions and his exposition becomes rather confusing (cf. Osa 1990:104). From his examples, we know that he had figured out that in content questions and answers the question word or the answer goes first, followed by the verb, and that everything else, often including the subject/topic if there is one, follows (much like in English). Azkue equated the most important element with the 'inquired element' (*galdegaia* in Basque and *elemento inquirido* in Spanish). He attempted to generalize this rule to other constructions, not without running into great difficulties.

Nonetheless, Azkue also recognized that sometimes there isn't any one element that is noticeably more 'important' than the others and for such cases he postulates

something like a ‘basic word order’ (Azkue’s Fifth Rule, p. 345). The basic order is, according to his fifth rule, either SOV or OVS, i.e. OV with freely invertible subject (ergative).²⁶

Although his generalizations were incomplete and even erroneous, Azkue planted the seeds that would bear fruition later on, particularly in the work of Altube (1921, 1929).

2.3.3 Altube 1920, 1929

The next major work on the subject of Basque word order is that of Altube (1921, 1929). Altube was a very perceptive syntactician, even though his work was marred by his desire to find *the* correct word order for Basque, which he believed was being subjected to the negative influence of Spanish. He was of the opinion that the ‘authentic’ word order of Basque had rheme-initial focus and was essentially left-branching, and that anything other than that was suspect of being a foreign import.

Azkue seems to have been worried about this as well, though later he became critical of purist excesses himself.²⁷ If one examines the compilation of Basque legends in Azkue 1989/1947 it is hard not to notice that the word order found there is almost exclusively verb-final, even though that is not what one finds in spoken Basque, or in the written Basque of most authors. It seems that he ‘corrected’ his consultants’ syntax, as per his own admission.²⁸

As Azkue’s version of these narratives with almost exclusive verb-final syntax shows, one can write (and speak) perfectly ‘grammatical’ Basque by making (almost) exclusive use of verb-final sentences, something that one could not do in any VO

language, for instance. Therefore, one might argue, non-verb-final sentences are at least suspect of being non-native, especially since those sentences look a great deal like those in the neighboring Romance languages.

Both authors, and in particular Altube, have been much criticized for their purist tendencies, manifested most strongly perhaps in the lexicon, the purpose of which seemed to be to make Basque as different from the surrounding Romance languages as possible. Still, their logic was not completely faulty when they said that Basque was ignoring its own resources, both lexical and grammatical, when it imported elements from other languages. This is particularly obvious in the lexicon, as no one will deny. As for grammatical borrowing, the problem is that the 'transfer' of grammatical elements, such as constructions, is not as direct as lexical borrowing is. Rather, it typically comes about through the modification of pre-existing native constructions, as we saw in Chapter 1.²⁹

Later Bascologists have criticized Altube for not recognizing that the constructions that he criticized were perfectly good and native constructions and not necessarily borrowings at all. Thus, the influential linguist Mitxelena has argued that there is nothing Spanish-like about placing the focus of the assertion after the verb, since this is a widespread practice and records attest that this is an old practice in the language. What Mitxelena failed to note, however, is that such a rule of 'focus extraposition', with marked intonation should be a marked construction (in any language), much more so than we find it to be in spoken Basque, for instance. Thus, it is true that Altube was too rigid and inflexible in his analysis of Basque word order, but perhaps it is also true that those who came after him have been too flexible and understanding.³⁰ Also, it seems to me that

a close and thorough reading of Altube's writings reflects that he was not as inflexible as a few out-of-context quotes have often made him appear to be.³¹

It seems clear to me that what Altube actually observed was not necessarily, as he thought, the importation of new constructions from Romance, but merely the exploitation of preexisting, alternative, Basque constructions to a degree and in ways which were new to Basque and which were likely to have been influenced by Romance structural models.

2.3.4 Narrow focus and Romance influence

Altube's study of 'focus' (*galdegaia* in Basque, literally "question matter") and word order in Basque is a great improvement over Azkue's initial observations and an example of quite sophisticated linguistic analysis in its own right. The basic concept for understanding Basque word order, according to Altube, is the 'inquired element', the 'point' of the assertion (statement or question, for example). Such an element is most obvious in content questions and their answers, as in (2.17a-c) (from Altube 1920:8).

- (2.17) Q: *Nok ekarri dau ori?*
who bring:PFV s/he.has.it that
"Who brought that?"
A₁ *Aitak ekarri dau ori.*
Father:DEF:ERG bring:PFV he.has.it that
A₂ *Ori aitak ekarri dau.*
that father:DEF:ERG bring:PFV he.has.it
A₃ *% Ori ekarri dau aitak.*
that bring:PFV he.has.it father:DEF:ERG
"Father brought that" (= "It was father who brought that")

According to Altube, only (2.17A₁) and (2.17A₂) are possible answers to (2.17Q). In these sentences the phrase which answers the question goes immediately before the verb (complex). All other phrases may come either before the assertion proper, or else after

the verb. Answer (2.17A₃), in which the ‘answer phrase’ is postverbal may also be heard, he argues, but it is not ‘good Basque’ and is characteristic of an “euskaldun *castellanizado*”, a ‘castilianized Basque speaker’ (Altube 1920:8). He argues that this ‘preverbal focus’ rule is rarely broken in spoken Basque (by good speakers) and that the major culprits are writers (Altube 1920:9-12, 1929:8-9).³² However, as we will see in Chapter 6, in present day Basque the ‘problem’ seems to be the opposite, since postverbal foci are much more likely to be found in speech than in writing.

As we can see Altube’s concern is not that the verb should be final, but rather that ‘focus’ should be preverbal. He does recognize that long “and complex” foci, such as complements clauses of speech verbs, may legitimately be postponed, as in (2.18) (Altube 1929:11 ex. a).³³

- (2.18) Amak *esan dit* «biyar lan asko daukala eta ezin etorri al izango dala»
 mother:ERG say:PFV she.has.it.to.me tomorrow work much she.has.it:COMPL
 and not.possible come:PFV able be:FUT she.is:COMPL
 “Mother told me that tomorrow she’s going to have a lot of work to do and that she won’t be able to come”

According to Altube, however, the “correct construction, though not the most common one” would have the complement clause in preverbal position (Altube 1929:11). In other words, such postposings should be treated as exceptional.

Altube also mentions the important distinction (valid both for ‘inquired’ and ‘non-inquired’ elements) between contrastive and non-contrastive elements, what he calls ‘oppositive’ vs. ‘absolute’ terms (cf. Altube 1920:36-42). He makes the interesting observation that with contrastive elements it is much easier to predict their pragmatic function (topic or focus) and their position in the clause, as either clause initial

(contrastive topics) or pre-verbal (contrastive foci) (Altube 1929:38). A contrastive topic or focus is quite easily recognized. The problem is with the informational classification of elements in clauses which do not have such contrastive elements. Altube seems to believe, however, that every clause has a focus (the preverbal element), something that has been rightly questioned by later authors.

2.3.5 Negative and emphatic clauses

As I mentioned earlier, in Basque the whole verb complex typically acts as a single word and counts as a unit for determining preverbal position. Altube notices that there are exceptions to this rule, namely asserted negative clauses (in all dialects) and in emphatic affirmative sentences in some eastern dialects of Basque, where the non-finite verb is postposed along with all other elements of the rheme when a focus element is placed in rheme-initial position (i.e. before the *finite* verb) (Altube 1929:14; cf. Chapter 6).³⁴

Inversion in affirmative emphatic constructions, however, is unheard of in the majority of modern Basque dialects, and in those in which it is possible it is associated with counter-statements to other negative statements, not with regular emphatic assertions. In asserted negative sentences, on the other hand, participle inversion is not only allowed, but required in all dialects of Basque.

Negation in asserted (main and completive) Basque clauses is formed by placing the negative particle *ez* before the finite verb and postposing all non-topical (non-setting) elements, including the non-finite verb. For some speakers some non-asserted finite subordinate clauses, such as conditionals and causatives, are also subject to this rule, at

least variably so. These seem to be the same subset of speakers for whom these clauses are not strictly word final. We will study this construction and its motivation in Chapter 6.

The negative-inversion construction is interesting, among other reasons, because it can be used as evidence for the syntactic unit status of the part of the ‘predicate’ which excludes the auxiliary. According to such a view, affirmative clauses with periphrastic verbs would have the structure in (2.19).

(2.19) [_S Subject [_{S/VP} ... V_{non-fin}] V_{fin}]

Thus negative clauses would be formed by postposing such a unit, creating a verbal brace with the complements of the verb, as in (2.20).

(2.20) [_S Subject *ez*-V_{fin} [_{S/VP} ... V_{non-fin}]]

As I already argued, I do not think that the syntactic structure in (2.19) is valid for present-day Basque clauses. It may have been so at one time, but the evidence indicates that ‘clause-union’ is complete in affirmative clauses. This is a problem only if we insist that the negative construction must be ‘derived’ from an affirmative underlying one. In a constructional approach there is no need to make such an assumption, however. Also note that the verbal brace in negative assertions has been heavily undermined in spoken Basque, as we will see in Chapter 7.

2.3.6 Verb and polarity focus

Altube recognized that besides complements of the verb, either the verb itself or the polarity could be the focus ('elemento inquirido') of the sentence (cf. Altube 1929, Chapter 2 and Chapter 3 respectively). In affirmative clauses this is marked primarily by placing the verb in rheme-initial position. In addition, if the verb is synthetic, this finite verb must be preceded by the affirmative particle *ba-*. (This is obligatorily done also when the rheme consists only of a verb.) In (2.21) below we see an example of a affirmative polarity focus. The filled underlining marks the major accent associated with an intonation unit. The dotted underlining marks a type of accent associated with topic nominals: a full accent, if dislocated in a separate intonation unit; a secondary accent, if integrated into the intonation unit; and no accent, if the nominal is fully integrated into the intonation unit, and even cliticized onto the verb, as in the case of pronouns.

- (2.21) Nik badakit erantzuna.
I:ERG EMPH:I.know.it answer:DEF
"I know / do know the answer."

If the verb is analytic no such marker is needed and simply placing the verb in rheme-initial position suffices, as we can see in (2.22).

- (2.22) Nik irakurri-dut erantzuna.
I:ERG read:PFV-I.have.it answer:DEF
"I have read the answer." / "I have read the answer." (ambiguous)

In some dialects, however, when the verb (not the polarity) is salient/focus, and the reason for the salience is that that is the major new information in the clause (and not that it is contrastive or emphatic), then a special construction may be used, as we will see in Chapter 6, the *egin* focus construction, in which the verb is pulled out of the verb-

complex, as it were, and replaced by the ‘dummy’ verb *egin* ‘do’, cf. (2.23). (The distinction between three different types of ‘rhematic salience’ or focus will be motivated in Chapters 3 and 7).

- (2.23) *Nik irakurri-egin-dut erantzuna.*
I:ERG read:PFV-do:PFV-I.have.it answer:DEF
“I have read the answer.”

Such a sentence would be an appropriate answer, for example, to the question: *How did you find out about the answer?* (new focus), but not as a rebuttal to a negative sentence (*You didn't read the answer*) (contrastive focus), in which in English the accent would be on *have* as opposed to *read*, or as an emphatic sentence (emphatic focus).

In negative sentences we have already seen that the negated finite verb is in rheme-initial position, cf. (2.24).

- (2.24) *Nik ez-dut erantzuna irakurri.*
I:ERG not-I.have.it answer:DEF read:PFV
“I haven't read the answer.”

Altube saw this peculiar ordering to the verb-fronting as being analogous to the fronting of the verb in emphatic affirmative clauses with *ba-*. According to him, the negation, by which he meant the negative particle, is the focused element in such clauses. We will return to this issue in Chapter 6.

As we will also see in Chapter 6, however, the brace is not always respected in speech, with exbraciation to post participial position being rather common, cf. (2.25).

- (2.25) *Nik ez-dut irakurri erantzuna*
I:ERG not-I.have.it read:PFV answer:DEF
“I haven't read the answer”

Also common is the ‘topicalization’ of what would otherwise would be a rhematic complement (here *erantzuna* ‘the answer’). What I mean by topicalization involves fronting the complement to pre-rheme position and thus turning it into a ‘setting’ element, outside the assertion (rheme) proper, cf. (2.26).

- (2.26) *Nik erantzuna ez-dut irakurri*
 I:ERG answer:DEF not I.have.it read:PFV
 “I haven’t read the answer”

This is an example of optional topicalization, which is quite common in negative sentences in Basque, as we will see. The topicalized element’s referent (or ‘idea’) doesn’t even have to be highly topical, although that helps (cf. Chapter 3). Such fronting is also possible in affirmative clauses, but is perhaps more common in negative ones, for reasons that we will see later on.

Altube admits that this type of clauses which display what we would call ‘verb focus’, and are thus ‘verb-medial’, are not always as marked (or emphatic) as they might seem, a fact that we will see is extremely interesting. Thus, for instance he mentions that the two sentences in (2.27a&b), the first one OV and the second one VO, are almost perfectly synonymous (cf. Altube 1929:69).³⁵

- (2.27) a. *Lau seme-alaba ditut: bi etxean eta bi kanpuan*
 four son-daughter I.have.them two home:DEF:LOC and two outside:DEF:LOC
 b. *Ba-ditut lau seme-alaba: bi etxean eta bi kanpuan*
 EMPH-I.have.them four son-daughter two home:DEF:LOC and two
 outside:DEF:LOC
 “I have four children: two at home and two away.”

In the first sentence, an object that would seem to be focal is preverbal and rheme-initial, as we would have expected it. In the second sentence the same, new and most

informative complement follows the verb and the verb receives the *ba-* mark needed by all rheme-initial finite verbs. In theory this sentence should be emphatic (cf. English *I do have four children*), and the referent of the complement should perhaps be accessible from the context. But this is not necessarily the case here. In other words this VO sentence is not as marked as we might have expected.

Altube also notices that, although for the verb to be rheme-initial (focus) in general is unusual and marked, there are some contexts, such as some types of action sequence clauses in narrative in which they may be unmarked (cf. Altube 1920:26). In these clauses, he argues, the focus is indeed the “affirmative quality of the verb”, i.e. the polarity. Altube doesn’t see anything foreign-like about these sentences, which he seems to see just as an extension of the verb/polarity focus construction. This use of the construction has been mentioned by most grammarians that followed Altube, typically as an exception to the OV pattern. In the following chapters we will uncover how common this pattern really is and some of the major contexts that it is found in.

2.3.7 The focus rule and subordinate clauses

As we have seen, Altube’s pre-verbal focus rule is applicable to asserted clauses, i.e. to all clauses which are not strictly verb-final, the same set of clausal constructions in which negative inversion applies. Altube 1920 argues that this rule actually applies in all subordinate clauses, although he hedges that claim somewhat.³⁶ His major argument for saying this is that intensive pronouns, which he argues are *always* foci, are indeed occasionally found in subordinate clauses.

As I will argue in Chapter 6, however, non-asserted clauses do not, and cannot, have assertion foci. It is true that such sentences may in some contexts have contrastive elements, and this explains the presence of ‘intensive pronouns’, which in asserted sentences are indeed foci.³⁷ The possibility of there being more than one intensive pronoun in a sentence also forces Altube to argue that a sentence may (occasionally) have more than one focus, a claim which also strikes me as mistaken (cf. Altube 1920:Chapter 5, 1929:Chapter 4).

Altube notices that completive (declarative complement) clauses are the only subordinate clauses that behave like main clauses for the purpose of the focus rule. Very perceptively he argues that such clauses “seem (at least in modern Basque) as if they themselves were main clauses, in such a way that the real main clause plays a secondary role, or even completely disappears” (Altube 1929:144).³⁸ This, of course, is why such clauses in so many languages display what has come to be called ‘main-clause phenomena’, which are nothing but focus-related, or information structure-related phenomena.

Altube also complains that what we might call postverbal ‘spilling’, placing constituents after the verb, as in asserted clauses, and more in particular, the negative inversion rule, is spreading to clausal constructions which should be strictly verb final, such as causatives, temporals and conditionals. Needless to say, he attributes this to a deplorable foreign influence.³⁹

Although Altube’s intuition about foreign influence is reasonable, I do not think that the extension of negative inversion to non-asserted clauses is necessarily a sign of foreign influence, although contact may have precipitated or rushed such an extension. In

fact, as we will see, inversion in asserted negative clauses is a highly motivated phenomenon when the negative polarity is the ‘focus’ and the proposition’s contents are given. This is a common context for negative assertions, as we will see, but not the only one possible. Inversion however, has become grammaticalized in asserted sentences and applies to all negative assertions, whether the polarity is salient or something else is. This reduces the motivation of the construction considerably and paves the way for its generalization to new contexts. We will return to this question in Chapter 7.

2.3.8 Basic word order and Basque

By now it should be obvious that the constituent order of Basque asserted clauses cannot be simply characterized in terms of grammatical relations, but, rather, that discourse pragmatic considerations (statuses and roles) are of utmost importance. Although it would not be impossible to argue that a certain order, such as SOV or SVO is basic, as long as we also explain the contexts in which such an order is basic, this won’t necessarily be very illuminating. As I mentioned in Chapter 1, however, most investigators who have pursued this issue have concluded that the basic order of Basque is verb-final (SOV) (cf., e.g., de Rijk 1969). Some, however, have also argued that the whole notion of basic word order is of dubious usefulness in Basque (cf. Mitxelena 1978; Brettschneider 1981).

Osa reviews earlier studies about this subject (Osa 1990:Ch.4.3). Osa believes that SOV, or rather, (following Rebuschi 1983) SCV (where C is another constituent, such as absolutive object, adverbial, non-verbal predicate), is indeed the basic order of Basque, and not because it is the most common, but because it is the neutral or unmarked order

(cf. Osa 1990:123). By this he means that this order is the most autonomous from the context, i.e. the order that is least dependent from a specific context, the one that can be most context neutral. This he identifies with the context in which the whole sentence is new information, that is sentences which answer to the question *What's happening?* or *What happened?*

It is indeed possible to argue that this is the most neutral context, the context which determines the basic word order of a language. However, there is something unsatisfactory about arguing that such an unusual, and rare, context is the most basic one, above all others. But even if we do, we have to admit that this order is not adhered to in contexts where the whole sentence, or at least the whole rheme, is 'new'. The rule is typically adhered to in writing, but not without exceptions, not all of which are clearly motivated. In speech the number of exceptions seems to be even larger, as we will see. The notion of basic word order is most appealing in cases where we can view non-basic orders as motivated departures from the basic one. But this is not always the case for SVC orders in Basque.

Another problem with Osa's argument is his claim that in these sentences the whole sentence is new information, and thus supposedly rhematic. In practice, in Basque any sentence which has more than one element before the verb has a topic-comment configuration, for as we saw, only one rhematic element may precede the verb. This, of course, is related to the fact that Basque, and OV languages in general, are typically topic-prominent (cf. Chapter 1). The topic in these cases is not necessarily 'discourse old' (or given), but it typically is not 'hearer old' either, and if it is new and unexpected it will ordinarily be left-dislocated, i.e. be in an intonation unit of its own (cf. Chapter 3).

There is a different way of interpreting what Osa means by neutral context, namely one in which there is no focused element, i.e. in which all the elements of the clause, or perhaps of the comment/rheme, are equally new or salient, or, at least, not significantly different in salience. This definition would seem to have greater chances of being successful. On the one hand it is obvious that many sentences do not have a clearly 'focused' element and that if there is a preferred order in such clauses, it might seem to be 'basic' or 'unmarked' somehow.

On the other hand, it is doubtful that there are many utterances in actual spoken language which meet these characteristics. In practice, there always seem to be differences in 'rhematic salience', or 'communicative dynamism', among the elements of a rheme (cf. Chapter 3), and word order may be attuned to such differences, even if they are not so great as to qualify the more salient element as the 'focus' of the sentence.

On the other hand, minor differences of this kind are less likely to be important enough to require coding, whether it is based on word order or not. In other words, this is an excellent context to look for a neutral order for a language. In Basque it might not be unreasonable to say that, at least in theory, such order is OV, i.e. with the complement (or *one* complement) in rheme-initial (focus) position (even if they are not strictly focal). In standard Basque and in out of context sentences this seems to be indeed the preferred order in these cases, but, as we will see, the picture can be rather different when it comes to speech and connected discourse and when we consider the variation among speakers.

Altube was ambiguous as to whether all (asserted) clauses had a focus element, although he seemed to imply that they did. Other investigators have questioned this possibility. It is obvious that in questions and answers to questions there is a rhematic

element which stands out, i.e. the focus, but it is not so obvious that this is also true in average statements, especially perhaps in narrative. Osa (1990), for instance, looked at word order in narrative and concluded that most clauses do not have a focus. He also found that the most common focus constituents are perhaps adverbials, especially modals and predicative adverbials. As to the core arguments, objects are the most common ones to fill that role, followed by ergatives, and finally by datives, which are hardly ever focused.

2.3.9 The preferred order of elements

Osa also found that in actual spoken narrative focused complements, just like non-focus complements, need not go in rheme-initial, immediately preverbal (focus) position (Osa 1990:222). According to him, “[f]or any type of complement, we can see that with the same frequency they can go either to the left or to the right of the verb when they form a communicative unit with that verb. Likewise, when the complement alone is the focus, it is perfectly common for it to go to the right of the verb” (in addition to being able to go to the left of the verb) (Osa 1990:222).⁴⁰

The likelihood that a focus element will go to the right of the verb, however, is not the same for all types of complements. Thus it is very likely for sentential complements to go to the right of the verb; for non-sentential (nominal) objects it is less likely; and this position is most rare for “predicate complements” and “manner adverbs” (*ibid.*). Furthermore, since the latter are optional complements, they are typically focal and in preverbal ‘focus position’.⁴¹ As I already mentioned and we will see later on, placing the focus after the verb is intonationally marked in Basque, but not as marked as we might

have expected. Osa concludes, without being any more specific, that, for the most part “constituent order is quite free in stories, since communicative strategy is realized primarily by prosodic factors, i.e. intensive accent, pauses, and intonation” (*ibid.*).⁴²

Some authors have also expanded the search for basic or default ordering beyond the subject, verb and object. Osa 1990, following Rebuschi 1984/1982, who himself follows Lafitte 1944, argues that, despite all the possible variation, the order specified in (2.28) is the ‘canonical’ or most ‘neutral’ order in Basque.⁴³

(2.28) Subject - Circumstantial - Dative - Object - Attribute - Verbal-Complex

It is not exactly clear what it means to say that this order is basic or unmarked. Sentences with that many elements are rare in discourse and the relative order depends a great deal on the context. It seems that what these authors have in mind is a context in which all these elements are equally new to the discourse, an even less common situation.

One thing that is clear, however, is that in any such sentence, only the element that precedes the verbal-complex will be part of the rheme, or assertion proper, and that all the other elements will be non-rhematic, or ‘setting constituents’, and that only one at most, the ‘clausal topic’ can be part of the same intonation unit as the rheme, the rest being ‘dislocated’ or placed in their own intonation units. The relative order of setting constituents seems to be a matter of ‘relevance’ or ‘scope’. It is possible that some relative orderings of setting elements are more ‘basic’ or ‘unmarked’ or easier to assimilate in minimal-context situations, which is how the claim embodied in the order preference expressed in (2.28) above would have to be interpreted. The relative order of the elements inside the rheme, which is subject to great variation, does not seem to be

based on the same principle of ‘scope’, but on principles such as rhematic salience, as we will see.

2.3.10 Basque as a left-branching language

Altube 1929 discusses at length (Chapter 6) the fact that Basque, unlike its neighbors, is a left-branching language (cf. Altube 1929:163-64). Altube attributes the left-branching nature of Basque, i.e. the fact that heads follow their complements, to the fact that the ‘relator’ in such constructions is placed at the end of the dependent element, and that relators want to be in the middle. In a footnote in the text of Chapter 6, Altube mentions that

the grammatical morpheme which relates or joins two elements ... typically likes to go between the two of them. In Basque (a postpositive language) said morpheme must go necessarily at the end of the governed element ...; and thus the governor must follow. In prepositive languages just the opposite is found: the governed constituent is preceded by the grammatical morpheme ... and for this morpheme to be in the middle it is necessary for the governor to occupy the first place in the sentence. (Altube 1929:165, *fn.1*; my translation, J.A.).⁴⁴

This, of course, is the iconic principle which I mentioned in Chapter 1, which, following Dik (1989), I called the Relator Principle.

Altube notes that its left-branching nature makes Basque run into problems when one attempts to translate long and complex sentences from right-branching languages, a major problem in the development of constructional and rhetorical ‘models’ for written Basque. Romance writing styles are notorious for their ‘run-on’ sentences with multiple subordination, a style which quite different from most speaking styles. Basque writers, which historically, before Altube, had been primarily translators, had attempted to copy

the right branching structure of their model languages by postponing or ‘right-dislocating’ complex elements when writing in Basque. The effect of such strategies have been very nefarious for Basque, according to Altube. This is, in part, because the ‘relator principle’ is systematically violated.

Altube would argue that classical Basque writing models which follow these extraposing tendencies should be avoided. Most modern writers would probably agree with Altube that a preferred solution to extraposing complex elements is, as any English teacher will tell us, to split the sentence into smaller units:

to solve these difficulties, in translations of foreign texts, it is recommended, as a guiding principle, to disarticulate, subdivide and reduce to several shorter clauses those Spanish phrases, sentences and elements which are excessively long before proceeding to translating them into Basque. (Altube 1929:166).⁴⁵

The problem of run-on sentences, i.e. multiple embedded right-branching sentences, produced on-line, is not a problem in speech, which is typically characterized by shorter sentences, coordinated structures, and parataxis as opposed to subordination. Altube seems to have identified this fact and also the cognitive difficulties introduced by left-branching structures, already mentioned in Chapter 1.

2.3.11 Some unanswered questions

Osa’s study and those of his predecessors, leave many questions with only partial or unsatisfactory answers. Thus for instance, we wonder whether Basque really does have a preverbal and a postverbal focus position. And if so, how does it differ from

Spanish and other VO languages which also make use of those two positions for different types of focus? When is the focus placed before the verb and when after?

Do all clauses have a focus, as Altube seems to claim, or is the focus the whole predicate in some cases, as Osa claims? What are the formal differences, and in particular word order differences, if any, between these two types of assertions.

These are many other questions require further consideration, especially by looking at different varieties of the Basque in some depth, in particular Basque texts and not the decontextualized examples which previous studies rely primarily on. That will be the goal of the following chapters, where I will return to these and other similar questions.

2.4 CHARACTERISTICS OF THE CORPUS

2.4.1 Introduction

Because of the variable nature of the word order in Basque, and because of the well known difficulties in introspecting about discourse pragmatic factors, I felt that in addition to introspective data, data of a different type would be needed in order to reach generalizations about the different ordering possibilities available to Basque speakers, their relative frequencies, and their functions, both in general and for different styles, genres, and media (basically, speech vs. writing).

Previous studies of Basque word order have relied for the most part on introspection, and those which have actually looked at data in context have been either impressionistic or based on limited written sources. No previous studies seem to have looked at data in context (discourse) systematically, and for this purpose I recorded

samples of narrative speech from 46 speakers. These speakers can be divided into two, not exactly homogeneous, groups: one composed of children with native and near-native degrees of fluency and the other composed of fluent and, for the most part, highly educated, adults. In a more complete and less preliminary study it would be desirable to include a much more acceptable sample of the population, but this will have to remain as a long term objective for the Basque linguistics community, given the amount of tedious work involved in analyzing recorded speech.

The topic of the sample narratives were controlled for since the speakers were asked to recount a short silent film that they had just seen. The listener, or audience, for each of these narratives was a school or work mate of the narrators who had not watched the film and did not know what it was about.

Two films were chosen for this study. The first one was the *Pear Film*, a 7 minute silent film made specifically for the purpose of obtaining narratives for linguistic analysis as reported in Chafe, ed., 1980. The second film was one which I made myself from clippings made from Charlie Chaplin's silent movie *Modern Times*. The whole sequence for this movie lasted 12 minutes. The several hours of recorded tapes were transcribed and turned into a database of intonation units, all of which were coded for characteristics and properties which might be of relevance to their analysis for this particular study on word order.

Having different speakers recount the same objective story has the added advantage of controlling for the characters (referents) and the action of the story allowing for a large number of inter-speaker generalizations, only some of which were taken advantage of in this study.

In addition to this corpus, which I will refer to from now on as the Spoken Basque Corpus, I collected a written Basque corpus, which I will refer to as the Written Basque Corpus. This second corpus consisted of the first chapters of two recent Basque books, containing narrative as well as dialogue: Atxaga 1991, a novel, and Elexpuru 1994, a travel diary. In addition, occasionally I will refer in this study to a larger corpus of written Basque, one composed of the following three novels: Amuriza 1984, Atxaga 1988, and Lertxundi 1994. Concordance searches were made in this extended written corpus for certain uncommon constructions.

In the next two sections I will present some general statistics about the clausal composition of the spoken and Basque corpora. Next I will explore the general word order characteristics of statement clauses in the corpora from a statistical point of view, comparing the spoken and written corpora, as well as the different subcorpora among each other where applicable.

2.4.2 The units of the Spoken Basque Corpus

The Spoken Basque Corpus was analyzed and divided into a total of 6760 intonation units of all types. This corpus contains 3,408 finite clauses and 632 non-finite clauses, as we can see in Table 2-10 below.

Finite clause type	<i>Pear</i>	<i>Chaplin</i>	Both	% Both
Sentence-body statements, affirmative	689	1990	2679	78.6%
Sentence-body statements, negative	22	109	131	3.8%
Complement clauses, affirmative	52	187	239	7.0%
Complement clauses, negative	4	25	29	0.9%
Head adverbial clauses	30	102	132	3.9%
Tail adverbial clauses	12	53	65	1.9%
Post presentative predicate	10	29	39	1.1%
Relative clause (finite)	18	48	66	1.9%
Other finite	8	20	28	0.8%
TOTAL	845	2563	3408	100.0%

Table 2-10: Finite clauses in the Spoken Basque Corpus.

Affirmative finite main statement clauses constitute the majority of the clauses. Negative main statements constitute a much smaller number. Other finite clauses include pre-clausal adverbial clauses, post-clausal adverbial clauses, post-presentative predicates, and relative clauses. Their relative numbers can also be seen in this table.⁴⁶

In Table 2-11 we can see the statistics for non-finite clauses, most of which are clause-internal complements and secondary predicates. Head and tail adverbials are the next most numerous group. Another type of head clauses consists of what I have called ‘head statements’, a peculiar Basque construction expressing a ‘preparatory action’ preceding a second, main one.⁴⁷

Non-finite clause type	<i>Pear</i>	<i>Chaplin</i>	Both	% Both
Predicate/complement clauses	93	279	372	58.9%
Statements	1	6	7	1.1%
Head adverbial clauses	14	32	46	7.3%
Tail adverbial clauses	49	89	138	21.8%
Head statements	7	29	36	5.7%
Relative clause	2	5	7	1.1%
Other	8	18	26	4.1%
TOTAL	174	458	632	100.0%

Table 2-11: Non-finite clauses in the Spoken Basque Corpus.

2.4.3 The units of the Written Basque Corpus

The Written Basque Corpus was smaller than the Spoken Basque Corpus. In Table 2-12 we can see the types and numbers of finite clauses categorized in this corpus. Statement clauses comprise $\frac{3}{4}$ of all main finite clauses. The other $\frac{1}{4}$ is composed of complement clauses, pre- and post-clausal adverb clauses, relative clauses, and others.⁴⁸

Finite clause type	<i>Behi</i>	<i>Kuba</i>	Both	% Both
Main clause statements, affirmative	237	415	652	52.4%
Main clause statements, negative	62	66	128	10.3%
Other main clause (questions, etc.)	83	63	146	11.7%
Complement clauses (statements)	41	32	73	5.9%
Head adverbial clauses	23	26	49	9.9%
Tail adverbial clauses	16	23	39	3.1%
Relative clause	21	53	74	5.9%
Headless relative clause	12	3	15	1.2%
Other dependent clause (questions, etc.)	20	48	68	5.5%
TOTAL	515	729	1244	100.0%

Table 2-12: Finite clauses in the Written Basque Corpus

We can see that in the written corpus there is a higher percentage of negative clauses, non-statement clauses, and adverbial subordinate clauses than in the spoken corpus. The spoken corpus, on the other hand, has a greater percentage of finite post-presentational predicate clauses (cf. Chapter 5). It is not clear to what extent these differences, other than perhaps the difference in the proportion of adverbial clauses, is due to typical differences between speech and writing (the medium) and which ones are just accidental about these particular corpora (having to do with style and genre).

Table 2-13 shows the numbers for the different types of non-finite clauses.⁴⁹

Non-finite clause type	<i>Behi</i>	<i>Kuba</i>	Both	% Both
Argument, complement	36	30	66	17.7%
Statements (head, fragment, etc.)	21	27	48	12.9%
Predicate	55	83	138	37.0%
Head adverbial clauses	30	26	56	15.0%
Tail adverbial clauses	19	22	41	11.0%
Relative clauses	7	10	17	4.6%
Other	3	4	7	1.9%
TOTAL	171	202	373	100.0%

Table 2-13: Non-finite clauses in the Written Basque Corpus.

The categories into which non-finite written clauses are slightly different from those used for the spoken corpus but give us a reasonable means for comparison. We see that the percentage and relative number of relative clauses is much higher in writing, for instance. And although there are more head (setting) adverbials in writing, the number of tail non-finite clause adverbials is slightly higher in speech. These, of course, are not central topics for this study and will not be mentioned further.

2.4.4 Constituent orders in main clauses in the Spoken Basque Corpus

All the clauses in the Spoken Basque Corpus and the Written Basque Corpus were coded for their word order and a number of other characteristics, both global ones and characteristics of their component parts. Word order coding was done in terms of a number of categories, both phrasal and clausal. The phrasal categories used are: verb (*V*), ergative 'subject' (*A*), absolutive 'subject' (*S*), absolutive object (*O*), dative object (*I*), and other constituent (*X*). Clausal categories are primarily finite completive clause (*Xla*) (with the ending *-la*), finite adverbial clauses, such as temporal ones (*Xnean*), causal ones (*Xlako*), and a variety of non-finite complement clauses (*Xten*, *Xtea*, *Xizeko*, etc).

From this coding, about 140 different orderings resulted in the Spoken Basque Corpus for affirmative statements, many of them obviously small variants of each other and most of them with very low frequency. In Table 2-14 we can see the types and frequencies of the 28 most common orders in three types of clauses: main, affirmative statements, dependent affirmative statements (completive clauses), and non-finite clauses. In addition, and for purposes of comparison, I have added in the last column, the statistics for those same orderings for affirmative statements in the Written Basque Corpus. The 28 orders account for about 83.4% of all the clauses in the first group, and about 90% of the second and third groups, and about $\frac{3}{4}$ of the written statements.⁵⁰ (The orders in Table 2-14 do not include clauses with completive complements which were split into more than one intonation unit.)

		Main Aff. Statements		Completive Aff. Statemts.		Non-finite clauses		Main Aff. Stats (Written)	
1	V	305	11.4%	20	8.4%	210	33.7%	14	2.2%
2	OV	137	5.1%	13	5.4%	151	24.2%	55	8.6%
3	VO	233	8.7%	9	3.8%	15	2.4%	11	1.7%
4	SV	146	5.4%	21	8.8%	5	0.8%	30	4.7%
5	VS	159	5.9%	9	3.8%	0	—	3	0.5%
6	VX	275	10.3%	8	3.3%	20	3.2%	9	1.4%
7	XV	234	8.7%	61	25.5%	99	15.9%	122	19.2%
8	AV	47	1.8%	12	5.0%	9	1.4%	5	0.8%
9	VA	17	0.6%	1	0.4%	0	—	1	0.2%
10	AVO	46	1.7%	6	2.5%	0	—	11	1.7%
11	AOV	14	0.5%	2	0.8%	4	0.6%	12	1.9%
12	VAO	9	0.3%	0	—	0	—	0	—
13	VOA	1	0.0%	0	—	0	—	0	—
14	OVX	30	1.1%	0	—	7	1.1%	33	5.2%
15	VOX	29	1.1%	0	—	5	0.8%	0	—
16	VXO	15	0.5%	1	0.4%	0	—	0	—
17	XVO	21	0.8%	0	—	2	0.3%	14	2.2%
18	SVX	88	3.3%	8	3.3%	0	—	21	3.3%
19	SXV	58	2.2%	11	4.6%	3	0.5%	44	6.9%
20	VSX	41	1.5%	1	0.4%	2	0.3%	1	0.2%
21	XVS	39	1.5%	24	10.0%	0	—	34	5.3%
22	VXla	78	2.9%	1	0.4%	7	1.1%	6	0.9%
23	XlaV	1	0.0%	0	—	1	0.2%	7	1.1%
24	VXtzen	90	3.4%	2	0.8%	1	0.2%	0	—
25	XtzenV	27	1.0%	3	1.3%	2	0.3%	7	1.1%
26	SVXtzen	19	0.7%	3	1.3%	0	—	1	0.2%
27	V:	52	1.9%	0	—	5	0.8%	0	—
28	XVX	23	0.8%	1	0.4%	4	0.6%	38	6.0%
	Total	2234	83.4%	217	90.8%	552	88.5%	479	75.3%
	Others	446	16.6%	22	9.2%	72	11.5%	157	24.7%
	ALL	2680	100.0%	239	100.0%	624	100.0%	636	100.0%

Table 2-14: The 28 most common word orders in the Spoken Basque Corpus, for main affirmative statements, dependent affirmative statements (completive clauses), and all affirmative non-finite clauses; and for main affirmative statements in the Written Basque Corpus.

As we can see in Table 2-14, the vast majority of clauses consist of only a verb or a verb plus one additional complement. For instance, in main affirmative clauses, 11.4% consist of just a verb, 13.8% consist of a verb and a direct object, 11.3% consist of a verb

and an intransitive subject, and 19% consist of a verb and one other complement (not S or O). These clauses constitute 55.5% of all the affirmative statements.

Many of the orderings are arranged in groups which allow us to see the relative frequency of preverbal vs. postverbal orderings for a particular subset of constituents. Concentrating first on the spoken corpus, we see in lines 2 and 3, for example, we see that VO clauses are more common than OV ones when they are finite, but less common when they are non-finite (see also lines 10-17 for the order of the O). VS order is slightly more common than SV order in main clauses, but much less common in non-main clauses (whether finite or non-finite). In other words, subject inversion is found almost exclusively in main clauses (see also lines 24-25). We also see in 8-9 and 10-11 that ergative subjects are much less likely to invert than absolutive (intransitive) ones.

Ordering preferences in the written corpus are somewhat different. Thus for instance, subject inversion is very rare. Here, XV is the single most common order by far, and, in general, verb-final clauses are significantly more numerous than in speech. Also, we can see that there aren't any examples of clauses with more than 2 complements after the verb.

As for the order of complement clauses with respect to the verb in speech, we can see two types of complement in Table 2-14: (a) *Xla*: finite completive clauses with the *-la* complementizer; and (b) *Xtzen* clauses: non-finite imperfective clauses, the most common type of non-finite clause, which can act as objects or adverbials/predicates, much like *-ing* clauses in English. In speech, the completive complements are almost exclusively postverbal, for all clause types (cf. 22-23), although very few dependent clauses have such complements. The non-finite clauses are also much more common postverbally in

main and finite dependent clauses (cf. 24-26). In writing, on the other hand, complement clauses are much more likely to be preverbal than in speech, particularly non-finite ones.

In Table 2-15 below we can see the different relative positions for the verb in finite and non-finite clauses in the spoken corpus.⁵¹ Post-verbal elements are found in a majority of main clauses: 60.2% overall (68% if we discount the 11.4% of clauses which are single verbs). Verb-final clauses, however, are a majority for all other clause types: completive clauses, with 56.9%, other finite dependent clauses, with 59.8%,⁵² and non-finite clauses, with 52.3%. If we add single-verb clauses, which naturally are also verb final, the numbers of verb final clauses increases to 65.3%, 68.1%, and 86.1% respectively (still only 39.8% of main clauses).

	Statements (fin.)		Completives (fin.)		Other finite		Non-finite	
V...	1164	43.4%	36	15.1%	38	14.3%	67	10.8%
...V...	451	16.8%	47	19.7%	47	17.7%	19	3.1%
...V	760	28.4%	136	56.9%	159	59.8%	325	52.3%
V	305	11.4%	20	8.4%	22	8.3%	210	33.8%
Total	2680	100.0%	239	100.0%	266	100.0%	621	100.0%

Table 2-15: Single-verb, verb-initial, verb-final and verb-medial clauses affirmative in the spoken Basque corpus by clause type: sentence-body (finite), Complement (finite), Other dependent finite, and Non-finite

As we will see in greater detail below, the larger percentage of verb-initial clauses in main ('root') clauses is due in part to the fact that existential-presentational and otherthetic clauses are often (though not always) verb-initial in speech (but not in writing), and these clauses are typically main clauses. Also, the relatively larger number of verb-final dependent clauses, especially non-finite clauses, is no doubt due in great part to the quite strong verb-final constraint that many of these clauses have, although it is not always

adhered to by all speakers. Some finite dependent clauses, however, are not strictly verb-final, even in the standard written language, as we have already seen.

A great number of further comparisons can be made from this data. In Table 2-16, for instance, we can see a comparison of the different frequencies for the different verb positions in the collection of *Pear* stories and in the collection of *Chaplin* stories.⁵³

Story	V...		...V...		...V		V		Total	
<i>Pear</i>	336	48.8%	107	15.5%	144	20.9%	102	14.8%	689	100%
<i>Chaplin</i>	828	41.6%	344	17.3%	616	30.9%	203	10.2%	1991	100%
Total	1164	43.4%	451	16.8%	760	28.4%	305	11.4%	2680	100%

Table 2-16: Verb position for all affirmative, main clause statements in the Spoken Basque Corpus according to the story used (across sentence types).

For no apparent reason, there are some discrepancies between the two subsets. Thus for instance the number of single-verb clauses is higher for the *Pear* stories, as is the number of verb initial clauses. The *Chaplin* set, on the other hand, has a higher percentage of verb-final clauses. Perhaps a higher percentage of presentatives and otherthetic clauses in the *Pear* story accounts for this discrepancy. This is important because it shows how significant the differences among seemingly comparable texts can be and urges us to be cautious about making hasty generalizations.

We may conclude, however, that the differences between speech and writing are significant. We may also conclude that classifying Basque as having a particular basic order cannot be done convincingly from a statistical perspective, since different clause types differ as to their preferences, as do different samples and, in particular in this case, samples from different media (speech vs. writing).

2.4.5 Differences among the subjects of the Spoken Basque Corpus

Another interesting comparison is the one among the different speakers, 45 in all, and between different groups of speakers, of which there were four (see next section). Four different groups of speakers used in this study, two of which were composed of children, *Hendaia* and *Ikasbide*, and two of adult speakers, *Deustu* and *Lur*. Not all the groups behaved similarly when it came to ordering preferences, even though the narratives were about the same topics and in identical proportions. Thus for instance, as we can see in Table 2-17, in the *Ikasbide* group, main clauses have postverbal elements in 74.6% of the (main affirmative) clauses (59.4 + 15.2), whereas in the *Deustu* group, this happens in only 52.6% (34.6 + 18).

	V...		...V...		...V		V		Total	
<i>Hendaia</i>	237	38.5%	95	15.4%	186	30.2%	97	15.8%	615	100%
<i>Ikasbide</i>	462	59.4%	118	15.2%	111	14.3%	87	11.2%	778	100%
<i>Deustu</i>	321	34.6%	167	18.0%	352	37.9%	88	9.5%	928	100%
<i>Lur</i>	144	40.1%	71	19.8%	111	30.9%	33	9.2%	359	100%
Total	1164	43.4%	451	16.8%	760	28.4%	305	11.4%	2680	100%

Table 2-17: Verb position for all affirmative, main clause statements in the Spoken Basque Corpus by subcorpora

The differences between these two extreme groups are greatest for verb-initial and verb-final clauses.

As I just said, the subjects of the Spoken Basque Corpus consisted of four different groups of Basque speakers, chosen both because of their characteristics as well as for pragmatic reasons having to do with availability. Ideally, a study of this kind would have as large and varied a sample as possible. In practice, however, this is not always possible. One of the goals which I had in mind when I undertook this study is to

determine possible variation among individuals, as well as among different groups of speakers, such as those from geographical dialects, different sociolinguistic backgrounds (degree of use of Basque), different psycholinguistic backgrounds (degree of 'fluency'), and different age groups, to name a few.

As it turned out, the sample was limited to 45 individuals, divided into the four groups mentioned above, three of which were fairly homogeneous. Despite the constraints on the sample, fairly suggestive preliminary conclusions can be extracted from it, as we will see. The four groups into which the subjects are divided are the following (profile summaries for each individual subject can be seen in Appendix 4.5):

- *Hendaia*: A group of 11 ten year olds from *Gure Ikastola*, a Basque language school in Hendaia, Lapurdi. This group includes children who are natives of Lapurdi, as well as children of refugees from the southern Basque country. These children have attended Basque school throughout their lives. The home background and the degree to which they use Basque in their daily lives outside school, and even in the school, varies from one child to another. Self-reports about their linguistic background and language use were used, together with reports from their school teacher, to determine their socio- and psycho-linguistic profiles.
- *Ikasbide*: A group of 12 thirteen year olds from *Ikasbide Ikastola*, a Basque language school in Donostia, Gipuzkoa. These children have also carried out their schooling in the Basque language and are equally of mixed linguistic backgrounds. They also provided reports of their own linguistic background and their teacher provided

additional information. They all live in the Donostia area, where use of the Basque language is quite limited.

- *Deustu*: A group of 12 twenty-two to twenty-four year old, fourth year students of Basque philology at the Donostia campus of the University of Deustu. Many grew up and/or live in towns outside the Donostia area where use of Basque is quite strong.
- *Lur*: A group of 6 adults, four of which worked at the *Lur* publishing house in Donostia, Gipuzkoa, and two adult relatives of a worker at the same center. They are all very fluent and use Basque throughout their daily lives. The ages of this group go from early 20's to 70's, making it much more heterogeneous than the other three.

The adult speakers in the last two groups were extremely fluent native speakers of Basque who use Basque in all facets of their lives. The younger speakers, on the other hand have more mixed, and some had quite complex, linguistic and sociolinguistic backgrounds. Thus, for instance, some of the children make rather limited use of Basque outside the school context, and some even use Romance with their peers at school as well.

2.4.6 Further notes on word order in the Written Basque Corpus

As we have seen, the written Basque has a greater proportion of verb-final utterances than spoken Basque. In Table 2-18 we can see the statistics for the different positions of the verb for different clause-types in the Written Basque Corpus.

	Affirmative statements		Affirmative Completives		Other finite		Non-finite	
V...	52	8.0%	0	—	0	—	0	—
...V...	253	38.7%	14	21.5%	8	9.4%	0	—
...V	333	51.0%	49	75.4%	66	77.6%	286	80.1%
V	15	2.3%	2	3.1%	11	12.9%	71	19.9%
Total	653	100.0%	65	100.0%	85	100.0%	357	100.0%

Table 2-18: Single-verb, verb-final, verb-medial, and verb-initial clauses in the Written Basque Corpus by clause type: main (finite) statements, complete statements, other dependent finite clauses, and non-finite clauses.

In main affirmative statements, the verb is clause-final more than half of the time, significantly more often than for any of the Spoken Basque Corpus groups (28.5% on average, cf. Table 2-17 above). The percentage of verb final clauses is more than ¼ in affirmative complete clauses. Moving on to other finite dependent clauses, but excluding relative clauses, which are all verb-final, and embedded questions, we find that the vast majority of them are verb-final or single-verb.⁵⁴ Finally, we can see that all non-finite clauses are verb-final or single-verb, for all types of clauses, whether affirmative or negative.⁵⁵

In Table 2-19, we can see the statistics for the different positions of the verb for affirmative main clauses in the Written Basque Corpus by subcorpus chapter. The only noticeable difference in word order between the two written samples is in the number of verb-initial clauses.

Sample	V...		...V...		...V		V		Total
<i>Behi...</i>	31	13.0%	86	36.1%	113	47.5%	8	3.4%	238
<i>Kuba...</i>	21	5.1%	167	40.2%	220	53.0%	7	1.7%	415
Total	52	7.9%	253	38.7%	333	51.0%	15	2.3%	653

Table 2-19: Single-verb, verb-final, verb-medial, and verb-initial affirmative main clauses in the Written Basque Corpus (two sources combined) by sub-sample.

In general the differences between the written and the spoken corpora are of the same general type as those among the fluent groups of adults and the less fluent groups of children in the spoken corpus: a higher percentage of verb-final clauses, a smaller percentage of subject inversion, more rigidly verb-final dependent clauses, and so on. In the next section we will look at these differences in more detail.

2.4.7 Additional comparisons between the spoken and written corpora

The general differences we have observed between the spoken and the written corpora in the direction of more verb-final utterances for the latter is most striking when we compare clauses made up of the verb and one other constituent, repeated here in Table 2-20 for convenience.

	Order	Spoken		Written	
1	SV	146	47.9%	30	90.9%
2	VS	159	52.1%	3	9.1%
3	*S*V*	354	55.8%	116	73.0%
4	*V*S*	280	44.2%	43	27.0%
4a	V*S*	232		5	
5	*A*V*	202	82.4%	56	78.9%
6	*V*A*	43	17.6%	15	21.1%
7	OV	137	37.0%	58	82.9%
8	VO	233	63.0%	12	17.1%
9	XV	234	46.0%	122	93.1%
10	VX	275	54.0%	9	6.9%
11	AOV	14	23.3%	12	50.0%
12	AVO	46	76.7%	12	50.0%
13	SXV	58	39.7%	44	66.7%
14	SVX	88	60.3%	22	33.3%
15	*Xla*V*	1	1.0%	12	34.3%
16	*V*Xla*	103	99.0%	23	65.7%

Table 2-20: Percentages for different pairs of clausal orderings in main affirmative statements in the written and spoken corpora.

We can see in rows 1 and 2 that the first alternation, SV ~ VS, is evenly distributed in the spoken corpus, but extremely uneven in the written corpus in favor of SV. We will see that this is due in great part to the tendency in spoken Basque, but not in written Basque, to postpose subjects in presentative sentences.

This doesn't mean that intransitive subject inversion is not common in writing. As we can see in rows 3 and 4, when there are other elements before the verb, be they topics or not, these subjects invert in writing as well, for about ¼ of all subjects. The percentage is higher, 44.2%, for the spoken corpus. Notice, however, that whereas 82.% of these clauses in the spoken corpus (232/280) are verb-initial, only 11.6% of those in the written corpus (5/43) are verb-initial (row 4a).

As we can see in rows 5-6, overt ergative arguments (A) are less likely to invert with the verb than are absolutive arguments in intransitive sentences (S). The difference is quite small for the written corpus, but significant for the spoken corpus.⁵⁶

As for the OV ~ VO alternation, Table 2-20 shows that whereas a majority of objects (in this particular configuration) are preverbal in writing, a majority are postverbal in speech. We will see later on that there is a tendency in speech to place objects which have given referents, i.e. which not 'important' enough to 'deserve' to be placed in rheme-initial position after the verb. For some speakers this extends even to objects with 'new' referents. In more 'careful' Basque, such as in writing, this only applies to cases in which the verb is noticeably more 'important' than the object, such as when it is contrastive, or when the whole assertion is emphatic.

The XV~VX alternation shows some of the same tendencies as the OV~VO one, but with a lesser degree of postposing. This seems to be due to the fact that the ideas that

X's represent (e.g. predicates and adverbials), are more likely to be the most 'important' part of the rheme than are the referents of objects. We will return to this issue in Chapter 6.

When a 'subject' is added to the previous two alternations we find some differences in the degree to which postverbal complements are found. In rows 11-12 and 13-14 we can compare the AOV~AVO and the SXV~SVX alternations for these triplets (not the only possible ones). First of all we see that the overall number of tokens is much smaller, due to the fact that overt subjects are relatively rare. Furthermore, we see that O's and X's are more likely to 'postpose' when there is a 'subject' than when there isn't one. The reason for the increase, especially noticeable in the written corpus, is that in many of these clauses, the 'subject' is actually a rhematic element, occupying the only preverbal rhematic slot, and this causes any and all other complements to be placed after the verb.

In the following chapters I will explore, and attempt to motivate, all these orders and the reasons for the differences among the corpora and among speakers. Chapter 3 runs through basic notions of discourse analysis or information structure, notions which are crucial for understanding Basque word order. Chapter 4 looks in more detail at the pragmatic role topic and its central role in Basque sentences. Chapter 5 examines sentences which appear to not have a topic, or else have inverted or otherwise unusual topics. Chapter 6 is about the counterpart of the topic pragmatic role, namely the focus pragmatic role and its preferred positions. This is the single most important factor in explaining Basque word order as well as the variation involved. Chapter 7 expands on

these notions and applies them to some major speech acts besides affirmative statements.
in particular negative statements and emphatic statements.

NOTES TO CHAPTER 2

1. The declension is fairly regular and in the definite column we can discern the definite marker *-a-* for most cases, seemingly related to the third-degree demonstrative. The synchronic irregularities in the locative series (such as the lack of *-a-*) has been accounted for diachronically, cf., e.g., Jacobsen 1977.

2. For some of the interesting properties of the genitive suffix *-ko*, cf. Trask 1985b; de Rijk 1988, 1993. Some nouns (not noun phrases) with the *-ko* ending, follow the noun, acting much like denominal adjectives in which the *-ko* suffix acts as a derivational suffix, but this is a rare phenomenon (cf. e.g. Agirre Berezibar 1991:632; de Rijk 1988:76-77). They may have been formed under the influence of Romance. Postposing of *ko*-genitive phrases is also possible, at least in some dialects, e.g. *animali txiki katuen antzerako bat* “a small animal of cat-like appearance”, cf. Agirre Berezibar 1991:632. Other derived adjectives also waver as to their position, but to a much greater extent than *-ko* suffixed nouns (cf. the suffixes *-tar-* and *-dun-*, mentioned in, e.g., de Rijk 1988:76-77).

3. In such cases, which perhaps were even more prevalent in the past, ‘definite marking’ in fact translates as ‘referential marking’ (as opposed to the marking of the indefinite declension, which is really more like ‘non-referential marking’). This coding neutralization is possible because the context *and* the word order of the phrase in question typically indicates whether the nominal is identifiable or not (cf. Chapter 3). Note too that definite case marking is the default marking used for predicate nouns with copula verbs in most dialects.

4. Perhaps the fact that most numerals precede the noun is motivated by the fact that complex numerals are essentially formed in right-branching and coordinate structures from greater to smaller number, e.g. *bi mila hirurehun ta berrogei ta hamabost* “2355” lit. “two thousand three-hundred and two-twenty and ten-five”, cf., e.g., Araujo 1975.

5. Basque relative clauses have been studied for instance by de Rijk 1972a, 1972b; and Oyharçabal 1985, 1986, 1987.

6. A common exception to this is found with verbs which take locative complements, such as *bizi* “live”, so that relative clauses such as *the house where I live* are possible using this construction. The crucial aspect seems to be recoverability of the noun’s role, which is greatest in the case of subcategorized arguments, cf. e.g. Keenan and Comrie 1977. There is an interesting construction, in which the *-ko* genitive suffix is attached to a regular finite relative clause modifying a time or place noun, *ama etorriko de-n-eko eguna* “the day when mother will come back.” The locative character of this genitive has been noticed by many before, e.g. de Rijk 1993.

7. In Spanish very clearly non-predicative modifiers precede the noun, cf. e.g. Bolinger 1967, 1972a.

8. The morphemes of finite verb forms are not always transparent and the morphophonemics are somewhat complicated, though we can often recognize argument morphemes (e.g. *-t/-da-* for the ergative). For discussion of some of the issues, cf. e.g. Heath 1977; Trask 1977, 1984. In the examples found here I will not gloss those morphemes.

9. The auxiliaries sometimes can also be used as main verbs, with the intransitive ones (those lacking the ergative argument) acting as copulas and the transitive ones as ‘possessive’ verbs (or ‘transitive copulas’), cf. *have*.

10. In addition, in most dialects, the verb may also have second-person ‘agreement’ with the interlocutor, with different forms depending on degree of ‘formality’ and the sex of the addressee (only in the informal conjugation). These *allocutive* forms do not appear in our corpora and will not be mentioned further. For more information, see, e.g., Rebuschi 1984/1982, Oyharçabal 1993.

11. In written Basque the two words are written separately. In the Spoken Basque Corpus I have transcribed the verb forms separately as well most of the time, except when the two are inextricably blended, as happens in some dialects. In the excerpts found here I sometimes unite the finite and the non-finite forms of the verb with an underline _ character and given them a joint gloss for simplicity’s sake.

12. The Basque linguist Mitxelena believes that the constraint against having a finite verb in rheme-initial position stems originally from the clitic character of finite verbs in Old Basque, though that theory has no relevance to the synchronic account and may not even be correct. Mitxelena argues that “Even today the tendency to avoid that a bare verb form begin a sentence is quite general, except in imperatives: usually it follows a nominal [*sic*] verb form (*nomen actionis*, participle, root), a noun, pronoun or adverb, or an affirmative or negative particle. It seems clear that this, which today is a syntactic mechanism in order to highlight the word which immediately precedes the verb, had its origins in the enclitic character of the finite verbal forms, at least as pertains to auxiliary verbs. That is, this should be interpreted in the same way Wackernagel proposed for similar facts in Indo-European in his famous article” Wackernagel 1892 (Michelena 1957:177, *fn32*; my translation, J.A.).

The similarities are suggestive, but actually I fail to see the relevance of these facts to the modern Basque structure or even its origins, which seems to have a logic all of its own. [“Aún hoy es general la tendencia a evitar que una forma verbal nuda inicie la frase, excepto en imperativos: normalmente sigue a una forma nominal del verbo (*nomen actionis*, participio, radical), a un nombre, pronombre o adverbio, o a una partícula afirmativa o negativa. Parece claro que esto que hoy es un precedimiento sintáctico para poner en relieve la palabra que precede inmediatamente al verbo tuvo su origen en el carácter enclítico de las formas verbales finitas, al menos en los auxiliares. Es decir, que debe interpretarse en el sentido propuesto para hechos indoeuropeos semejantes por

Wackernagel en su famoso artículo de *Indogermanische Forschungen* I, 333 ss.; cf. también R. Thurneysen, «Zur Stellung des Verbuns im Altfranzösischen», *ZRPh* 16, 289 ss. Concuerdan con esto otras particularidades del verbo finito: frecuencia de *d-* (cf. § 6), escasez de oclusivas aspiradas, apócope (*-d*, *-t*, característica de primera persona sg., ante sufijo *-da-*, *na-iz* «soy» frente a *iza-n* «sido», sul., *n-iza-te* «sería», etc.)”].

13. Others have disagreed who would like to derive behavioral differences between ‘subject’ and objects from structural considerations.

14. (a) Lower verb with no complements: 7 <V>Behar + Aux; 3 <V + egin>Behar + Aux; 2 S<V>Behar + Aux; (b) Lower complements precede lower verb: 2 <OV>Behar + Aux; 4 <XV>Behar + Aux; 1 S<XV>Behar + Aux; 1 A<XV>Behar + Aux; (c) Affirmative broken lower clauses: 1 <V>Behar + Aux<O>; 1 <V>Behar + Aux<SVtzen>; 1 <V>Behar + Aux<Vtzera>; 2 <V>Behar + Aux<X>; 1 <V>Behar + Aux<VtzenX>; 2 <V>Behar + Aux<XX>; 1 O<V>Behar + Aux<X>; 1 S<V>Behar + Aux<X>; (g) 1 FOC<V>Behar + Aux; 1 QnWrd<V>Behar + Aux.

15. (a) Main clause: 1 *-la_Pred*, 1 *-rik*, 1 main; (b) Lower clause: 2 OV, 6 XV; Main clause type: 5 main, 2 *-an_Adv*, 1 *-la_Compl*; (c) Main clause type: 1 *-la_Pred*, 1 *-lako*, 1 *-na_HRCL*, 1 *zeatik*; (d) Lower clause: 3 OV, 3 XV, 1 XXV, 1 XOV; Main clause type: 1 *-la_Compl*, 7 main; (e) Lower clause: 1 VI, 5 VO, 1 VOX, 2 VX, 2 VXX; Main clause type: 1 *-nean*, 2 *nola...*, 8 main; (f) Lower clause: 2 OVX, 1 XVX; Main clause type: 2 main, 1 HLRCL; (g) 1 AXVVtu.

16. Other attested verbs with *-tzen* complements are (total/postverbal): *jarri -tzen* “start, continue”, 2/1; *segitu -tzen* “continue”, 2/1; *jarraitu -tzen* “continue” 5/5; *utzi tzen* “allow”, 2/2; *bukatu -tzen* “finish”, 1/0. Similar verbs with other types of complements: *hasi Adv* “begin Adv.”, 4/2; *gelditu -ra* “get, start” 7/5. With a nominalized complement: *Espero -tzea* “hope”, 1/1; *egoki -tzea*, 1/1. Speech verb complements: *esan -tzeko* “tell to”, 8/8. Clause internal purpose clauses: *joan -tzera* “go to” 6/5; other *-tzera* 1/1; other with *-tzeko* complement: 12/11.

17. Two others have a special function (transitive weather verb) and one has an adverbial *-ka* complement.

18. Completive clauses often seem to derive from adverbial clauses and thus completive subordinators (‘complementizers’) often derive from adverbial markers. Notice that adverbial clauses are sometimes used in English, Spanish, and also Basque, for example, as completive clauses, e.g. *My sister said how she was going to come to visit me*.

19. The *-tu* verbal form is also used as the unmarked form of the verb in most dialects of Basque and is even used as the non-finite version of the imperative (see below for more details), is strictly verb final in the written corpus, but not so in the spoken corpus.

20. Unit Type = *Fin_Cl_Main*; Speech Act = *Body_Statement*; Polarity = *Aff*; Predicate arguments = *Abs / Abs_Erg / Abs_Dat / Abs_Dat_Erg*.

21. Two other functions which are often expressed through detransitive constructions, reflexives and reciprocals, are formally fully transitive in Basque, except in some dialects and under the influence of Spanish.

22. As I argued in Chapter 1, Basque does not need an object promoting passive, since all it needs to do to make the object the topic of the sentence is to place it in topic, clause initial, pre-rhematic position. Notice however, that spoken Spanish is much like Basque in this respect. The Spanish promotional passive, analogous to the English one, is almost completely absent from spoken Spanish. The most common passive in spoken Spanish by far is the demotional or reflexive passive, which doesn't necessarily or typically entail the promotion of the object to topic role. As I mentioned in Chapter 1, in Basque deverbal adjectives can be used in ways which are reminiscent of English-like passive constructions. The predicate in these constructions is stative, however, unlike in true passives (cf., e.g., Osa 1990:241; Euskaltzaindia 1993:428).

23. Unit type = *Fin_Cl**; Ergative ellipsis = *No*; Other codes = **Erg_Syncretism**.

24. Azkue 1891 has a Basque and a Spanish version side by side. Azkue's Basque version of this statement: "Esapide araturik *ezdaukala* euskaraak' batek baino geiagok esan daue; nik aldaitan ondoen' euskaraak daukan esapidea baino arautuagorik ezin egon leikela erakutsiko daut." Azkue's Spanish version says: "Más de uno han dicho que el eúskara *no tiene* construcción regulada; yo, de la mejor manera que me sea posible, he de manifestar que no puede haber construcción más regulada que la que tiene el eúskara" (Azkue 1891:334).

25. The original quote says: "se coloca por delante el elemento que más imporancia tenga, el que ménos tenga por detrás". This is very reminiscent of an idea that Givón has argued for in recent years. Azkue credits Astarloa 1883 with this basic idea (*Discursos filosóficos*, 755), though I was not able to identify this idea in this reference, an extremely obscure and linguistically rather uninspired work.

26. Azkue's 5th rule says that "[w]hen the elements are near equal in importance, first comes the verb, next the verb (or non-verbal predicate) and the auxiliary, if needed; the being [i.e. the subject] may well be placed before the object or after the verb and the auxiliary", e.g. "*Etse+bat egin dau Andresek=Andresek etse bat egin dau*", i.e. house one he.made.it Andrew (OVS) = Andrew house one he.made.it (SOV) (Azkue 1901:344-45; my translation, J.A.) ["QUINTA REGLA. Cuando los elementos tienen casi idéntica importancia, primeramente se dice *el objeto*, á continuación *el verbo* (ó el atributo) y el auxiliar si lo necesita; *el sér* puede muy bien decirse antes del objeto, ó detrás del verbo y *del auxiliar*."].

27. Cf. Azkue 1969/1923-5, and Altube's 1934 response, and Azkue's 1935 rejoinder.

28. Azkue confesses to ‘correcting’ the stories he collected when they had “Spanish syntax” (cf. Azkue’s *Euskaleñaren Yakintza*, vol. I, p. 18; mentioned in Villasante 1980:23-24).

29. The same thing can be said about the lexicon, where foreign influences have been studied more deeply and are better understood. It is true that purists were trying to do away with words which had been part of the Basque vocabulary for centuries and substitute them with newly coined words, and this has been properly ridiculed. It is also true, however, that in some cases lexical borrowings were not totally entrenched and were perhaps influencing the loss of perfectly good Basque words, something that speakers, and particularly writers, are perfectly entitled to recognize and do something about.

30. Villasante argues that “the mentioned book by Altube, based primarily on observed tendencies in today’s spoken language, establishes very precise ordering rules for the elements of the Basque clause. But it makes these tendencies too absolute, making them to be rigid, unchangeable and universal. Thus it gives too narrow a picture, which doesn’t correspond to the reality in the written language, nor even to the spoken language overall” (Villasante 1980:232; my translation, J.A.) [“El citado libro de Altube, basándose principalmente en tendencias observadas en la lengua hablada de hoy, establece unas reglas muy precisas de ordenación de los elementos de la frase vasca. Pero absolutiza con exceso las dichas tendencias, convirtiéndolas en normas rígidas, inmutables y universales. De este modo da una imagen demasiado estrecha, que no responde ni a la lengua escrita ni aun a la misma lengua hablada en su totalidad”].

31. Altube’s focus positioning rules were criticized sharply by Basque linguist Mitxelena, going back to Michelena 1953. Altube rebutted Michelena 1953 in Altube 1956. Villasante, who had lauded Altube’s 1929 work, sided with Mitxelena in Villasante 1956 (though more diplomatically than him). Mitxelena’s arguments against Altube were several: (1) Altube’s rules may work for some dialects (more or less), but they are not general enough; (2) Even in those dialects where they work, they are not universal; (3) Good classical writers, like Axular, do not follow those rules, because they are not the rules of speech either; (4) Even if those were the rules of speech, they may not work for writing; and (5) Written sentences following these rules strictly make for very difficult reading; (6) When the ancient authors introduced new constructions in their translations they were enriching the Basque language, not corrupting it. After reading Altube’s insightful original work one gets the impression that Mitxelena was overreacting, or else, that he was reacting against to the excessive zeal with which some authors were using Altube’s rules (cf. Irigoyen 1980 for an interesting account of the debate).

32. In Altube’s words: “Fortunately, in the people’s Basque language the rule I have just expounded is rarely broken. No doubt because the life of Basque literature is so weak, the sickness of *erderism* [cf. *erdera* = foreign language] causes greater havoc (devastation) there” (Altube 1920:9, emphasis in the original; my translation, J.A.) [“Afortunadamente, en el lenguaje popular vasco pocas veces se infringe la regla que

acabamos de exponer. No podemos decir otro tanto del euskera escrito. Sin duda por ser muy débil la vida de la literatura euskérica, hace en ella mayores estragos la *enfermedad del erderismo*”]. The 1929 version says: “In the popular Basque language rarely is the rule just mentioned broken, except in some Basque speaking areas which, unfortunately, are starting to feel already the invasion of erderism that we’re discussing here” (Altube 1929:7; my translation, J.A.) [“En el lenguaje popular vasco pocas veces se infringe la regla que acabamos de exponer, fuera de algunas zonas euskaldunas que, desgraciadamente, empiezan a sentir ya la invasión del erderismo que nos ocupa”].

33. According to Altube “we must admit that in sentences with a long and complex inquired element, even in the spoken language the mentioned rule is frequently broken when the verb of the sentence expresses ideas such as *say, answer*, etc.” (Altube 1929:11) [“Hay que reconocer que tratándose de oraciones con miembro inquirido largo y complejo, aun en el lenguaje hablado se infringe con frecuencia la regla establecida cuando el verb de la oración expresa ideas tales como *decir, contestar*, etc.”].

34. Altube mentions that “Eastern Basques frequently use a special, very expressive, turn to emphasize even more the inquired [focus] character of the parts of the sentence. It consists of taking apart the periphrastic verb, inverting its parts, with the result of placing the non-finite verb after the auxiliary verb” (Altube 1929:14; my translation, J.A.) [“Los vascos orientales se valen con frecuencia de un giro especial, muy expresivo, para recalcar aún más el caracter inquirido de los miembros de oración [*sic*]. Consisten en desarticular la flexión perifrástica invirtiendo sus componentes a fin de hacer seguir el nombre verbal al terminative auxiliar”].

35. He argues that here “synonymy is almost perfect” and “the difference in meaning is barely perceptible” (Altube 1929:69; my translation, J.A.) [“la sinonimia parece casi perfecta” and “la diferencia de significación apenas es perceptible”].

36. According to Altube, “all the rules proposed in the previous chapters dedicated to the formation of ordinary sentences, the use of intensive terms, etc., etc., *are in general applicable, more or less rigorously*, in the composition of subordinate sentences (especially when they form part of an inquired phrase), with the additional requirement that in such sentences the verb must be the inquired element” (Altube 1929:143; my translation, my italics, J.A.) [“todas las reglas expuestas en los capítulos precedentes dedicados al estudio de la formación de las oraciones ordinarias, el uso de los vocablos intensivos, etc., etc., son en general aplicables, con más o menos rigor, en la composición de las subordinadas (muy particularmente cuando constituyen o forman parte de un sintagma inquirido) con la adición imprescindible de que en estas, repetimos, debe ser su verbo el elemento pospositivo”].

37. Altube does notice that the emphatic *ba* particle also typically not found in non-asserted subordinate clauses when the verb is the single element in the clause, a particle that is obligatory in such cases in asserted clauses, cf. Altube 1929:143, *fn*1.

38. The original reads as follows: “oraciones subordinadas declarativas cuyo verbo se caracteriza por el sufijo *-la* ... se nos presentan a veces (al menos en el euskera actual) como si constituyeran oraciones principales, en tal forma que la principal auténtica, desempeña en ellas un papel secundario o desaparece por completo” (Altube 1929:144). By disappearing, Altube is referring to the possibility that there is no overt main verb and that these completive clauses be ‘orphans’ (cf. Altube 1920:58-59).

39. According to Altube, “in towns where Basque is giving way to Spanish, [negative inversion] reaches almost all subordinate negative clauses” (Altube 1929:152). (Cf. “[e]n los pueblos donde el euskera va cediendo terreno ante el erdera, la irregularidad que nos ocupa alcanza a casi todas las oraciones subordinadas negativas”). His examples involve negative inversion in a variety of subordinate clauses: conditionals (*ba-*), causatives (*-lako*), temporal (*-nean*), and adverbial (*-larik*).

40. The original quote says: “Edozein osagai delarik ere, ikusten dugu maiztasun berdintsuz betetzen dutela bai aditzaren ezkeraldea eta bai eskubialdea, aditzarekin batera multzo komunikatibo bakar bat osatzen dutenean. Halaber, osagaiak soilik duenean fokutasuna ere, normaltasun osoz doa aditzaz eskubialdera” (Osa 1990:222).

41. According to Osa, “[w]hen the complement is long, such as when it is a subordinated clause, even if it is the focus, it almost always goes to the right of the verb” (Osa 1990:222; my translation, J.A.) [“osagaia luzea denean, demagun klausula subordinatu bat, nahiz eta fokua izan, aditzaz eskubitara doa ia beti”]. And “[p]redicate complements and modal adverbials ... almost always go in preverbal focus position” (Osa 1990:222; my translation, J.A.) [“osagai predikatiboena edo moduzko adverbioena, ... ia beti fokuaeren posizio erregulatuan baitoaz”].

42. According to Osa, “[i]t is obvious that the constituent order is quite free in storytelling, the communicative strategy being expressed through prosodic factors, that is intensity accents, pauses and intonation” (Osa 1990:222; my translation, J.A.) [“Dena dela, bistakoa da sintagma-ordenak oso libre jokatzeko duela ipuinketan, estrategia komunikatiboa batipat faktore prosodikoetan, hots, azentu intentsiboan, etenean eta intonazioan gauzatzen delarik”].

43. There are some caveats, adds Osa, such as that the Circumstantial cannot be a modal and that time and place adverbs go before the subject (Osa 1990:126). Lafitte’s example for a long neutral Basque sentence has a time adverbial behind the ‘subject’: *Aita sainduak / atzo / bi erresumeri / aphezpiku bat / ararteko / igorri dute* “Yesterday the Holy Father sent a bishop to the two kingdoms as mediator.”

44. The original version says “[e]l morfema gramatical que relaciona o enlaza a dos vocablos ... gusta de situarse, generalmente, en medio de los dos. En euskera (idioma pospositivo) dicho morfema debe ser colocado necesariamente al final del vocablo regido ... luego el regente habrá de seguir a este ... En los idiomas prepositivos ocurre lo contrario: el vocablo regido va precedido del morfema gramatical ... y para que este [*sic*]

quede en medio es necesario que el regente ocupe el primer lugar de la frase” (Altube 1929:165-166).

45. The Spanish original says: “Para obviar esas dificultades, en las traducciones de textos erdéricos, es recomendable, como primera providencia, desarticular, subdividir y reducir a varias clausulas [*sic*] más cortas los sintagmas, frases y períodos castellanos excesivamente largos, antes de proceder a su versión al euskera...” (Altube 1929:166). Some authors, such as Villasante, have called for the emulation of classical models, and in particular those of the classical writer Axular, who made extensive use of complex sentences and right branching structures, as a way of enriching the language.

46. Unit type: *Fin_Cl_Main* vs. *Fin_Cl_Sub*; Speech Act: *Body_Statement** vs. *Head* vs. *Tail* vs. *Predicate* vs. *Rel_Cl*; Polarity: *Aff* vs. *Neg*; Story: *Pear* vs. *Chap*. A small number of questions and other non-statement (main-clauses) speech acts were not counted (Speech act = *Content_Qn*, *Y/N_Question*, *Imperative*). Most involve questions on the part of the interlocutor and most are not full sentences. Subordinate (embedded) questions were counted as complement clauses.

47. Unit Type = *Non_Fin_Cl*; Speech Act = *Complement* vs. *Predicate* vs. *Body_Statement** vs. *Head* vs. *Tail**, vs. *Head_Statement* vs. *Rel_Cl*. ‘Complements’ are subjects or objects or complements of speech verbs, for instance. ‘Predicates’ are all other complements and ‘secondary predicates’.

48. Unit Type = *Fin_Cl_Main*, *Fin_Cl_Sub**. Speech Act (main clauses) = *Body_Statement*; Speech Act (dependent clauses) = *Body_Statement** vs. *Head* vs. *Tail* vs. *Rel_Cl*.

49. Unit Type = *Non_Fin_Cl**; Speech Act = *Argument_Cl*, *Body_Statement**, *Complement*, *Predicate*, *Head*, *Rel_Cl*, *Tail*.

50. Main affirmative statements: Unit Type = *Fin_Cl_Main*; Speech Act = *Body_Statement*; Polarity = *Aff*. In addition to these 28 possible orders for main clauses, there are about 112 other minor orders in the sample, which account for 25.4% of all the main, affirmative clauses. Dependent statements (completive clauses): Unit type = *Fin_Cl_Sub*; Speech Act = *Body_Statement**; Polarity = *Aff*. Other finite dependent clauses: Unit type = *Fin_Cl_Sub*; Speech Act = Not *Body_Statement**. Non-finite clauses: Unit type = *Non_Fin_Cl*; Polarity = *Aff*. These templates do not include conjunctions or other intonation unit internal setting elements, such as *and*, *then*, *suddenly*. Written corpus statements: Unit type = *Fin_Cl_Main*; Speech Act = *Body_Statement*; Polarity = *Aff*. Not counted are 98 main finite clauses which consist of a quote, followed by the verb and the other elements (the verb always follows the quote): Unit type = *Fin_Cl_Postquote*.

51. Unit type = *Fin_Cl_Main*; Speech act = *Body_stat**; Polarity: Not *neg*. Word order = *V* vs. [*V** and not *V*] vs. [**V* and not *V*] vs. [**V** And Not *v* And Not *v** And Not **v*].

52. Non-complement, dependent clauses (non *-la* clauses) have the following subordinators: *-la (Adv.)* (41 ea.), *-la eta* (1), *-lako* (23), *-larik* (1), *-n* (embedded question, not counted in these statistics) (8), *-na* (headless relative clauses) (10), *-nean* (62), *-nik* (1), *-nez* (2), *ba-* (3), *ba- ... bezala* (3), *-n bitartean* (6), *nola ...* (42), 'Subjunctive' (1), *zeren ...* (13), *zergaitik ...* (9).

53. Word Order codes: (a) Verb-initial = *V** and not *V*; (b) Verb-final = **V* and not *V*; (c) verb-medial = **V** and not *V* and not *V** and not **V*; (d) single-verb: the rest.

54. Unit type = *Fin_Cl_Sub*. Speech act = *Head** Or *Tail** Or **complement** Or **Predicate**. Polarity = Not *Neg*.

55. Unit type = *Non_Fin_Cl*.

56. The tokens meeting these characteristics had to be counted by hand, since the template causes multiple mismatches, due to the fact that there are other *a*'s in the code, e.g. *Xla*.

Chapter 3

Discourse pragmatics and linguistic form

3.1 DISCOURSE PRAGMATICS AND THE DOMAIN OF WORD ORDER

3.1.1 Introduction

It is widely recognized that the formal expression of referents (entities) and other 'ideas' expressed in asserted and other propositions, including their 'size', prosody, and relative ordering, reflects in various ways the pragmatic, or contextual, characteristics of those 'ideas', that is, their 'discourse pragmatics'. This relationship between form and pragmatics is often rather iconic, following universal cognitive and communicative principles, but it is typically mediated by language specific defaults and 'preferences' deposited as 'sediments' in more-or-less arbitrary constructions, which accounts for the differences found among different languages.

In this chapter I will look at some of the pragmatic categories and principles which influence linguistic form to varying degrees in all languages, and in particular the constituent order in assertion constructions cross-linguistically. I will look at the pragmatic properties of propositions ('propositional ideas') and their component ideas, such as referents and other ideas. Within asserted propositions, I will look at two types of pragmatic characteristics of referents and other 'ideas' expressed by elements in them: their inherent **pragmatic-referential** (cognitive) **status** in the minds of speakers and

hearers, and the **pragmatic-informational roles**, or **functions**, that they play in asserted propositions.

Linguistic units are used to express 'ideas', or concepts, of different types, primarily entities, properties, events and states (cf. Chafe 1994a). Nominals are typically used to express ideas of entities, real or imaginary, individuals or sets, although they can also be used to express properties when they are not referential (cf. She is *a thief*). Ideas of entities, or referents, are the most 'permanent' and easy to form, and retain, a mental picture of, and are the ones which are used to 'peg' information to, as we will see. (All references to *referents* and *entities* in this volume are meant to be referenced to the ideas of those referents; that is, to their mental representations in the speaker's and the hearer's discourse models.)

States and events are more complex, 'propositional' ideas, which involve other simpler ideas. Some of their cognitive properties, however, are not unlike those of simpler and concrete ideas, such as referents. Verbs express the central component of events and states and represent abstractions over such events and states and thus are more abstract and less 'pictureable' than nominal referents or propositions.

These units of language are deployed as speech by means of constructions which put the basic units of language together for some communicative purpose. Linguistic utterances consist primarily of what we may call asserted propositions, assertions, or speech acts, that is, propositions used for a communicative purpose, or function, such as informing about states of affairs and requesting information. Assertions, thus, are the fundamental unit of communication. Other constructions are used to express non-asserted propositions, which perform other functions relative to the assertion clause.

What we call a discourse is nothing but a sequence of linked, or interconnected, speech acts.

Some propositions are not used to advance the communication, but rather as 'backgrounds' or 'settings' for asserted clauses, or as modifiers of referents (relative clauses), or as internal 'arguments' or 'secondary predicates' inside another proposition. In other words, propositions (clauses) can be used as modifiers (adverbials and adjectivals), arguments, and predicates inside assertions, in addition to being used as assertions. Clauses with dependency markers are typically not asserted, but rather act as complements or modifiers inside an assertion. There are exceptions to this, however, and completive complement clauses and some clauses which follow main assertions, for example, seem to function as assertions which are dependent on the main assertion. That is, they are both part of other assertions and assertions in their own right (cf. below).

Asserted propositions (such as statements, emphatic statements, questions, imperatives, and so on) are typically uttered in a single intonation unit (cf. Chapter 1). In other words, there is a strong correlation between the clause unit and the intonation unit. Of course, not all intonation units consist of clauses: non-clausal 'settings' are also typically in their own intonation unit, as are 'asserted fragments'. Also, there are motivated mismatches, such as when a proposition is divided into more than one unit, which happens occasionally with complex propositions. Also, sometimes multiple propositions can be united under a single, complex intonational contour. Each intonation unit has one major accent or peak of its intonational contour. As we will see, where that accent lies depends to a great extent on pragmatic properties of the assertion. All these issues will be elaborated on below.

3.1.2 The structure of simple assertions

Each asserted proposition contains typically at least two ‘ideas’, but rarely more than three, all of them united by the state of affairs being expressed. In practice, typically only one of these ideas actually advances the communication (in the context of the whole asserted proposition), which may be the verb itself, or some other complement *in addition to or together with* the verb. Below I will call this the **focus** of the assertion, which is the element which receives the main accent of the assertion’s intonation unit. In other words, there seems to be a strong cognitive constraint on the size and complexity of information units found in natural language. This principle is what Chafe (1980, 1992b, 1994a) calls the “one-new-idea-at-a-time constraint”, what Givón’s (1984) calls “one chunk [of new information] per clause” principle, and what Du Bois (1987) calls the “One Lexical Argument Constraint”.

This principle is also related to the notion of ‘preferred argument structure’ or ‘canonical sentence schema’ put forth by Du Bois, Lambrecht, Duranti, Ochs, Bentivoglio, and others.¹ Du Bois’s (1981, 1985, 1987) concept of **preferred argument structure**, for example, refers to the typical or least marked structure of clauses in natural languages, according to which (proto-)typical, unmarked clauses consist of a verb and at most one nominal argument. In addition to having a single most-salient element, the vast majority of asserted propositions are also ‘grounded’ on a single ‘idea’, one privileged element among those involved in the event or state, which in connected discourse is typically a ‘continuous’ referent. I will refer to this element as the **topic** of the assertion. We will return to these issues below.

The basic asserted clause and its satellites, whether clausal or not, placed in preceding and following intonation units, as we shall see in the next section, constitute what is traditionally known as the sentence. Elements which precede the assertion proper in their own intonation units constitute what I will call ‘settings’ and are outside of the scope of the assertion: they lay the ground for the assertion proper. Sentential elements which follow the assertion proper are extensions of the assertion and are themselves asserted. The structure of the asserted sentence can be seen at a glance in Table 3-1.

(Setting *)
Asserted clause (Assertion proper)
(Extension *)

Table 3-1: Asserted sentence and its parts

Obviously, the existence of this basic sentence structure, and the existence of a constraint on the number of units inside the asserted clause proper, both of which are valid crosslinguistically, have important consequences for the analysis of word order. The basic sentential structure determines, to a very large extent, the domains of word order, primarily the asserted sentence and the asserted clause. The preferred argument structure constraint, on the other hand, determines the number and type of units whose ordering needs to be accounted for. As we will see in the next section, the information units mentioned here are intrinsically tied in with the intonational units of speech and the cognitive limitations on speech production (cf., e.g., Chafe 1994a).

3.2 THE INTONATIONAL AND INFORMATIONAL STRUCTURE OF THE SENTENCE

3.2.1 Intonation structure and information structure of the sentence

In my analysis of spoken Basque I have found, as others have before me, that speech proceeds in 'spurts of verbalization', each under a single intonation contour. The two major intonation contours separating these intonation units are the final or period contour, and the continuous, or comma intonation, with some more specialized variations (cf., e.g., Crystal 1972, Cruttenden 1986; Bolinger 1986, 1989; Chafe 1994a and elsewhere).

Each intonation unit has one word which receives the unit's main accent. Occasionally, however, intonation units have a more complex structure and may even be blends of simpler units. The nature and diversity of intonation units and contours is a complex field of study, but their most basic characteristics and their relationship with information structure are quite clear and are applicable crosslinguistically.

In my study of spoken and written Basque, I have found the sentence to be a valid unit of analysis from an informational and intonational point of view. The sentence is a unit which consists of a core, asserted clause and additional optional elements preceding or following it: **assertion settings**, which are outside the assertion proper and typically precede it, and **assertion extensions**, asserted, but dependent elements, which follow the asserted clause. Thus, from an information structure perspective, sentences are not strings of elements, but rather are structured units which center around the asserted clause and which have specific functions. In the transcripts used in this study I divide speech into intonation units (cf., e.g., Edwards and Lampert, eds., 1993 and the articles therein). In practice the indeterminacy in breaking up the stream of speech into such units was

rather minimal. Written text can also be divided into such units fairly easily, as Chafe has argued based on the intonational contours resulting from reading such texts aloud and from the functions performed by the parts of the sentence (cf. Chafe 1986b, 1988b, 1992a, 1992b).

I have found that ‘substantive intonation units’, that is, fully-formed ones, those which are not ‘regulatory’ (e.g. *mhm*, *well*), or aborted (false starts, and the like), fall into three categories, depending on their function in the sentence:²

1. **Sentence-heads:** ‘preposed’ or ‘left-dislocated’ optional units which precede the asserted clause. These units, whether sentential or not, provide background information or **settings** for the assertion proper. They include setting adverbials, which may be clausal, as well as so-called ‘left-dislocated’ arguments of the predicate.
2. **Sentence-body:** the core of the sentence, the asserted clause itself. As I mentioned earlier, there are different types of assertions, such as statements, questions, commands, etc., as well as different variations of each one of these.
3. **Sentence-tails:** like the sentence heads, these are also optional units. Here we find what we might call addenda, extensions, or closings which are added to the assertion, and thus form part of the assertion. We also find what we can call ‘dependent assertions’, such as, e.g., post-assertion causatives, are also found here. These are clauses which have the form of dependent clauses, but which behave as paratactic assertions.

In (3.1) below we can see an example of a sentence containing all three parts of this structure: one asserted clause with three sentence heads and three sentence tails.

- (3.1) The other day,
while I was still in bed,
my little brother,
he came to see me.
At seven in the morning.
To tell me about the accident.
Cause he thought I should know.

Most sentences, of course, are not this elaborate. The vast majority of sentences consist of single asserted clauses. In the next section we will look at the statistics about sentence composition in the Spoken Basque Corpus.

3.2.2 The intonation units of the Spoken Basque Corpus

A little over $\frac{1}{2}$ of the approximately 6,000 intonation units into which the Spoken Basque Corpus is divided are sentence-body intonation units. The vast majority of these are statements (declarative assertions), although there are a handful of questions and imperatives (about 1% of each). Less than $\frac{1}{6}$ are sentence-head intonation units. About half as many ($\frac{1}{12}$) are sentence tail units. Over $\frac{1}{6}$ are regulatory units or unfinished units.

Thus, most sentences consist only of the core assertion, uttered in a single intonational contour. In addition, about $\frac{1}{3}$ of all assertions are preceded by one or more sentence-head intonation units in which orientation, setting, or other ‘grounding’ elements are placed. Finally, about $\frac{1}{6}$ of asserted clauses are followed by sentence tail intonation units containing ‘afterthoughts’, ‘clarifications’, ‘rectifications’, ‘justifications’, and other ‘addenda’, which add to, or otherwise modify, the assertion. Unlike sentence-head elements, sentence-tails are part of the assertion or constitute additional, dependent assertions. The type of dependent assertions which are most

commonly found here in the corpus are causatives and post-presentational dependent assertions on a new referent.

In the excerpt in (3.2) we can see a typical setting intonation unit in (3.2:78), a typical main intonation unit (3.2:79), and a typical tail intonation unit (3.2:80).

(3.2) 93C2B01

- 74 Geo poliziak <etorri> ta <ixistu jotze-de> <jatea juteko>. ..
 then policemen:DEF come:PFV and whistle blow:IMPFV-they.have.it
 eat:NOMIN:ALL go:NOMIN:GEN
- 75 Ta juteie jatea, (MAIN)
 and they.go eat:NOMIN:ALL
- 76 ta asteie jaten, (MAIN)
 and they.begin eat:IMPFV
- 77 <BREATH>
- 78 ta an komedorin jaten ai-dile, - (SETTING)
 and there dining.hall:DEF:LOC eat:IMPFV busy/doing-they.are:ADV
- 79 bi tipo etorri-die, (MAIN)
 two guy they.have.come
- 80 beste baten bille. (TAIL)
 other one:GEN in.search.of

“Then the police come and they blow the whistle to go eat. / And they go eat, / and they start eating, / and while they’re there in the dining room, / two guys come, / looking for someone else.”

Immediately before this sentence in (3.2:78-80), which contains a head and a tail in addition to the asserted clause, we can see two asserted clauses, in (3.2:75) and (3.2:76), without any settings or additions.

In this excerpt we can also see, however, that occasionally some intonation units are more complex than the ideal. Thus in the intonation unit in (3.2:74), which was uttered quite fast, we see sentence which is a combination of three clauses under what seems to be a single intonational contour. This sentence seems to be a composite of two intonation units, as its two main accents suggest, and indeed it could be uttered with two separate intonational contours. It is actually not uncommon for sentence-heads and

sentence-tail intonation units to form a single macro-intonation unit together with the main assertion intonation unit, something which is most noticeable in rapid speech. The accentual and intonational properties of those super-units, however, leave no doubt as to their complex structure. This is perhaps the most common cause of indeterminacy when subdividing speech into informational-intonational units.

The first clause in (3.2:74) is a common non-finite type of clause in which a state of affairs which provides background (setting) information for the assertion is expressed. In English and other languages this type of function is typically performed by what looks like a main assertion. The second unit of this sentence in (3.2:74) is also composed of two clauses (*they blow the whistle | to go eat*), the main assertion and a ‘bound’ purpose clause, which is not accented.

3.2.3 The ordering of constituents in the sentence

Because of this structure, inherent in the sentence, we can see that we talk about the ordering of the elements of the sentence, we are not talking about the relative ordering of all the elements of the sentence with respect to each other. From a discourse pragmatic perspective the ordering relationships seem to be the following:

1. The order of setting elements with respect to the asserted clause. Settings typically come before, but, as we will see below, they may also come after, in an inverted order, when the assertion is emphatic or has a very salient focus.
2. The order of setting elements among themselves (when there is more than one). The main ordering parameter involved here seems to be ‘scope’, since each setting element reduces the context in which the assertion applies.

3. The order of fragment assertion extensions, including dependent assertions, with respect to the main assertion. Naturally, extensions always follow the main asserted clause.
4. The order of extensions among themselves (when there is more than one). There may be some ordering preferences among extensions which are planned.
5. The order of elements inside the main, asserted clause. This is the domain of word order which I will be concerned with primarily in this study. This domain also has structure of its own, a structure which is also of an informational, and not merely syntactic in nature. I will return to this issue below.

The importance of distinguishing between the different informational and intonational parts of a sentence and not referring to the sentence as a continuous string of equivalent elements has been noticed, for instance, by Chafe (1984b), who argued that the principles which determine the order of satellites (settings and extensions) with respect to the main clause are different from the principles which determine the order of elements inside the main clause.

Chafe's study deals with the position of adverbial clauses in samples of spoken and written English. Chafe argues that adverbial clauses may appear in several positions, depending on their pragmatic function. In Table 3-2 we can see the distribution that Chafe found for adverbial clauses in a sample of English.

Position of adverbial clause	Speech	Writing
a. Preposed free (= sentence-head)	40%	39%
b. Preposed bound (= sentence-body initial)	2%	1%
c. Postposed bound (= sentence-body final)	27%	37%
d. Postposed free/unbound (= sentence-tail)	31%	23%
Total	100%	100%

Table 3-2: Position of adverbial clauses in spoken and written samples of English (as reported in Chafe 1984b).

Chafe found that these adverbial clauses are twice as common in writing as they are in speech, and he also found that the principles which govern their order is different when they are in inside the sentence-body (bound) than when they are detached (free or unbound).

Chafe argues that whereas ‘familiarity’ (‘new’ vs. ‘given’) is an important factor on the location of adverbial clauses inside the main clause, that principle is irrelevant for the unbound adverbials clauses.³ Although this intuition is correct to a large extent, in this chapter I will argue that the differences that we observe here have to do with the different **discourse functions** that these clauses, as well as any other sentence elements, perform in the sentence. Elements which are outside the asserted clause perform either a setting function or an assertion extension function. Elements which are inside the asserted clause, on the other hand, may or may not perform a number of other pragmatic functions as well, as we will see. I will return to this issue after looking in some detail at the nature of the different parts of the sentence: head, body, and tails.

3.2.4 Sentence-head intonation units

Sentence-head intonation units contain elements which provide the ground, stage, or setting for the main assertion, including optional setting adverbials, as well as

'dislocated' elements from the main clause. I will typically refer to these elements as **settings**. Settings are separated from the following units by a comma intonation and each one receives a full accent. In (3.3) below we can see some examples of setting elements (cf. also Chapter 1).

- (3.3) a. *When everyone left, I washed the dishes*
b. *In the morning, I washed the dishes.*
c. *In the morning, when everyone left, I washed the dishes.*
d. *Unfortunately, I didn't wash the dishes.*
e. *The dishes, I didn't wash them.*
f. *My mother, she didn't wash the dishes either.*

When there is more than one setting element in the sentence, as in (3.3c), the relative order of the different setting intonation units seems to depend on familiar notions of 'scope' or 'relevance'.

The notion of setting is equivalent to Chafe's (1984b) 'guidepost' which he identifies for what he calls "preposed free adverbial clauses" and other types of expressions.⁴ It is also equivalent to the notions 'topic' and 'background' information, as used by, e.g., Haiman (1978a, 1986) and Thompson and Longacre (1985),⁵ and to the notion 'theme', as used by Halliday (1985). Settings are often used to 'ground' the assertion either *thematically* ('left-dislocated topics') or *circumstantially*, which means primarily in space (by a *locative*) or in time (by a *temporal*), but may have a number of other functions as well.

Of the nearly 1,000 intonation units in the Spoken Basque Corpus which I classified as sentence heads, about ¼ are actually (non-clausal) false starts (e.g. *eta orduan ...* "and (then) ..."). Of the remaining ¾, about ¼ are nominals, and about ½ are other non-clausal adverbials. The remaining ¼ are clausal settings, ¾ of which are finite

clauses and $\frac{1}{4}$ of which are non-finite. We may divide all the setting elements into thematic settings (the $\frac{1}{4}$ which are nominals) and circumstantial settings (the remaining $\frac{3}{4}$ adverbial and clausal ones).⁶

The most common settings in the spoken corpus by far are *(eta) orduan* “(and) then”, *(eta) gero* “and then”, and variations such as e.g. *gero kartzelan*, “then in jail”, as well as other temporal adverbials, e.g. *Ta bapatean* “and suddenly”, *Ta azkenean* “and finally”. The most common, stereotypical setting adverbials, which express temporal continuity (e.g. *and then*), are very often not in intonation units of their own, but are rather found, attached to the beginning of the sentence body intonation unit, and thus unaccented.

3.2.5 Sentence-body (main) in tonation units

The main, and only obligatory, part of the sentence is the asserted clause, typically a finite clause, an instantiation of some type of speech act construction. Asserted clauses and their informational structure and ordering properties will be the main topic of this and following chapters.

As I said earlier, sentence body assertions constitute about half of all intonation units in the Spoken Basque Corpus. The vast majority of these, or about $\frac{4}{5}$ of them, consists of finite, affirmative statements. In addition we also find finite negative statements ($\frac{1}{25}$).

Finite completive (complement) assertions in their own intonation units are about $\frac{2}{25}$ of all asserted clauses. About 10% of them are negative. These right detached completive clauses seem to act as true assertions rather than as part of other assertions (cf., e.g., Thompson and Mulac 1988, 1991a, 1991b).

In the spoken corpus we also find insignificant numbers (2/25 of all) of non-finite, but asserted, clauses with an elided verb or auxiliary, as well as asserted clause fragments, where all the elements of the clause other than the assertion focus are 'elided' (see below).⁷ Questions, commands, and exclamations (emphatic assertions), are a small minority of the assertions in the spoken corpus.

3.2.6 Sentence-tail intonation units

Sentence-tail intonation units contain primarily extensions or modifications to the preceding assertion. These elements, which are always accented, as they are in independent intonation units, can be seen as part of the assertion, or even as (focus) fragments of implicit assertions. In the spoken corpus there are about 500 such tails. Of these, about 3/5 are non-clausal. A little over 1/5 are non-finite clauses, and somewhat over 1/10 are finite clauses.

Tails are different from what are sometimes called right-dislocations. These so-called right-dislocations are elements found at the end of the asserted clause, with a greater or lesser degree of integration but without constituting a separate intonation unit. Therefore they are unaccented. We will return to this issue below.

Tail elements are sometimes termed **afterthoughts**. Quite often, however, these are not afterthoughts in the sense that they do not contain elements which the speaker 'forgot' to mention during the assertion. Typically they are **additions**, or **elaborations** of the assertion, and they may be 'planned' or 'unplanned'. Sometimes they are **clarifications**, which expand on the verbalization of a particular element to make it more clear, or **rectifications**, which change or replace the verbalization of an idea

expressed inside the asserted clause, typically to make it more specific. We can see an example of what I have in mind in (3.4).

- (3.4) a. I washed the dishes.
b. All of them.
c. Last night.
d. Before I went to bed.
e. (Be)cause I felt like it.

The tail in (b) elaborates on the object of the assertion in (a). The tail in (c) adds a temporal fragment, equivalent to the assertion *I washed them last night*. The clausal tail in (d) is similar to the one in (c), but adds a more specific time. Finally, the tail in (e) can also be seen as a fragment assertion equivalent to *I washed them cause I felt like it*. It can also be seen, however, as an assertion in its own right, what I will call a ‘dependent assertion’.

Lakoff (1987), for instance, has shown that clauses in **sentence-tails** are assertions in their own right, since they display properties associated with asserted clauses, that is, ‘main sentence’ phenomena (cf. Emonds 1970, Hooper and Thompson 1973), as we can see in the examples in (3.5) from Lakoff 1987 (Case study 3, p. 473-4).

- (3.5) a. I’m leaving, because here comes my bus.
b. We should go on a picnic, because isn’t it a beautiful day!

As Lakoff notes, these inversions are not possible when the clauses are in sentence-heads (i.e. when they function as settings) (e.g. **Because here comes my bus, I’m leaving*, Lakoff 1987:474). Nor are they possible with constructions which have an intrinsic setting function, such as conditionals (e.g. **I’m leaving, if here comes my bus*). The similarity in form between tail causatives, for instance, and non-tail causatives thus

proves to be deceptive. In Basque, for instance, there is a specialized type of causative construction which can only appear in tails.

The intonation units in which such assertion expansions are introduced are sometimes tacked on to the end of the preceding intonation unit, duplicating the intonation unit final contour, and with only a minimal pause or, seemingly, no pause at all. This is particularly true of corrective afterthoughts. These elements, however, should not be viewed as part of the assertion proper, especially for purposes of word order. Strictly speaking they are additive elements which follow the assertion. It remains to be seen, however, whether some such sentence tail elements can be incorporated into an assertion construction proper.

In the following excerpt we find an example of an ‘instant addition’ of this kind in (3.6:224). The following two units (3.6:225-226) contain ‘elaborational’ tails.

(3.6) 93C2B01

224 kanpora jun-de. Kartzelatik. -
outside:DEF:ALL he.has.gone. jail:DEF:ABL

225 Kartzelako patiyoa.
jail:DEF:GEN patio:DEF:ALL

226 Kartzelatik kanpo.
jail:DEF:ABL outside

“he goes outside. of the jail. / to the patio of the jail, / outside the jail”

Although many tails seem to contain ad hoc additions to an assertion, some other tails seem to contain more specialized elements which are more tightly integrated with the preceding assertion. Thus we find for instance that clauses in which a new referent is introduced are very often followed by tail elements the function of which is to predicate some property of the newly introduced entity. Such predicates may be of any type, including finite clauses. In English such finite clauses take the form of relative clauses

and thus they are often thought of as extraposed modifying clauses. Their function, however, is not to modify the nominal in question, but to predicate some property of it. We will return to such clauses in Chapter 5.

3.2.7 Conclusion: The domains of word order

To summarize, the sentence, cross-linguistically, can be divided into three separate domains, a core domain, the main asserted clause, plus an initial and a final domain, both of which can have an indefinite number of units, each one occupying a full intonation unit. This division is crucial for the study of word order. The ordering of the domains with respect to one another is eminently iconic and seems to be universal. In addition, we also find ordering relationships exist among setting units, among tail units, and among the elements inside the asserted clause. It is this latter domain that I will concentrate on in the remainder of this chapter as well as in the following chapters.

The matter is somewhat more complicated for, as we will see, sometimes we find that elements which, under normal circumstances, we expect to find inside the main asserted clause, are found 'dislocated', either in a setting or a tail intonation unit.

Some studies of word order, among them some studies of Basque word order, fail to make the proper distinction between the three domains and treat all the elements of a sentence as if they were a string of elements all of which are ordered with respect to each other in a single ordering domain. I believe that this is a mistake, for the principles which account for the ordering of elements in one domain are not necessarily the same as those which apply in a different domain.

The relative ordering of setting elements, for example, has to do with scope relationships. The relative order of the settings and the main assertion has to do with the

function of the settings with respect to the assertion. The ordering of elements inside the assertion has to do with the function of the elements and some choices about the position of elements with certain functions. Even within the asserted clause there seems to be more than one domain in which different ordering principles rule. Finally, the order of assertion extensions with respect to the main assertion is also, like the order of setting elements, quite iconic and universal.

I will return to these issues later on in this chapter, as well as in the following chapters. Before discussing the structure of the assertion, however, I would like to discuss earlier approaches to pragmatics and word order, approaches which are based primarily on the pragmatic and cognitive statuses of different ideas expressed by the different constituents in the sentence, as opposed to the functions that they perform.

3.3 DISCOURSE PRAGMATICS AND WORD ORDER

3.3.1 Introduction

A number of (intrinsic and extrinsic) pragmatic properties of the ideas expressed by the constituents of sentences, and in particular asserted clauses, have been said to determine, to one extent or another, their relative ordering. I am referring to properties such as ‘givenness’, ‘accessibility’, ‘familiarity’, ‘identifiability’, ‘importance’, ‘predictability’, and so on. These are all concepts which are very important to the form of linguistic expressions, including their position in the sentence. Below I will argue, however, that their effect is to a great extent mediated by pragmatic functions ‘topic’ and ‘focus’.

The pragmatic properties which I will review in this section are highly variable cognitive properties of the ideas in the events expressed by asserted clauses, including ideas of referents (real or imagined), properties, and propositions. The pragmatic functions, on the other hand, would seem to be an attempt to discretize and automate those differences by the grammar of languages.

The pragmatic properties of ideas, such as *given* vs. *new*, are often associated with the Prague School of linguistics, even though the Prague School had well-known predecessors in this area. The Prague School, however, has put forth what seem to be all the major variations of the relevant pragmatic concepts and I will review them briefly in the next section.

3.3.2 The Prague School approach

It has often been argued that the pragmatic or cognitive properties of ideas expressed in asserted clauses, and not (only) their semantic or syntactic properties, determine or influence their form in a variety of ways, such as whether a pronoun or a full nominal is used, as well as their relative ordering. As I said, one well-known precursor of more recent attempts to explain facts of linguistic form, and in particular word order, according to pragmatic principles is the Prague School of linguistics and the Functional Sentence Perspective (FSP) approach (cf., e.g., Mathesius 1928/1964, 1929, 1983/1929; Vachek, ed., 1964; Firbas 1966, 1986, 1987, 1992). This approach itself had some less well-known precursors (cf., e.g., Bean 1983, Sgall 1987 for summaries of earlier related studies).⁸

In the Prague School writings one can find many of the relevant ideas that are still being used today, as well as the seeds of much of the vagueness and confusion that

continues to surround these matters. There are two major versions of the Prague School approach to pragmatics, one having to do with what we could call pragmatic *statuses* of the ‘referents’ and other ideas, that is, *new* vs. *given* “information” (i.e. ideas), and another one having to do with pragmatic *functions* or *roles*, that is, *theme* vs. *rheme*, or *topics* vs. *focus*. (In this study I use these latter terms for the pragmatic functions. I also use the term rheme to refer to the whole assertion minus the topic constituent, if there one, or any other clause or sentence initial ‘setting’ elements.)

The information status approach to constituent order is based on the notion that constituents expressing ‘new’ and ‘old’ (‘given’) ‘information’ (i.e. ideas) tend to be found in different places in the sentence. In the Prague School approach ‘given’ typically means that it has been used in the prior discourse (‘discourse-old’ vs. ‘discourse-new’). It has also been used in the sense that the addressee is ‘familiar’ with a particular entity (‘hearer-old’ vs. ‘hearer-new’, cf. below). Prague School grammarians have claimed, for instance, that in languages such as Czech ‘given information’ (i.e. elements of a sentence expressing given referents or ideas) come before ‘new information’ (‘old before new’). An ‘inverted’ version of this normal order (‘new before old’) is allowed under exceptional circumstances, such as for emphasis (‘inverted order’). This principle doesn’t seem to have a great amount of predictive power and it has even been claimed that for some languages the opposite order is actually found. In addition to information, we must contend with the difference between information. Other investigators have found interpreted the old vs. new dichotomy in terms of whether the idea is in the hearer’s mind at the time or not (see below).

The second version of the Prague School approach involves what we could call *information roles or functions*. In this view, the sentence has two privileged constituents, namely the theme and the rheme, or focus, which come at opposite ends of the sentence, with possible 'transitional' material between the two. Each of these three parts of the sentence has elements with greater 'communicative dynamism' than those in the preceding one. Firbas 1966 has argued that 'communicatively dynamic' material, i.e. more informative, typically 'new', material comes after less dynamic material.

The notion of theme is, of course, not unlike like the notion of topic or setting in other linguistic traditions and, in practice, it is often the same unit as the subject. The rheme, or the most 'communicatively dynamic' element of the theme's complement, which is also sometimes referred generically as the rheme (or 'comment'), is equivalent to what some investigators refer to as the focus of the assertion.

These two approaches for accounting for word order differences within a language, the one which is based on pragmatic statuses and the other which is based on pragmatic functions, are necessarily contradictory. The pragmatic status of the referents or other ideas seems to have very important formal consequences and to be at the heart of the choice of elements to fill the different pragmatic roles.

The major problem with the Prague School approach is that it mixes different domains of word order and thus runs into unsolvable contradictions about ordering preferences within any one language and is unable to explain the variation in constituent order found among the languages of the world. This 'monostratal' approach mixes three major different domains of world order: (a) the domain of the ordering of the theme/topic and other settings with respect to the assertion proper (the 'rheme' or 'comment'); (b)

the domain of the relative ordering of different setting elements (including the topic) with respect to each other; and (c) the domain of the ordering of elements inside the ‘rHEME’. These three different domains follow different principles, as we shall see.

In the following sections I will review the major pragmatic characteristics (statuses) of ideas, in particular those of ideas about entities (including abstract ones), which are expressed by nominals. All these properties of ideas have been said to have formal repercussions of one type or another. The properties which I will be discussing are listed in Table 3-3.

1	Contextual givenness
2	Cognitive givenness: activation, accessibility
3	Identifiability
4	Referentiality
5	Specificity
6	Accommodation (inferential accessibility)
7	(Topic) importance
8	Inherent topicality
9	Empathy

Table 3-3: Commonly discussed pragmatic statuses of ideas in discourse.

These notions are all connected to the overarching notion of ‘topicality’, ‘accessibility’, or ‘thematic salience’. These notions are important for instance in the choice of elements for filling the different pragmatic functions, and in particular the topic function. They are also important in determining the exact formal realization of these roles.

In addition, other pragmatic notions have also been put forth which go beyond the pragmatic status of ideas and which are more akin to the notion of ‘information’ and ‘communication’. These notions, which are listed in Table 3-4 are involved, together with the notion of topicality, in the choice of and element for the focus role.

1	Newsworthiness
2	Noteworthiness
3	Unpredictability
4	Contrast
5	Surprise
6	Emphasis

Table 3-4: Other pragmatic notions or statuses of ideas.

They are also important in the actual realization of this role in a sentence, as we will see. These pragmatic notions, together with the notion of topicality, are different aspects of what I will refer to as the ‘focality’ of an idea. Focality goes beyond the status of ideas in the minds of hearers (as assessed by the speaker) to include aspects of how the speaker wants to present these ideas in a sentence.

3.3.3 Givenness: Activation, accessibility, identifiability

As I mentioned above, there has been some confusion in the literature about what exactly is meant by saying that the referent or idea represented by a linguistic expression is ‘given’ or ‘old’, as opposed to ‘new’. Prince (1981b) noticed that these terms had been used in at least three ways by different investigators, namely:

1. **Cognitively:** the ‘old/given’ vs. ‘new’ dichotomy has been defined in terms of their referents’ activation in the speaker and/or hearer’s minds (e.g. Chafe 1974, 1976). Prince 1992b refers to these cognitive statuses, as pertain to the hearer, as *hearer-new* and *hearer-old* (since they are ‘in the hearer’s head’);
2. **Textually:** this dichotomy has also been defined in terms of the preceding discourse. Thus, given ideas are those which have been mentioned in the preceding discourse and new ones are those which have not. Prince refers to

these two concepts as *discourse-new* and *discourse-old* (since they are in the “in the discourse model”).⁹

3. **As pragmatic functions:** The terms ‘old’ and ‘new’ have also been used to describe the pragmatic functions of the ideas in a sentence or construction. That is, they have been used to refer to the notion of topic and certain aspects of focus-presupposition constructions.

I reject this third approach to ‘givenness’, which I will argue belongs to a different, though interconnected realm, which should be kept nonetheless separate. The textual approach to givenness doesn’t seem to be particularly useful, since anything which has been in the prior text is also presumably in the speaker and the hearer’s mind, and there is not compelling reason to separate what gets into the mind from the co-text from what gets there from other aspects of the context.

Thus I believe that the only valid approach to givenness is a cognitive one, one having to do with sharedness of ideas between speaker and hearer and related to both long term memory (how ‘familiar’ the hearer is with an idea and how ‘accessible’ it is in a particular context) and short-term memory (whether the hearer is thinking about that idea at the present moment or not).

Ideas are ‘given’ to different degrees, for different reasons, and in different ways. Givenness, which I will later refer to as ‘accessibility’, refers to the ease with which an idea can be incorporated by the hearer into the discourse at a particular point. In this section I will present an analysis of these concepts which borrows heavily from those of Chafe (e.g. 1974, 1994a) and Givón (e.g. 1983a, 1983b, 1988, 1989). For both Chafe and

Givón, what goes on in the speaker and hearer's minds is the relevant parameter of 'givenness', with numerous repercussions for the form of linguistic expressions.

Chafe (e.g. 1994a) distinguishes primarily between (a) 'ideas' that are taken (by the speaker) to be currently activated in the listener's consciousness (**active** or **given**), (b) ideas that are not active at the time of speaking (**inactive** or **new**), and (c) ideas that are in 'peripheral consciousness', either because they have been mentioned recently or for a variety of other reasons. The fundamental concept for Chafe is 'activation cost', i.e. how 'costly' is it—cognitively speaking—for the speaker to conjure, or activate, a certain idea. If the idea is already active, the cost is zero, if it is inactive, the cost is higher. In actuality there is a whole range of possibilities since accessibility of a concept depends on a variety of contextual factors.

Givón too has put forth a cognitively based approach to 'givenness', based on the concept of 'topicality' or **referential accessibility**. Referential accessibility refers to the ability of the hearer (as assessed by the speaker) to conjure up a particular referent-idea (a 'referent') from a mental file. When a referent was mentioned in the previous clause, the referent is said to be *continuous*, which is equivalent to Chafe's notion of *active*. How accessible a referent is depends primarily on two factors: how recently it was last mentioned (if it ever was), and whether there are other accessible referents in the discourse which could interfere with the one the speaker has in mind ('potential interference'). The notion of accessibility thus is a gradient one which depends on a variety of factors which conspire to make a referent or other idea more or less accessible in a particular context.

Total lack of givenness (inaccessibility) is found when the idea in question does not exist in the listener's long term memory. Such ideas, and in particular referents, are said to be **unidentifiable** ('brand-new' in Prince's 1981a terminology). Identifiability is the cognitive category which is often grammaticalized in languages as the linguistic category 'definiteness' (cf. Chafe 1976:38ff), with unidentifiable referents being coded as indefinites and identifiable ones as definites. Despite its cognitive basis, the actual use of the definite category varies somewhat from language to language depending on a variety of factors.¹⁰

Givón's concept of referential accessibility (or topicality) is similar to Chafe's concept of activation cost. Givón's notion emphasizes the gradience that exists from full activation all the way to unidentifiability, whereas Chafe's model emphasizes the existence of three linguistically relevant values (active, inactive, and semiactive). The addition by Chafe of the intermediate category ('semiactive') is problematic and will be discussed below.

In general I believe that the two approaches are not incompatible. I believe that the notion of accessibility is a gradient notion, but some points in that continuum are more crucial than others, from a linguistic perspective. Also it is important to realize that different factors are involved in the determination of an idea's accessibility, of which recency of mention, which is Givón's main 'measure' for accessibility, is but one factor. Accessibility, or ease of activation, doesn't depend solely on prior mention, but also on many other contextual factors, such as the physical and frame-semantic context, the 'world of discourse', as well as more inherent (less contextual) properties about the idea

in question itself, such as its salience or importance in the speaker's and the hearer's system of ideas, its degree of 'concreteness' ('pictureability'), its specificity, and so on.

3.3.4 Some complicating factors for the notion of givenness

The notion of givenness or accessibility is straightforward enough in most cases, but in some cases it doesn't seem to be enough to account for the facts. That is, in the simple cases it is enough to say that something is either in or out of long- or short-term memory. In some other cases, however the matter is more complicated.¹¹

For example, ideas that are hearer-given (identifiable) differ as to how 'entrenched' and how accessible they are. Some referents, such as *the moon*, can be said to be 'omnipresent' and easy to bring up to short-term from long-term memory, whereas others have may be 'buried' in long term memory and may need to be 'exhumed', as it were, before they can be treated as given.

Another variable involves how clearly defined the referent in question is in long-term memory? Some referents may be vaguely identifiable, but the speaker or hearer may have but a very vague notion about such entities. When the ideas are sets of individuals (e.g. *the children*), the members may also be more or less clearly defined as entities in the speaker's mind.¹² The notion of **specificity** has been used by some investigators to refer to these types of differences existing among identifiable ideas (cf., e.g., Myhill 1992).

Some ideas are **abstract** and they seem to be harder to conjure ('visualize') and retain in short-term memory than more concrete ones. Ideas referring to properties, as opposed to entities, are of this kind. Thus verbs and adjectives are less 'accessible' as ideas to be talked about in discourse. Even when such ideas are nominalized, they

remain abstract and are said to be **non-referential**. Ideas about more concrete entities can also be non-referential in certain contexts if they do not refer to actual entities in the 'real world', but in some 'possible world'. In many languages, such ideas are coded like non-identifiable ones (e.g. English *a book*), whereas other languages have a special way of coding them. A final example of abstract ideas involves **generic** concepts, that is, those which encompass all the members of the category, all individuals of a kind, e.g. *elephants*. Such concepts are typically quite accessible and topical and are usually coded in the same way as identifiable ideas in many languages.

Another important consideration is that not all unidentifiable concepts are cognitively equivalent. In a particular context, some unidentifiable ideas (or identifiable ones, for that matter) may be harder to 'assimilate' or incorporate into the discourse model (the world of discourse) than others. Some ideas seem to need a more conspicuous introduction than others which are easier for the hearer to assimilate. Other ideas, equally unidentifiable, because of the context, or because they are in some kind of relation with an identifiable idea, especially if it's a stereotypical semantic relation, are easier to 'assimilate' or **accommodate** into the discourse model (e.g. *my brother's girlfriend* vs. *my brother's pet elephant*).¹³ In other words, the amount of 'energy' or 'cost' associated with processing a new (unidentifiable) referent is not equal for all referents.

Sometimes ideas which are not, strictly speaking, identifiable, are treated by the speaker as if they were, because some aspect of the context allows the hearer to 'infer' their existence after the fact, as it were. In other words, these ideas are treated as 'indirectly accessible', or easy to accommodate into the world of discourse by the hearer. The source of this 'induced accessibility' may be any aspect of the context, such as the

physical context, the discourse context, or stereotypical associations from the world of discourse, or the 'semantic frame' (cf., e.g., Fillmore 1976, 1982, 1985a). Thus we find, for instance, that entities which are not identifiable to the hearer receive definite marking if they are 'modified' or 'described' by information which uniquely identifies a single entity and makes it easy to accommodate. Thus someone may refer, for example, to *Mary's boyfriend* or *the guy that Mary's dating*, even if the hearer did not know that Mary had a boyfriend or was dating anyone. The description itself makes the referent in question easy to accommodate and makes it easy for a hearer to 'create a file' for them, whether he or she was aware of their existence or not and allows the speaker to code the idea as definite.

Stereotypical entities of stereotypical situations, or 'frames', such as drivers of buses, menus in restaurants, the keys of a jailer, and so on, are also often 'accommodated' as being given once the context (the 'semantic frame') has been brought up, even if, strictly speaking those entities are 'brand-new'. Thus nominals expressing such ideas are often coded as definite and are treated as if they were accessible. Stereotypical elements of more general contexts, such as *the street*, *the weather*, *(the) school*, *(the) jail*, *the hat* (in contexts in which it is common for people to wear hats).

In Basque, or at least in some dialects of Basque, we find that the so-called definite declension is used in contexts in which a referent would not seem to be easy to accommodate and in which other languages would not use definite descriptions at all. This is probably due to the fact that the context, as well as the pragmatic role of the nominal, are enough to determine identifiability. Thus the contrast between the definite and indefinite declensions in Basque sometimes seems to code not identifiability, but

referentiality, with the discourse context and the pragmatic role (and thus the word order) being decisive in determining the referent's identifiability.¹⁴

These are some of the complications involved in ascertaining the accessibility of ideas in discourse. These considerations are important because they show that a simplistic coding of ideas as given or new is not sufficient for determining correlations between these labels, on one hand, and pragmatic roles and word order, on the other. A particular idea or referent may be 'new', but treated as 'given', or 'given' and be treated as 'new'. We will see below that there are additional factors involved in the accessibility of particular ideas in a particular context or in most contexts.

Investigators have dealt with these complexities in different ways. Givón's assumption that accessibility is a gradient notion, and not an all-or-none or categorical distinction, goes a long way towards helping us deal with the differences we find in actual discourse, although it does little to help us understand the different factors which determine an idea's accessibility in context. Chafe's 'semiactive' category is also intended to account for the middle ground between full accessibility (activation) and inaccessibility. This notion is problematic however for, as Lambrecht for instance has argued, what seems to be relevant is the 'potential for activation' or the 'ease of activation' of an idea, and not whether the idea is in a particular activation *state*.¹⁵

It may very well be that there is such a thing as a 'semiactive' cognitive state in consciousness for ideas which have recently been active and which are still in the 'periphery' of short term memory. On the other hand it doesn't seem that this 'state' can be used to describe the situation of relative enhanced accessibility of ideas which come easily to mind in a particular context, such as those which belong to an active, or evoked,

semantic frame and which require but very little effort on the part of the hearer to accommodate into the discourse model.

3.3.5 The domain of givenness

Givenness or accessibility is primarily a property of the ideas of *referents* or *entities* in people's minds. Givenness doesn't seem to be very relevant to certain abstract ideas, such as ideas about *properties*. But givenness and the issues just described in the previous section are also relevant for ideas, namely ideas of states and events ('states-of-affairs' or 'propositions'), another relatively highly 'pictureable' type of idea.

The relevance of accessibility to more abstract concepts, such as those describing properties or abstract states of affairs, ideas which are typically lexicalized as verbs or adjectives, seems to be much more limited, since these ideas are not 'permanent' enough to have continuity in discourse. The ideas expressed by verbs and adjectives are much less specific and less subject to be 'thought about', stored, and referred to in discourse, although this is indeed a possibility, just as the 'reification' ('nominalization') of such ideas is also a possibility in probably all languages (cf. Chafe 1994a).

But, as I said, givenness seems to be a relevant concept to describe the mental representation of *concrete* states and events ('propositions'). In fact, the phenomenon which is usually referred to as 'pragmatic presupposition' seems to boil down to the accessibility of a proposition, and not to any facts about its truth or falsity. Some constructions, such as so-called 'focus-presupposition' constructions and content question constructions, require that a certain proposition be given, *or treated as given*, i.e. accommodated, much like in the case of ideas about referents. Also, the phenomenon of ellipsis is intimately tied to the accessibility of propositions in context.¹⁶

3.3.6 Thematic importance and inherent topicality

The ease with which an unidentifiable referent is assimilated by the hearer and the degree to which it is introduced prominently into the discourse by the speaker, and is accessible to the hearer throughout the discourse, also seem to be correlated with the **importance** or centrality that this idea has in the discourse. Unimportant referents, minor players, or ‘props’, may be introduced ‘casually’ anywhere in the sentence once, even unaccented, even if they are unidentifiable. Such referents tend to be very unspecific or vague and speakers can easily ‘accommodate’ them into the discourse model. More important referents, which tend to persist over longer stretches of discourse once they’ve been introduced, tend to be coded in a more salient manner when first introduced, always accented, thus indicating to the hearer that this idea is worth opening an ‘file’ for.

In (3.7a) we see an example of an unimportant new referent, which is not likely to attract a lot of attention from the hearer, and which need not be accented.

- (3.7) a. Then he ran into a rock, and fell to the ground.
b. Then he ran into a woman, and asked her for directions.

The new referent in (3.7b), on the other hand, would seem to be more likely to be accented and to call for a prominent place in the hearer’s consciousness.

Both Chafe and Givón recognize this third relevant cognitive property of referents in context (besides activation and identifiability), which is said to affect how ideas are verbalized. They recognize that not all new referents are equally important in a story, whether they are hearer-new or hearer-given. Some referents are mere “props” which are used to build the context, for instance, whereas others are major ‘discourse topics’ (or ‘protagonists’), and still others stand somewhere in between. Givón (1983a) has

operationalized the concept of 'thematic importance' by the measure of 'topic persistence', which refers to the number of clauses after a particular (re)mention in which the referent continues to be mentioned. Givón's thematic importance is roughly equivalent to Chafe's 'referential importance' (1994a) and what Cumming (1995), for example, calls 'protagonism'.¹⁷

Wright and Givón (1987) found that speakers and hearers are rather consistent in their ranking of referents by degree of importance (cf. Chafe 1994a:89).¹⁸ It is not clear, however, whether one should postulate discrete degrees of importance or whether only the two poles (protagonist vs. prop), or even merely relative importance is relevant. Hopper (1986b), for example, has argued for a discrete, three-way distinction between primary, secondary, trivial importance from structural considerations.

Thematic importance is presumably correlated with other pragmatic and semantic properties of referents. For instance we know that human and other animate referents tend to be important topics of conversations ('discourse topics') more often than inanimate ones. But this is not an absolute and it is to some extent dependent on the context. Thus in a lecture about the solar system, animate topics are not likely to be very common or salient, whereas planets and other inanimate elements probably will be. In other words, importance must always be determined with respect to the context at hand.

However, there is also no doubt that stereotypical, intrinsic or inherent 'topicality', or 'accessibility', across most contexts, can have important grammatical repercussions. Topicality hierarchies have been known to be relevant to the grammars, a grammaticalized reflection of topicality related tendencies and preferences (cf., e.g., Silverstein 1976). As is well-known, languages have strong correlations between

inherent topicality and grammatical and pragmatic roles. Semantic roles can also be placed in a hierarchy from most inherently topical (agents) to least inherently topical (circumstantials) (cf., e.g., Gruber 1965, 1976; Fillmore 1968/1967, 1969/1966). Referents are also more likely to be important, and thus more 'topical', or *thematically salient*, to the extent that they are close to the speaker, i.e. to the degree that the speaker 'empathizes' with them (cf. Kuno 1976, 1987). An empathy hierarchy thus can be established in which the speaker and hearer probably rate highest. Humans and animates also rate high, as compared to non-humans and inanimates, respectively. Such hierarchies of inherent topicality have been found to correlate with formal linguistic phenomena, including perhaps word order (cf., e.g., Tomlin 1986).

Thus we can say that, in addition to the context and the discourse model, this kind of 'intrinsic topicality' is one of the factors of what I call thematic salience. Some referents are easier to conjure up, or accommodate into the discourse model than others. In the abstract we can say that human topics are more inherently topical, across contexts, as are the discourse participants, the speaker and hearer. A particular concrete context, however, may provide its own set of important, and thus likely, discourse topics.

3.3.7 Thematic salience and the content of linguistic expressions

As I already mentioned, all these cognitive 'status' differences have been said to have formal consequences for the way ideas are verbalized. The formal consequences can be of two types, one related to the **form/size** of the expression coding the 'idea', and the other to the **position** or **order** of the expression coding that idea with respect to the surrounding units in a clause.

It has often been noted that **activation** or lack of activation of referents determines to some extent how they are verbalized (cf., e.g., Chafe 1976, Gundel 1977/1974, Prince 1981b). **Active** referents tend to be coded in “a more attenuated way” than non-active, accessible, or new information: with a “weakly accented pronoun”, with ‘agreement’ markers on the verb, or with no overt expression (zero), such as when the referent’s function can easily be recovered from context (as in Japanese and other languages for instance). **Non-active** accessible referents and new referents, on the other hand, tend to be coded with “accented full noun phrases” (Chafe 1994a:75). The less accessible and hardest to conjure up an idea is, the more ‘complex’ its verbalization will be. Active referents may also be coded in non-attenuated ways for a variety of reasons, such as to avoid possible ambiguity in context or when they are contrasted with other referents (see below).

The referentiality, or rather **non-referentiality**, or perhaps even non-specificity, of a nominal’s idea, may have other formal repercussions. In some languages, for example, non-referential nominals, such as non-referential objects and subjects, are known to gravitate towards the verb, and in some languages they are even incorporated into the verb. Non-referentiality is also said to have repercussions on the ordering preferences of a nominal. Thus in Hungarian, for example, non-referential objects must be placed in preverbal (focus) position, whereas referential ones would under similar circumstances have to appear in postverbal position (cf., e.g., Sasse 1995b:172), definite ones obligatorily so (unless they are ‘foci’), and indefinite ones preferentially (unless there is another definite nominal in the sentence) (cf. Sasse 1995b:172).¹⁹ (We will return to this issue below.)

Identifiability seems to be an even more important determiner of form than activation (cost), with unidentifiable referents receiving the most different marking (cf., e.g., Chafe 1994a:76). In languages with a definite marker, this marker is used with identifiable nouns.²⁰ Languages which do not have a morphological or lexical way of marking identifiability (i.e. ‘definiteness’) sometimes have ways of expressing identifiability status by means of word order or at least may have some restrictions on the position of the noun-phrase based on identifiability. (As we will see below such restrictions seem to be mediated by pragmatic functions.)

Important, non-identifiable referents sometimes have a special way of being identified, such as in the demonstrative pronoun *this* in spoken English (cf., e.g., Prince 1981c; Wright and Givón 1987; Givón 1990) as well as by using special ‘presentative’ constructions (see below).

It has also been noticed that identifiable referents may receive different types of coding (such as pronouns, demonstratives, proper names, or a modified common noun), depending on how ‘familiar’ to the listener they are perceived to be, i.e. their degree of accessibility or givenness (cf., e.g., Prince 1981b, Chafe 1994a). There seems to be a clear negative correlation between a referent’s accessibility and the size and complexity of the nominal needed to code it, although a speaker can always choose to be more specific than the language requires for a variety of reasons (cf., e.g., Gundel, Hedberg and Zacharski 1990, 1993).

3.3.8 Thematic salience and the order of linguistic expressions

The relative order of nominals has been argued to depend on pragmatic characteristics of those nominals. One type of ordering influence is related to the size

and complexity of the nominal in question, which, as we just saw, is in turn correlated with the accessibility of the nominal's referent. It has been argued that for reasons very likely having to do with processing constraints or ease of processing, complex nominals tend to prefer either clause (or intonation unit) initial or final position. The former ('fronting') is often found in strictly verb final languages, whereas the latter ('backing', 'postposing', or 'extraposing') is a tendency found perhaps in all other languages. It is not always clear to what extent such processing-motivated, complex-phrase 'fronting' or 'backing' can be kept separate from other cognitive and pragmatic factors influencing word order, but at least sometimes this may be seen as an independent factor.

More interestingly, we saw that the FSP school has found correlations between given nominals and clause-initial position and new nominals and clause-final position. The old-before-new principle is presumably most relevant to 'free word order' languages such as Czech, though it is also said to apply to languages with more rigid word order, such as English, at least in some contexts. Thus this principle has been used for example to account for the use of constructions such as the dative shift (Erteschik-Shir 1979) and the passive construction (Ross 1975) in English.

The old-before-new principle has been strongly criticized by Givón, for instance, who has serious reservations about its relevance. Givón proposes that in languages with truly free word order, such as Ute and other Native American languages (cf., e.g., Creider 1977, Tomlin and Rhodes 1979), actually the opposite of what this principle predicts typically obtains, that is, the preferred order is **new-before-old**. According to Givón's word order principles, 'new' material tends to come early (first) in the clause in actual (contextualized) speech. NPs verbalizing highly topical/active referents, when mentioned

at all, i.e. when not elided, are often ‘postposed’ (cf., e.g., Givón 1989:224-5). The new-before-old principle has also been proposed in the Basque grammatical tradition as a principle of word order, at least in some contexts, for instance by Villasante (1980).²¹

One problem with this principle, as well as with Givón’s critique, is that they mix different word order domains, such as the sentential domain and the (asserted) clausal domain. As we will see below, they also mix pragmatic properties of the referents of nominals, such as accessibility, and their pragmatic functions in the assertion, such as topic and focus. Finally, another major reason for the discrepancies is that the Prague School approach uses almost exclusively sentences out of context for its data, where overt subjects, and thus topics, which tend to be given, predominate, whereas Givón uses narrative data in which topics are predominantly elided. (Furthermore, as we will see later on, Givón looks at languages with preverbal focus position, whereas Prague School investigators have looked primarily at languages with postverbal focus position.)

Givón has argued that languages use different coding devices for marking points in the topicality continuum (e.g.. Givón 1990:Chapter 20, 1992). As we have seen, two important correlators of topicality are size (all the way from complex noun phrase to pronoun or verbal coreferencing), and accentuation (from full accent to lack of any accent) of the verbalized nominal. In addition, he argues, there is, cross-linguistically, a more or less noticeable tendency to place nominals in different places in the sentence depending on the referent’s accessibility (or, rather, with his measure of accessibility, namely ‘referential distance’).

Givón argues that ‘**continuing topics**’, i.e. when textual referential distance is minimal (and there is no possible interference), are coded by “minimal gap devices”, i.e.

by ‘zero anaphora’ or by an ‘unstressed pronoun’, and tend to appear **late in the clause**.²² (This contradicts the Prague School’s predictions, but Givón argues it is the situation found in truly free-word order languages such as Ute.) **Non-continuing but recently mentioned** (within 2-3 clauses) topics are ‘brought back’ by means of “small gap devices,” such as a stressed pronoun or, for objects, ‘**fronting**’ (or ‘Y-movement’, as Givón calls it, by which he means non-dislocating topicalization). More **inactive**, or even **non-identifiable**, referents are coded by “long gap devices,” such as ‘**left dislocation**’ or by means of a ‘repeated definite NP’.²³

In addition, for Givón ‘continuity’ is not the only factor involved in the positioning or accenting of nominal referents. For him, continuity (accessibility) is only one type of high predictability, just like non-continuity is a type of low predictability. There are other types of information which display low-predictability, such as contrastive and ‘important’ information (cf. Givón 1990:753-4). For Givón, low predictability is always associated with early position (‘first things first’) and high predictability (accessibility) with late position, cf. Table 3-5.

Low- predictability	←————→	High- predictability
Non-continuous		Continuous
Contrastive		Non-contrastive
More important		Less important

Table 3-5: Ordering relations between low predictability items, such as non-continuous referents, and high-predictability items, such as continuous referents, according to Givón.

Givón, like the Prague School investigators before him, tries to devise a single parameter which can account for the ordering preferences of clausal constituents. Givón uses the low predictability vs. high predictability continuum, where the Prague School

used the new vs. old parameter, and the communicative dynamism parameter. I believe that the single parameter approach used to account for ordering preferences is mistaken and that we must look at the discourse *functions* that the different ideas play in assertions.

Despite all the interesting correlations and generalizations which the single parameter approach has availed us, I believe that without a pragmatic function based approach we will not be able to account for the exceptions to the correlation, the apparent differences we find among languages, and the apparent differences among constructions within any one language, such as the differences between continuous narrative clauses and out-of-the-blue, or decontextualized, statements, or that between emphatic and non-emphatic statements.

In other words, I believe that a great part of the problems and contradictions that previous approaches have come up against are due to the failure to bring into the picture the 'informational', and thus pragmatic, roles or functions (**topic** and **focus**) that the ideas expressed by nominals play inside the clause. The correlations found between cognitive status (accessibility, predictability) and word order are actually correlations that are mediated by those roles. That is, (1) pragmatic functions correlate with different formal expressions (such as positions within the asserted clause); and (2) pragmatic functions correlate with the cognitive statuses of the ideas involved. In addition, however, pragmatic roles never have a single possible formal realization, and pragmatic status does play a role in determining the actual form (including order) of the linguistic units verbalizing ideas with different pragmatic roles. Pragmatic status, however, can never, in an of itself, determine the form and order of expressions, which is why the correlations between status and form are never perfect.

But pragmatic functions are not needed just to account for the imperfect correlations between pragmatic status and linguistic form. There is much evidence which suggests that pragmatic (information) roles are necessary to understand the structure of assertions, as well as the connection of assertions to other assertions and to the world of discourse.

In the remainder of this chapter, starting in Section 3.4, I will motivate the pragmatic functions of **topic** and **focus**. In that section I will argue that **topic** is a relevant linguistic category, one that is closely related to the category subject in languages which display such a formal category, even though a minority of subjects are not topics and some topics are not subjects. A prototypical topic is both the predication base for the assertion (that is, most assertions are predications in natural language), and a link to the world of discourse, grounding the assertion to it. The topic is nothing but a specialized type of setting. This accounts for the fact that overtly expressed topics are typically clause-initial in most languages, regardless of their pragmatic status. This also accounts for the fact that topics (and thus subjects) are typically quite accessible or easy to accommodate.

In Section 3.5, I will discuss the other major pragmatic role, namely that of **focus** of the assertion. In some assertions it is quite clear and obvious that there is one element which stands out, or is the most communicatively salient, or noteworthy, part of the assertion (such as in answers to content questions). In other assertions it is less clear that one element stands out, but phonological (accentual) properties, as well as ordering characteristics, always seem to code one element as special. This element is the most informative, newsworthy, or surprising element in the assertion, and hence the correlation

between focus and newness and unpredictability. The existence of a single privileged idea in the comment is, of course, fully compatible with the notion of 'preferred argument structure' discussed earlier in the chapter. As to the preferred position of focus nominals, languages differ. There is a strong tendency for it to come next to the verb: either before (in so-called OV languages) or after (in so-called VO languages), although assertion final *and* 'extraposed' position is also at least a possibility in perhaps most languages.

3.3.9 The coding of nominals

Before turning to the study of pragmatic roles, I would like to mention the attempt that I made early in this study to code nominals according to topicality associated factors. For the reasons I mentioned above, determining the exact pragmatic status of the idea of a nominal expression in the text turned out to be a rather difficult task. I soon learned that it was relatively easy to determine when an idea is active or when it is discourse-new or hearer-new, but not so easy to determine degrees of *accessibility* and topicality in context. Furthermore, I soon found that determining accessibility was not as important as determining the pragmatic role of the ideas in question.

For the purpose of this investigation at first I utilized a number of categories, which are summarized in Table 3-6 below (they were coded in the Pragmatics field for each nominal). These categories were applied only to ideas expressed by nominals. Predicates and modifiers were not coded with respect to topicality, but, rather, received different pragmatic (functional) coding (such as *Manner, Purpose, Reason*). The pragmatics field associated with each phrase in the database was also occasionally used to record other information, such as contrast and other salient focus, cf. below.²⁴

	Category	Characteristics
1.	Active	the referent can be presumed to already be in the mind of the speaker and hearer (talking about it right now, 'continuous')
2.	Accessible	the referent cannot be presumed to be active, it is accessible from the context or can presumably be easily accommodated given the context
3.	Inactive	the referent is presumably identifiable, but cannot be presumed to be accessible to the hearer
4.	New Main Hum	unidentifiable, important referent (all are human)
5.	New_Prop1_Hum/Anim/Inan	unidentifiable, unimportant referent that is only mentioned once, 'props' (most are inanimate)
6.	New_Prop+ _Hum/Inan	unidentifiable, relatively unimportant referent that is mentioned more than once 'props' (most are inanimate)
7.	New Nonspecific Hum/Inan	
8.	Non_Specific	Idea with a non-specific referent or no referent: it is either generic (e.g. <i>the police, the street</i>), abstract, or non-referential (e.g. <i>(have) hunger, (in) movies</i>), or other non-specific (e.g. <i>everything, something, (eating) apples, (have) drugs</i>)
9.	Proposition New	new idea that is not a referent, but a proposition

Table 3-6: Categories used to code for ideas expressed by NPs in the spoken Basque corpus (coding in the pragmatics field for each sentence position). (Fields: 'Clause initial pragmatics', 'Preverbal pragmatics', 'Postverbal pragmatics', 'Postverbal 2 pragmatics'.)

The referents that most detailed information was collected for were new ones. They were divided into major ones (*New_Main*), one-time props, and props which lingered a bit longer in the discourse. It was also noted whether they were human, animate, or inanimate referents. Accessible referents are those that have been mentioned in the text or have been clearly evoked by the context. Generic and non-referential nominals were coded as *Non_Specific*.

The purpose of this detailed coding of the pragmatic properties of the referents of nominals was to ascertain any possible correlations between pragmatic properties and the ordering of those nominals. As we will see later on, some correlations turn out to be important, although I will argue that they are not direct correlations, but rather are mediated by the pragmatic functions topic and focus.

3.4 TOPICS: THEMATIC GROUNDS AND PREDICATION BASE

3.4.1 Topics and topicality

Some linguists have argued that the topicality-related pragmatic characteristics of referents discussed in the previous section are the major or only area of discourse-pragmatics which determine, to one extent or another, and quite directly, grammatical phenomena such as word order. Others however, in a more or less explicit manner, have argued for the need to add a second, mediating layer that looks at the pragmatic or informational structure of asserted clauses and the roles which the different ideas play in it. From this perspective it has been argued that the vast majority of sentences in discourse have a **topic-comment** structure, or configuration, and that 'comments' themselves have a single, most important element, namely the **focus** of the assertion. (I use the term 'rheme' for the part of an asserted clause which excludes the topic and other settings in clauses which have them, and for the whole clause in clauses which don't; see below.)

The idea that assertions typically have a topic-comment configuration is very old and is basically the idea behind the classical division of clauses into **subject** and **predicate**, a view going back at least to the Greek philosophers. This view has been enshrined in modern grammar as a basic *syntactic* division for all clauses (NP-VP). Other traditions have recognized a pragmatic division, such as in the **topic-comment** division of Hockett (1958), Lyons (1968), and others, or the **theme-transition-rheme** of the Prague School and Halliday (1967a, 1967b, 1968). An revised version of the traditional view recognizes that not all sentences display such a bipartite structure, an

idea revived in recent years by Kuroda (1972) and Sasse (1987), for example, and which goes back to the 19th century philosophers Brentano and Marty.

I believe that the approach to discourse pragmatics which is centrally based on the pragmatic roles 'topic' and 'focus' is the correct one. Topicality and accessibility are very important notions, but in the end we must recognize the existence of information roles which mediate between pragmatic statuses and form. The variable notion of topicality, for example, cannot substitute for a category topic, even though it is the case that topics tend to be high in topicality. But topicality alone cannot explain the phenomenon of topics.

Approaches such as those of Chafe and Givón do often suggest the importance of a topic-like role in discourse pragmatics, although they tend to downplay it. In fact, Chafe's notion of 'guidepost' is obviously related to the notion of 'topic' which I will be developing here. Firbas (1964) and Halliday (1985) have used the notion of "point of departure" or "theme", for example, which also boils down to the same principle.

Problems with earlier definitions for the terms topic and focus have led many investigators to coin new terms or to give up in despair. Grimes (1982) for instance finds "the terms 'topic', 'theme', and 'focus' to be no longer usable, not because they are not getting at something valid, but because so many linguists have used them in so many different ways for different reasons" (Grimes 1982:383). Schlobinski and Schütze-Coburn (1992) are also quite pessimistic about finding the "correct" interpretation for these terms.²⁵

The causes for the confusion, however, are not all that mysterious, once we begin to separate the universal from the particular and the prototypical from the accidental, and

the functional basis of the phenomena from the grammaticalized versions of these categories in individual languages or classes of languages. It is also important to realize that languages must make some arbitrary choices in some contexts, and that some of the variation we find is context dependent. In non-prototypical situations languages may make different choices as to whether to have a topic or not, or as to where to place the phrase which codes the topic. We should also always keep in mind that that topics can be covertly expressed and that some clause types may not have topics or they may not have prototypical ones.

3.4.2 Topic-comment and all-comment information structure

As I said, the idea behind a bipartite structure for sentences into subject (topic) and predicate (comment) goes back to the ancient Greeks. This structure was taken for granted by grammarians and philosophers for two thousand years until Frege dispensed with it in the field of logic in the mid-19th century. Later on, Brentano and his student Marty argued that there was a type of ‘judgment’ (i.e. declarative assertion or statement) in language which was indeed bipartite since was “constituted of two successive acts: naming an entity and making a statement about it” (Sasse 1987:512). They called this a ‘categorical judgment’. At the same time they argued that some statements (‘judgments’) in language are not double or categorical, but rather simple or ‘thetic’. Weather expressions for instance are (typically) examples of thetic statements, a subtype which Sasse calls ‘event-central’, as are many existential-presentative assertions, which for Sasse are ‘entity-central’ thetic statements. In other words, in natural language the vast majority of statements and other assertions (though not all) are not only assertions of a proposition, but are assertions *about* a referent or idea, namely the topic.

It isn't only somewhat peculiar sentences such as weather expressions and existentials which havethetic structures, however.²⁶ The most interesting evidence for the distinction betweenthetic statements and categorical (grounded) ones, and, more generally, for the existence of a topic-comment configuration in most statements in natural language, comes from sentences which can appear in both configurations (as we will see, under normal circumstances only some types of clauses have the possibility of having both structures). When this happens, the two assertion types differ at least intonationally, as Bolinger (1972b) and Chafe (1974) have noted, and very often in word order as well.

Let us compare two pairs of sentences, first the ones in (3.8a&b), from Chafe 1974 (examples 5b and 6b, respectively), and, second, the ones in (3.9a&b), from Bolinger 1972b (p. 642).

- (3.8) a. My SISTER's DYing (categorical)
 b. My SISTER died (thetic)
- (3.9) a. Why are you coming indoors?
 —I'm coming indoors because **the sún is shining** (thetic)
 (= because of the sún)
- b. What is the state of the sun?
 —**The sun is disappéaring.** (categorical)

In categorical (topic-comment) sentences, the initial topic/subject may receive a more or less noticeable accent, depending on its activation/accessibility, but the crucial thing is that the predicate must receive an accent. If the topic is very accessible, as in (3.9b), it typically isn't fully accented (and thus Bolinger doesn't mark it as such). If it is active, in English it will typically be coded by an unaccented pronoun, and in other languages it will be elided. If it is relatively inaccessible and unexpected, as in (3.8b), it may be fully

accented and if it is very 'inactive', it may be even 'dislocated', in its very own intonation unit.

In other languages such pairs may have other formal differences. In many VO languages with flexible subject position, for instance, athetic sentence equivalent to (3.8b) would have subject inversion (*Died my sister*) and in Japanese the subject of a categorical sentence would have the *wa* suffix, whereas the subject of athetic sentence would have the *ga* suffix.

3.4.3 Other explanations for thetic structure

Some linguists have tried to explain the differences between thetic (topicless) and categorical (topicful) assertions by recourse to semantic factors of the predicate or the subject or topicality related pragmatic factors of the subject (the subject is central here because it is the default topic-candidate). With respect to **semantic factors**, it is quite clear that there are correlations between the likelihood (or even the obligatoriness or impossibility) of the use of the thetic structure (i.e. of the subject not being the topic) and the semantics of the predicate. Thus, although almost any predicate seems to be able to appear in a topic-comment structure, only some may be used with athetic structure, typically so-called 'unaccusative', or 'stative' predicates.

The likelihood of a clause displaying thetic properties, such as 'subject inversion' in languages in which the subject inverts in such clauses, has also been correlated with the semantics, or the 'thematic role', of the subject. Likely candidates are said to be, for instance, those with a 'theme' or 'patient' role (i.e. subjects of unaccusative/stative predicates). It is obvious that there is a close connection between the semantic properties and thetic structure, but that doesn't mean that the semantics determine the formal

differences. Rather, they simply influence the choice of information structure. That is, clauses with this type of predicates and subjects only display a *preference* for thetic structure, at least in most contexts, though they may also have topics, as we saw in (3.8) and (3.9) above.²⁷

In Romance linguistics, for example, it is well-known that there is an association between subject inversion and certain types of predicates (e.g. appearance and disappearance, and in general so-called unaccusative predicates with patient-like or ‘undergoer’ subjects) and types of events (surprising or unexpected events) (cf. Sasse 1987:566-7). Also, it has been noticed for instance that in many such (inverted) sentences the information is often rather trivial or stereotyped. Thus, for instance, inversion obtains in a sentence about a dog barking, but not in one about a person barking (cf. Sasse 1987:569).

As for the **discourse pragmatic factors** which correlate with thetic constructions, it has been noticed that typically the subjects in such constructions are either non-referential, in which case the construction is obligatorily thetic, or referential but not (discourse) ‘given’, also obligatorily thetic, or even contextually accessible but not active, in which case they are optionally thetic (cf. Sasse 1987:570).²⁸

Despite the semantic and pragmatic correlations with thetic assertions, I believe, along with Sasse and others, that these correlations are indirect and that the true reason for the formal differences between categorical and thetic constructions has to do with information structure, namely that whereas the former have topics, the latter do not, that whereas in the former the subject is the topic, in the latter it is not. The semantic and pragmatic differences account for the likelihood of such predicates and such nominals

being found inthetic constructions (topicless assertions), but do not themselves determine or explain the formal differences between such constructions.

3.4.4 What are topics?

The functional category topic is somewhat complex, but it is not as hopelessly intractable as it is sometimes claimed to be.²⁹ Its study is complicated by formal differences in different languages stemming from the different ways in which this functional notion is grammaticalized. The existing strong correlations between topics and topicality on the one hand, and between topics and the major argument of a verb (the main ‘topic candidate’), have also often further muddied our understanding by trying to reduce topics to either topicality or to subjecthood. Also, we must keep in mind that topics may have different characteristics depending on a variety of factors, such as their accessibility.

The major obstacle to a correct definition of this functional notion, however, seems to have been a one-sided interpretation of its nature. Several major basic definitions of topic have been proposed, and in fact they are probably all relevant to a proper characterization of this function. There are two major definitions, however. First of all, in asserted clauses a topic is certainly the **predication base** of the assertion, what the assertion, i.e. the predicate, is ‘about’ (cf. Reinhart 1981, 1982; Li and Thompson 1981). In this sense topics are properties of asserted clauses.

But a topic is more than that. A topic in discourse is typically the element which **grounds** the proposition to the discourse, a privileged type of **setting** element, which links the proposition to the universe of discourse, very often being the major element of continuity between an assertion and the prior assertions in the discourse. It is an element

which is both inside and outside the assertion. In this sense the notion of topic can be extended to non-asserted clauses as well as be analogized with other types of privileged settings which are not thematic, but circumstantial in nature, as we shall see. In other words, an analysis of the notion topic in terms of grounding allows us to:

1. understand the double role performed by the topic in asserted clauses as ‘thematic ground’, and as ‘predication base’;
2. explain the pragmatic requirements on topics;
3. relate this kind of thematic grounding to other, non-referential types of grounding, such as circumstantial grounding, as well as to clause external grounding (in the sentence-head position); and
4. extend the explanation to non-predicational types of sentences, such as background sentences and reduced complement sentences.

To repeat, in the vast majority of asserted propositions in discourse, the referent of one of the participants of the event, one of the major arguments of a predicate, typically a default one, as determined by its semantic role, performs (primarily) two simultaneous roles:

- **Discourse link or ground.** Depending on the accessibility of the referent, it grounds the assertion in one of two different ways. If it is active or highly accessible it connects or *grounds* the assertion to the present universe of discourse (a ‘continuous topic’). If it is not as accessible, it selects a new ground for the discourse from the set of available referents in the larger universe of discourse (a ‘new topic’). The external or ‘discourse role’ has been emphasized by different investigators, who have

concentrated on different aspects of this role: such as ‘staging’ (Grimes 1975), ‘empathy perspective’ (Kuno 1975, 1976, 1987), ‘scene perspective’ (Fillmore 1977), ‘viewpoint’ (DeLancey 1981), ‘vantage point’ (Dik 1989), ‘starting point’ (Chafe 1991), ‘point of view’ (Chafe 1976, 1994a), etc.

- **Predication base or ground.** The topic is also *what the sentence is about*, the base of the assertion, which is thus a predication (cf., e.g., Kuno 1972, Dik 1978, Reinhart 1981, 1982). This is the function emphasized in the traditional grammatical literature for the notion (psychological) subject.

This double articulation of grounds, schematized in (3.10) below, seems to be a crucial fact of discourse structure.

(3.10) Context (World of discourse, discourse model) ← [Ground] → Assertion

On the one hand the topic looks outwards, towards the context, that is, the world of discourse, the discourse model. But it also looks inwards, towards the state of affairs expressed by the clause.

The discourse link function is related to the notion of ‘interpropositional coherence’. As many functionalists (notably Givón), have repeatedly emphasized, discourse, which is characteristically multi-propositional, is eminently coherent and this coherence is obtained by means of elements of continuity among subsequent propositions. Furthermore, the major type of continuity in language is undoubtedly ‘referential continuity’, or what I call thematic grounding, and Givón calls *thematic coherence* (cf., e.g., Givón 1990).³⁰

Thematic grounding is a more general concept than thematic continuity ('topic continuity'). Continuous clauses, which are indeed the majority of main clauses in narrative, as we will see in the next chapter, are grounded by the ongoing, continuous referent.³¹ Non-continuous clauses, unless they are groundless (topicless, orthetic), are also overwhelmingly found to be grounded on an accessible referent, typically the argument of the verb which is most inherently topical, but perhaps some other. This is true even in minimal context situations. Even in mono-propositional discourse, such as linguists' sentences with minimal context, rare as they are in natural language, there is a tendency to connect the proposition to the world of discourse, by grounding the proposition on a referent that is part of the universe of discourse (given), or can be accommodated as such.

To repeat, we are not dealing here with 'old' and 'new' parts of the proposition. It is not 'old parts' of assertions which provide the thematic coherence to an assertion, but rather it is the topic role that does. The topic is in a way a part of the asserted proposition, but at the same time it is outside the assertion. That is, it is not under the 'illocutionary operator' (cf. Lambrecht 1994), or is outside the 'assertion scope' (cf. Givón 1990).³²

The notion topic clearly correlates with the notion of topicality, since the more topical the topic the better it serves its linking function, but it is independent from it. The notion topic is also closely related to the notions of 'perspective', 'point of view', 'empathy', 'starting point', and the like in the functionalist literature. It is, of course, also directly related and based on to the notion of (non-grammaticalized) topic discussed in Li and Thompson 1976a for topic-prominent languages.

3.4.5 Topics and subjects

In theory any referential element (event participant) of an asserted proposition could be the thematic ground of a clause. In practice, it seems that only core arguments are topics, and overwhelmingly a particular argument of a predicate is typically the topic, i.e. is the default topic: the so-called 'subject argument'. Even languages which do not have a formal subject category, such as Basque, display such a topic default phenomenon. The choice of default topic depends on the semantic role of the arguments, with *doers* and *experiencers* having priority over *patients* and *themes* (cf., e.g., Fillmore 1968).

Other arguments of the verb may also be the topic of the predicate, however. In some languages voice operations on the predicate are needed for a major argument other than the default topic to become the topic. In other languages, simply moving a non-subject argument phrase to clause-initial position, with the proper intonation, is enough to code it as the topic.³³ Recipients and benefactive arguments (coded in Basque by the dative case and dative agreement) are also very often used as topics.

The strong correlation between the pragmatic category topic and the formal category subject, formally coded by a certain case marker or a position in the clause, for instance, in those languages where one exists, is certainly not accidental. As many have argued before, and it certainly seems like the most reasonable explanation, I believe that the category subject is nothing but the grammaticalization of the category assertion topic on the default argument of the verb.

On the other hand, we know that not all subjects are topics. That is, the category subject, its formal coding, is not dispensed with when the subject's referent is not the topic, as happens inthetic assertions, for instance, or in unasserted sentences where the

notion topic is less relevant (see below). It does seem to be the case, however, that subjects which are not topics are not coded exactly like those which are, and that there are always some differences, which may be accentual (as in English), positional (as in Spanish), or morphological (as in Japanese).

3.4.6 Covert vs. overt topics

Statistically speaking, in narrative the typical topic is active or highly accessible in context, and thus it is not coded by a full noun phrase, or an autonomous word at all. It may be fully elided, or ‘understood’ (as in Japanese), or there may be some reminder of its existence in the form of so-called verb agreement or coreferencing (as in Spanish), or in the form of clitic or attenuated (and rather fixed) pronouns (as in English).

The ubiquitousness of continuous topics in actual discourse would seem to be the primary motivation for the existence of subject coreferencing in so many languages.³⁴ It is also the functional motivation for the grammaticalization of same subject clause chaining constructions, reduced complement clauses, and same-reference marking systems in other languages (cf., e.g., Longacre 1983, 1985; Givón 1990).

In a minority of cases in actual narrative discourse, and perhaps a higher percentage in other genres, the topic is not fully active, or even readily accessible, or there is interference from other referents, and thus, since it is not predictable, the topic is covertly expressed by a full nominal. To be a ‘proper ground’, a referent doesn’t have to be readily accessible or highly topical itself, or even identifiable. It is enough for it to be itself grounded on an accessible referent from the universe of discourse, as in the statements in (3.11) below.

- (3.11) a. My second grade teacher just called me.
b. Someone I met the other day just called me.

If the referent is very inaccessible it may even be placed in its own intonation unit, i.e. it will be ‘dislocated’ in a sentence-head intonation unit or introduced in a separate assertion, as in (3.12).

- (3.12) (Remember) my second grade teacher, she just called me.

A speech act participant, or other highly salient referent in the world of discourse can also readily become a ground. The physical context can also be the source of valid topic (e.g. *That man is wearing a red tie*). It seems that only ungrounded unidentifiable referents cannot easily be topics, since they are not very good at grounding the proposition, though languages may differ in the rigidity with which they adopt this statistically valid principle, and the availability of exceptions in particular contexts.³⁵

3.4.7 The position of overt topics

A topic is in a sense both **outside** the assertion, as both the link to the context and the base for the assertion/predication, as well as **inside** it, since it is semantically part of the proposition. In a sense, it is nothing more than a special case of *setting*, a setting that may be coded by a ‘dislocated’ phrase, as settings typically are (i.e. may be placed in a sentence-head, setting intonation unit), but which, if it is not extremely inaccessible, may attach itself to the sentence-body intonation unit where the clause it grounds is located, much like stereotypical, unaccented settings are (e.g. *then, later*).

The fact that the topic is a special case of a *setting* accounts for the fact that topics, when they are overt, typically go at the beginning of the clause, before the

predicate, just like all other settings do. The clause-external character of topics is shown by their being placed separate from the assertion proper, just like other settings. Their clause-internal character is evidenced by their being in the same intonation unit as the clause (unless they need their own intonation unit to become a proper ground). This is, I believe, the reason why by default the vast majority of the world's languages have their typical subjects in clause initial position (at least when they are topics). This is also the functional motivation leading to grammaticalization of the 'subject position' in the majority of languages.

It has been argued that a minority of the world's languages do not have their (supposedly) topic subjects in clause-initial position (VSO, VOS, and OVS languages). This may be related to the fact that even subject-initial languages under some circumstances allow the postponing of subjects which seem to be topics (or 'antitopics', cf. Chafe 1976; Lambrecht 1986a; see below). On the other hand, even non-subject-initial languages seem to allow for the left-dislocation of certain topics, or even fronting in the same intonation unit, under some circumstances (cf., e.g., Aissen 1992). Also, such languages are known for making great use of 'voice alternation' constructions which allow them to front a topic when necessary. Remember also that clauses with covert topics tend to be a small percentage of those used in actual speech, particularly in narrative. In other words, labels such as 'subject-initial language' or 'non-subject-initial language' do not tell the whole story about word order and in fact may even be irrelevant categories.

3.4.8 Split in the topic's sub-roles

I have argued that the topic is both outward-looking and inward-looking. This fact has been the source of some confusion as to the nature of the privileged element of the clause, with some stressing its outward-looking continuity aspect (e.g. Givón) and others stressing its status as predication base, that is, its inward-looking status.

It would seem, however, that occasionally these two roles are split between two different referents between which there is some semantic or pragmatic relation. The well known examples of 'double-subject constructions' in languages like Chinese might seem to be examples of this phenomenon ('Chinese topics', cf. Chao 1968; Chafe 1976:50-1). We can see an example of such a construction in (3.13) (from Li and Thompson 1976a:468, ex. 23).

- (3.13) *Nèike shù yèzi dà*
that tree leaves big
"That tree (topic), the leaves are big" ["That tree has big leaves"]

In this example, the initial nominal sets up an entity which links the assertion to the universe of discourse. But the assertion itself has its own 'topic', a *part of* the tree, which is what the predication that follows is about. What makes this split possible is the part-whole relation holding between the two 'topics', two referents connected semantically by the *part of* relation.

In (3.14a) below we can see a schematization of the prototypical case in which a topic (thematic ground) performs both the discourse linking role and the predication base role. In (3.14b) we see a schematization of the less common situation in which the two roles are performed by two different referents which are related between themselves.

- (3.14) a. Context ← [Ground] → State of affairs
b. Context ← [External Ground ↔ Internal ground] → State of affairs

Languages differ as to the extent and the exact ways in which they sanction such a construction. In the Basque corpus I have found that in narrative it is relatively common for the external linking role to be fulfilled by the referent of covert argument and the internal predicand role to be fulfilled by another, overt major argument (cf. Chapter 4).

3.4.9 Other types of assertion grounding (non-thematic grounding)

The thematic grounding of asserted propositions, whether continuous (covert) or non-continuous (overt), seems to be a rather special and primary way of connecting propositions to the discourse context or to the world of discourse in general. On the other hand, thematic grounding of propositions, however privileged it may be, it is not the only possible type of grounding found in human language. Other types of grounding are also important and in some contexts they can be primary.

The similarity in function, and even sometimes in form, between topics and other settings has been often noted in the past. But setting elements are typically ‘dislocated’, in their own intonation units, preceding the asserted clause, whereas topics are only sometimes dislocated, when the pragmatic characteristics of the idea in question so warrant it. On occasion, however, we find certain settings, such as locatives, in clause and intonation-unit initial position. The situation typically involves clauses which have no topic, that is, *thetic* clauses, such as in existentials and presentatives (and perhaps also sentences which do not have *overt* topics).

It has been noticed before, for instance, that there are formal similarities between ‘subjects’ (i.e. default thematic grounds) and the ‘preposed locative’ in ‘locative

inversion' sentences in English, a term which includes locative phrases as well as other circumstantial elements, e.g. *On the bench sat an old lady* (cf., e.g., Bresnan 1994 and references therein).³⁶ Some have gone even as far as suggesting that these locatives are 'subjects'. Although I disagree with this particular analysis, I think the intuition it expresses is correct. I would argue that these locatives are a privileged type of grounding element, much like topics are. Instead of thematic grounds these are locative grounds, but the pragmatic grounding function is exactly the same in both cases.³⁷

The idea that there are different types of grounds is very similar to Halliday's (1985) notion of multiple (overt) themes, though Halliday goes further than I would in what he is willing to call a 'theme' (what I would call 'ground'). Halliday argues for the existence of textual themes (continuative, structural, relative, and conjunctive), interpersonal themes (modal, polar, and vocative), and ideational themes (topical theme) (cf. Halliday 1985).

3.4.10 Thematically grounded and ungrounded sentences

Until quite recently, in the traditional Western grammatical tradition it has been assumed, or taken for granted, that all sentences have a bipartite subject/topic-predicate structure. The analog of this in modern formal linguistics is the formal division of the clause into NP and VP. Of course, this is but an approximation to the actual situation in the most common type of assertion constructions in natural language.

Viewing the division of asserted propositions as a pragmatic division, and not as a formal one, allows us to account for exceptional cases which do not fit into this mold. As Kuroda (1972) has argued, not all asserted sentences have a topic-comment configuration. Some assertions are indeed assertions of a full proposition, and not

predications of a (privileged) element of that proposition (cf. Kuroda 1972, Sasse 1987, Lambrecht 1994).³⁸

Before proceeding, I should make clear that not all sentences in which the subject is not the topic are what is typically known as a *thetic* assertion. To begin with, something other than the subject could be the topic, as in (3.15), where the object is obviously the topic and the subject is postverbal and part of the assertion.

- (3.15) Ese libro lo leyó mi madre ayer
that book it she.read my mother yesterday
“That book, my mother read it yesterday”

Secondly, sometimes the subject is inside the predicate because it is the single new element, as in (3.16aA), as in the answer to a content question about the subject role, or when the subject is contrastive, as in (3.16bA), in other words, what I will call below a very salient focus.

- (3.16) a. Q: Who came?
A: My mother came.
b. Q: Did your father come?
A: No, my mother came.

Notice that in English both of these ‘narrow subject focus’ sentences are indistinguishable from the *thetic* sentence *My mother came*, uttered as an announcement for instance, for the simple reason that in both cases the subject is the focus and not the topic of the assertion (see below.)

Thetic assertions are indeed a minority of sentences in natural languages. These are sentences which have special semantic and pragmatic characteristics. There are two types of *thetic* assertions, according to Sasse: referent central *thetic* sentences and event central *thetic* sentences. Referent central *thetic* sentences, such as existentials, do not

have a topic because the only candidate for that role simply is not topical enough (it is not a 'proper ground'). Event centralthetic assertions are those in which a referent is not involved, i.e. impersonal expressions, such as weather expressions.³⁹ In fact, as we will see in Chapter 5, 'referent central'thetic clauses often have non-thematic grounds (cf. above), and in many languages 'event central'thetic clauses often do have topics, just not the argument of the verb which is expected (i.e. the 'subject').

Languages differ as to how they codethetic assertions differently from thematically grounded ones. This always seems to involve treating the would-be topic differently from a true topic. This can be accomplished by giving it different coding, such as by means of a different morphological marker, as in Japanese (cf. Kuroda 1972; Sasse 1987, 1995a, 1995b).⁴⁰ Another way, perhaps more common, consists of placing the would-be topic inside the assertion. This may be accomplished by inverting the subject with the verb and typically having it bear the rheme's main accent, a common strategy in VO languages with a flexible subject, such as Spanish. Placing the subject inside the assertion can also be accomplished by simply placing it in assertion-initial position inside the intonational contour of the predicate, a strategy which is common in flexible OV languages, such as Basque, or rigid word order languages, such as English (cf., e.g., Bolinger 1954).⁴¹ In Table 3-7 we can see some of the different possibilities for topic-comment andthetic assertions in Spanish and English.

Assertion type	Lg.	Example	Characteristics
Topic-Comment (categorical')	Eng.	The <u>teacher</u> <u>spoke</u>	Two accents or accent only on the predicate
	Span.	El <u>profesor</u> <u>habló</u>	Two accents or accent only on the predicate
Thetic (Topic-less)	Eng.	(Then) the <u>teacher</u> spoke	One accent (on the subject); ambiguous: a) THETIC (all new) b) focal, contrastive subject c) focal, new subject
	Span.	(Luego) <u>habló</u> el <u>profesor</u>	One accent; subject inversion; ambiguous: a) THETIC (all new) b) focal, contrastive subject c) focal, new subject
Other Topic-less	Span.	(¿Quién <u>habló</u> ?) El <u>profesor</u> <u>habló</u> .	One accent; no inversion; ambiguous: a) focal, contrastive subject b) focal, new, emphatic subject

Table 3-7: Examples of topic-comment and topicless sentences in Spanish and English.

As we can see here, topic-comment sentences are very similar in English and Spanish. Thetic sentences, on the other hand, are quite different: SV for English and VS for Spanish. In English the thetic sentence is identical to a subject focus construction. In Spanish the VS construction can also be a narrow focus construction, but the SV construction may also fulfill that role (though it cannot be a thetic sentence).

Actually, the reason for the positional differences of subjects in thetic in categorical sentences is not that the language strives to code a non-topic subject 'differently' from a topic subject. As I will argue below, the reason has to do with the fact that in thetic sentences such as the ones we've looked at, the subject is indeed the focus of the assertion, and as such it must be placed wherever focus subjects are placed in the language. In Spanish a focus argument of any kind, whether a subject or object, must be postverbal (and accented) and hence the inversion. In English, as we saw in the examples in (3.16), the position for a focus subject, though not of focus objects, is preverbal, bearing the assertion's accent, and thus that is the position of the subject in a thetic assertion. In Basque, on the other hand, the main position for a focus subject, or

object, is preverbal, and thus Basquethetic sentences are on the surface very similar to English ones. I will return to this issue in Section 3.5 below.

3.4.11 Obligatory vs. optional thematic grounding

The vast majority of asserted propositions in discourse are thematically grounded propositions, that is, they have a topic-comment configuration, and are thus predications. Indeed most of assertions cannot be expressed asthetic assertions at all, especially if they have additional focal elements. So the question is, when can an assertion have an all-comment configuration instead of a topic-comment one, and when, if ever, is such all-comment configuration required?

Kuroda (1972) has argued, for instance, that what he calls “generic” sentences, i.e. those which make statements “about a general, habitual, or constant state of affairs of affairs” (Kuroda 1972:160), are always thematically grounded and can never havethetic structure, e.g. *People eat, John reads*.

Sasse’s view, which I believe is essentially correct, is that assertions expressing (non-caused) events, or “happenings”, are amenable tothetic interpretations,⁴² but not so predicates representing actions, and in particular those expressing states or descriptions, that is situations in which there is a salient autonomous entity of which the assertion must be interpreted as a predication (cf. Sasse 1987:565). Clauses with non-verbal predicates and copula verbs, for example, such as, e.g., *Mary is a carpenter*, must have a topic-comment informational configuration and may never bethetic (although the subject may be the ‘narrow’ focus under some circumstances). The same thing is true of actions, the more so the more prototypical the action is.⁴³ Assertions expressing events, or happenings, on the other hand, in which the subject’s role is not a prototypical topic role,

are more amenable tothetic expression, though they may also have a topic-comment structure if the referent is topic-worthy enough or the speaker chooses to make it so.

Some assertions exhibit a preference forthetic interpretation. Existentials in particular often do because there is no topic-worthy referent in such assertions. Other assertions that exhibit this preference are those containing verbs of appearance, psych-verbs, meteorological conditions, and the like which can be given an event-assertion interpretation, as opposed to a predicate-assertion interpretation (cf. Sasse 1995a:23-24) and whose subjects, if any, are not very topic-worthy. However, most of these predicates can also be found in thematically grounded clauses.

Sasse argues that the closer a 'predicate', such as a verb, is to being a prototypical predicator the less possible it will be for that predicate to be found in athetic assertion. Better predicators, such as actions and states/descriptions, are those in which the predicate is most autonomous from the predicand. Worse predicators are those in which the (potential) predicand forms an intrinsic part of the event predicated by the predicator. Thus predicates which are less prototypical predicators can be used in non-predicative (i.e.thetic) functions.

Besides the semantics of the predicate, the **aspect** of the assertion is also important in determining the possibility that the assertion will bethetic (i.e. that the subject will be the focus), since the aspect is also a determinant of the possibility of the assertion having an event interpretation. Thus stative predicates, such as *know* and *like*, when used with imperfective aspect, always require a topic-comment articulation, as I mentioned earlier. When a stative predicate is in a perfective aspect, however, an event

interpretation may be possible. (Notice that some languages may use totally different predicates in the two situations, e.g. *know* vs. *figure out*.)

A thetic configuration is possible with a stative copula verb under some circumstances, such as in example (3.17a) below, with an existential/presentative function. Notice that the active location, and not just any location, is required in this case, which is why the sentence in (3.17b) doesn't seem to be pragmatically well-formed as an announcement, since it has a topic-comment structure but not a topic-worthy subject. The assertion, of course, would be perfectly acceptable if it was an answer to the question *Who is in the house?*, in which case the subject is a very salient, or 'narrow', focus).

- (3.17) a. A man is here (to see you). (= There's a man here)
b. ?A man is in the house. (<> There's a man in the house)

In (3.18a) we see another copula sentence with a thetic structure. Notice that this is a stereotypical event/happening (a baby crying). When the assertion doesn't express a stereotypical event, as in (3.18), a topic-comment configuration is required, which requires a more topic-worthy subject than the one here.

- (3.18) a. A baby is crying (= There's a baby crying)
b. #A baby is barking.

Contrary to what is sometimes assumed, out-of-the-blue type statements answer to what's-the-matter type questions are not necessarily topicless any more than any other statements are (cf. Osa 1990, Lambrecht 1994). A topic-comment configuration seems to be required of this type of statement unless the propositions and the referents of the relevant arguments meet the criteria mentioned above. In example (3.19) we see several possible answers to a question which requires an 'all-new' statement.

- (3.19) Q: What's the matter?
 A₁: My father came. (thetic)
 A₂: My father came. (topic-comment)
 A₃: My father read my book. (topic-comment)
 A₄: #My father read my book. (non-topic-comment)

With the predicate *come* and an identifiable subject, as in A₁, thetic structure seems to be the most likely possibility. In a more 'complex' context, such as if the arrival of the addressee's father was an accessible proposition, a topic-comment configuration, such as the one in A₂, would be found. On the other hand, with a more complex predicate, such as *read my book*, a topic-comment configuration, such as in A₃, would seem to be required. A topicless configuration, such as the one in A₄, is a possibility only if the proposition is contrasted with another accessible proposition which is different from it only in having a different subject. In other words, when all the other elements of the assertion are very accessible (non-focal).

Lambrecht 1994 seems to imply that in contexts with low presuppositionality, where *all-news* sentences are often uttered, these sentences should be thetic. This is however not the case. His main (real life) example *My CAR broke down*, uttered by a person entering a bus in a rather boisterous manner, is obviously a topic-less assertion, but in exactly the same context a person could have uttered a thematically grounded sentence, such as *I LOST my wallet!* I will return to these questions in Chapter 5.

3.4.12 Conclusions

To conclude, I find that the notion topic, and the split between topic and predicate are key notions for understanding the structure of the majority of assertions, as well as the minority of assertions which have a different structure. Pragmatic status notions such as

topicality or accessibility, and syntactic or semantic notions such as subject and predicate (NP-VP), are insufficient for accounting for the facts of language, including the ordering of elements within the clause.

In the following section I discuss the other major pragmatic function, the focus function. This category accounts for many of the properties of language, including ordering preferences. Unlike the topic category, which has a clear preference for clause initial position, at least in statements, the most common type of assertion, the focus category seems to have a single preference, namely being next to the verb, the core of the predicate/assertion. Thus one possibility is for it to be preverbal, as in so-called OV languages, and the other is for it to be postverbal, as in so-called VO languages.

Complications arise in oddball languages, such as English, in which the order of constituents is quite rigid and associated to a large extent with syntactic categories. The unusualness of English in the very important respect constitutes a very strong argument for rejecting the use of English categories and patterns for the analysis of most languages, a practice which has been common in many formalistic approaches to linguistic analysis in recent years.

3.5 FOCALITY AND FOCUS

3.5.1 Introduction

In the previous sections I have argued that not all elements of an asserted proposition are equal from an information perspective. In the vast majority of asserted clauses, one element is selected as the ‘thematic ground’ or ‘topic’. This element grounds the proposition to the world of discourse and is what the assertion, now a predicative assertion, is about. In a way we can say that this element is outside as well as inside of the predication proper. Which element of the proposition will be the topic is typically determined by default, but occasionally this default can be overridden. The topic is coded positionally and well as intonationally, and occasionally by means of topicalizing operators. In actual discourse, the topic is more often than not predictable and thus is coded minimally, or even ‘elided’. We also saw that not all asserted propositions have a thematic ground, though the vast majority do.

In this section I want to explore another way in which an element of an asserted proposition may differ from other elements in that proposition. This time the element in question is inside the assertion proper, i.e. inside the predicate when there is a topic, but it stands out in some way above all other elements of the assertion. This element is somehow the most important, or noteworthy element of the rheme. In other words, it is the most focal element, and thus the ‘focus’.

3.5.2 Assertion focus

In (3.19A₄) above, repeated in (3.20) below for convenience, we saw an example of a declarative assertion in which an element of the assertion stands out over the rest.

The asserted proposition is [My father read my book], but this particular asserted clause, with the only single accent on *father*, can be found only in a context in which the open proposition [X read my book] is highly accessible, such as after the question *Who read your book?*

- (3.20) My father read my book.
(~ “It was my father who read my book”)

The rhematic constituent *my father* in assertions such as this is often said to be the focus of the assertion. In this case the focus is the subject and thus there is no topic, but the focus may be an object, for instance, as in (3.21).

- (3.21) Q: What did your father read?
A: He read my book. (Accessible: “Your father read something”)

The background for the assertion in (3.21A) is the open proposition [Your father read X/something].⁴⁴ The element which fills in the gap is said to be the focus of the assertion. To be sure, it is the whole sentence (minus the topic) that is asserted, all of it that is ‘under the scope of the assertion’, but there is something special about this one element since it is what turns an accessible open proposition into a meaningful assertion. Notice that the major accent of the assertion always falls on this focus element. (The accent could actually fall on some internal part of the focus phrase, the part which makes it salient, e.g. *My father read my book*, but the whole object, or rather its referent’s idea, is the focus of the assertion.)

The existence of a focus(ed) element in assertions is quite obvious in a number of contexts, such as in content questions and answers to such questions, as well as in cleft constructions for example (cf., e.g., Chomsky 1971/1970). The question is whether this type of structure inside the assertion can be generalized to other types of assertions

besides those in which the open proposition consisting of the asserted proposition minus the focused element is highly accessible. I believe that this is indeed the case, namely that the notion of focus can be generalized to all types of assertions. In other words, I would argue that *in all assertions there is one element which stands out above the rest, one single element which is the most salient element in the rheme. This is also invariably the element which receives the single accent that typically falls on the rheme part of the assertion.*

The actual degree of salience of the focus(ed) element, what I will call 'newsworthiness' or 'communicative dynamism', but which I will call focality for consistency, however, varies from one sentence type to another. Not all focus elements are as salient as those involved in an assertion in which the open proposition which excludes the focused element is very accessible or even active. But it is true that all assertions have a single element on which the accent falls and, very importantly, in languages in which the focus occupies a single consistent position, such as Basque or, to a great extent, Spanish, all assertions have a constituent filling that position.

Thus it is not surprising that the assertion in (3.21A) above could equally be an answer to the question *What did your father read?*, *What did your father do?*, or even *What happened?* The accent would still fall most naturally on the object, which would in every case be the focus of the assertion. This is so even though in the answer to the first question the focus would be more focal, it would stand out more over all the other ideas involved in the assertion.

Any element of the assertion can be the focus, including the verb itself, the polarity, complements, and even parts of complements. Even when the whole assertion

(predicate) is 'new', there is always an element which stands out and receives the rheme's main accent, a motivated, default focus, just like there is a motivated, default topic. If the rheme consists only of a verb, as is the case in about 10% of clauses in narrative, the focus is on the verb. If the verb has any (non-topic) 'new' complements, which more often than not are single complements, then by default that complement, and not the verb, is the focus of the assertion, even if the verb is also 'new', especially if the verb expresses a stereotypical or basic-level action, such as *read*. On the other hand, when the object is active or very accessible, as in (3.22a), or when the verb is semantically complex, and the idea expressed by the complement is easy to accommodate or not very important (i.e. low in focality), as in (3.22b), the verb may then be the focus.

- (3.22) a. He read it.
b. He perused a book.

Finally, when the whole assertion is emphatic (and the complements are not too focal), then the finite verb, standing for the assertion, and thus its affirmation or negation, behaves as the focus constituent, as in (3.23a&b) below.

- (3.23) a. Guess what? My father read a book.
b. My father did read a book.

These assertions involve the emphatic affirmation, or negation, of the whole rheme, and it is found typically in cases in which the proposition and all the elements are easy to accommodate. A typical place where we find this type of assertion is in answers to polar (yes-no) questions, which is why they are sometimes called assertions with polarity focus. Notice that the polarity is always instantiated on the finite verb itself. I will look at these assertions in greater detail in Chapter 7.

Notice that the idea that every assertion has a focus is fully consistent with the claim, mentioned above and proposed by many previous investigators, that in the overwhelming majority of cases in actual discourse asserted clauses contain one and only one 'new' or 'newsworthy' element, a single overt constituent: either a verb or a complement together with a 'stereotypical' (not very focal) verb. It is possible, though not common, as we will see below, for an assertion to contain other 'new' elements. Still, unlike the focus element, these other elements consistently seem to be treated as given, accommodated, or setting-like elements, and cannot be accented, unless they are dislocated to the sentence tail. Under no circumstances do they have the same 'informative status' or 'salience' as the accented element when they are inside the clause's intonation unit.

To summarize, the focus is that part of the assertion where the information is most concentrated, where the main point lies. In Prague School terms we can call it the most *communicatively dynamic* element (cf. Firbas 1966), or in Mithun's (1987) terms, the most *newsworthy* (or 'noteworthy') element, and also typically the 'newest'. In Givón's term we could say that it is the *least predictable*, the *most unexpected*, or *most surprising* (cf., e.g., Givón 1990). Unlike other investigators, however, I believe that this is a unique pragmatic *function* or *role* associated with every assertion.

The focus notion, just like the topic notion itself, cannot be reduced to the pragmatic, or cognitive, status of the idea expressed by focus phrases. It is true that just as topics are often given, foci are often new, but the correlation is independent of the pragmatic function. Just as a topic need not be given, as long as it can be made identifiable, a focus need not be not-given (new), as long as it is the most informative

element in the assertion. In other words, the two sets of notions belong to different planes. The correlation between roles and statuses is indirect, although it is definitely more likely in discourse that an assertion will be grounded by accessible elements and that the most salient part of the assertion will be a new idea. As Bolinger has argued, “old information may well **tend** to be unfocused, and new to be focused, but that is the speaker’s business. Old information is focused if it is still of sufficient interest” (Bolinger 1992:294).

Also, as I said earlier, the pragmatic status (as well as other pragmatic properties) of the idea filling the focus role is important with regards to the actual formal realization of the focus, much as we saw the pragmatic (accessibility) status of a referent was important for the choice of topic and its formal realization.

The actual nature of the focus depends greatly on the variety of assertion that is involved. It seems that three major types of assertion are relevant and have formal consequences in many languages, which are summarized in Table 3-8.

Assertion Type		Focus Type
Average assertion	↔	New focus
Contrastive assertion	↔	Contrastive focus
Emphatic assertion	↔	Emphatic focus

Table 3-8: Major assertion types and focus types.

The first type of assertion type is the all-new assertion, the most common type by far. The type of focus which corresponds to this assertion I will call new focus. The second type of focus is the contrastive focus, which corresponds to a contrastive assertion. Finally, an emphatic focus corresponds to an emphatic assertion. These three types can be found in different types of speech acts, such as statements and questions. I will

describe these different types of assertions below and, in more detail, in the following chapters.

3.5.3 New (regular) assertions and new focus

The first type of assertion is the most common one by far in natural language. It applies to most statements (declarative assertions) and questions, for instance. It applies to out-of-the blue statements (most of which, as we saw, are grounded), to questions and to answers to questions, such as those in (3.24)

- (3.24) a. What did you do?
b. What did you buy?
c. I bought a book.

This is also the type of assertion found in the vast majority of continuous sentences in narrative, as in (3.25), which involve new ideas in the context of new, asserted propositions.

- (3.25) a. I was bored this afternoon.
b. So I went to {for a walk / to a bookstore}
c. and I bought a book.

As we can see in the example in (3.24c), there doesn't seem to be any formal difference between assertions in which all but the focused element is accessible from the context, as when it follows the question in (3.24b), and those in which the whole predicate is new, as when it follows the question in (3.24a), although in the former case the focus element is more pragmatically salient, or focal. Thus, in either situation, in the assertion in (3.24c) *I* is the topic ('given') and *bought a book* is the predicate, or rheme, with the object *a book* receiving the focus accent. As an answer to (3.24b) the focus is the *single* truly 'new' or 'informative' idea in the predicate (the single new, or informative element

in the assertion). As an answer to (3.24a), on the other hand, all the elements in the predicate are new, and thus the focus is simple the *most* informative element. The two answers are indistinguishable from the point of view of information structure, however, and the rheme's accent falls on the object *a book* in both cases.

Actually, the assertion in (3.24c) would have the same pragmatic structure even if it was a reply to the question *What happened?* Thus the final asserted clause in the mini-discourse in (3.25) (*I bought a book*) has the same form and structure as the sentence in (3.24c). All the ideas in the predicate (*buy* and *book*) may be new, such as when the context is *I went for a walk*, or they may be somewhat inferentially accessible, such as when the context is *I went to a bookstore*. In either case, and just like in the preceding contexts, the object is the focus of the assertion.

The questions in (3.24a&b) are themselves assertions. As in most content questions, the focus element is the question (*wh-*) phrase, which in English, and perhaps in most languages, is always clause initial, preceding the finite verb and with the subject inverted with the finite verb (i.e. the default focus position for OV languages and the marked focus position for VO languages, as we will see). An interesting question is whether these inverted subjects are topics, the way preverbal subjects are typically topics in statements. Typically, and from a pragmatic or informational point of view, we can identify the referents of these subjects as the topic of the assertion, what the assertion is about, though this is not always the case. (As we will see later on, this question can be generalized to all cases of subject inversion in Basque and some types of inversion in languages such as Spanish.)

The key questions then are: why are these subjects inverted, and why does the difference between topic-like subjects and those which are not topic-like neutralized? Later on I will argue that this is due to a tendency, which is obviously grammaticalized in these constructions in English, for assertions that have a very marked (focal) focus, such as emphatic and contrastive assertions, assertions which are disruptive and non-continuous, to display subject inversion, since the notion of topic and grounding tends to be less crucial or necessary and since there is greater urgency in verbalizing the (very focal) focus. Clause-initial position seems to be more ‘marked’, or ‘salient’, than other possible focus positions, which is why in languages in which there is more than one option for the position of focus constituents in statements, initial position is reserved for more marked ones, such as those that are contrastive. I will return to these questions throughout the rest of this study.

3.5.4 Contrastive assertions and contrastive focus

Contrastive assertions are those in which an element of the rheme, the focus, which may be any element or even the whole rheme, contrasts with some element in a similar (or identical) proposition which is accessible from the context. The two propositions differ in the one element that is contrasted, or in the whole rheme, in which case the finite verb is treated as the (contrastive) focus constituent of the assertion. We can see an example of such a contrastive assertion in (3.26A).

- (3.26) Q: Did you buy an umbrella?
A: No, I bought a book

Here the asserted proposition [*I bought a book*] contrasts with the proposition [*I bought an umbrella*] (or, alternatively, the predicate [*bought a book*] contrasts with the predicate

[*bought an umbrella*]). For an assertion to be contrastive the open background proposition need not be active, as it is here, but it may be merely accessible through contextual inference. (In the case of the question in (3.26Q), either the polarity or the object could be accented and thus be the focus, as the differences on where the accent falls, but not any differences in word order in this case, show; cf. Chapter 7.)

In the assertions in (3.27B), the whole predicate contains given ideas (*buy* and *a book*), and the whole assertion (rheme) contrasts with a given one.

- (3.27) A: Why didn't you buy a book?
B: I did buy a book. / I bought a book.

In this case, as we can see the element which receives the focus accent and is treated as the focus is the bearer of the 'polarity', in other words, the whole affirmation (affirmative assertion) is contrastive.

In standard English a contrastive focus is coded much the same way as a new focus: accentually, although the accentual prominence may be typically higher in this case. In some dialects of English, a contrastive focus may be placed in a different place from where a new focus would be expected, as we will see.

So far we have seen two possibilities for the positioning of a focus constituent in English: clause-initially, for subjects and question words; and postverbally, elsewhere. In the case of rheme or assertion focus, when the finite verb is treated as the focus constituent, the verb is in rheme-initial position, just like when it is clause-initial, the only difference being that here the topic-subject is not inverted. It has been noted that in Yiddish English and other dialects of English, a contrastive focus complement may be placed in assertion-initial position as well (cf. Prince 1981a).

A contrastive focus idea is even less strongly correlated with the pragmatic property ‘new’ than a new focus is. As Chafe argues, “contrastiveness is independent of activation cost [i.e. accessibility, topicality]. That is, a contrastive referent may be given, accessible, or new” (Chafe 1994a:77). Plain new foci, on the other hand, are much more likely to be rather new. Givón has argued that both contrastive and non-contrastive (informative) foci display an identical type of pragmatic property, which is what makes them both focal, namely “low informational predictability”, or “unexpectedness” (cf., e.g., 1983a, 1995:Chapter 7). In the case of contrastive focus, unexpectedness actually means counter-expectedness, which is why contrastive foci are so ‘salient’. As I said, I will use the term focality for the gradient property which underlies of foci, which is analogous to Givón’s notion of “low informational predictability” and Mithun’s notion of “newsworthiness”, for example.

We cannot, however, identify contrast with focus. Contrastiveness is an independent parameter from focus. For one thing, as is well-known, topics too may be contrastive, as we can see in (3.28).

(3.28) My mother bought a book,
and/but my father (bought) an umbrella.

In this pair of asserted clauses we see that the topic of the second clause (*my father*) contrasts with the topic of the first one (*my mother*), much like the focus of the second clause (*an umbrella*) contrasts with the focus of the first one (*a book*). (With different verbs in the second clause’s predicate it would be the whole second predicate which contrasts with the first one, e.g. *My mother bought a book and my father washed the car.*)

Furthermore, an element may be contrastive which is not either a topic or an element of an asserted clause. Thus in (3.29) below we can see that an element in a

setting clause in (3.29c) contrasts with another element in the setting clause in (3.29a), making the two settings contrast with each other.

- (3.29) a. When I went to the bookstore,
b. I bought a book,
c. but when I went to the bakery,
d. I didn't buy anything.

An embedded contrastive element is not the focus of an assertion either, though it makes the larger constituent which is a sister of the asserted verb the focus of the assertion, as in (3.30).

- (3.30) I read [a book with green covers], not blue.

Here the focus constituent is [*a book with green covers*], although the contrastive element is only the noun phrase's modifier (*green*). Focus, much like topic, is primarily a pragmatic relation which is relevant in asserted clauses, whereas contrastiveness may be relevant elsewhere as well.

Authors such as Bolinger have argued that the difference between contrast and new focus is not such a clear-cut one. In theory, when we say that an idea is contrastive the idea is that it contrasts with a limited set of clear-cut alternatives, and when we say that it is simply new (non-contrastive) it means, in fact, that it contrasts with all other possible (contextually compatible) ideas, or at least a non-specific set of ideas, all of which were equally unexpected. In real life, however, a closed set may be somewhat open ended and unspecific, and an open set is may be quite constrained by the context.

I too believe that contrast is not an actual discrete pragmatic category and that contrastiveness is a gradient notion depending on whether there are more or less explicit and clear-cut alternatives in the context or not. What is relevant is that the more clear and

obvious the contrast between a focus element and another element from the context, the greater the degree of **intrinsic focality** of the focus, and it is this increase in degree focality that may have formal consequences, such as the coding of the focus in some special way, such as by placing it in a special position or in a special (marked) focus construction, or simply the greater intensity of the accent.

Thus, *ceteris paribus*, the more contrastive a focus element is the more salient (focal) it will be and its salience will be intrinsically derived from the context in which the assertion and the focus are found. There are other situations in which a focus element has enhanced focality which derives from the context, such as in content questions, in which the focus is typically quite salient (with respect to the other elements in the assertion, which are treated as accessible), a fact that is reflected in the (grammaticalized) form of content questions in English, for instance. There are other situations, however, in which a focus element is very focal because the speaker chooses to make it so. These are what I call (extrinsically) emphatic foci and emphatic assertions, to which I now turn.

3.5.5 Emphatic assertion and focus

Emphatic assertions, such as those found in exclamations, are not a different type of assertion from contrastive and regular (non-contrastive) assertions. They are simply assertions in which the focus, which may be the whole rheme, is more salient than usual, typically because the speaker chooses to emphasize it, for some other reason. This also makes the whole assertion, and not just the focus, more salient than an average assertion. Both contrastive and non-contrastive assertions can be emphatic. We also find that emphasis, like contrast, is a matter of degree, and it has to do with the ‘forcefulness’ with which an assertion is made, the degree to which it is not open to challenge, the degree to

which the speaker feels strongly about it. When the whole rheme is presented emphatically, but no particular element of the rheme is focal, the finite verb itself, instantiating the assertion (e.g. affirmation or negation) is treated as the focus constituent.

Emphatic assertions are always coded by a higher degree of intonational 'intensity', a more marked accent on the focus element, with a more noticeable accent, and a more noticeable pitch contour. These iconic coding correlates of emphasis are found in all languages. In addition, it is very common for emphatic constructions in many languages to alter the ordering of their focus and/or topic constituents from those typically found in non-emphatic assertions. This kind of ordering is what is often known as **inverted order**. The characteristics of inverted order are the same for OV and VO languages:

- **Focus fronting:** That is, the placement of the focus element in rheme-initial, preverbal position. This, of course, is the normal position for foci in "OV languages", as well as the order found in content questions in English. Focus fronting is not common otherwise in English, but it is quite common in VO languages such as Spanish. Focus fronting is found in certain minor emphatic constructions in English, as well as with very salient foci which are not arguments of the verb, such as adverbials (e.g. *Tomorrow I'll come*).
- **Topic inversion:** That is, the placement of an overt topic towards the end of the assertion, along with other non-focus assertion-internal complements. Like those other complements an inverted topic must be unaccented. For this to happen the topic has to be fairly accessible. Inversion with emphatic assertions is not restricted to topics, however, and it applies to other setting elements as well. These inverted

topics seem to be the same thing as Chafe's 'antitopics' and the right-dislocated subjects of English (e.g. *She's great your sister!*).

The outcome of these two 'operations' is that the highly salient assertion, as well as the highly salient focus, the most salient part of the assertion, are placed in **clause initial** position, as well as **sentence initial** position when there are no setting elements or when they are postposed.

These ordering preferences are associated cross-linguistically with emphatic assertions, i.e. with assertions with a very focal focus, and with constructions which express such assertions (or which *used to* express such assertions, since constructions may, to some extent, change their pragmatic properties while retaining their form, though not without the system reacting in order to correct the mismatch between form and function). These 'reorderings' (initial focus and inverted subject) are associated with assertions (or assertion constructions) of any type in which the focus element is significantly more salient than the rest of the elements of the rheme, i.e. which have very focal foci. This would seem to account for the fact that the question phrase in the English question construction is placed in clause initial position and the fact that the subject (whether it is pragmatically the topic or not) is inverted.

These iconic ordering principles are reflected, and grammaticalized, in other minor constructions in English as well, such as in exclamations of different types, e.g. *Never will I do that again!*, and with quote complements, *'I will go,' said my sister*. In other languages, however, the same principles manifest themselves much more freely, as we will see is the case in Basque.

In Spanish it is well-known that an emphatic assertion with a contrastive or non-contrastive focus element can trigger both characteristics of inverted order: object fronting and subject backing, as in (3.31) below (cf. Silva-Corvalán 1983, 1984).

- (3.31) Un libro compré yo! (OVS)
one book I.bought I
“I bought a book!” (“A book I bought!”)

Note that this in this context we often find overt subject pronouns which would typically be elided in normal (non-emphatic) assertions. Sometimes it would seem that these pronouns are contrastive (topics), the most common situation in which overt subject pronouns are used in languages like Spanish. Other times, however, it doesn't seem that these pronouns are contrastive at all. They rather seem to be superfluous pronouns, the only purpose of which might be to show that they are inverted and thus help code the assertion as emphatic.

Focus fronting in emphatic assertions seems to be a universal strategy and it is found even in English in cases in which word order constraints allow it. Take for instance the following pair of statements in (3.32A'&A'') from Hannay 1991 (p. 132).

- (3.32) Q: Did you get wet?
A': Wet? I was bloody soaking.
A'': Wet? Bloody soaking I was.

We see here that the focus constituent can be either rheme-initial or postverbal even in English, with the former position being associated with a more salient (focal) focus. In this case the focus is also contrastive.⁴⁵

The issue of the status of the **inverted topic** seems to be more complex than was intimated above. Although there seems to be little doubt that emphatic assertions may be thematically grounded, i.e. may have topics, just like regular assertions can, sometimes

emphatic assertions with inverted topics seem to be less like predicate-assertions and more like proposition-assertions of a type similar tothetic (ungrounded) assertions. This is not surprising, after all, since inverted topics behave just like (non-focus) assertion-internal complements and, like other such complements, they are not accented, and are typically quite accessible. (In flexible VO languages, such as Spanish, both clause types involve subject inversion, but keep in mind that inthetic sentences the subject is the focus, whereas in emphatic constructions the inverted subject is never the focus. I will return to this issue below and in Chapter 5.)

In (3.33) below we can see an example of such a sentence with an inverted subject in Spanish, one which also involves an overt dative argument in postverbal (assertion-tail) position.

- (3.33) Un libro le ha comprado mi padre a mi madre! (OVSD)
one book to.her he.has bought my father to my mother
“My father bought a book for my mother!”; “My father bought my mother a book!” (“Can you believe it?”)

Is the referent of the inverted subject the topic of this assertion? Or, in other words: Is this a predication about the subject’s referent the same way that the canonical version of this sentence, shown in (3.34), would be?

- (3.34) Mi padre le ha comprado un libro a mi madre! (SVOI)
my father to.her he.has bought one book to my mother
“My father bought a book for my mother”

It is not easy to tell whether the subject is the topic in this assertion. It would seem, however, that this type of emphatic assertion is not as much an assertion *about* a referent (the topic) as about the whole event. This impression may be related to the fact that these

sentences are disruptive and non-continuous, unlike normal (non-emphatic) statements, in which the grounding requirement is more crucial.

As to the actual position of the inverted subject, here it is immediately after the verb, but just as with other unfocused rhematic elements in Spanish rhemes, the order is quite free and it could be clause final, cf. *Un libro le ha comprado a mi madre mi padre!*. There doesn't seem to be any discernible difference between the two assertions. In other words, the position for inverted subjects is the assertion-tail of the clause.

We should keep in mind that there is no objectively ascertainable degree of emphasis that causes an assertion to display emphatic properties. In some genres and styles, for instance, focus fronting and topic inversion seem to be more common than in others. This happens, for instance, in certain Basque journalistic styles, as we will see in Chapter 4. What is interesting is that the more common these formal characteristics are the less marked, and thus the less (objectively) emphatic, they seem. In other words, how emphatic these reorderings are felt to be depend on their frequency, a fact which I will argue may play an important role in word order change.

Finally, I should say thatthetic assertions may also be emphatic. In other words, if the focus of athetic assertion (the subject) is very focal, athetic clause may also display properties of emphatic clauses, such as preverbal focus. This is what we find in Spanish, for example, where focus subjects are typically postverbal, as in (3.35)

- (3.35) *Vino mi tía.*
she.came my aunt
“My aunt came.

But in an emphaticthetic assertion, with a very focal focus, the subject focus may be preverbal, as in (3.36) below.

(3.36) ¡Mi tía vino!
my aunt she.came
“My aunt came.”

On the other hand, the full topicless assertion (the rheme) may be emphatic, such as when the ideas expressed in it are not particularly focal, which means that the assertion itself is focal and, as we saw earlier, the finite verb is treated as the focus, as in (3.37).

(3.37) ¡Está sonando el teléfono!
it.is sounding the telephone
“The phone's ringing.”

Here it is possible for the phone (the subject's referent) not to be the focus of the assertion if it is an accessible idea, or one that is easy to accommodate in context (phone's in general are quite accessible). This contrast between subject focus and rheme focus is merely coded accentually in Spanish (and it cannot be coded at all in English, since the subject cannot be inverted normally in this language). We will find later, however, that this distinction has very interesting repercussions for word order in Basque. It turns out that normalthetic assertions in Basque have SV order, since the subject must be placed in preverbal (focus) position. On the other hand, when the wholethetic assertion (the rheme) is focal, or emphatic, and the finite verb is treated as the focus, the subject inverts with the verb and is placed in the assertion-tail, much like a postposed topic (an 'antitopic'), thus neutralizing the form of the two types of assertions: comment-topic assertions with rheme-focus, and rheme-focusthetic assertions, both of which have VS order.

3.5.6 Default focus

One very interesting aspect of the focus role concerns the extent to which an element of the rheme is 'focused' by default and the extent to which the speaker has something to do with it. Discussions of focus and similar notions often seem to imply that the focused element is the one which is, objectively speaking, the least predictable and most newsworthy, noteworthy or informative element in the assertion. Others, such as Bolinger, however, have emphasized the speaker's subjective choice in such matters.

While remaining open to the possibility that it is the speaker, and not the context, which occasionally determines the choice of focus, there seems to be little doubt that in most cases it is the context that 'chooses' which element will be focused, and thus will receive the major accent of the rheme and be placed in focus position. We should remember that the choice in fact is quite limited, rhemes in context rarely containing more than one complement inside the main clause (as opposed to complements which are added in sentence tails). Furthermore there seem to be defaults which most of the time make the choice of focus a rather automatic one.

With marked (very salient) focus constructions, such as content questions, as well as answers to such questions, the focus element is always the question or answer phrase. In the content question construction, the construction, by definition, codes which element is the focus. With answers to questions the situation is very similar, since the open proposition in question is active (it has just been uttered) and the constituent which (co)responds to the question phrase is naturally the focus. In fact, typically only the focus 'fragment' is uttered, although it stands for the whole assertion.

There are other situations in which the choice of focus is quite straightforward. The most basic one is when there are no complements and thus the verb, or the whole rheme, is the focus, as in (3.38) below.

(3.38) He came.

When the idea expressed by a complement, or by the verb, is active it is typically not the focus either, as in examples (3.39a&b), unless it is contrastive, i.e. unless there is something else that makes it focal or informative. In (3.39a) the object is active. In (3.39b) it is the verb that is active.

(3.39) a. I saw a nice book at the bookstore, and I **bought** it.
b. I felt like singing when I got up, so I **sang** an aria.

On the other hand, when all the elements in the assertion are relatively discourse-new, or more or less equally newsworthy or informative, the determination of the focus element would seem to be somewhat more problematic.

When all the elements of the rheme are 'new' and equally unexpected, if the verb is a copula, or a stereotypical verb, such as *have*, *see*, *say*, *buy*, which are easy to accommodate, then the focused element is almost surely a complement of the verb and not the verb (the verbal idea) itself. The example we saw earlier with the predicate *buy a book* in an all-news context is of this type, for instance. This makes sense since the complement in these clauses would seem to be 'naturally' more focal.

In fact, it is quite possible that a major motivation behind the use of cognate complements (e.g. *sing a song* vs. *sing*), predicate decomposition (e.g. *give a kiss* vs. *kiss*), and telic path predicates (e.g. *go in* vs. *enter*), all cases in which a verb alone can express the same semantic meaning as a verb plus complement, is to make a complement,

representing a ‘tangible’, referential idea, instead of the verb, the focus of the assertion. Languages differ as to the extent that they exploit these three different ‘decomposition’ strategies. English and many other languages typically locate telic path predicates outside the verb, whereas the Romance languages do so much less frequently and less systematically, as we can see in (3.40) (cf. Talmy 1985, Aske 1989). Basque on the other hand makes extensive use of predicate decomposition (cf. Aske 1987).

- (3.40) a. He bent down.
b. Se inclinó

When the whole rheme is emphatic or contrastive, but no particular element inside it is more focal, then the whole rheme is focal and the finite verb acts as the focus element. I will refer to this as ‘rheme focus’, or as ‘polarity focus’, since it is often found in polarity questions and answers to such questions. We can see an example of such a rheme-focus assertion in (3.41b).

- (3.41) a. I was very bored,
b. so I bought a book!

Notice that here the object (*a book*) is new and in a regular assertion it would normally be the accented (focus) element. But this is not a regular assertion, this is an emphatic assertion in which the whole rheme is focal, and since the verb and the object are equally new, the accent falls on the finite verb, which stands for the rheme. Accenting the object in this context would almost imply that it was contrastive.⁴⁶ Notice too that in this context it is clear that the particular book itself is not important, rather the fact that the speaker bought a book, any book, is.

The difference between **verb focus**, in which the verbal idea is focal, and rheme or **polarity focus**, in which the whole rheme is, is neutralized in cases such as the ones

we have seen in which the verbal idea and the ‘finiteness’ are instantiated by the same word. In the case of periphrastic verbs in English, for example, the two focus types can be easily differentiated, as we can see in the pairs of assertions in (3.42).

- (3.42) a. I have bought it (verb focus) vs. I have bought it (rheme/polarity focus)
b. I was reading it (verb focus) vs. I was reading it (rheme/polarity focus)

I have mentioned that when both a verb and a complement are equally new, the tendency is for the focus accent to fall on the complement, at the expense of the verb. With **non-stereotypical verbs**, that is, those which are semantically more complex and harder to accommodate, the verb is intrinsically more focal and thus it is more likely to be the focus and thus bearer of the focal accent. This would include verbs such as *possess* (as opposed to *have*) and *whisper* (as opposed to *say*), as in the example in (3.43b).

- (3.43) a. I bent down
b. and whispered a secret in her ear.

It should be noted that whenever an element is chosen as the focus of an assertion which is not intrinsically (contextually) more focal than the other possible focus candidates in the assertion, this focus is automatically interpreted as being either contrastive or emphatic. In other words, in such situations it is automatically assumed that an element that is not felt to be the focus on its own merits, it must have had some salience added to it.

Finally, I should mention that complement focus is not the default in all types of assertions. **Imperatives** and desideratives, for instance, are similar to rheme-focus assertions in that in the default case, when the whole proposition is ‘new’, the accent, and

thus the focus, is on the verb and not the complement, even if the complement is new, and thus intrinsically focal, as in (3.44).

(3.44) Buy a book!

The complement could indeed be the focus element but then it would be interpreted as contrastive, as in (3.45B), unless the verbal idea was itself active, as in (3.46)

(3.45) A: I don't like this book!
B: Get another one.

(3.46) Q: What should I buy?
A: Buy a book.

This has important repercussions for the default constituent order in such constructions in languages, such as Basque, in which differences in focus constituent always have word order repercussions, as we will see in Chapter 7.

3.5.7 A typology of focus positions

In the previous sections we have seen that in English the focus of the assertion may come in different positions in the sentence. We also saw that occasionally a marked focus may be placed in a different position from where it would be placed in a canonical sentence, especially in the case of emphatic complement-focus. We have also seen some examples which show that in languages with greater word order flexibility, a focus constituent is much more likely to move from canonical position and to be found in a different position in the clause. What exactly are those positions?

The positions in which foci are found cross-linguistically, and in any one language, even a fixed word order language, are quite limited, given that in practice

asserted clauses (as opposed to whole sentences) rarely have more than one rheme-internal complement in addition to the verb. We saw that in statements in English, a subject focus is rheme-initial and preverbal. Also rheme-initial and preverbal are all question words which, in the default case, are the focus of content questions. I also mentioned that contrastive objects in Spanish and some dialects of English are rheme-initial and preverbal. In English focus is always marked by the assertion's main accent, but there is no single position which all focus constituents must occupy. But the positions which different elements occupy are the same positions which are focus positions in one language or another, as we will see.

In statements, focus elements other than the subject come after the verb in English. But what exactly is the position of a postverbal focus? All the examples we have seen so far involved a single complement and thus the issue didn't come up. But what if there is more than one postverbal complement? It seems that the primary position is right after the verb, that is, on the opposite side of the verb from the other major focus position, as in (3.47a&b).

- (3.47) a. I read a book yesterday.
b. I gave a book to her.

But the requirement that direct objects come next to the verb in English allows us to discover a third possibility. If the adverb in (3.47a) or the dative in (3.47b) were to be the focus, they could (seemingly) remain in place, as in (3.48a&b) (although if they were very focal they could also be fronted, cf. *Yesterday I read a book*, *To her I gave a book*, and the dative can be moved to postverbal position).

- (3.48) a. I read a book_(,) yesterday.
b. I gave a book_(,) to her.

But this is not just an example of *in situ* ‘focalization’. This is in fact an example of rheme-final focusing, with rising comma intonation right before the focus, and even extraposition to a separate intonation unit. This position seems to be the third and last possibility for placing focus constituents cross-linguistically. I will call it extraposed focus position. Although such rheme-final foci are not necessarily fully extraposed into full, independent intonation units, the clefting is significant enough so that typically there are two accents in the rheme, one on the rheme-initial verb and one on the focus constituent itself. Focus extraposition is typically associated with long and complex focus constituents, in which case the focus is indeed in a separate intonation unit, but it may be used for less complex foci under a variety of circumstances, as we will see.

An analysis of additional, more complex examples suggests that assertion-final focus is in reality, or at least in origin, a dislocation or extra-position ‘strategy’. That is, in these cases there may be a full intonational break before the focus, i.e. the focus may be ‘dislocated’ into the assertion tail, and then the verb receives a full accent as well. As I said, assertion-final focus is common with long and complex focus phrases, and in these cases the intonational break is quite obvious and the assertion is clearly split into two intonation units, as in (3.49) for example.

(3.49) I read yesterday, a book about the yearly migrations of penguins.

Such sentence-tail focus constituents are quite similar in their properties to extensions of the assertion proper which are placed in sentence-tail intonation units.

Dislocated foci in English are not found only when there are intervening elements between the verb and the focus, however. Focus dislocation seems to be a way of emphasizing the focus by separating it from the rest of the rheme, which may consist of

just the verb. Thus the simple sentence in (3.50) can be uttered with two accents and an intonation break, and even a pause for extra markedness.

(3.50) I didn't read a newspaper, I read, a book.

Notice that here the verb is accented, even though it is quite accessible. We can see the contrast between plain postverbal focus and assertion-final focus in the pair of examples in (3.51).

(3.51) a. I gave a book to her. (postverbal focus)
b. I gave her (,) a book. (extraposed focus)

In the assertion in (3.51a) the postverbal object receives the single accent of the rheme. In the assertion in (3.51b) the focus is still the same constituent, but here it is placed at the end of the rheme, with a more complex intonational structure.

I mentioned earlier that the rigidity of English word order would seem to be the reason why focus constituents are typically found in their 'canonical' position. We should not forget, however, the construction by which an accusative and a dative object may switch positions in English, namely 'dative shift'. Objects may take advantage of the dative-shift construction to alter the position of the focus constituent.

(3.52) a. She gave her mother the book (postverbal dative focus)
b. She gave her the book. (assertion-final accusative focus)

Thus in (3.52a), a dative focus is placed next to the verb, whereas in (3.52b) an accusative object is placed assertion-finally.

Non-arguments are also more flexible than arguments as to where they are placed when they are foci. Focus adverbials may be postverbal, if there is no object, assertion-

final, if there is one, or clause initial if it is very salient, as in (3.53) below, where it is a contrastive focus.

(3.53) Yesterday I read a book (not today).

Notice that, here too, the focus is not actually right before the verb, since the subject pronoun *I* separates it from the verb. In other words, this fronting would seem to be to clause initial and not to rheme-initial position. However, it is also true that the only subjects which can intervene between a fronted focus and the verb are light subject pronouns which are in fact unaccented and cliticized onto the verb and thus act very much like ‘agreement’ markers in other languages (cf. *?*Yesterday my mother read a book*).

A cross-linguistic review of focus constituent positioning reveals that the three possible focus positions mentioned so far fully exhaust the possibilities for positioning a focus constituent in other languages as well. The three positions are summarized in Table 3-9.

	Focus position	Code	Schema
1	Assertion-initial, preverbal (post-topic)	A1	(T) [<u>F</u> V ...]
2	Assertion-second, postverbal	A2	(T) [V <u>F</u> ...]
3	Assertion-final (dislocated)	AL	(T) [V ... _(s) <u>F</u>]

Table 3-9: The three possible cross-linguistic positions for a focus constituent in asserted constructions.

English, as we have seen, uses all three positions. What makes English unusual cross-linguistically, and this is no doubt associated with its unusually rigid word order, is the fact that the choice of location for a focus constituent is predetermined to such a large extent by its syntactic role. As we will see below, Spanish also allows all three positions,

but grammatical function has very little, if anything, to do with the choice of position except for in the case of verbs. And we'll see that Basque allows positions 1 and 3 (but not 2). Also, Basque, like other OV languages, is most consistent in placing all focus constituents in the very same position, independently of grammatical function, including verbs.

Preverbal (rheme-initial) and postverbal (rheme-second) positions are the two major positions for focus elements cross-linguistically, with the rheme-final alternative, which may perhaps better be seen as an extraposition alternative with lesser or greater reintegration, being a more marked alternative in perhaps all languages. Languages differ as to which of the two positions are available, which one is the main one, and in which contexts and in which constructions each of the positions is used.

In fact, I would go as far as to argue that this constitutes the major difference between so-called OV and VO languages and among different OV or VO languages themselves. I believe that the major word order difference among languages is related to the preferred position for a unmarked focus complement. Preverbal, rheme-initial focus position is the primary or only focus position in OV languages and postverbal focus position is the primary, though rarely the only one, in VO languages. Assertion-final or extraposed focus, on the other hand, seems to be an added possibility for all languages which are not rigidly verb-final.

In some languages a single position is used for all focus constituents. This seems to be found in more or less strictly verb-final languages, and perhaps also in strictly verb-initial languages. Verb-medial languages, on the other hand, tend to make use of both positions, as we have seen is the case in English and Spanish, though they may vary as to

how the two positions are exploited. One thing is fairly clear, however, and that is that when there is a choice between preverbal and postverbal position, preverbal position is reserved for the more focal foci, such as contrastive and emphatic ones. Less focal, new foci, on the other hand, especially when both the verb and the complement are relatively new, are placed in postverbal position.

Besides preverbal and postverbal position as the two default focus positions we have to contend with the ‘extraposed’ position. As I said before, this is an intonationally marked option, which might even derive from a more complex, two-part construction, that is from a dislocation construction which in certain contexts may have been reintegrated into a complex intonation unit

In (3.54) below we can see an example from Spanish in which what is basically the ‘same’ assertion, with a focus subject, may have the focus in any one of the three positions.

- (3.54) a. ¡Yo pago la cuenta!
I pay the bill
b. Pago yo la cuenta(!)
I pay I the bill
b. Pago la cuenta (,) yo.
I pay the bill I
“I’ll pay the bill.”

The focus constituent can go in preverbal and rheme-initial position, which is reserved for very salient foci, as in (3.54a). It can also go in postverbal position, the most common location for the least marked type of assertions, as in (3.54b). And finally it can go assertion-finally and dislocated (to a greater or lesser extent), as in (3.54c). All three positions would be equally available to an object focus constituent for instance.

3.5.8 Assertion final focus (focus extraposition)

The special rheme-final focus position, which we may also call ‘focus extraposition’, ‘focus dislocation’, ‘focus cleft’, or ‘delayed focus’, is used in a variety of contexts. Its use may be optional, in order to achieve a rhetorical effect, or it may be triggered by some other factor, such as the pragmatic characteristics of the focus element itself or of other parts of the assertion.

The first type of use for the assertion-final focus is what we could call **suspenseful punch-line delay** effect. Keeping the hearer’s attention suspended until the end of the assertion has a clear rhetorical effect of emphasizing the focus, of drawing the hearer’s attention and expectation towards it. In extreme cases the focused element can be placed in an intonation unit of its own, following the main clause’s intonation unit, as we can see in (3.55).

(3.55) My mother gave (to) my grandmother, a book.

Separating the two units in this case we find full ‘comma-intonation’ and even a pause, which, in writing, is sometimes represented with a comma as well. The intonation break may be less ‘severe’, however, as I have already said. (If it turns out that the source of assertion-final focus is full focus dislocation to sentence-tail position, reintegration of the two parts might be due to restructuring, a type of grammaticalization.)

The second context in which a focus constituent is often assertion-final is found when the **focus constituent is formally complex and ‘heavy’**. Thus, for example, an accusative object focus would seem to be more likely to appear extraposed (in a dative shift construction), the more complex it is, regardless of rhetorical effect, as we can see in the example in (3.56).

(3.56) My mother gave (to) my grandmother (,) the book that my uncle had recommended.

Of particular importance here is the case of completive clauses, complement assertions of evidential verbs, such as *see*, *say*, *seem*, etc. Such ‘complex’ sentential complements are typically the focus of the overall assertion, as well as assertions in their own right. For this reason this type of complement is often placed in its own intonation unit and the accentual pattern is that of an extraposed focus, even if they are not ‘moved over’ any other constituents, as in (3.57).

(3.57) My uncle said (to me the other day), that he didn’t want to read that book.

A third situation in which focus extraposition is found is when the assertion has **more than one highly focal element**, as in (3.58), where the dative is contrastive.

(3.58) I don’t know what he said to you, but my brother said to me, that he wasn’t going to come.

In other words, it would seem that here the ‘second focus’ element is treated as an ‘addition’ to (or ‘extension’ of) the assertion in sentence tail position.

A similar thing happens when the assertion contains a focal complement and a verb which is not a stereotypical, ‘basic level’ verb with low semantic content (cf., e.g., Lakoff 1987), but is rather a semantically complex, or non-stereotypical verb. This seems to indicate that stereotypical, basic level verbs are accommodated in simple assertions, whereas more specific verbs need a full accent for their assimilation into the discourse. In (3.59a&b) we see two examples of speech verbs with completive clauses. In the former the verb is the basic level, stereotypical speech verb, *say*, whereas in the

latter it is a more specialized speech verb, *shout*, in which the manner of speech is also part of the semantics of the verb.

- (3.59) a. I said (that) I would read a book.
b. I shouted (,) that I would read a book.

In other words, the complexity of the assertion, having more than one very focal element, may be a motivating factor for postposing part of it to a sentence-tail intonation unit as an assertion extension. If so, an assertion of this type could perhaps be seen as being split in two and having two foci. This would also help account for the often noted impossibility of extracting elements from complement clauses with manner of speech verbs, as we can see in (3.60) below.

- (3.60) ?# What did you shout (to your mother) that you would read?

The reason would be that the manner verb, being very focal and hard to accommodate, wants to be the focus of the assertion, whereas the construction says that the question word must be the focus.

Finally, the assertion-final focus strategy is also used in speech when **the speaker is looking for the appropriate verbalization** for the focus idea and thus pauses before it. This typically also entails the lengthening of the stressed syllable of the (accented) verb and a pause, and even an intonation break, before the focus element, as in example (3.61).

- (3.61) I rea=d .. a mystery book.

In a language like English this may seem like a mere assertion internal pause without structural consequences, but as we will see below, I have found that assertion-final focus

is common in spoken Basque in just such situations. In Basque the dichotomy is between preverbal (rheme-initial) focus and rheme-final focus and thus it is not a matter of mere pausing inside the rheme. I will look at this matter in some detail in Chapter 6.

To recapitulate, there seem to be four major contexts in which a focus complement is often placed in assertion-final position with an accent on both the assertion-initial verb and the assertion-final focus, as summarized in Table 3-10.

1	Suspenseful focus delay
2	Complex focus postposing
3	Semantically complex predicate in addition to focus complement
4	Focus verbalization hesitation

Table 3-10: Major reasons for postposing a focus complement to assertion-final position.

In other words, focus extraposition can be motivated by either (1) online restrictions on speech production ('delayed verbalization'), (2) principles of preferred information structure ('one focal idea at a time'), (3) the desire to facilitate processing ('complex constituents last'), and (4) rhetorical considerations ('suspenseful delay').

We can make one final observation with respect to early vs. delayed focus positioning, at least in those cases in which the language allows a positioning choice. It is obvious from iconic considerations that early positioning will be preferred when there is some urgency in expressing the focus, the main point of the assertion. That is why in emphatic or exclamatory assertions, when the focus is very salient, the focus is placed in initial position, even in languages where the main focus position is postverbal. In other words, placing the focus there, especially if that is not its canonical position, iconically codes its urgency and salience.

The advantage of a delayed focus, on the other hand, is that the context for the 'punch-line' has already been laid out by the rest of the clause, which means that rheme-final is also a very salient position. As long as there is a choice, we can say that 'early focus' is used in more 'emotive' contexts, whereas 'late focus' is used in more 'descriptive' and less 'emotive' contexts.

3.5.9 Focus typology in two VO languages: English vs. Spanish

In the preceding sections we have seen that the pragmatic functions topic and focus, and information structure more generally, has very important repercussions for constituent order. In English, with its rigid, grammatically based constituent order in declarative assertions (statements), the influence of information structure and focus positioning seems to be minimal on word order, though not non-existent. In other languages, the influence of pragmatic roles on word order are much stronger, however. In this section I would like to compare, as a preview to the analysis of focus structure of Basque, the differences between the information structure characteristics of two VO languages: English and Spanish.

As I said, like most VO languages, and perhaps all SVO languages, both English and Spanish make use all three focus positions, and when there is a choice, the implications are quite similar, with rheme-initial position being more marked (salient) than postverbal or even assertion-final position. On the other hand, whereas the position of a focus element in English depends strongly on its grammatical role, and primarily on whether it is a subject or not, in Spanish grammatical role hardly plays a role.

In Table 3-11 below we can see the different possibilities for the three different focus positions in English statements (declarative assertions).

Position	Type of focus
Preverbal (rheme-initial)	Subject; verb; polarity (finite verb); emphatic adverbials
Postverbal (rheme second)	Non-subject complements; adverbials (when there is no object)
Extraposited (rheme-last)	Non-subject complements; adverbials; dislocated complements.

Table 3-11: Exploitation of the 3 possible focus positions in English statements.

Other assertion constructions follow different rules. The content question construction in English typically places the question constituent, which is almost always the focus of the assertion, in rheme-initial position (and inverts the topic or would-be topic, i.e. the ‘subject’, with the auxiliary verb). In some special contexts, however, the focus constituent in a question may be in rheme-last position, as we can see in (3.62), so-called *in situ* focusing.

(3.62) She came when?

In Spanish, another so-called SVO language, the exploitation of these three different slots is very different, as we can see in Table 3-12, with examples for each possibility. Spanish allows complements of all types much greater access to all three potential focus positions than English does.

Position	Type of focus	Example
Rheme-initial	Verb, polarity, emphatic complements or modifiers (including subject)	(Mi <u>madre</u>) <u>vino</u> . ((<i>topic-</i>) <i>comment</i>) <i>My <u>mother</u> <u>came</u>.</i> <i>¡<u>Vino</u> mi madre! (<i>antitopic</i>)</i> <i>She <u>came</u> my mother!</i> <i>¡Mi <u>madre</u> vino!</i> <i>My <u>mother</u> came.</i> <i>¡El <u>viernes</u> vino!</i> <i>On <u>Friday</u> she came.</i> <i>¡Un <u>libro</u> ha leído!</i> <i>She read a <u>book</u>.</i>
Rheme-second	Non-emphatic complement or modifier (including the subject)	(Mi <u>madre</u>) vino <u>ayer</u> . <i>(My <u>mother</u>) she came <u>yesterday</u>.</i> (Mi <u>madre</u>) leyó un <u>libro</u> <u>ayer</u> . <i>(My <u>mother</u>) she read a <u>book</u> <u>yesterday</u>.</i> Vino mi <u>madre</u> . <i>My <u>mother</u> came.</i>
Rheme-last	Any complement or modifier (including the subject)	(Mi <u>madre</u>) vino _(,) <u>ayer</u> <i>(My <u>mother</u>) she <u>came</u> <u>yesterday</u>.</i> (Mi <u>madre</u>) leyó <u>ayer</u> _(,) un <u>libro</u> <i>(My <u>mother</u>) <u>yesterday</u> she read a <u>book</u>.</i> <u>Vino</u> <u>ayer</u> _(,) mi <u>madre</u> . (<i>contrast</i>) <i>My <u>mother</u> came <u>yesterday</u>.</i>

Table 3-12: Exploitation of the 3 possible focus positions in Spanish statements.

As we can see, these two languages are quite different as to where they place focus constituents in statements (whether emphatic or not). The most noticeable difference is perhaps the asymmetry found in English between subject and non-subject complements, which contrasts with the symmetry found in Spanish. The major inconsistency found in Spanish is between complement focus and verb/polarity focus, an inconsistency not found in 'OV' languages, since in these languages focus position is identical for all elements of the assertion.

In both English and Spanish the topic, and in particular the default topic, the subject, is always placed clause initially (or dislocated). The two languages differ, however, as to the position of subjects (would-be topics) which are not topics, that is.

primarily, those which are foci. In English they are placed assertion-initially, without changing its order with respect to the verb, so-called *in situ*. Intonationally, a topic subject (in a topic-comment assertion) and a focus subject are very different, but in terms of word order they are identical. Not so in Spanish, where focus subjects (as well as antitopics) are placed after the verb. In that position they are accented since they are foci. We can see the two versions of the samethetic assertion in (3.63) (in both cases the sentences are ambiguous between a contrastive focus and athetic, event interpretation).

- (3.63) a. My mother came.
b. Vino mi madre.

In the case of emphatic assertions, we also find that the subject is much more likely to invert to postverbal (non-focus) position than in English. Subject inversion in English, what is known as ‘subject extraposition’ (although such subjects are in the assertion-tail and unaccented), is used in a rather limited number of contexts in which emphatic assertions are made, as in examples (3.64a&b). (Note that despite the comma used in writing, and despite the intonational break, the ‘extraposed subject’ in these sentences is not accented and is not extraposed, i.e. it is not in a separate intonation unit.)

- (3.64) a. He’s a great guy, your brother!
b. It’s very scary, that movie!

In Spanish such inverted subjects are much more common, as I have already mentioned, and as we have seen in several examples above. The Spanish versions of the English examples in (3.64a&b) can be seen in (3.65a&b).

- (3.65) a. ¡Es fabuloso tu hermano!
b. ¡Es muy de miedo esa película!

In this type of emphatic assertion (with an extra-salient focus), the focus, of course, may also be rheme-initial, as I have already mentioned, something which codes the focus as being more focal, as we can see in (3.66) below.

- (3.66) a. ¡Fabuloso es tu hermano!
b. ¡Muy de terror es esa película!

These sentences may seem odd out of context, but they are perfectly common in conversational speech. Notice that extraposing the focus sometimes has a somewhat similar effect of enhancing the focus idea, though the effect is not emotive, but analytical, as in (3.67a&b).

- (3.67) a. Tu hermano es, ¡fabuloso!
b. Esa película es, ¡muy de miedo!

To conclude, I have suggested that (S)VO languages are those which place focus constituents primarily in postverbal position. This is true both of Spanish and English. Both languages also allow focus constituents with certain characteristics to appear in rheme-initial and rheme-final position. However, these two languages use these positions quite differently. Thus, whereas English is a rigid SVO language, with the subject being almost exclusively preverbal and the object almost exclusively (immediately) postverbal, Spanish belongs to the class of SVO languages with a 'flexible' subject (and object), i.e. one in which a subject may be placed in postverbal focus position or in postverbal (assertion-final) non-focus position. The object too may be placed in rheme-initial position in Spanish, in situations in which the focus is very focal, such as when it is contrastive or otherwise emphatic.

The only thing that prevents Spanish from being fully consistent in its focus placement is that when the verb or the 'polarity' are the focus they, unlike complements,

are always in rheme-initial position. It seems that a language can only be fully consistent if its major focus position is the rheme-initial preverbal position. That is the situation that we find in Basque and perhaps most other so-called OV languages, as we will see in the next section.

3.6 FOCUS STRUCTURE IN BASQUE

3.6.1 Main focus position in Basque

In Basque, like in OV languages in general, the main position for focus constituents is rheme-initial, preverbal position. This applies, for instance, to content questions, much like in English, but also to answers to such questions, for example. Like in Spanish, the position of a complement focus constituent is independent of its grammatical role. In (3.68a&b) we can see examples of an ergative subject question constituent and answer constituent.

- (3.68) a. Nork erosi-du liburu hau?
who:ERG buy:PFV-s/he.has.it book this
“Who bought this book?”
- b. Nik erosi-dut liburu hau.
I:ERG buy:PFV-I.have.it book this
“I have bought this book.”

The object goes in the assertion tail in these examples, but it could very well be topicalized and placed before the assertion proper (cf. *Liburu hau nork erosi-du?*, *Liburu hau nik erosi-dut*). A very similar pair of sentences can be seen in (3.69a&b), a content question and an answer to such a question, with the absolutive argument as focus.

- (3.69) a. *Zer erosi-du aitak?*
 what:abs buy:PFV-he.has.it father:DEF:ERG
 “What did father buy?”
- b. *Liburu bat erosi-du aitak.*
 book one buy:PFV-he.has.it father:ERG
 “Father bought a book.”

Here too the additional complement, the ergative argument is placed in the assertion tail, but, as with the object, it could just as well be placed outside the assertion proper (‘topicalized’) (cf. *Aitak zer erosi-du?*, *Aitak liburu bat erosi-du*).

As we can see, the focus constituent goes in assertion-initial pre-verbal position. Other arguments of the verb may precede the focus element, in which case they are outside the assertion proper, i.e. they are ‘topicalized’ or ‘settings’, or they may follow the verb, as unaccented, non-focus assertion-tail complements which are given or accommodated. This is equally true of questions or statements. In Basque, as we have seen, subject inversion is not required in questions. Subjects (would-be topics, or antitopics) are quite freely postverbal in Basque in assertions of all types when they are not foci. Although subject postposing in questions is common in Basque, it is not grammaticalized.

From these examples we can also discern an important fact about the Basque verb, namely that when the verb is periphrastic or analytic, the whole verb counts as a unit for determining preverbal position. This is true only of affirmative verbs, however. (As we will see in Chapter 7, negative clauses—along with one marked affirmative construction in some dialects—present a more complex situation.)⁴⁷ Thus, when the verb is affirmative and periphrastic, both elements, always in the order *V+Aux*, count as an information unit, and, in fact, as a phonological word as well. This also means that,

unlike with analytic verbs in English or Spanish, in Basque we do not have the option of accenting the finite (auxiliary) or the non-finite (main verb) parts of the verb in different focus situations (i.e. polarity vs. verb focus).

3.6.2 Unmarked focus in all-new assertions

In the examples in the preceding section, the foci are significantly more salient than the rest of the assertion, which is either accessible (in the case of the answer to a content question), or coded by the construction as being meant to be accommodated (in the case of the content question itself). The question is, what happens in other types of situations, such as the ones which I described above for English and Spanish, when all the elements of the assertion are relatively new?

When the assertion has a complement in addition to the verb and both are relatively new, i.e. when one is not significantly more salient than the other (as was the case with content questions and answers), in the unmarked case the complement is the focus, which is placed in rheme-initial position and receives the focus accent, as we can see in (3.70).⁴⁸

- (3.70) *Amak liburu bat erosi-du.*
mother:DEF:ERG book one buy:PFV-she.has.it
“Mother (has) bought a book”

This (topic-comment) statement could be either an out-of-the-blue statement or an answer to the question *What happened?* or *What did mother do?* (in addition to being a possible answer for the question *What did mother buy?*, in which the focus is more focal). The end result of this is that in these sentences the default position for the (single) complement of a predicate in Basque is preverbal position. In other words, in these cases the object is

the focus by default, just like in the other languages we have seen. This is fully consistent with claims that the default word order in Basque is SOV.

3.6.3 Verb focus

When all the elements in the predicate are new and unexpected, by default the focus is on a complement, especially if the verb is a semantically simple, basic level verb, which is easy to accommodate. Sometimes, however, the verb is more salient than any and all of the complements. This happens, for instance, when the **verb** is **contrastive** (which causes the whole predicate to be contrastive) or when the **complement** is quite accessible, or even active. In such cases, as we could have expected, the accent falls on the verb, which is also placed in assertion-initial position with any and all non-topicalized, unaccented complements following it. We can see an example of such a situation in (3.71a).

- (3.71) a. *Ikusi-nuen liburu hori,*
 see:PFV-I.had.it book that
 b. *baina ez-nuen erosi*
 but not I.had.it buy:PFV
 “I saw that book, but I didn't buy it”

In this clause, the object's referent would have to be quite accessible, and the verb (*see*) contrasts with the verb in the following clause (*buy*). Note that the object could conceivably, and optionally, be fronted or 'topicalized', in which case it would be outside the rheme, as in (3.72) below. This type of fronting would seem to be less restricted in Basque than it is in English. Also notice that it doesn't leave any 'resumptive' elements in the assertion proper beyond the already existing argument coding on the verbal inflection.

- (3.72) Liburu hori ikusi-nuen, ...
 book that see:PFV-I.had.it
 “I saw that book, ...” (? “That book I saw it”)

Notice that ‘verb focus’ assertions are indistinguishable from those in which the whole rheme is focal (as in emphatic, or ‘polarity’ assertions).

In the previous examples, the verb is analytic or periphrastic. When the verb is synthetic, however, the fact that the verb is in assertion-initial position and that the verb (or the whole rheme) is the focus must be marked by means of the affirmative particle *ba-* prefixed to the finite verb, as in example (3.73). This particle is related to the word *bai* “yes”.

- (3.73) Nik badaukat liburu hori.
 I:ERG EMPH:I.have.it book that
 “I have that book.” (also “I do have that book”)

It is hard to find clear-cut examples of such focal verbs, as opposed to rheme focus, however, since synthetic verbs are always very easy to accommodate and thus not very focal, as semantically more verbs are.

3.6.4 Rheme/polarity focus

In Basque, as in English and Spanish, when the polarity is the most salient element of the clause, it is the finite verb that is treated as the focused element. Polarity focus is found when all the elements of the rheme (and the whole proposition) is accessible, either because it has been mentioned, or because it has been suggested by the context. Polarity focus is also found (typically) in polarity questions and answers to such questions.

As I said in the previous section, since periphrastic verbs form a single phonological word, polarity focus clauses are indistinguishable from verb focus constructions in Basque. Thus, all the clauses we have seen since (3.71a) could also be used in polarity focus contexts. The English versions of these clauses would, in some contexts, call for *do*-support if the verb is non-periphrastic, or, in the case of periphrastic tenses, the accent would fall on the finite element of the verb.

In (3.74) we can see two examples of assertions with polarity focus. One is an out-of-the-blue polarity (*yes/no*) question, and the other is (the verbose version of) the reply to that question.

- (3.74) Q: *Erosi-zenuen liburu bat?*
 buy:PFV-you.had.it book one
 A: *Bai, erosi-nuen liburu bat.*
 yes, buy:PFV-I.had.it book one
 “Q: Did you buy a book? A: Yes, I bought a book” (= “I did buy a book”)

Notice that the complement is new, and thus, under normal circumstances we would expect it to be the focus of the assertion. The object *would* be the focus in a polarity question or answer such as these in a context in which the idea of buying (something) was in any way accessible or if the object was contrastive, as in the English example *Did you buy a book?* However, an out of the blue polarity question, unlike declarative assertions, displays a ‘preference’ for the finite verb, and not a focal complement, to act as the focus of the assertion, thus being placed in rheme-initial position and receiving the assertion’s main accent.

The polarity (the whole rheme) may be focal because it is contrastive, as is the case in denials, for instance, e.g. *Erosi-nuen liburu bat!* “I did buy a book” (= 3.74A). A context for such an polarity focus assertion could be someone else asserting, or

presupposing, the contrary proposition, e.g. *You didn't buy a book*, or *Why didn't you buy a book?*, or *Why didn't you buy anything?* This polarity focus assertion could also be uttered in a context in which the speaker wants to dispel the possible inference of the contrary state of affairs, as in the context: *I didn't buy a present for mother, but I did buy a book*. In all of these situations the added salience whole assertion overrides the salience of the new referent in the absolutive (object) role.

Polarity focus assertions, in Basque as in other languages, are often emphatic assertions, since very often the polarity is the focus because it is contrastive, or otherwise very salient (the proposition is typically active or very accessible). And, as we saw earlier, emphatic assertions are often accompanied by subject (would-be topic) inversion (antitopics), at least when the assertion is rather disruptive and non-continuous.

In continuous narrative would-be topic inversion it is actually quite rare, but that is presumably because in narrative emphatic assertions are rare, topics (subjects) are overwhelmingly elided, and inversion in an assertion is quite disruptive to the flow of discourse. In the Spoken Basque Corpus topic inversion was only common in identificational clauses with copula verbs, for instance (e.g. *He was the thief!*, *There was Chaplin!*). In lively dialogue, where a larger percentage of assertions are not continuous and which is likely to contain more emphatic assertions, subject inversion would seem, impressionistically speaking, to be more common. In some specialized contexts, such as some journalistic styles this type of inversion is quite common, as we will see in chapters 4 and 5.

3.6.5 Delayed or extraposed focus in Basque

As we have seen, unlike in English, the position of the focus constituent in Basque does not depend on its grammatical category or other grammatical characteristics of the focus. In that sense, Basque is much more like Spanish. On the other hand, Basque does not display the double option found in VO languages like Spanish, of placing focus elements in either preverbal or postverbal position depending (primarily) on the focality of the focus. In Basque, like in all other OV languages, the focus element goes in what seems to be the most salient of all focus positions, rheme-initial position, followed directly by the verb (unless, of course, the verb itself is the focus).

However, as has sometimes been noticed by Basque linguists, Basque does allow focal elements to be ‘postverbal’. What I believe has not been sufficiently appreciated in those studies is that postverbal focal elements in Basque do not have the characteristics of the most basic postverbal type of focus in VO languages, i.e. postverbal focus, but rather they display all the characteristics of the other type of focus, the assertion-final focus: (1) the intonation break and double accent found in such assertions and (2) the possibility of inserting unaccented unfocused rhematic complements between the verb and the focused element (although this happens rarely in real discourse). In (3.75) we can see an example of such a sentence.

- (3.75) Aitak erosi-dio (amari) _(s) liburu bat.
father:ERG buy:PFV-he.has.it.to.her (mother:DEF:DAT) book one
“Father bought (mother) a book.”

The contexts in which this focusing strategy is used, as opposed to the regular rheme-initial focusing strategy, are the same as the contexts in which it is used in English or Spanish, namely (1) to add suspense to (and thus highlight) the focused element, (2)

when both the verb (or another complement) and the extraposed focus are salient. (3) when the complement is complex, and (4) when the speaker is looking for the right verbalization of the focused idea.

Although the contexts for the use of focus extraposition/delay are not unlike those in other languages, the degree to which the option of extraposing the focus constituent is used in Basque seems to be subject to a lot of variation. In some contexts, however, focus extraposition is very common in both writing and speech, and equally common for all speakers, such as when the focus is a completive (finite complement) clause. In speech, as opposed to writing, the focus delay strategy is quite common when the speaker hesitates about the verbalization of the focus constituent. Some speakers, however, also use this strategy in cases where it doesn't seem to be justified, as commonly as it would be used, for instance, in a verb-medial language, such as Spanish.

3.6.6 Summary of focus positions in Basque

Earlier we saw that although both English and Spanish use the same three focus positions, the way in which they are used is very different in the two languages. In particular, we saw that focus position in English is highly dependent on the grammatical role of the focus element. In Spanish we saw that focus position was much more independent of grammatical role than in English, though there was a difference between complement focus (with unmarked postverbal position) and verb and polarity focus (with unmarked assertion-initial position).

Ignoring assertion-final focus positioning for the time being, we can see that in Basque, focus positioning is even more consistent than in Spanish, since there is but a single focus position: rheme-initial position, which applies to both complements and the

verb itself. This property is probably shared by most OV languages. Basque, however, differs from more rigid verb-final languages in that it does make use of the assertion-final position, in some cases to a significant extent. In Table 3-13 we can see a summary of the focus types in Basque.

Position	Type of focus	Example
Rheme-initial	Verb, polarity, any complement or modifier (including subject)	<i>A<u>ma</u> etorri-zen (subject focus)</i> <i><u>M</u>other came.</i> <i>(A<u>ma</u>k) lib<u>ur</u>ua irakurri-zen.</i> <i>(<u>M</u>other) she read the/a <u>book</u>.</i> <i>(A<u>ma</u>k) irakurri-zen liburua.</i> <i>(<u>M</u>other) she {<u>did read/read</u>} the/a <u>book</u>.</i> <i>E<u>t</u>orri-zen ama. (antitopic/emphaticthetic)</i> <i>She <u>came</u>, my mother.</i>
Rheme-final	Any complement or modifier	<i>(A<u>ma</u>k) irakurri-zen (_(s) liburua</i> <i>(Mother) she <u>read</u> the <u>book</u>.</i> <i>E<u>t</u>orri-zen (_(s) ama.</i> <i>My <u>mother</u> came.</i> <i>(= The one who came was my <u>mother</u>)</i>

Table 3-13: Exploitation of the 2 possible focus positions in Basque. The topic in parentheses may also be postposed (as an antitopic).

As we can see Basque is more consistent than either one of the VO languages we compared earlier: Spanish and English. Let us now look at some more examples of Basque information structure in more detail.

3.6.7 Further examples of focus structure types in Basque

3.6.7.1 Copula constructions with two overt complements

All the above considerations reveal that Basque constituent order is completely determined by the notions of topic and focus, and that the choice of topic and focus depends on a variety of gradient properties of the ideas in the assertion, as well as properties of the whole assertion itself. Common, or default, correlations between pragmatic functions with grammatical functions, that is, subject = topic and object = focus, account for the correlations between word order and grammatical role, but the underlying ordering principles rely solely on the pragmatic functions.

The flexibility of Basque word order, based on the relative flexibility of the choice of elements for topic and focus roles, can be nicely exemplified by means of the ordering variations of an assertion with a copula verb and non-verbal predicate. The variations depend first on whether the subject (the main topic candidate, with the most 'topical' semantic role) is the topic, the focus, the antitopic, or none of the above (an extra argument in the assertion-tail). The second parameter is the position of the focus constituent: either rheme-initial or rheme-final. In Table 3-14 we can see a number of the different resulting possibilities.

	Information structure	Basque	Gloss	English
1	T F V	Lapurra hori da	thief:DEF that she.is	The thief is that one
2		Hori lapurra da	that thief:DEF she.is	That one is the/a thief
3	T V F	Lapurra da (s) hori	thief:DEF she.is that	The thief is that one
4		Hori da (s) lapurra	that she.is thief:DEF	That one is a/the thief
5	T ba-V X	Hori ba-da lapurra	that yes-she.is thief:DEF	That one is a/the thief
6	T ₁ T ₂ ba-V	Hori lapurra ba-da	that thief:DEF yes-she.is	That one is a/the thief
7	F V T/X	Hori da lapurra!	that she.is thief:DEF	That one's the thief!; The thief is that one!
8		Lapurra da hori!	thief:DEF she.is that	That one's a thief!

Table 3-14: Possible focus structures for a copula proposition with two arguments.
Ba = polarity focus marker for synthetic verbs.

The assertions in (1) and (2) are both topic-comment rheme-initial focus constructions with a nominal predicate as focus. The only difference between the two has to do with the semantics/pragmatics of the two nominals. In the sentence in (1) the topic is a nominal expressing a role (i.e. *the thief*), an underspecified entity, and the focus indicates the 'identity' of that underspecified entity (i.e. *that one*).⁴⁹ Thus we could say that the focus and the assertion have an 'identificational function'. In the assertion in (2) further information is provided about an already identified referent. Informationally, however, the two assertions have the same structure. The second type was much more common in the corpus. Non-verbal predicates can also consist of an adjectival or adverbial phrase.⁵⁰

In (3) and (4) we see the versions of (1) and (2) with identical meaning and information structure, but with the focus constituent in rheme-final position. In this particular set of examples the choice between preverbal and final focus doesn't seem to have any obvious motivation. Further studies are needed to ascertain any correlations

which might exist between the different focus positions and rhetorical properties, for example. I will return to this issue in Chapter 6.

In (5) and (6) either the verb or the polarity (the whole assertion/rheme) is the focus of the assertion. Given the meaning and the aspect of the verb, verb focus is hardly a possibility in these sentences and the default interpretation is that these are assertions with polarity focus. Note that in (6) the predicate nominal (*lapurra* thief:DEF “the thief”), is topicalized, something which would barely be possible in English, but which is quite acceptable in Basque, as long as the idea it expresses is quite accessible, or even active (cf. *This guy, a thief, of course he is*).

In emphatic or ‘exclamatory’ versions of the assertions in (1) and (2), the topic, or would-be topic, of the assertion may be inverted and placed in assertion tail position, as we can see in (7) and (8). In fact, identificational sentences such as the one in (7) are probably the most common context in which topic inversion occurs. Pragmatically speaking, one could often argue that the referent of this postverbal NP is indeed the topic of the clause, what the assertion is about, especially with this type of identificational focus. On the other hand, sometimes it could also be argued that such emphatic assertions are not really predications, or at least not prototypical ones, but rather assertions about a state-of-affairs. The distinction between polarity focus with topic inversion and polarity focus is neutralized with predicates which do not require a topic, so thatthetic sentences with rheme/polarity focus may display subject inversion in Basque, just like in VO languages (cf. Chapters 4-7 for further discussion).

On occasion, something that looks like an antitopic construction seems to be used to introduce (or bring up) *expected*, or *given*, entities into the discourse. Thus this type of

assertion can be somewhat similar tothetic presentatives, which introduce *unexpected*. or *new* entities, with the introduced element in focus position (see Chapter 5). In this type of ‘antitopic construction’ the presented entity must be quite accessible, just as all assertion-tail elements must be, and unaccented. These pseudo-presentational clauses typically have a salient locative phrase in the role of focus. We can see an example of this type of sentence in (3.76) below, with an uninverted equivalent assertion in (3.77).

(3.76) Hor dago lapurra! XVS
 there she.is thief:DEF
 “There’s the thief.”

(3.77) Lapurra(,) hor dago. SXV
 thief:DEF there she.is
 “The thief is there”

The topic in (3.76) is postposed, given the high salience of the focus element and the ‘urgency’ associated with such an exclamatory identification (*There’s the thief!*). Although these two assertions seem to be equivalent in terms of the pragmatic roles of their elements, they do not seem to be equivalent in the sense that they are used in identical contexts, or for identical (rhetorical) purposes. As we can see, with a copula verb the Basque inversions almost always have corresponding inverted versions in English. In Basque too, as I said, clauses with copulas are by far the most common ones displaying this kind of inversion, perhaps because non-verbal predicates are so clearly focal (at least in comparison to the copula verb).

The extent to which the referent of the (non-focus) postverbal subject can be viewed as the topic, that is, the extent to which it is like a prototypical topic, depends somewhat on its accessibility. (As I already mentioned, unaccented assertion tail elements must always have accessible referents.) A sentence like the one in (3.76) could

be uttered while looking for the thief in question, i.e. when the thief is quite accessible, or even active in the minds of the hearer and the speaker. Interestingly, however, this construction is occasionally used with subjects with less accessible referents, but which can be contextually accommodated. Thus, walking in the woods, someone might say: *Look, there's a frog*, though perhaps this would be less likely if the speaker was at a library, where one does not expect to find frogs. In this case, however, it doesn't seem that the frog in question could be the topic of the assertion (even though it must be easy to accommodate to be in the assertion tail). The reason is that the phrase coding this entity could not be a pre-rhematic topic, as regular antitopics can (cf. * *A frog is there*).

The assertion in (3.76) above should not be confused with the rheme-final focus assertion in (3.78) below, for instance, in which the locative is a setting/ground element and the 'subject' is the extraposed focus. This assertion answers to the (explicit or implicit) question: *Who is there?*, whereas the one in (3.76) answers to the question: *Where is the thief?*

- (3.78) *H_{or} dago_(s) lapurra.*
 there she.is thief:DEF
 “There (there) is the/a thief”. (= “The one who's there is_(s) the thief”)

The assertion in (3.78) is equivalent to the rheme-initial focus assertion in (3.79).

- (3.79) *H_{or} lapurra dago.*
 there thief:DEF she.is
 “There is the thief”.

I should emphasize that in emphatic (or 'exclamatory'-like) assertions such as the one in (3.76) above, it is not just the (would-be) topic that is placed in the assertion tail, but also any other setting-like element the sentence might have, as we can see in (3.80).

- (3.80) (Eta) hor zegoen lapurra sartu ginenean!
 (and) there she.was thief:DEF enter:PFV we.were:WHEN
 “(And) there was the thief when we walked in.”

The reason for this is, of course, that the whole idea in these assertions with a very focal focus is to place the focus in initial position, thus postposing all setting elements with ideas that are quite accessible or easy to accommodate.

3.6.7.2 One-argument copula constructions

In the previous section we looked at two-argument copula constructions, all of which contained a topic or antitopic (and perhaps neither). In this section I will look at the possibilities found with a single argument copula construction, which in practice means existential constructions. A copula construction with a single overt complement, of course, may also be any one of those we saw in Table 3-14, in which the topic is covertly expressed, but I will not be discussing those cases here, since they do not differ in their properties from their counterparts with overt topics.

In Table 3-15 we can see the different ordering possibilities for an existential proposition with a copula verb and a complement. All the different sentences have the same (semantic) meaning.

	Information structure	Basque	Gloss	English
9	<u>F</u> V	Lap <u>ur</u> bat dago	thief:DEF it.is	There is a <u>thief</u>
10	V <u>F</u>	Dago(,) lap <u>ur</u> bat	it.is thief:DEF	There is a <u>thief</u>
11	T <u>ba-V</u>	Lapur bat <u>ba</u> -dago	Thief one yes-it.is	There <u>is</u> a thief
12	<u>Ba-V</u> T/X	Ba-dago lapur bat	it.is thief:DEF	There <u>is</u> a thief

Table 3-15: Possible focus structures for a copula proposition with one argument (existential).

Typically existential constructions arethetic, that is, they do not have a topic, cf. (9) and (10). On the other hand, a topic-comment configuration is also possible (cf. *Electrons exist*), cf. (11) and (12). But (12) can also be the structure of athetic assertion with rheme focus, in which the subject's referent is also somewhat topical, but not the topic, as we have already seen.

When the subject is the focus we find two ordering possibilities for the focus subject: rheme-initial and preverbal (9) or rheme-final (extraposed) (10). As in the case of the two-argument copula sentences, rheme-initial focus is the main possibility found in writing, but the rheme-final alternative is quite common in speech, as we shall see, at least for some speakers.

In addition, we find the possibility that the verb or the polarity is the focus of the assertion. (In this case, with a copula verb, polarity focus is the only possibility, since the verb is relatively content-less.) In such cases the nominal may be either the topic, as in (11), or either antitopic (inverted topic) or a non-topic/non-focus constituent, as in (12).

Interestingly, these two topic-comment and comment-topic alternatives to thethetic clauses are not as marked in Basque as we might have expected them to be. As we will see in Chapter 4, north-eastern Basque dialects often utilize a topic-comment structure for existentials where other dialects (and English, for that matter) would use athetic structure. The only requirement seems to be that the idea or referent in question be minimally accessible from the context (cf. Oyharçabal 1985a). And in other dialects the rheme-focus alternative for topic-less sentences (with $\underline{V}S$ order) is also quite common in speech. Whether the subject here is seen as an inverted topic or a non-topic, however, is not always clear.

In other words, a simple Basque existential assertion may have 4 possible forms, and the differences don't always depend on the identifiability properties of the nominal involved. In Table 3-16 below we can see the 4 possible versions of the sentence *There is a girl*.

Structure	Example	Properties	Comments
<u>F</u> V	<u>Neska</u> bat dago girl one she.is	Rheme-initial focus	Standard existential
V <u>F</u>	Dago <u>neska</u> bat she.is girl one	Extraposed focus	Common in speech
T <i>ba</i> -V	<u>Neska</u> bat <u>badago</u> girl one EMPH:she.is	Polarity focus; topicalized nominal	Accessible referent; presentative in north- eastern dialects
<i>ba</i> -V T/X	<u>badago</u> neska bat EMPH:she.is girl one	Polarity focus; subject in assertion tail (thetic or antitopic)	Stereotypical narrative beginning; common as <i>presentative</i> for some speakers

Table 3-16: Possible realizations of a simple existential assertion (*There is a girl*).

In the first two versions the nominal is the focus, which can be rheme-initial or rheme-final. In the last two versions, the polarity is the focus of the assertion and the nominal can be either an initial topic or it can be in the assertion tail (unaccented). In this latter case, the sentence can have two interpretations, since it could be either a comment-topic or a thetic sentence. We will return to this issue in Chapter 5.

3.6.7.3 Impersonal assertions

Finally, I would like to touch upon one other type of copula construction, namely so-called impersonal statements, such as *It's three o'clock*, *It's cold*. In English such clauses have an unaccented subject pronoun, said to be a *dummy*, non-referential pronoun. In Basque such statements never have an overt topic, and they seem to be thetic (topicless). In (3.81) we can see one such sentence with an rheme-initial complement

(there are 4 in the corpus) and in (3.82) one with a postverbal (rheme-final) complement (there are 2 in the corpus, both displaying ‘postverbal hesitation’) (*Other Codes = Impersonal*).

(3.81) 93C1A05
61 jateko ordua da,
eat:NOMIN:GEN hour:DEF it.is
“It’s time to eat”

(3.82) 93C2A09
94 eta= da= bazkariya juteko ordua.
and it.is lunch go:NOMIN:GEN hour:DEF
“and it’s time to go eat”

Although these sentences seemthetic, one might also argue that they are linked to the context in some vague way, and might be thought of as identifying something about the context. This is the reason why Bolinger (1977b) has argued that the *it* found in the English versions of these sentences is not really a syntactic ‘dummy’, but is actually referential, even if the referent is quite vague or non-specific. The same linking function would be performed in the Basque sentences by the covert verbal ‘agreement’ marker.

It is interesting to note in this respect that many Basque weather expressions, another type of so-called impersonal sentence, take ergative-absolutive verbal agreement, with the phrase expressing the atmospheric effect in the absolutive case. The ergative ‘argument’, however, is never expressed overtly.

3.6.8 Conclusion

As we have seen in this section, the word order of Basque sentences and asserted clauses is eminently a pragmatic order and not determined in any way by the grammatical roles of the constituents, for instance. In the sentence the default order is that of settings

followed by the assertion proper, followed by additions and extensions to the assertion. When the assertion is emphatic (with a very focal focus), however, setting elements can be placed at the end of the asserted clause, in the same intonation unit, and unaccented.

The order of elements inside the assertion proper depends on the notions topic and focus and the positions reserved for elements bearing these two pragmatic roles. Complications arise primarily from the possibility of having foci either in rheme-initial or rheme-final position (extraposed). Marked emphatic assertions, some of which may be in the process of becoming unmarked, are also responsible for some further complications.

In the following section I return to the topic of Basque word order and in particular the topic of basic word order. Among other things, I will review previous analyses of Basque word order in light of the conclusions which I have reached so far.

3.7 FOCUS STRUCTURE AND WORD ORDER IN BASQUE

3.7.1 Introduction

Inside the asserted clause, the order of constituents depends on pragmatic notions, such as the pragmatic roles topic and focus, and the pragmatic statuses of topicality and focality, that is, respectively, the degree of accessibility and the degree of informative salience of the different ideas expressed in the assertion. Although topicality and focality are in at least some ways two different perspectives on the same pragmatic continuum, the same is not true of the pragmatic functions topic and focus, which are roles in the information structure of assertions. The major different possible structural configurations for Basque sentences with an asserted clause can be seen in Table 3-17.

<p>1. Standard assertion w/ rheme-initial focus (T) F V (X)</p> <p>(Setting*) (Topic) Focus Verb (Assertion_Tail*) (Sentence_Tail*)</p>
<p>2. Standard assertion w/ extraposed focus (T) V (X) (s) F</p> <p>(Setting*) (Topic) Verb (Assertion_Tail*) Focus (Sentence_Tail*)</p>
<p>3. Emphatic assertion w/ postposed settings and (anti)topic: F V (X)</p> <p>Focus Verb (Assertion_Tail*) (Sentence_Tail*)</p>

Table 3-17: The three major possible focus structure configurations in Basque asserted sentences.

If the topic is covert it may (1) be dislocated (in a setting intonation unit), or (2) receive a secondary accent, or (3) be unaccented. If there is no topic, the would-be topic (subject) is the focus element by default (as inthetic sentences). If the subject is a contrastive focus, for example, there may or may not be a topic (some other argument).

The focus structure described here has very definite consequences for the word order found in Basque asserted clauses:

- In standard assertions, elements that are part of the assertion but are not the focus are always postverbal, i.e. are found in the assertion tail (unless they are topicalized and turned into settings).
- In standard assertions, if the verb or the polarity is the focus of the assertion, all non-topicalized elements of the assertion are postverbal.
- In emphatic assertions, setting elements, including the topic (or would-be topic) may be placed in the assertion's tail.

- In assertions with an extraposed focus, all non-setting elements will be postverbal, including the focus.

Thus we can see the complications involved in attempts to classify the word order of Basque, and most other languages, on the basis of grammatical relations. All these are obstacles to the claim that Basque is an SOV, or verb-final language. In the following section I will review how some other investigators have dealt with these issues in their attempt to define a basic word order for Basque on the basis of grammatical relations.

In Table 1-18 we can see a summary of the consequences for word order of difference choices of topic and focus, or of there being no topic at all. (All the examples may have occasionally an additional non-focus, non-topic complement in the rheme, after the verb. Brackets indicate rheme domain.)

	<i>Topic Status</i>	<i>Focus Status</i>	<i>Word Order</i>
1	S = topic	O = focus	S [<u>O</u> V]
		V/Rheme = focus	S [<u>V</u> O]
2	No topic	S = focus	[<u>S</u> V]
		V/Rheme = focus	[<u>V</u> S]
3	O = topic	S = focus	O [<u>S</u> V]
		V/Rheme = focus	O [<u>V</u> S]

Table 1-18: S = subject, O = object, X = any non topicalized complement of the verb; underline = focus.

In addition to these basic choices, there are additional ones, summarized in Table 1-19, for both transitive and intransitive clauses, which complicate the picture somewhat. Notice that intransitive antitopicalization has the same form as a topicless (thetic) sentence with polarity focus.

	<i>Variations</i>	<i>Explanation</i>	<i>Example</i>
1	Focus extraposition	The focus is placed in a separate intonation unit after the clause; or clause finally with intonation cleft.	S [V, O] [V, S]
2	Topicalization	Complements may become settings by moving to the pre-rheme area (only one can be undislocated).	S O [V]
3	Antitopicalization	Topic backing if focus is quite salient: formally and pragmatically (status-wise) like any other postverbal complement	[O V S] [V S]

Table 1-19: Pragmatic operations which complicate the basic, pragmatically motivated word order choices in Basque.

I turn now back to some previous studies of Basque word order and will try to shed some light on some of the questions raised in them from the perspective outlined in this chapter.

3.7.2 Osa 1990: Basic order and markedness revisited

As I mentioned earlier, Osa (1990), in his study of Basque word order from a functional perspective, implies that announcements, or all-news sentences, sentences which answer to the question *What happened?*, are topic-less, something which I have already argued against. He also argues that these are the most neutral and unmarked (and thus 'basic') sentences in natural language, since they display the least amount of presuppositionality, i.e. they are the ones which require the least amount of context.

The claim that these are the most basic and neutral sentences results from the bias of viewing sentences with minimal context, isolated sentences, as the most basic ones in language ('imagined sentences' with minimal context). But this is only *one* of the many possible contexts in which we find assertions, and it certainly is not the most common

one in actual language use. They are the most common ones in the context of linguistic elicitation, whether from someone else or from oneself (introspection).

The pragmatic properties of the referents of the complements in these out-of-context sentences are also in part responsible for their interpretation. Thus, for example, in most sentences, even those with minimal context, there is a relatively accessible topic involved. Notice, for instance, Osa's own examples in (3.83) and (3.84) (cf. Osa's examples 27-30, pp. 123-4), which have a personal name as the 'subject', thus representing an identifiable and familiar idea, which must be interpreted as the topic, among other reasons because the preverbal complements prevent an interpretation in which the 'subject' is inside the rheme.

(3.83) Sarak liburua irakurri-du. (SOV, "unmarked")
Sara:ERG book:DEF read:PFV-she.has.it
"Sara read the/a book"

(3.84) Sara etxean gelditu-da. (SXV, "unmarked")
Sara house:DEF:LOC stay:PFV-she.is
"Sara stayed home"

Osa argues that these two sentences with **Subject-Complement-Verb** (SCV) order are unmarked with respect to their alternatives with either Complement-Subject-Verb (CSV) order, cf. (3.85) and (3.86), or those with Subject-Verb-Complement (SVC), cf. (3.87) and (3.88), which are said to be marked.

(3.85) Liburua Sarak irakurri-du (OSV, "marked")
book:DEF Sara:ERG read:PFV-she.has.it
"Sara read the book"
= "The book was read by Sara"
= "It was Sara who read the book"

- (3.86) Etxean Sara gelditu-da (XSV, “marked”)
 house:DEF:LOC Sara stay:PFV-she.is
 “Sara stayed home”
 = “It was Sara who stayed at home”
- (3.87) Sarak irakurri-du liburua (SVO, “marked”)
 Sara:ERG read:PFV-she.has.it book:DEF
 “Sara read the/a book” (ambiguous: verb or polarity focus)
 OR: “Sara read the book.” (ambiguous:thetic or contrastive S)
- (3.88) Sara gelditu-da etxean (SVX, “marked”)
 Sara stay:PFV-she.is house:DEF:LOC
 “Sara stayed home” (ambiguous: verb or polarity focus)
 OR: “Sara stayed home” (ambiguous:thetic or contrastive S)

In the CSV sentences in (3.85) and (3.86) the C[omplement] is the topic/ground of the clause, being outside the assertion proper. The SVC clauses in (3.87) and (3.88), on the other hand, are ambiguous between four interpretations, although Osa does not explicitly mention this:

- The ‘**subject**’ is the **topic**, and the **verb** is the **focus** of the assertion: S V C.
- The ‘**subject**’ is the **topic**, and the **polarity** is the **focus** of the assertion: S V C.
- The ‘**subject**’ is the **topic**, and the **complement** is a rheme-final or **extraposed focus**: S V C.
- The ‘**subject**’ is the **focus** and the complement is rheme-internal, but not the focus: S V C.

I believe that Osa is essentially correct in saying that the order SCV is in some ways less marked than SVC order, in the sense that when the predicate contains a verb and a complement, it is typically the complement that is the focus, since it is usually the

most focal element of the assertion, and thus in placed in focus position. But to say that this order is basic or *the* unmarked order misses the point that order depends on the pragmatics of the assertion, i.e. on the context and the pragmatics of the ideas involved in the assertion. One could equally argue that assertions with an overt topic are marked, given their rarity. It is only our bias towards out-of-context assertions, or discourse initial assertions that makes us think that there is something unique about them.

The claim that SCV structures are unmarked, or basic, even when the Complement is the focus, also ignores the fact that focus extraposition is not as uncommon in spoken narrative as it is in writing, in fact for some speakers focus extraposition may even be unmarked in many contexts. This would lead to the equally erroneous conclusion that the basic order in these varieties of Basque is different from the basic order in other varieties.

The Subject-Verb-Complement order is ambiguous between a topic-comment structure interpretation and athetic structure interpretation, but only if we discount the information and intonation structure of the assertion. The SVC order is also ambiguous in many other languages, such as English for example. As to the frequency of CV vs. VC order, Osa himself admits elsewhere that SVC order is probably more common than SCV order (cf. Osa 1990:224).⁵¹ However, Osa attributes SVC order to ‘verb fronting’, which he argues is marked, although he also admits is relatively more common in speech than in writing, and in some writing styles than in others. Unfortunately, Osa doesn’t discuss what verb-fronting is and what the different contexts in which it is more common are, although he seems to be referring to those cases in which the verb (or the polarity) is more focal than the complement and thus is the focus of the assertion.

3.7.3 Marked and unmarked orders: A classical example

There is a well-known sentence in the Basque word order literature, taken from the first sentence of the Bible's parable of the prodigal son (Luke 15:11) (cf. Villasante 1956, 1980). In this example the whole proposition is new and, unlike in Osa's examples above, the referents involved are not identifiable. The sentence in question is *A (certain) man had two sons.*⁵²

Villasante (1956:16) noticed that different speakers use different orders for this sentence when telling this story. Subsequently, Mitxelena noticed that different Bible translators have produced different orders through the ages (Michelena 1978, 1987).⁵³ A survey of 9 translations give the following results: 4 SVO, 3 SOV and 2 OVS sentences. This is obviously an all-new proposition with no prior context. In (3.89) we can see the different versions of the sentences found by Mitxelena. As we can see, among the clauses which Mitxelena identified as SVO, there are two possible variants.

- (3.89) a. Gizon batek bi seme zituen. (SOV, 3 ea.)
man one:ERG two son he.had.them
b. Gizon batek zituen bi seme. (SVO, 2 ea.)
man one:ERG he.had.them two son
c. Gizon batek bazituen bi seme. (SVO, 2 ea.)
man one:ERG EMPH:he.had.them two son
d. Bi seme zituen gizon batek. (OVS, 2 ea.)
two son he.had.them man one:ERG
“A man had two children.”

Because of the lack of context and presumably because the referents involved are not identifiable, some investigators have assumed that this is athetic, topic-less or all-rheme assertion (cf., e.g., Goenaga 1980:206, Osa 1990:223). It doesn't seem to me, however, that such pragmatic structuring is possible in this case in which there are two strong focus

candidates.⁵⁴ The spoken sentence's intonational properties also suggest that this cannot be athetic sentence. The SOV sentence in (3.89a) in particular obviously cannot be thetic, since only one rhematic element may appear in preverbal position.

The SVO sentence in (3.89b) could conceivably have a focus subject, with the subject receiving the assertion's main accent, but the only possible interpretation of such a sentence would be one in which the focus subject is contrastive (cf. *It was a man (the one) who had two sons*). The normal intonation, the only one that is possible in this context, would have the ergative as the topic and the absolutive as an extraposed focus (cf. Goenaga 1980, Osa 1990:159).⁵⁵

The SVO sentence in (3.89c) cannot be thetic since the *ba* marker on the verb identifies the verb or the polarity as the focus and the preceding nominal must be the topic. Given the semantics of this verb it is clear that the only possibility is that the polarity is the focus and that this is an emphatic assertion with a given or accommodated object. Since the object's referent (*two sons*) is new in this context, the motivation for the use of this emphatic (polarity focus) construction is that this is a stereotypical story beginning, just like the ones we saw earlier in this chapter with a copula verb. (Notice that if the verb had been periphrastic, it would have lacked the formal marking indicating that the verb or the polarity was the focus of the assertion, and thus, the written sentence would have been ambiguous without intonational cues.)

The OVS sentence in (3.89d) is also not thetic even though there is no clause-initial topic. I believe that the subject here is an antitopic, which, again, is sanctioned by the traditionally emphatic nature of this story's initial assertion. Inverting this topic, of course, is a rather marked phenomenon in this context, given the fact that it is so

inaccessible. On the other hand, the stereotypical nature these new referents may be enough to warrant their being treated as accessible.

The SOV sentence in (3.89a), although it is in the minority, is the one which is said to be unmarked in Basque, with a clause-initial subject (topic) and a preverbal object (focus). The fact that only a minority of the sentences in this sample have this order is related, I believe, to the marked nature of this context, which in some styles calls for emphatic-like polarity focus (4 out of 9 authors). It is also very interesting to see that almost as many authors chose to extrapose the focus constituent as those who kept it in preverbal focus position.

The four sentences in (3.89) do not seem to exhaust all the possibilities. According to Rebuschi (1983), in the Lower Nafarroa-Lapurdi dialects this story could also begin with the VSO sentence in (3.90).

(3.90) Bazituen aita batek bi seme (VSO)
EMPH:he.had.them father one:ERG two son

Furthermore, he argues that in these dialects, the version in which the object is the focus, such as in (3.89a), is not possible for, as we will see in Chapter 4, with this verb (transitive *izan* “have”), just as with the copula verb (intransitive *izan* “be”), the absolutive NP in question would be interpreted as a predicate nominal, and not as an object (cf. Oihartzabal 1984). In other words, the object must be in assertion-tail position or must be topicalized.

I believe that this sentence in (3.90) is a variant of the one in (3.89c). In Rebuschi’s sentence the subject-topic of this emphatic (polarity-focus) assertion has been inverted, that is, it is an antitopic. This is not surprising, however, given that topic inversion is quite common in emphatic assertions, as we have seen. Topic inversion is

also seen in (3.89d), the version chosen by two out the nine authors. Again, topic inversion is unusual in a topic as inaccessible as this, but the context of this sentence may very well be what accounts for the suspension of the normal topicality requirements for inverted topics.

All these different examples are quite interesting in that they show that different varieties of the language, including different dialects and different styles, may chose somewhat different ways of making what would seem to be the very same assertion. Although all these different versions are well motivated by principles which hold throughout Basque, a speaker may chose one to apply one principle or another based on a institutionalized context-related preference or in order to achieve different rhetorical effects.

It is clear, however, that some of the options which are available in this marked context (the beginning sentence of a traditional story), are not available under normal circumstances and in most contexts. Some of the options, however, such as whether to extrapose the focus constituent, seem to be much more free, and the rhetorical differences associated with these options are much less marked, or noticeable. This doesn't necessarily mean that extraposition is never marked. Extraposition with a very marked intonation break and long pause *is* indeed marked, but extraposition in which the intonational cues are less conspicuous, seems to be much less marked.

The possibility that different speakers may display different preferences in rhetorical packaging of identical information has extremely important consequences for the relative markedness of the different focus structure possibilities and, consequently,

for the markedness of different constituent orderings. I turn now to some of these consequences.

3.7.4 Some speculations on markedness and word order change

As I just mentioned, it is very interesting that all those different focus structure possibilities are available in the context of the first sentence of the prodigal son story, especially since this is the type of context in which (supposedly) unmarked SOV (or Topic-Focus-Verb) sentences are said to be found. On the other hand, as we have seen, this particular context (a traditional story) seems to be quite marked and some of the variation seems to be related to the fact that emphatic assertions are often used in the first sentence of traditional stories.

On the other hand, some of the options which are available in less marked contexts, such as the option to extrapose a focus constituent, account for much of the normal word order variation that is found in the speech of any one speaker and in the different varieties of the language, as we will see in the following chapters. Options which are supposedly marked may become less marked through time, originally only in some contexts, and eventually in more and more contexts. Another option which accounts for the variation between (S)OV and (S)VO sentences in Basque, besides focus extraposition, is the one to make the verb the focus in situations in which the object is somewhat accessible (Osa's 'verb fronting').

As I suggested in Chapter 1, marked options may become relatively less marked, and eventually unmarked, due to language internal considerations, but they may become unmarked due to, at least in part, influence from other languages, through a process of convergence. The fact that Spanish has a different pattern of focus expression from

Basque (although they are partially overlapping), and the fact that most speakers of Basque are also speakers of Spanish, may have something to do with the variation we encounter in Basque and with changes which seem to be taking place. I will return to these issues in the following chapters, in which I will deal with the issues of topic and focus realization, or non-realization, in the case of topics, in greater detail.

NOTES TO CHAPTER 3

1. For some references about the history of this principle, see, e.g., Ochs 1988:106.
2. Speech Act code for sentence heads: *Head*. For sentence tails: *Tail**. For sentence bodies: *Body_Statement* vs. *Content_Question* vs. *Y/N_Question* vs. *Imperative*, etc. Other codes: *Regulatory*, *Indecipherable*, *Unclear*.
3. According to Chafe, within the sentence-body what determines the relative ordering of main and adverbial clause is ‘familiarity’, i.e. information that has not been already mentioned or evoked, or which is not inferable. Since adverbial clauses typically express unfamiliar information, they typically go last, following the theme-to-rheme ordering typical of English and other languages (about which “some American Indian languages are conspicuous exceptions”, Chafe 1984b:440). But familiarity doesn’t seem to be relevant to the placement of the adverbial clauses outside the sentence-body. Adverbial clauses are placed in such positions to fulfill specific functions, setting or closure, and not because of the particular informative status of these clauses (‘new’ or ‘given’).
4. According to Chafe (1984b), a “preposed free adverbial clause” “serves as a kind of ‘guidepost’ to information flow, signaling a path or orientation in terms of which the following information is to be understood. The same function is served by expressions like ‘however’, ‘anyway’, ‘for example’, ‘on the other hand’, and the like. A guidepost par excellence is ‘meanwhile, back at the ranch’. Preposed free adverbial clauses, I am suggesting, do the same kind of thing, providing a temporal, conditional, causal, or other such orientation for the information in the upcoming main clause.
“Guideposts, then come before the information to which they are guides. This is a different principle from the familiar to unfamiliar progression discussed earlier, though it seems to be another manifestation of a more general strategy of providing the contents of the frame” (Chafe 1984b:444-45).
5. Haiman used the notion ‘topic’ in a very general sense, and not in the narrower sense that I will use below. As Haiman shows, in some languages ‘topic’ clauses receive the same type of marking as nominal topics do, which is why he suggests, for instance, that ‘conditional clauses are topics’. See also Givón 1990:845-46, and Jacobsen 1992, about one conditional construction in Japanese which uses the *wa* marking of topics. All settings, including ‘assertion topics’, perform a ‘grounding’ function for the main assertion. It is thus not surprising that, at least some of them receive equivalent treatment, including positioning and even coding in some languages. Conditional clauses, for instance, provide a world-ground for the assertion. A similar thing is accomplished in some contexts by questions. Thus Haiman 1986 has noticed that some languages display “interchangeability or identity of conditional and interrogative markers” (Haiman 1986:215). In other words, in these languages the function of the “protasis” or sentence-head clause, and thus the grounding of the main clause, is performed by a question, and this commonality of function translates into common marking.

6. Unit Type code for sentence heads (Speech Act = *Head*): Nominal settings: *NewTop*; Finite clause: *Fin_Cl_Sub*; Non-finite clause: *Non_Fin_Cl*; others: *Non_Cl*.

7. Sentence body statement intonation units. Speech act = *Body_Statement**. Unit type: *Fin_Cl_Main* vs. *Fin_Cl_Sub* vs. *Non_Fin_Cl_Chain* vs. *Non_Fin_Cl* vs. *Elided_Verb_Cl* vs. *Elided_Aux_Cl*, vs. *Non_Cl_Bipartite* vs. *Fragment*, vs. *Non_Cl*. Polarity: *Aff* vs. *Neg*.

8. Mathesius' ideas were not new, they have a long history. They have had a very important influence on the study of discourse pragmatics in the US, where they are still often cited. However, as Sgall reminds us, if Halliday (1967a, 1967b, 1968) had not rescued Mathesius' ideas, we probably would have never heard of them. As to the history of these ideas, Sgall mentions that they "came from Germany to Bohemia, and from France to Germany. The importance of these issues, both for the system of language and for the process of communication, has been known since Weil (1844). The 'psychological subject and predicate' were included in the investigation of the general properties of language systems by Gabelentz (1868; 1891), and Wegener (1885:esp. 29, 31) was the first to state that stress is the main means of expression of this dichotomy. Marty (1897) introduced these debates to Prague. Jespersen (1924:145ff.), as well as Ammann (1928), and later Loepfe (1940), Boost (1956) and others studied the relevant phenomena under one or another heading without any real influence from Czech linguistics. In Russian studies too there is an important tradition of analysing the issues of 'logical stress', i.e. of the interplay of the placement of the intonation centre and the word order" (Sgall 1987:47). It should also be remembered that there was much variation in the beliefs about 'information structure' among Prague School linguists: "Czech linguists never were united in their views of what we would like to call the topic-focus articulation (TFA)..." (Sgall 1987:48). For a critical analysis of Prague School theories, see, e.g., Hopper 1985, Givón 1989:208ff, Chafe 1994b.

9. 'Old' or 'given' ideas are always old, both in the discourse and in the hearer's mind. The problem is with new information, which could be either *hearer-new* (Prince's *brand new*) or *hearer-old*, i.e. *identifiable* (Prince's *unused*) (see below).

10. Lambrecht for instance says that "the correlation between the cognitive category of identifiability and the grammatical category of definiteness is at best an imperfect one" and that we must "distinguish between a discrete (grammatical) and a non-discrete (cognitive) category (Lambrecht 1994:79). While the definite/indefinite contrast is in principle a matter of yes or no, identifiability is in principle a matter of degree" (Lambrecht 1994:84). Thus, it is better "not to think of definiteness as a universal linguistic category. What is presumably universal is the COGNITIVE category of IDENTIFIABILITY, which is imperfectly and non-universally matched by the grammatical category of definiteness" (Lambrecht 1994:87).

There is often a “coding difference between specific unidentifiable referents which are meant to become topics in a discourse and those which play only an ancillary narrative role” (Lambrecht 1994:83), cf. *this/a guy*.

11. Hawkins warns us of this problem, in particular in the context of attempting to match pragmatic categories to the formal category of word order, which is what concerns us here: “shared information and mutual knowledge between speaker and hearer represent a more subtle phenomenon than this and involve various inferencing procedures, different types of community knowledge about entities and their relationships to one another, situationally given information, etc., in addition to the simple previous mention of entities ... and there has been little systematic investigation of word order permutations in relation to these more fine-grained conceptions of ‘old’ and ‘new’ information. Second, there appear to be other pragmatic functions performed by word order variants which (if they have been discussed at all) have been lumped together under a simplifying Theme-Rheme rubric. Green (1980) discusses some of these other functions in English” (Hawkins 1986:47).

12. Thus the three boys who help the pear thief in the Pear Story film used in this study are consistently treated as a set and rarely as individuals (cf. *the three boys*). Sometimes the set is much more less specific since we do not know the exact composition of the set, as in the case of ‘the police’ often mentioned in retellings the Modern Times Chaplin film clip also used in this study. In this latter type of case noun phrases are typically marked as identifiable, i.e. as definite (cf. *the police*, *the bad guys*).

13. The term and concept of contextual ‘accommodation’ goes back to Lewis 1979, cf. Werth 1993. Propositions in discourse can be asserted, whether in main or dependent clauses, (semantically) presupposed to be true (e.g., *John used to smoke* in *John stopped smoking*), opaque (e.g. *John used to smoke a pipe* in *John denied smoking a pipe*), or backgrounded, i.e. used as required background for the assertion. The latter type can be either accessible (from having been asserted or mentioned earlier in the discourse) or they can be accommodated by their simple mention in a sentence head (setting) or some other specialized non-rhematic construction, such as *since* constructions in English.

14. Take the following example from the spoken corpus *ir*, which a new referent (*a rock*) is coded by means of the definite declension (literally *the rock*).

(i) 93C2A08

64 **arrixa** seuan parian,
stone:DEF it.was side/next:DEF:LOC
“there was the/a rock next to him,”

The rock mentioned in this assertion is not clearly accessible and only remotely could one see it as easy to infer from the fact that the context is a country road. In cases such as these one gets the impression that the definite declension simply marks that the expression is referential, that is, that the rock does exist.

15. Lambrecht has the following to say about this: “I would like to argue that accessibility (semi-activeness) of a referent, in particular accessibility of the ‘inferential’ or ‘situational’ type, does not have to entail that the accessible referent is somehow present, indirectly or peripherally, in the hearer’s consciousness, as Chafe seems to assume. Rather what seems to make a referent accessible is the fact that, due to the existence of certain semantic relations within an invoked schema, due to presence in the situational context, or due to other contextual factors, the referent is easier to conjure up in the addressee’s mind than a referent which is entirely inactive. I suggest, then that we think of cognitive accessibility as a POTENTIAL FOR ACTIVATION rather than as the STATE OF A REFERENT in a person’s mind” (Lambrecht 1994:104).

16. The cognitive similarities between entities and propositions is underscored by the fact that both of them may be referred to by identical pronominals (e.g. *it*). Pro-verbs, on the other hand, such as *do so*, seem to have much stricter requirements for their use, such as that the verbal idea be active in very specific contexts (constructions).

17. Cumming 1995 proposes three pragmatic (topicality) related parameters: accessibility, referentiality, and protagonism. The property of protagonism refers to the importance of a referent in the story as a whole.

18. According to Chafe, weight is a product of both activation cost and referential importance (Chafe 1994a:92).

19. In languages which tend to be verb-final and which prefer pre-verbal positioning of ‘focus’ constituents (see below), these non-referential nominals are overwhelmingly drawn to preverbal position (this agrees with the tendency to place nonreferential nominals in preverbal position in Hungarian). In Russian, which like Hungarian, doesn’t have morphological means of marking definiteness, on the other hand, it is postverbal subjects that are interpreted as indefinite or non-referential. So if a subject is definite, it must be preverbal (whether the sentence is ‘thetic’ or not, see below, Sasse 1995a:22). Although these differences between Russian and Hungarian seem arbitrary they have a perfectly good explanation in terms of the preferred placement of focus constituent, as we will see later on in this chapter. Hungarian has preverbal focus, whereas Russian has postverbal focus.

20. Details of their use may vary however. A language may or may not allow definite markers with proper nouns or with generic nouns for instance.

21. Givón’s rule, of course is not novel. Basque grammarian Villasante for instance says something very similar, and equally vague: “In general, the speaker tends to prepose the element which, for any reason, he or she considers most interesting” (Villasante 1980:237; my translation, J.A.) [“Por lo general el hablante tiende a anteponer el elemento que por cualquier motivo estima como más interesante”]. Besides this marked order, he also recognizes the importance of accent and intonation: “preposing of the element which is felt to be most important is not the only means we may utilize to express its greater interest. With the same ordering certain words certain words may be

emphasized by reinforcing their accent of intensity, raising their intonation or slowing down the tempo with which they are uttered. Any of these phonological means, or all of them at once, can be used for example in a sentence such as *Mi padre llegará esta noche* [“My father will arrive tonight”], so that either [the verb] *llegará* or [the temporal] *esta noche* will stand out, so that the listener will perceive them as having a greater expressive prominence” (Villasante 1980:238; my translation, J.A.) [“la anteposición del elemento que se siente como más importante no es el único medio que podemos emplear para significar su mayor interés. En igualdad de ordenación cabe realzar determinadas palabras reforzando su acento de intensidad, elevando su entonación o retardando el *tempo* con que se articulan. Cualquiera de estos medios fonológicos, o todos ellos conjuntamente, pueden hacer, por ejemplo, que en la oración «Mi padre llegará esta noche» se destaquen «llegará», o «esta noche», de modo que el oyente perciba su mayor relieve expresivo”]. Unfortunately, in the end Villasante seems to throw up his hands in despair when he says that “finding the appropriate ordering of constituents in each situation is a stylistic art, about taste and expression. It is not about the mechanical application of rules conceived ‘more geometrico’” (Villasante 1980:240; my translation, J.A.) [“Es todo un arte estilístico, de gusto y expresión, el buscar en cada caso la ordenación adecuada. No se trata de la aplicación mecánica de unas reglas concebidas «more geometrico»”].

22. Givón does not mention other factors which influence ‘topicality’ or accessibility other than recency of mention, i.e. textual topicality. Late mention of topical arguments, at least in languages which follow this pattern, may be related to what Chafe calls ‘antitopics’, though I believe that antitopics (as evidenced by the examples provided by Chafe and Lambrecht for instance) are a different sort of phenomenon, something which I will discuss later on in this chapter and in later chapters.

23. Finally Givón 1992b mentions ‘gap irrelevant devices’ which “code non-continuing topics whose antecedence is rather heterogenous [*sic*], probably a mix of situational, generic, and textural sources”, i.e. ‘definite nouns’ and ‘modified definite NPs’ (Givón 1992b:22).

24. Very focal elements (see next section) are not coded for their cognitive properties either, but merely as rhematically salient (*Rhem_Salient*). Other categories found in the Pragmatic coding sections are: Deictic (a location accessible from the context), Placeholder (e.g. ‘whatchamacallit’) and Rate (adverbials like *again*, and noun phrases like *two baskets*, *more*, as well as numerals).

25. Even among functionalists there tends to be a rather varied use of terminology without there always being a clear attempt to harmonize them. Thus Dryer 1992b, in a review of Payne 1990, complains that “[w]hin the field as a whole, research in word order pragmatics is currently rather diffuse, and the relationships between the claims of different approaches are often unclear. While the text-based nature of Payne’s approach clearly resembles the recent tradition associated with Givón, ed., 1983, the relationship between her notion of pragmatic markedness bears a clear resemblance to the notion of

focus assumed by many linguists and to the notion of newsworthiness proposed by Mithun (1987), but Payne does not discuss the relationship of her notion to these other sorts of notions” (Dryer 1992b:83-84).

Some linguists, of course, while recognizing that there is something interesting in there, would just as well dispense with all the confusion, and are interested only in “clear formal distinction[s]”. Unfortunately, most, probably the greatest amount, of language regularity is not of that kind. I believe Bickerton’s words are typical of such attitude: “If we approach the topic/focus difference from a functional viewpoint, we might quickly get bogged down in a discussion of new information versus old information, presupposition versus assertion, and other notoriously slippery issues. Fortunately, there is a clear formal distinction between the two” in Creoles, namely that (dislocated) topics have resumptive pronouns whereas focus ones don’t (Bickerton 1993:192).

26. In English for instance they involve dummy ‘subjects’, signaling the fact that they do not have regular subjects (i.e. topics).

27. One and the same predicate can sometimes display either one or the other information structure, even in the same context, which one is used being the speaker’s choice, that is, a matter of perspective or communicative strategy, though sometimes one or the other is given preference in the language and is thus conventionalized. This has been noted by Bolinger 1954 for instance and many others. Lambrecht for instance argues that “There are well-known exceptions to the unconscious tendency of language users to construe the subjects of isolated sentences as topics. These exceptions have to do with the lexical nature of certain predicates, the propositional content expressed by the sentence, and the semantic role of the subject argument.” e.g. “sentences like *Her father died* or *My car broke down* tend to be interpreted as event-reporting sentences, whose subjects will naturally be construed as being in focus. This is due to the fact that in the minds of speakers and hearers certain propositional contents are strongly associated with certain types of discourse contexts. ... It is also well-known that certain experiential predicates with strongly non-agentive subjects, as well as certain types of passive constructions, favor non-topic status of the subject” (Lambrecht 1994:133). This doesn’t mean that Lambrecht’s two examples cannot have a topic-comment interpretation in the right context though, that is, when the subjects can be interpreted as Topics, which could happen as long as there is the slightest background to connect *her father* or *my car* to the context (such as having been speaking of relatives or means of transportation, for instance), just enough to ground them to the discourse context as topics.

28. Others, of course, have adduced syntactic factors to account for these constructions, such that the subjects in question are ‘underlyingly’ objects. While it is true that the subjects in these constructions share more semantic similarities with typical objects than with typical subjects, I believe the correlations are indirect and not causal. Of course, I do not believe either that there is an underlying syntactic structure involved here at all.

29. For a discussion of the different ways in which topics have been defined, see, e.g., Reinhart 1982:4-5; Brown and Yule 1983.

30. Chafe also emphasizes this point: “Certainly speakers do not simply verbalize one focus of consciousness after another, with no concern for larger coherences. Producing language involves much more than adding new beads to a string. The larger coherences, of course, are crucial to the determination of activation cost, starting points, referential importance, identifiability, and similar features” (Chafe 1994a:120).

31. In discourse there is a rather natural type of sequencing of clauses, which are united by some factor which gives the sequence ‘thematic coherence’. Such thematic ‘paragraph’ sometimes goes by the name of clause chaining (cf. Longacre 1985, Givón 1990, 1991b:343-46). Such chains tend to display the following three factors, which contribute to this coherence: common topic (‘referential continuity’), tense-aspect-modal continuity, which is why some of the clauses are often non-finite, as well as ‘sequential unity’ (cf. Givón 1991c). Referential or thematic grounding (continuous or not) seems to be the most basic type of grounding and is based on the fact that sentences are about referents, which are the participants of events and the ‘bearers’ of states.

32. Lambrecht uses the term focus for that part of the assertion which is under the ‘illocutionary operator’. This differs from the use of this term in other traditions and in this study. In one tradition, focus has been seen as the complement of presupposition in certain types of constructions (Chomsky, Jackendoff), in particular clefts. In a not unrelated tradition, focus has been seen as a concept related to contrast as well as, more generally, new information. Although focus is highly correlated with contrast, not all contrastive elements are focus, as they may also be contrastive topics. Thus Ross’s contrastive focus construction is really a contrastive topic construction, as has been rightly pointed out before.

Givón refers to the scope of assertion: “With or without contrastive focus, a portion of most propositions (‘clauses’) tends to fall under **assertion scope**. Further, clauses in natural discourse tend to have, on the average, *one chunk* of asserted information per clause, while the rest of the information is not asserted” (Givón 1990:701). Then, “the chunk of information that falls under assertion scope is not necessarily under contrastive focus. However, if a constituent is under contrastive focus, it also falls under the scope of assertion” (Givón 1990:702).

33. As is well-known, languages with word order freedom are less likely to have such constructions, since they can freely turn an overt argument or complement into the ground by fronting it.

34. For speculations as to how these constructions might come to be see, e.g., Givón 1979, Haiman 1991.

35. Numerous studies have identified the groundedness requirements of subjects. A modifier containing a readily accessible referent is usually enough to convey groundedness to a non-accessible referent (cf., e.g., Ward 1988 and references therein).

36. As Bresnan notes, “[t]he term LOCATIVE will be used to subsume a broad range of spatial locations, paths, or directions, and their extensions to some temporal and abstract locative domains,” including VPs containing a LOCATIVE (e.g. *Crashing through the woods came a boar, Coiled on the floor lay a ...*) (Bresnan 1994:75).

37. Bresnan 1994, in her analysis of these constructions, discusses the interest in this construction in the transformational literature: “What makes locative inversion particularly revealing of the architecture of UG is its mismatches of role, function, and category. Throughout the history of generative grammar, roles and inner relations have been represented in the vocabulary of ‘outer’ structures (syntactic phrase structure categories and configurations), and the correspondences between the levels have been presented as transformational (phrase-structure dependent) operations. Within this general framework, locative inversion has seemed to defy analysis. ...), the literature on locative inversion is filled with conflicting proposals that the postposed subject is a subject, a demoted subject, or an object, and that the preposed locative is a topic, a subject, and adverbial adjunct, or a pseudo-subject. The recent generative literature on locative inversion in English continues this divergence, with the inverted subject treated as an unaccusative object in Coopmans 1989 and Levin 1991, a subject in Rochemont & Culicover 1990, a small clause complement subject in Hoekstra & Mulder 1990, and a demoted subject in Levin & Rappaport Hovav [1995]” (Bresnan 1994:74).

A particularly interesting construction in this regard is the English *there*-construction. In this construction, the erstwhile locative *there* has been claimed to be a subject in English, since it has some of the formal characteristics of subjects (such as subject-verb inversion). However, it seems to me that just because it (sometimes) quacks like a duck, there is no need to rush to the conclusion that it is a duck.

38. As I mentioned above, Kuroda credits Franz Brentano and Anton Marty for the claim that natural language actually has both types of what we could call information structures: thematically grounded ones (what they call *categorical judgments/sentences*) and thematically ungrounded ones (what they call *thetic judgments/sentences*).

39. In both cases English grammar demands a ‘dummy’ ground, the dummy, or non-specific locative ground *there* in the former and the dummy thematic ground *it* in the latter. Although these grounds do not seem to be referential in modern English, they must have been real grounds at one time.

40. In Japanese, a strictly verb final language, which does not allow subject inversion, on the other hand, thematically-grounded and non-thematically-grounded sentences are distinguished by the marking on the noun phrase. The *wa* suffix is employed for subject-topics of thematic ground sentences and the *ga* suffix is used for non-thematically-grounded sentences. Interestingly the ‘topic’ suffix *wa* is used not only for referential,

but also for other types of grounds, thus hinting at the underlying informational unity of all these elements.

41. Lambrecht 1994 has claimed that in these cases what is happening is that the subject is receiving a treatment analogous to the treatment of narrow focus subjects. I will argue that this is not the case and that the reason whythetic subjects are accented is that they are assertion-initial and thatthetic sentences, which are always all-rheme sentences, tend to receive a assertion-initial accent, in English as in other languages. More on this below.

42. According to Sasse: “Events are not prototypical predicates to the same extent [as entities are prototypical predication bases] but may become predicates in the presence of an independent entity with which they are linked. This entity is normally the most autonomous—that is, event-independent—member of their typical sets of participants” (Sasse 1987:565).

43. It seems that the larger the number and the strength of characteristics associated with transitivity (cf. Hopper and Thompson 1980), the less likely a predicate will fit into athetic informational frame.

44. The focus phenomenon is independent of the information structure of the assertion. That is, an assertion has a focus whether it has a topic or not (isthetic). Lambrecht has argued that in sentences with contrastive subject focus, such as the one in example (3.20), the open proposition, which is ‘presupposed’ (i.e. what I call accessible), is what links the new proposition to the discourse context. I do not believe that that is the correct analysis for these sentences. I believe that it is topics, i.e. referents, that ‘ground’ asserted propositions to the discourse model (the ‘universe of discourse’), and not propositions. (Such referents may, of course, be nominalized propositions, such as when a subject is sentential, e.g. *His coming surprised me.*)

45. Hannay argues that “such formal differences” should be analyzed “in terms of message content rather than in terms of stylistic choice”, which, if I understand it correctly, I think is mistaken, cf. Hannay 1991:131.

46. Accenting the object in this emphatic assertion would imply that there was a significant reason for accenting the object and not the verb, and a good reason, if not the only one, in this context would be if it was contrastive, although the speaker could conceivably choose to place the emphasis on the object to make it more salient.

47. As we will see, in some north-eastern dialects, however, in the context of a very marked (salient) focus, it is possible for the analytic verb to be split up and for the focus complement to be placed before the finite verb, with the non-finite verb being postposed along with any other (non-topicalized) complements, cf., e.g., Euskaltzaindia 1987b:491, ex. 17, =1993:436, ex. 42.

48. If the assertion contains only a verb, of course, the verb is, trivially, the focus of the assertion. It is in assertion-initial position (and assertion-final position as well, for that matter) and it bears the assertion’s accent.

49. Other Codes = *Identificational*. An identificational focus may be more or less contrastive, depending on the number or actual possibilities. There are 26 identificational main copula sentences in the Spoken Basque Corpus. Of these, 5 are negative ‘dis-identificational’ sentences, which actually deny an identification (e.g. *That man wasn't the thief*) and an additional 14 sentences are questionable identificational sentences. Of the 24 identificational copula sentences among main clauses in the data, 5 were negative.

50. Other Codes = *Non_Verbal_Pred*. In predicational copula sentences the focus is a non-verbal predicate. In the Spoken Basque Corpus there are 125 certain predicational copula sentences, plus another 14 which are questionable.

51. Osa says clearly that SVO order, and SVC[omplement] order in general, is very likely the most common order in speech. According to him, “there is no doubt that, although it is farthest from the canonical order, [SVO order] is also closest to the actual language use of the common speaker” (Osa 1990:224; my translation, J.A.) [“Dudarik ez dago, kanonetarik urrutiago bai baina hiztun xeheen erabileratik askoz ere hurbilago dabilela ordena hori [S.A.O]”]. Osa attributes this higher frequency to the extensive use of optional “verb fronting” (“aditza aitzinera”) in speech, particularly in spoken narrative, though also found in some written narrative styles (such as in Agirre’s bible translation).

52. As Mítxelena adds in a footnote to Villasante 1980:25, the original Latin and Greek versions were SVO: “the Latin version has *homo quidam habuit duos filios*, calqued from the Greek *ánthrōpos tís eĩkhe duo huious* [sic], which has itself been imitated literally in Germanic, Romance, and Slavic languages” (Mítxelena’s footnote in Villasante 1980:25; my translation, J.A.) [“el latín dice *homo quidam habuit duos filios*, calcado del griego *ánthrōpos tís eĩkhe duo huious*, que ha sido a su vez imitado literalmente en lenguas germánicas, románicas y eslavas”]. In other words, Mítxelena seems to think that the SVO order in the translations might be influenced by the SVO order in the original version(s), cf. also Michelena 1987:469.

53. Mítxelena notices that a certain translator, Lardizabal 1855, is more likely to produce verb-final sentences, whereas a contemporary from the same general geographical area is more likely to produce (S)VO clauses. For more examples see Euskaltzaindia 1987a:43-44.

54. For Goenaga 1980, in this context the whole sentence must be rheme, since the whole sentence is new. Although a thetic sentence with these characteristics is perhaps possible (cf. the verb-initial versions below), it is not the most common.

55. Goenaga (1980) very perceptively notices that in this sentence there is an intonation break between the verb and the object. This is indeed a clause with an extraposed focus constituent and thus two main accents on the predicate, one on the assertion-initial verb and one on the assertion-final focus.

Chapter 4

Settings, topics, and antitopics

4.1 THE NATURE AND VARIATION OF THE TOPIC NOTION

4.1.1 Introduction

In this and the following chapters I will discuss, motivate, and exemplify the functional categories topic and focus as they express themselves in assertions of different types in natural language. Chapter 4 deals with the topic role. Chapter 5 is about assertions which either do not have topics, in particularthetic sentences, as well as assertions with inverted topics. Chapter 6 concentrates on the focus role and the pragmatic notion of focality, or rhematic salience. Finally, Chapter 7 looks in some detail at some specialized types of assertions, such as negative statements, questions, commands, and emphatic assertions in general in which the focus is significantly more focal than other elements in the assertion.

The notion of topic, by that or any other name, such as ‘subject’ or ‘psychological subject’, is quite old, as is the division of the sentence between that topic and the rest of the sentence, i.e. the predicate, rheme, comment, VP. As I said in Chapter 3, the vast majority of asserted clauses have a topic and a rheme, the former of which is expressed covertly more often than not, at least in spoken narrative. Overt topics are perhaps more

common in other genres, as they certainly are in sentences out of context, which is why most grammarians have seen sentences with overt topics (and objects) as the most 'basic' type of sentence, even in languages in which covert expression of topics as light or incorporated pronouns, 'agreement', or even 'zero', is allowed and even statistically more common.

The strong correlation between the pragmatic notion topic and the grammatical notion subject has long been recognized. However, the fact that some grammatical subjects do not always act like topics and the fact that other arguments (non-subjects) sometimes do act like topics, have led to the realization that the two categories, the pragmatic one and grammatical one, must be kept separate since they belong to separate realms, albeit highly interconnected ones. Thus the pragmatic notion 'topic', by that or any other name, such as 'psychological subject', was developed.¹

There has been much confusion in the literature about the actual meaning and formal realization of the topic function. This is related, no doubt, to the fact that the formal and non-formal characteristics of topics depend on other notions such as the pragmatic status of the referents and the information structure of each assertion. The formal and semantic similarities between topics and other grounding (setting) elements, as well as the interaction between the pragmatic role and the more or less grammaticalized counterparts, have all been the source of confusion and disagreement in the literature.

As I argued in Chapter 3, I view a topic as a special type of setting or ground for the assertion, one which is both inside the asserted clause, as a core argument, as well as outside it, one which links the assertion to the context and which can be seen as what the

assertion is about. We also saw that not all assertions have topics, however, and not all topics have an identical realization. Thus, for instance,thetic assertions are topicless and the topic associated with emphatic assertions often has different pragmatic and formal properties. Furthermore a topic in a continuous assertion, for example, is typically coded minimally (it can be ‘understood’ or ‘elided’, or coded by ‘agreement’ or an unaccented pronoun), or, if inaccessible (or unexpected) it can be dislocated into its own intonation unit, or something in between, depending on the cognitive characteristics of the topic idea in the discourse model.

Another major issue, as I mentioned, is that of the relationship between the pragmatic notion topic and the grammatical or formal notion subject. I believe, along with Li and Thompson (1976a), that subjects are grammaticalized topics, or would-be topics, the argument of the verb which is the default candidate for the topic role, given a number of inherent topicality related semantic considerations. Languages differ, however, as to how they extend this grammatical notion to include rather un-topic-like elements, with some languages, such as English, more willing than others to have non-topic or low topicality topic subjects.

4.1.2 Topic vs. subject prominence

Basque seems to be what Li and Thompson (1976a) call a ‘topic prominent language’.² In other words, in Basque most sentences are best described as having a topic-comment configuration, rather than a subject-predicate one. In fact there is no formal category subject from a formal (morphological or syntactic) viewpoint in Basque, which is a (morphologically) ergative language without any true splits. This doesn’t

mean that predicates do not have an argument which is the default candidate for becoming the assertion's topic, or a default and single candidate to be the controlled argument of reduced clauses. If that is what is meant by subject, then certainly Basque does have them. However, these behavioral properties have more to do with the intrinsic topicality of the different semantic roles than with any grammatical (syntactic or morphological) aspect of the language.³

The grammaticalization of the pragmatic topic category, or perhaps rather the *default topic* category, into a formal category subject, morphologically and/or positionally coded, in languages like English has as a major consequence the fact that the topic must express itself typically in the subject role. This often leads to the development of constructions for promoting non-subject topics to subjects, which is for instance the major (though not the only) use of the passive construction in English. A high degree of grammaticalization for this category also results in the types of semantic roles which are allowed to appear as subjects to be greater than in languages in which topicality is less directly connected to grammatical role (cf. Fillmore 1968).⁴

A consequence of having an obligatory formal category subject is that subjects which are not topics, such as inthetic sentences, look very much like default topics, as is the case in English, even though there are always some differences, either intonational or word order differences, which distinguish subjects which are topics from those which are not.

Another characteristic of strongly subject-prominent languages, which seem to be a small minority of the languages of the world, and among which English is curiously

found, must have an overt subject constituent in all clauses, including those which when asserted are *thetic*, such as the impersonal clauses mentioned in Chapter 3.⁵

Basque, unlike subject-prominent languages, (1) has no morphological or syntactic (word-order defined) category subject, (2) has no voice constructions to speak of (cf. Chapter 2), (3) codes ('coreferences') all three major types of arguments on the verb, without privileging any one of them, and (4) displays great freedom of 'topicalization' (see below). Thus, Basque seems to qualify as a topic-prominent language. In addition, as we shall see, Basque has a tendency to treat certain elements as topics which other languages might treat as assertion internal complements, either focus or non-focus complements.

4.1.3 Inherent topicality and topics

As is well known, semantic hierarchies can be established which, in subject dominant languages, determine which argument of the clause will be the subject by default (cf., e.g., Fillmore 1968). Such a hierarchy seems to involve the intrinsic, or typical relative topicality of the different semantic roles. Thus, for instance, 'actor'-like, or 'agentive', arguments are more intrinsically topical than 'undergoer'-like, or 'patient-like', ones (cf., e.g., Foley and Van Valin 1984, Van Valin 1993b). For arguments with lower roles in the hierarchy to become topics specialized grammatical construction may be used in some languages, although clausal fronting and sentential fronting (left-dislocation) are also two possible ways in which arguments with certain pragmatic characteristics can become assertion external settings. In topic prominent languages that may the main topic promoting strategy.

Constructions which promote arguments to subject are not all identical crosslinguistically, however, and they may betray their historical sources, or even have acquired functions which are not topic related. Thus the standard English passive is derived from an adjectival (stative) construction, and we still find a high correlation between this construction and perfective aspect. Other passive constructions (such as the English *get* passive) can only be used when the object/patient is affected by the action in a certain way. Also, a passive subject need not be the topic of the assertion at all, since it may even be the focus in some contexts (e.g. *The door was opened, not the window*).

Still, the correlation between the English passive and promotion to topic is quite strong. Thus we find that this construction is often used when what would be the 'subject' in an active sentence (the 'actor') is actually less topical than what would be the 'object' (the 'undergoer'), such as when it is non-specific (e.g. *I was arrested (by a cop)*) or less important (e.g. *A child was run over by a car*). Fronting, on the other hand, is associated with topics which are contrastive and accessible (e.g. *That book I don't like*). Left-dislocation is associated with topics which are less accessible and thus need to be 'set up' (i.e. made into settings or grounds) in their own intonation unit (e.g. *A book I wouldn't buy; That book, I wouldn't buy (it)*).

In addition to semantic role hierarchies such as those investigated by Fillmore (1968) and others, other inherent-topicality hierarchies have been proposed by linguists, Silverstein's (1976) hierarchy. Silverstein suggested that certain types of referents are more inherently topical than others, so that, for example, all things being equal, humans are more inherently 'topical' than, say, rocks, i.e. they are—under normal

circumstances—of more interest to humans, more easily remembered, more noticeable, and more often talked about.

Similar to this is the concept of empathy hierarchy, discussed by Kuno (1976, 1987), who argued that some entities are easier for humans to identify with and more likely to be talked about, i.e. to use as topics, than others. Thus friends and relatives outrank strangers or inanimate objects in a Kuno-style empathy hierarchy. It would seem that the higher an element is in these inherent-topicality hierarchies the more likely it is that they will be used as assertion topics (and conversation topics, for that matter).

Although the choice of topic in topic-prominent languages is more ‘fluid’ and less dependent on grammatical roles and constructions than in subject-prominent languages, the fact remains that the most common choice of topic in the former is typically identical to the obligatory choice of subject in the latter. This is, of course, because the same inherent topicality hierarchies apply in all languages, even though differences are bound to exist from one language to another, as in all other aspects of grammar, with some choices occasionally being less motivated and more arbitrary than the norm perhaps.

The fact that default topic assignment in Basque, an ergative and topic prominent language, follows the same pragmatic/semantic hierarchy that an accusative and subject prominent language such as English follows for subject selection has been taken by some to mean that Basque is ‘underlyingly’ accusative and only superficially ergative. Although this view makes sense at a certain level, I believe it is mistaken in that it mixes levels of analysis. Topic selection in Basque and subject selection in English are similar because they are both based on similar semantico-pragmatic considerations, not because Basque is a *syntactically accusative* language. On the other hand, formal ergative

marking has its own motivations which have nothing to do with information structure considerations, as we saw in Chapter 2.

4.1.4 Thematic ground or topic as a special case of ground

The recent Basque linguistic tradition uses the term *mintzagaia* (“speech topic”) to translate the term topic (cf., e.g., Mitxelena 1981). Interestingly, the Basque Language Academy, in its study of Basque word order, uses this term to refer to all setting elements which provide the grounding for the predication/rheme, be they thematic (topic) or circumstantial, thus recognizing the underlying similarities between the different types of settings, although at the same time it minimizes the differences which do exist (cf. Euskaltzaindia 1987a:28).

Others too, such as Halliday (1985), have tried to generalize the notion of setting, what he calls **themes**, even further to apply to basically anything that can appear in sentence initial position before the assertion proper, including conjunctions and other assertion connectors. Although I do think that it is important to notice the similarities between the different types of setting elements, it is also important to realize the special status of the topic in language, an element which is always a major argument of the verb.

4.1.5 The basicness of the topic-comment configuration

As I mentioned in Chapter 3, it is often assumed that ‘all-news’ sentences, such as answers to the question *What happened?*, do not have a topic, i.e. that they are all rheme sentences. Thus, for instance, Oyharçabal, while admitting that the question and the

answer in (4.1), have a topic, namely *jendea* “the people”, he seriously doubts that the *jendea* in (4.2A) is a topic at all (cf. Oyharçabal 1985:108).

(4.1) Q: *Jendea nondik ateratzen da?*

people:DEF where:ABL come.out/exit:IMPFV it.is
“Where are the people coming out from?”

A: *Jendea zinematik ateratzen da.*

people:DEF movie.theater:DEF:ABL come.out/exit:IMPFV it.is
“The people are coming out of the movie theater.”

(4.2) Q: *Zer gertatzen da?*

what happen:IMPFV it.is
“Whats happening?”

A: *Jendea zinematik ateratzen da.*

people:DEF movie.theater:DEF:ABL come.out/exit:IMPFV it.is
“The people are coming out of the movie theater.”

Oyharçabal assumes that since the referent of this nominal is ‘new’ it must be inside the assertion or rheme.⁶ This, however, confuses the role of topic with the status of topicality, which are two rather different dimensions. As I argued in Chapter 3, I do not believe that these all-news sentences are topic-less, orthetic, assertions.

The two identical assertions in (4.1A) and (4.2A) (*Jendea zinematik ateratzen da*) have identical information structure: both have the same topic and the same focus, which is why they have the same formal structure and intonational properties. It is true that in (4.1A) the topic is more topical (given, accessible) and the focus is more focal (salient), but that doesn’t change the facts about the pragmatic *roles* or *relations* involved. The differences in the pragmatic *status* of the different ideas could, and normally would, have repercussions on the manner that assertion is verbalized, that is, intonational properties, but not on its pragmatic *roles*. So, for instance, the assertion in (4.1A), but not the one in

(4.2A) could be verbalized by means of the single constituent representing the focus: *zinematik* “from the movies”, given that all the other ideas involved in the proposition are already active. Such a fragment assertion would still involve the whole proposition, however, with the topic and other parts of the predicate exclusive of the focus constituent being elided.

The answer to an ‘all-news question’ such as the one in (4.2Q) need not have a topic-comment information structure configuration. Whether the subject can be the focus, however, depends on the ease with which its referent can be accommodated and the existence of a better focus candidate, such as *zinematik* “from the movies” in this case. If the subject’s referent is accessible (or easy to accommodate), it can also be presented as an antitopic in what I have been calling an emphatic assertion.

The most likely version of these sentences in Spanish, for example, would depend on the accessibility of the subject’s referent and the speaker’s choice of whether to postpone it. If the speaker didn’t see the referent of *the people* was accessible enough, then it could be expressed as an overt topic, and even dislocated, as in (4.3)

- (4.3) La gente _(,) está saliendo del cine.
the people it.is exiting from.the cinema
“People are coming out of the movies.”

If the referent of the nominal *the people* was deemed to be accessible enough (e.g. if the question was prompted by a large number of people turning around the corner), it could be postposed, cf. (4.4) (the subject could also be assertion final without any apparent semantic or pragmatic change, cf. *Está saliendo del cine la gente*).

- (4.4) Está saliendo la gente del cine.
it.is exiting the people from.the cinema
“People are coming out of the movies.”

In addition, in Spanish there is another construction which could be used in such a context, a construction which we can call the double focus construction. In this construction, a new referent is introduced in its own intonation unit, but not so much as the topic for the assertion, but as the (fragment) answer to the question *What happened?*. The new referent then becomes the topic in a second assertion, which follows in the next intonation unit and which is coded as a completive clause, as in (4.5).

- (4.5) (Es) la gente. Que está saliendo del cine.
(it is) the people. that it.is exiting from.the cinema
“It’s the people. Coming out of the movies.”

In the most common version of this construction, in the initial assertion typically does not have an overt (copula) verb, although one is possible.

In Basque the very same possibilities are found. In southern/western Basque the imperfective *ari* construction would also be preferably used, which makes the lower clause (*leaving the movies*) a complement of an imperfective aspectual verb, as in (4.6). Notice that in this case the aspectual verb and the finite auxiliary would normally be elided, something which emphasizes the topic-comment structure of this assertion, including its intonational structure.

- (4.6) Jendea (_i) zinematik ateratzen (ari da).
people:DEF cinema:DEF:ABL exit:IMPFV (busy/doing it.is)
“People (are) coming out of the movies.”

Basque also has a parallel construction to the Spanish construction exemplified in (4.5), seen here in (4.7).

- (4.7) Jendea. Zinematik ateratzen (ari dela).
people:DEF cinema:DEF:ABL exit:IMPFV (busy/doing it.is:COMPL)
“People. (They’re) coming out of the movies.”

Notice that here too there are two quite distinct and independent intonation contours, that is, two different intonation units and thus two different assertions integrated and combined into a single construction.

As in Spanish, topic inversion would also be possible in an emphatic antitopic construction, such as in (4.8).

- (4.8) Zinematik ateratzen ari dela jendea.
cinema:DEF:ABL exit:IMPFV busy it.is:COMPL people:DEF
“The people are coming out of the movies.”

In a situation in which there was no better focus candidate than the subject, on the other hand, the subject could optionally be the focus, as in (4.9) below.

- (4.9) Jendea ateratzen (ari dela).
people:DEF exit:IMPFV cinema:DEF:ABL (busy/doing it.is:COMPL)
“The people are coming out.”

4.1.6 Multiple topics

The question often comes up as to whether assertions may have more than one topic, especially given the fact that more than one argument phrase may be topicalized. I mentioned in Chapter 3 that an assertion may have more than one overt setting constituent, all in separate intonation units preceding the asserted clause proper. I have also argued that a topic is a special type of setting, a partially grammaticalized setting which is instantiated by a major argument of the verb and which, when overtly expressed,

may form part of the same intonation unit as the asserted clause. The question is: could an asserted clause have more than one topic?

In Chapter 3 I mentioned the possibility that two different entities could fill what seem to be the two different subroles performed by topics: the linking role and the predication-base role. Thus it is possible, though not common, for two major arguments of a predicate to be topics, i.e. thematic grounds of the assertion, with the external topic being active and coded minimally. This often happens when an argument of the clause is active and/or contrastive, which motivates its contrastive topicalization to clause-initial position, what Givón calls Y-movement, when there is already an active external topic. That kind of situation can be observed in the following pair of examples taken from Hannay 1991.

- (4.10) Q: Have you thought of going to London?
A': No, I hadn't considered London actually.
A'': No, London I hadn't considered actually.

The assertion in (4.10A'), which has the speaker as its topic, shows that the active referent *London* need not be topicalized despite its being active in this context. But, as we can see in (4.10A''), such a referent could be topicalized to clause-initial position. When this happens it is typically the case that the default topic is a minimally coded very active referent, in this case the speaker coded by the unaccented ('cliticized'?) pronoun *I*.

Thus it seems that double-topic constructions are possible, but they are always exceptional, and subject to strong constraints. One thing that seems to hold is that only one nominal setting may be in the same intonation unit as the predicate part of the

assertion. This also seems to be true of the Chinese-type double-topic construction which I mentioned in Chapter 3.

4.1.7 Non-topic topical elements

The possibility of there being more than one topic, topic-like, or topical element in a clause has caused some investigators to do away with the notion of *topic role* and to reduce the relevant parameters to the notion of *pragmatic* or *cognitive status*, including topicality (accessibility) and focality (unpredictability), for the different ideas expressed in an assertion, or to otherwise minimize its importance (cf. Givón 1983a).

It seems to me that this reductionistic approach will not work and that we must indeed distinguish elements which are part of the assertion proper from those which are outside it, and in particular the ubiquitous role of assertion topic which is found in the vast majority of enunciative and other speech acts. Topical, and even setting-like, elements may be part of the predicate without being topicalized and without being the focus of the assertion either as we saw in the assertion in (4.10A') above, for example.

The Basque Language Academy has claimed that topical, non-focus, postverbal elements in Basque are postposed topics/settings, as in the sentences in (4.11) and (4.12). The Academy classifies these postverbal locative and dative elements as postposed 'topics' or setting elements (*mintzagai* elements) (cf. Euskaltzaindia 1987a:42).

- (4.11) Neuk agintzen dut **etxean**
I:EMPH:ERG order:IMPFV I.have.it house:DEF:LOC
"I am the boss in the house."

- (4.12) Sagarra nik eman nion **haurrari**
apple:DEF I:ERG give:PFV I.had.it.to.him child:DEF:DAT
“The apple, I gave it to the child.”

Notice that in both of these cases there is a contrastive (and thus probably emphatic) focus constituent in focus position and that the postverbal, non-focus elements must be quite accessible (possibly through accommodation). It could be that these are postposed settings associated with emphatic assertions, though it could also be that these are just plain non-focal parts of the predicate with very accessible referents and which are redundantly mentioned but could be easily be omitted.⁷

The somewhat superfluous overt mention of referents that are quite accessible but not totally unpredictable has been noticed before, e.g. by Du Bois (1987:830). Such mentions can be made in setting intonation units, but also as unaccented nominals inside the predicate/assertion. Du Bois claims that such overt mentions of given referents “is especially common for inanimate referents, which are often mentioned lexically even just one or two clauses after a previous lexical mention” (Du Bois 1987:830). This is supposedly due to the fact that the low degree of topicality of such referents causes them to decay more rapidly, that is, more evanescent, as it were.

This same phenomenon has been mentioned by Grimes (1982), for instance, who notes that non-central, given elements often get coded as nouns where a pronoun would also be appropriate. Grimes alludes to a possible explanation for this phenomenon, suggested by Reichman (1978) before him, which has to do with what she calls “the centrality factor”. According to this principle, “the less central in the reference space, the more explicit the type of mention of a referent that needs to be made, even if this principle must override the ordinary principles of pronoun placement” (Grimes

1982:387). In other words, the less inherently topical an element is the more it takes to bring it up and maintaining it in the hearer's consciousness. These accessible elements, however, whether overt or covert, need not be topics or settings, unless they meet the definitional requirements we saw in Chapter 3.

4.1.8 Topic-comment structure with non-verbal comments

The bipartite division of clauses into topic and comment/predicate is most obvious when the comment contains a prototypical predicate, such as a non-verbal predicate (especially if the copula is missing), than when it contains a less-prototypically predicational event predicate (cf. Sasse 1987:565).⁸ We can see an examples of such a clear-cut topic comment structure with an elided (copula) verb in (4.13:54), for example (*and the other guy on the side*). (The assertion in (4.13:53) has an extraposed focus.)

(4.13) 93C1A10

53 Eta- - gizona dago <or eserita>,
and man:DEF he.is there sit:PFV:ADV

54 eta beste gizona ondoan,
and another man next:LOC

“And the guy is there sitting, / and the other guy next to him”

The bipartite structure of a sentence is also quite obvious when the predication is a ‘pro-verb’ standing for a whole predicate or when the topic is contrastive, as in (4.14:2) (*well, a demonstration not*).⁹

(4.14) 93C2A03

1 Ba asieran dago= manifa=-
well beginning:DEF:LOC it.is demonstration-

2 bueno manifa ez,
well demonstration no

3 kriston - ba= e- jende pilla trafikoa ta ola,
bunch well uh people pile traffic:DEF and such

“Well at the beginning there is a demonstration= / well not a demonstration [lit. a demonstration, not], / a bunch of people, traffic, and stuff”

Some imperatives too display verb elision, such as when the complement is a directional and the verb is easily recovered. Here too the bipartite nature of the clause is particularly clear, as we can see in (4.15) (*You / (go) inside*).

(4.15) *Behi euskaldun baten memoriak*, p. 78

—Ez, ttiki, hi ikuilu barrura—esan zidan Genovevak
no little you stable inside:ALL say:PFV she.had.it.to.me Genoveva:ERG
“No, little one, you inside the stable,’ Genoveva said to me.”

In other words, it seems to me that in addition to the pragmatic/cognitive notions of topicality (and focality), we need to distinguish in language the pragmatic role or function topic. As we have seen, this notion is often coextensive to a high degree with the syntactic or formal notion subject in many languages, but the two must be clearly distinguished so as to deal with cases in which a subject is not a topic and in which a topic is not a subject.

4.1.9 The formal expression and position of grounds

As we have seen, the formal expression, or verbalization, of the pragmatic topic depends in part on the pragmatic/cognitive status of the referent, including its degree of givenness and contrastiveness. These characteristics typically determine the degree of

intonational integration of the topic with the predicate, going from full dislocation all the way to full integration (agreement or unaccented clitic-like pronoun).¹⁰

I have also shown that the location of the topic with respect to the predicate may depend also on the degree of salience (focality) of the focus, and thus of the assertion, i.e. whether the assertion is 'emphatic' or not. What qualifies as emphatic, and in what contexts and to what extent topic inversion is possible (or likely), differs somewhat from language to language and even from one variety to another within a language. The emphasis motivation for topic inversion, however, seems to be universal and quite iconic in nature. The English version of topic-inversion, i.e. right-dislocation, is quite constrained and is associated with exclamations (emphatic assertions) of a certain type. A similar function has been proposed for the Japanese version of this phenomenon for instance (cf. Givón 1990).¹¹ In some languages topic inversion may even become quite unmarked in some contexts.

As I mentioned in Chapter 3, the relative loss of markedness of the antitopic construction may be the motivation for 'basic sentences' in verb-initial languages. In other words, the topic inversion (emphatic assertion) construction becomes relatively unmarked in these languages and, thus, the default position for overt topics is postverbal position. This would account for the 5% or so of languages which do not have subject-initial basic order (SVO or SOV). It turns out, however, that that verb-initial languages also have constructions for fronting more marked topics. Aissen, for instance, reports that Mayan languages, most of which are said to be VOS languages, and a few of which are said to be VSO, use clause initial topics (as well as marked preverbal foci) (Aissen 1992:43).

Aissen's account of Mayan languages is particularly interesting because it reveals that Mayan languages differ as to how well integrated this initial positioning of the topic is into the grammar of the language. Thus while all of these languages, including Tzotzil and Jakaltek, allow left-dislocation in main clauses (though supposedly not of third-person pronouns), some languages, such as Tz'utujil, appear to have what we could call clause-internal initial topics (Aissen 1992:44).¹² In other words, all languages allow clausal arguments to be external (sentence-head) settings, under rather similar pragmatic circumstances. The placement of assertion topics in clause initial position is a more grammaticalized phenomenon which, although pragmatically motivated, can be suspended in some languages under some circumstances. We should not forget, in any case, that clauses with overt subjects are a minority and that perhaps only about half of all overt subjects are topics.

4.1.10 Clause-internal vs. clause-external topics

I have argued that a topic is a special type of setting (ground) element for an assertion. Unlike other settings, a topic may be part of the same intonation unit as the rest of the assertion, despite being, as a setting, outside the assertion. However, a topic may also be 'dislocated', in its own intonation unit, just like any other setting. And a topic may also be elided, or covertly expressed, if it is active enough. Topics are special also in that in most types of assertions they are required by the grammar and must be a core argument of the verb.

But what exactly is the relationship, or the difference, between a dislocated setting topic and one which is more integrated into the clause, including a grammaticalized one?

In some languages internal topic placement and external ‘dislocation’ can be distinguished formally by more than just the presence or absence of an intonational break. In some languages, for instance, and at least under some circumstances, an assertion with an external (dislocated) topic, but not one with an internal one, has a ‘resumptive’ anaphor inside the clause. Another possible difference may be that in the case of external topics the case marking of the topic nominal is neutralized for the different cases (cf., e.g., Givón 1990:758).

Although the differences between a dislocated and a non-dislocated (clause-internal) topic seem clear in extreme situations, in actuality they aren’t always so clear-cut. As Givón argues, in some languages, for instance, “the structural distinction between L[eft]-dislocation and Y-movement [same intonation unit fronting] may be partially neutralized” (Givón 1990:759).¹³ In other words, none of the formal differences are categorical.

To begin with the intonational differences between dislocation and clause-internal topic, as Mitxelena has noticed for Basque, aren’t always clear-cut and we may find more of a continuum between unaccented clause-initial nominal to fully accented and dislocated nominal (cf. Michelena 1987). The more unexpected and harder to process a topic is, the more likely it is it will be ‘dislocated’, and the more accessible it is, the more likely it is to be unaccented and integrated in the clause. But as Mitxelena says, the distinction between clause-internal topic and dislocated topic may sometimes seem “more illusory than real” (Michelena 1987:470).¹⁴

With respect to the presence versus absence of (clause-internal) resumptive anaphors, notice that in languages with grammaticalized verbal coreferencing of the

subject and perhaps also other arguments as well, such as in Basque, there isn't such a distinction, since these anaphors are always present, whether or not there is an overt topic of any kind or not, whether or not the topic is dislocated, and whether or not there is a topic at all.

In Spanish, for example, there is always an anaphor present for the subject argument (verbal 'agreement' or 'coreferencing') as well as, typically, for most dative objects, whether there is a fronted dative nominal or not, as well as for most fronted accusative objects (as long as the object is referential). In other words, in Spanish 'coreferencing' is fully grammaticalized for subjects, almost fully grammaticalized for datives, and partially grammaticalized for accusative objects. In Basque, on the other hand, 'coreferencing' is fully grammaticalized for all three arguments.

Finally, case-marking neutralization is also a less than ideal cue for topic type. By neutralization of case marking it is meant that all topics are marked identically. This often applies to dislocated topics but not to internal topics. Givón mentions, however, that in Japanese, where topics bear the topic suffix *wa* and no case marking, regardless of case, case neutralization is equally found for dislocated ('L-dislocated') and clause-internal ('Y-moved') topics. According to Lambrecht, detached fronted topics in German, which do not count as being part of the clause for verb-second purposes, may *optionally* lack case marking (Lambrecht 1994:194). Being case marked, however, is not a sure clue that the topic is internal.

In Basque too we find that occasionally a dislocated nominal may not bear the 'correct' case marking. This typically involves an absolutive marked dislocated nominal which fills the ergative role inside the assertion. This, however, doesn't seem to be an

instance of case neutralization. This phenomenon is very likely attributable to the fact that often when speakers set up a topic in a sentence-head intonation unit, sometimes they haven't yet formulated the assertion and thus they do not yet know what role the topic will have in it. In such situations, the absolutive case might seem to be the 'default' or least-marked case, though mismatches in the other direction are also found.

4.1.11 Pragmatic characteristics of topics and settings

The topic role, or function, is to some extent independent of the notion of topicality, which refers to the cognitive-pragmatic status of ideas. There are, however, correlations between topicality and the likelihood that an argument will be the topic of an assertion, as well as constraints on how non-topical a topic can be and still be a 'proper thematic ground' for an assertion.

Thus we find that topics (and thus subjects) predominantly have active referents in discourse, or else they are highly accessible, or, at the very least, they are identifiable (i.e. 'definite') (cf., e.g., Givón 1979; Chafe 1994a:105-106). In languages in which the correlation between the formal category subject and the pragmatic category topic is greater than in English, identifiability (and thus definiteness) may indeed be a minimal requirement for subjects. In other languages, such as English, this is only a strong preference, though it is not a requirement. Thus, in English, occasionally we find exceptions to the identifiability condition on subjects, as Givón and Chafe, for instance, concede.

However, the (rare) existence of unidentifiable subjects may be well motivated. First of all, as we have seen, occasionally subjects (topics) are new, but they appear in

their own intonation unit, i.e. are ‘dislocated’ (cf. Chafe 1994a:106). Furthermore, Chafe argues, “[m]ost of the subjects that appear superficially to express non-identifiable referents actually express generic ideas or are nonreferential, and thus fall outside the domain of identifiability” (Chafe 1994a:106). However, although generic subjects may be topics, that is not the case with non-referential ones. Thus I believe that non-referential subjects are not topics at all, but foci.

Chafe also mentions the case of new, but trivial subjects, which he argues is a motivated exception to the identifiability requirement. He argues that placing a trivial new subject in such a position is perhaps not so burdensome since its reference doesn’t become an issue. Furthermore, he argues, that this is actually a rhetorical trick in order to “express surprise” (Chafe 1994a:91). He mentions the following example in (4.16) with the accented new subject *this ván*.

(4.16) Then one afternoon, ... *this ván* pulls in *there*, ...

Of course, if the subject received the main accent of the assertion in this sentence, as it very well could, it would definitely be the focus and thus not a topic at all. On the other hand, if the rheme internal complement *there* receives the main accent, then the subject’s referent, which is new, must indeed be the topic, although this assertion does strike me as odd since for there to be the focus it would most certainly have to be contrastive (since it is so accessible).

To summarize, the identifiability constraint on topics, and thus on subjects, is very strong in all languages for obvious reasons. Languages in which subjects are not allowed to be unidentifiable have the strongest (presumably unidirectional) identification between

the formal notion subject and the pragmatic notion topic. In languages in which subjects may be unidentifiable, this happens most likely in cases in which the subject's referent is not the topic of the assertion, but rather the focus.

4.1.12 Pragmatic characteristics of topicalized elements

In addition to this research on constraints on what can be a subject (topic) in different languages, there has also been much interest in the literature as to what exactly can and cannot be topicalized, i.e. fronted, whether dislocated or not (cf., e.g., Reinhart 1982/1981; Gundel 1985, 1987, 1988; Lambrecht 1986a, 1988a; Ward and Prince 1991).

As Ward and Prince (1991) conclude in their study of previous proposals on constraints on topicalization, "these previous analyses have all attempted to account for the same general observation, namely that NPs which represent entities brand-new to the discourse are not felicitously topicalized" (Ward and Prince 1991:168). By this they mean that they cannot be entities which are brand new *and* are not themselves formally grounded on entities which are not brand new.

In other words, they argue that the proper formulation of this principle is that the minimal condition on preposing (adapted from Ward 1988/1985) is that "[t]he entity represented by the preposed constituent must be related, via a salient partially ordered set relation, to one or more entities already evoked in the discourse model" (Ward and Prince 1991:173). Thus, for instance, the nominal *A man I met the other day* could indeed be topicalized even though the entity in question is brand new since the pronoun *I*, referring to the speaker, grounds the entity in question. This constraint, of course, makes perfect sense given the function of the topicalized element, namely grounding the sentence

thematically. To this, of course, one must add the additional possibility of non-identifiable topics which are salient entities in contexts other than the textual context, e.g. [pointing] *That man, he was here the other day too.*

Some investigators have argued that this minimal constraint is really not enough in most cases and that stricter constraints are needed. Indeed, the vast majority of topics are much more accessible than just merely identifiable or, worse yet, grounded unidentifiable. Thus, for instance, Reinhart (1982), Gundel (1985, 1987, 1988) and Lambrecht (1994), all propose that ‘accessibility scales’ predict the likelihood of something being a topic, and that the more accessible referents are the better topics they make, and the least accessible they are, the more marked and unusual they are too.¹⁵

Occasionally, most inaccessible referents can be topics too, but there are constraints on their formalization. Gundel, for example, mentions that what she calls the *topic-familiarity principle* can be “suspended, under certain conditions” and only in some constructions, such as ‘left-dislocation’, but not topicalization or right-dislocation, and only in assertions (not in questions or commands).¹⁶ Lambrecht too believes that there are certain types of exceptional topics must be accounted for by means other than the scales. That is, he believes “that violations of the Topic Acceptability Scale [such as those in Reinhart 1982] can be explained as genre-specific instances of the principle of the PRAGMATIC ACCOMMODATION of presuppositional structure” (Lambrecht 1994:196).¹⁷ And, of course, different constructions might then have different requirements as to the type of topic it is likely to have, as Gundel proposed.

Other constraints on topicalization and left-dislocation have been proposed for specific situations, such as Reinhart’s (1983) suggestion that ‘arguments’ cannot be

dislocated in questions whereas ‘adjuncts’ can (cf. also Bresnan 1994:82). Thus the version of the sentence in (4.17) with the verb *sat* seems much worse than the version with the verb *knitted*.

(4.17) On the platform among the guests of honor, who knitted/*?sat?

Reinhart claims that the locative phrase is a complement of *sit*, and thus its ‘sister’ in the ‘underlying’ structure, whereas it is an adjunct of *knit*, and thus not a sister but a higher adjunct. Thus the constraint in topicalization would be a formal one, with something blocking the topicalization of VP-internal adverbials but not VP-external ones.

It seems to me, however, that the differences involved here are not structural or syntactic at all, but that they rather have to do with the information structure of these predicates. The activity verb *knit* is quite different semantically and informationally speaking from the stative verb *sit*. With the verb *knit*, the sentence-head locative in (4.17) is not a fronted or topicalized complement at all, but simply an optional setting. The verb *sit*, on the other hand, much like a copula verb, has little semantic content and is typically used with a predicate-like complement.

When *sit* is not accompanied by a rhematic complement, which by default fills the role of focus, there is a very strong tendency for the verb *sit* to be interpreted as contrastive (focus). In other words, this verb, when uttered alone, has a very strong tendency to be the focus of the assertion, and that may be what makes it sound odd in this context, especially since the question pronoun *who* is the default focus in this question construction. In other words, for the verb *sat* to be the (contrastive) focus the question

word must be 'given', i.e. the question must come after a similar question (e.g. *On the platform, who sang?*).

Notice that the presence of some other complement of the verb *sat* which is able to fill the focus role is all it takes to make this sentence acceptable with minimal context, such as in the example in (4.18) below.¹⁸

(4.18) On the platform among the guests of honor, who sat cross-legged?

It is obvious that, for some reason, perhaps because it expresses a stereotypical state, the idea expressed by the phrase *sat cross-legged* is easier to accommodate in this content question than the idea expressed by the phrase *sat*. It would seem that the verb *sit*, when used by itself and imperfectively, must be interpreted as contrastive and thus very focal (unless, of course, the idea of sitting, as opposed to standing, for example, was active in the context).

To summarize, the more accessible a referent is the better it performs its grounding function. Thus relatively inaccessible topics are rare. When they do occur they are typically dislocated, something which supposedly aids the hearer process the referent and allow it to become the ground for the asserted clause which follows. Unidentifiable referents are made much more topic-worthy by being themselves grounded by an entity which is itself quite accessible (such as the speaker). Constraints on the topicalization of the verb's complements (arguments and modifiers) may also depend on informational characteristics of the verb itself.

4.1.13 Information structure vs. syntactic structure: topic-comment vs. NP-VP

I have suggested above that there is a strong parallelism between what I call the default information structure of a sentence (topic-comment) and the syntactic structure which has been proposed for many languages in the form of the grammatical units NP (subject) and VP (predicate). In this section I would like to follow up on an argument I started in Chapter 1, according to which such syntactic structure is a mirage, nothing but the reflection of the information structure, as well as semantic bonding among the elements of the clause, which also motivate the default information structure.

The fact that the typical sentence has a topic-comment pragmatic configuration and that in English a subject must always be overtly expressed (even if it is by an unaccented pronoun) and that it must always be preverbal, have been used as motivation for the putative existence of an underlying *syntactic* structure in English, and other languages, in terms of the *syntactic* units NP and VP. However, looking beyond English and beyond surface structure, allows us to see that the units in this case are primarily pragmatic, as well as semantic, and that there is no need to posit any syntactic structure for asserted clauses (sentence-body units) beyond their functional structure.

It may be argued that the VP is the formal analogue of a semantic unit, the semantic predicate. When the assertion expresses an event or action (not at state), it is said that the default topic (the subject) is typically the element which is most independent from the event (cf. Sasse 1987). This might account for the fact that default predicates, e.g. *(X) ate hamburgers*, seem to form better conceptual or semantic units than, say, the verb with the subject, e.g. *John ate X*. On the other hand, what seems to account for the conceptual unity of a verb and its complement is the specificity of the complement's

referent and not so much its semantic content. This can be observed in the three assertions in (4.19), in which the objects differ as to their specificity.

- (4.19) a. John ate hamburgers.
b. John ate a hamburger.
c. John ate my hamburger.

Although the semantic bonding between *ate* and *hamburgers* seems to be greater than the bonding between *ate* and *John*, the same is not necessarily true when the object is *my hamburger*. The difference is probably even less when the two complements are equally topical, e.g. *Mary kissed John*. In other words, the ‘semantic predicate’ *eats hamburgers* would seem to be a tighter conceptual unit than *ate a hamburger*, which in turn is more unit-like than *ate my hamburger*. Thus it doesn’t seem that conceptual semantic considerations will be of much use in motivating a verb phrase formal unit.

One might argue that the syntactic NP-VP division of the clause proposed by formal theories is real because it has been grammaticalized, at least in some languages, the so-called ‘configurational’ languages (cf. Hale 1982, 1983, 1989, 1992). Such languages, which, of course, include English, use word order for grammatical purposes and have also grammaticalized the pragmatic role topic into the grammatical category subject, often to such an extent that it has taken a life of its own (cf. Li and Thompson 1976b, Thompson 1978). That is, one might say that the subject-predicate structure is nothing but a grammaticalization of the prototypical, or most common, type of information structure: the topic-comment structure.

In English, a language with perhaps the most rigid type of word order, such an argument might seem plausible. However, I do not see that we actually gain anything by

adopting such a position. To begin with there are clause types that this structure doesn't apply to, i.e.thetic and other topicless clauses. This is especially obvious in VO languages in which the focus subject is in postverbal position, such as Spanish. Secondly, the degree of conceptual bonding (the basis of all syntactic structure) within the clause, is variable and depends on characteristics of the complements of the verb, as we have seen. Third, there are other competing, conflicting, and intersecting bonds within the clause, such as those between elements that form the 'verb-complex' which cannot be ignored.

Finally, English is an extreme case of a configurational (rigid word order) language, but other (less 'configurational') languages display a greater degree of freedom of word order, so that even formal theories which hypothesize a verb phrase admit that some languages may not have such a constituent. But the fact is that most languages display some degree of pragmatically motivated word order freedom, but the verb phrase is an all-or-none structure and cannot reflect minor differences and have been said to be non-configurational, but what are we to do with all the languages in between, most of which display word order flexibility to some degree or another?

Of course, one of the main arguments for the NP-VP and other spurious syntactic structure in formal approaches to linguistic analysis has to do with asymmetries of one type or another between subjects and other elements of the clause, such as objects, and which are said to be daughters of the VP and not the S. It seems to me, however, that any account of such asymmetries should begin by taking into consideration the existing pragmatic role asymmetries and that a purely structural account is but a poor substitute

for the real cause of the asymmetries, which is obviously related to those differences in pragmatic functions.

4.2 COVERT AND INITIAL OVERT TOPICS IN THE SPOKEN BASQUE CORPUS

4.2.1 Introduction

As I mentioned in Chapter 3, and as we can now see in Table 4-1 below, the majority, at least 4/5, of all the affirmative declarative assertions (main and dependent finite statements) in the Spoken Basque Corpus have topics, i.e. are thematically grounded. The majority of these topics in this narrative corpus, 7/10 of all, are **covert topics**, i.e. they are elided (in non-finite clauses) or coded merely by a verb-internal anaphor. About 1/9 of all clauses have **overt initial topics**. Only about 1/8 are clearlythetic clauses with (new, non-contrastive) focus subjects, and less than 1% seem to have **antitopics**.

Information Structure	Information Structure	All	All%	Main	Main%
Covert topic	<i>Covert_Topic_*</i>	2563	69.3%	1826	65.3%
Overt, initial topic	<i>Overt_Topic_*</i>	410	11.1%	369	13.2%
Overt, final antitopic	<i>Invert_Topic_*</i>	47	1.3%	27	1.0%
Thetic (no topic)	<i>Non_Topic_*</i>	452	12.2%	385	13.8%
Other ¹⁹		228	6.2%	190	6.8%
TOTAL		3700	100.0%	2797	100.0%

Table 4-1: Number of finite affirmative statements by topic status in the Spoken Basque Corpus.²⁰

If we restrict ourselves to main affirmative statements, as we can see in the *Main* and *Main%* columns in Table 4-1, which are ¾ of the total, we see that the proportions do not change significantly.

As we can see in Table 4-2, about half of topics in the corpus are ergatives, with almost as many absolutes in intransitive clauses ('intransitive subjects'). Less than 4% of all topics are datives and less than 4% are absolutes in transitive clauses (objects). The percentages are almost identical whether the clauses are main clauses or main and dependent statement clauses combined.

Topic	Info. Str.	All	All%	Main	Main%
Ergative	<i>*Topic_A*</i>	1540	49.7%	1112	48.5%
Absolute (subject)	<i>*Topic_S*</i>	1324	42.7%	1014	44.2%
Dative	<i>*Topic_I*</i>	121	3.9%	87	3.8%
Absolute (object)	<i>*Topic_O*</i>	114	3.7%	81	3.5%
TOTAL		3099	100.0%	2294	100.0%

Table 4-2: Topics in the Spoken Basque Corpus classified by case role, for all affirmative clauses and all affirmative, finite statements in main clauses.

In other words, more than 90% of all these assertions have as topic the same argument which would be the subject in an accusative language such as English. The exceptions to this rule, however, are very important since they relationship between the category topic in discourse and the (semantic or syntactic) category subject (however it is defined) is not an absolute, identity relationship. The exceptions show that despite the high correlation, we cannot equate these two categories.

In Table 4-3 we can see the number of topics for each role, broken down by the argument structure of the predicate. Here we can see that that, although A and S are the overwhelming favorites for the topic role, in clauses which have a dative/benefactive argument, which are a minority of clauses, the referent in this role is also quite likely to be the topic, much more so than the absolute argument.

Predicate Args.	A top.	Covert	S top.	Covert	O top.	Covert	D top.	Covert
Abs_Dat_Erg	307	(88.9%)	n/a		1	(0%)	54	(87%)
Abs_Erg	1220	(87.7%)	n/a		113	(74.3%)	n/a	
Abs_Dat	n/a		16	(81.3%)	n/a		58	(94.8%)
Abs	n/a		1006	(73.7%)	n/a		n/a	

Table 4-3: Choice of topic for all possible combinations of arguments in the Spoken Basque Corpus in affirmative statements.²¹

Although the vast majority of these topics are covert topics, there is a slight difference as to relative proportions of covert vs. overt topics for the different argument types, as expressed in the percentages in parentheses next to the absolute numbers. In general we can say that absolutive topics are slightly more likely to be overt than ergative and dative ones.²²

In transitive clauses with a dative argument the dative is the topic about 15% of the time and the ergative about 85%. Absolutive objects are topics in less than 1% of these clauses. However, about 1/5 of these dative topics (10/54) are really absolutives with human referents which have been ‘promoted’ to dative status, much like in (the Basque dialect of) Spanish.²³ In transitive clauses with no dative argument, the ergative is the topic about 91.5% of the time and the absolutive about 8.5% of the time. As we can see in Table 4-4 there are only about ¼ as many transitive clauses with dative arguments than without them.

Predicate Args.	A top.	S top.	O top.	D top.	All	%
Abs_Dat_Erg	307	0	1	54	362	13.0%
Abs_Erg	1220	0	113	0	1333	48.0%
Abs_Dat	0	16	0	58	74	2.7%
Abs	0	1006	0	0	1006	36.3%
TOTAL	1527	1022	114	112	2775	100.0%

Table 4-4: Topic-comment clauses by argument structure and topic choice.

In intransitive clauses with a dative argument, on the other hand, the dative argument is more likely to be the topic (about 4/5) than the absolutive argument (about 1/5). Clauses with this argument structure, however, are but a small minority of all clauses, as we can see in Table 4-4 (2.7%). Transitive clauses with a dative argument are also a small, though more significant, percentage of the total (13%).

4.2.2 Non-‘subject’ topics

As I have mentioned, there is a clear topic hierarchy in Basque, as in all other languages, as to which argument is the most likely one to become the topic, the default topic, just like there is a clear connection between semantic roles and grammatical roles. The topic preference, or ‘unmarked topic’ hierarchy might look something like the one in (4.20), given the facts that we have just reviewed for Basque.

(4.20) A (Ergative) / S (Absolutive) > I (Dative) > O (Absolutive)

Some have argued that direct objects are consistently more topical than indirect objects (e.g. Givón 1990:901).²⁴ It seems to me that the matter is more complicated. This discrepancy may be due, in part, to the fact that Basque dative arguments code optional benefactives as well as ‘true’ dative arguments. It is interesting in this respect that, as I mentioned earlier, in Basque (like in some dialects of Spanish) there is a growing tendency to code the most topical of direct objects as indirect objects.

Let us now look at an example of an assertion with a non-subject as topic. The following excerpt in (4.21) contains a ‘topic chain’ of four clauses all of which share a topic. Actually, in the first clause in (4.21:117) the referent, the girl who stole the bread

at the bakery in the Chaplin movie, is introduced in athetic sentence with VS order (see Chapter 5). This girl is the topic of the remaining 3 assertions. In the second clause, in (4.21:119), the girl is the referent of the elided ergative argument. In the fourth clause, in (4.21:121), the girl is the referent of the elided absolutive (intransitive) argument. But in the third clause, in (4.21:120), she is the referent of the dative (a human object coded as a dative), with an overt ergative argument with a new referent, and she clearly seems to be the topic of the clause as well. In other words, there doesn't seem to be any doubt that that assertion is *about* the girl.²⁵

(4.21) 93C2A01

117 Eta ortikan pasatzen-da neska bat trajea puskatuta,
and there:ABL:LOC she.passes.by girl one dress:DEF torn:PFV:ADV

118 bueno soinekoa puskatuta eta,
well dress:DEF tear:PFV:ADV and,

119 ta artzen-du bat- e ogi bat.
and she.takes one uh bread one

120 **Ēta andre batek ikusten-dio.**
and woman one:ERG sees.her

121 Ta ateatzen-da korrika. ez?
and she.comes.out running. no?

“And then this girl comes by with her dress all torn and stuff, / well, her dress [synonym] all torn, / and she takes a loaf of bread. / **And a woman sees her.** / And she runs out, right?”

Notice that in this assertion the ergative argument could conceivably be the topic, but the intonation would give this away immediately. Given the new status of the ergative argument's referent, this nominal would have to be dislocated or at the very least would require a full accent, as in (4.22a), whereas if the ergative is part of the predicate it receives the main accent of the predicate, as in (4.22b).

(4.22) a. Eta andre batek (,) ikusten-dio.

b. Eta andre batek ikusten-dio.

The sentence in (4.21:120) with AV order reflects the standard position for non-topic ‘subjects’ in Basque, namely rheme-initial, preverbal, focus position. In spoken Basque, however, inversion of a non-topic ergative arguments is also possible, and relatively common: about 1/3 of non-topic ergatives invert with the verb in the Spoken Basque Corpus. A search of the Spoken Basque Corpus shows that there are 5 sentences with a continuous object topic and an inverted ergative argument, such as the one in (4.23) below, corresponding to the same scene in the movie as the sentence in (4.21:120).²⁶

(4.23) 93C1B06
 208 ikusi-du andre batek,
 she.sees.her woman one:ERG
 “a woman saw her,”

These assertions must be interpreted as emphatic (surprising) assertions in which the polarity is the focus, with the ergative referent accommodated in postverbal (non-focus) position. Even though the ergative-idea is new, it is not important in this context, and thus it is (relatively) low in focality, which makes verb/polarity focus possible (cf. Chapter 7). (The ergative could be the focus if it was extraposed, something which would be clearly coded intonationally and accentually.)

In addition to the 5 clauses with ergative inversion describing this particular scene, we find in the spoken corpus an additional 6 sentences with a continuous dative topic and an inverted ergative argument. The number of cases in which non-topic overt ergative arguments do not invert with the subject seems to be exactly twice as large: 10 each with covert absolutive topics, such as the one in (4.21) above; and 12 each with covert (continuous) dative topics.²⁷

The fact that ergative argument inversion is also possible in similar contexts in Spanish, for instance, doesn't mean that this strategy is borrowed from this language however. In Spanish subject inversion in a sentence such as this is structurally ambiguous between subject focus and salient (emphatic) verb/polarity focus with subject postposing, as we can see in (4.24a&b).²⁸

- (4.24) a. Le ve una mujer. "A woman sees her." (thetic, subject focus)
b. Le ve una mujer! "A woman sees her!" (emphatic, 'polarity focus')

In English, on the other hand, such contrast is not possible since subject postposing is limited to antitopics in certain limited-use constructions ('right dislocation'). The closest parallel to the sentence in (4.24b) would be a passive construction, cf. *She's seen by a woman!*

4.2.3 Covert topics in the Basque corpus

As we saw in the previous section, ellipsis is very prevalent for all topics in the spoken Basque narrative corpus (cf. Table 4-1 and Table 4-3). The percentage of covert topics is highest for ergatives (almost 90% of all) and for dative topics (87% when there is also an ergative and 94.8% when there isn't one). Absolutive object (O) topics are also for the most part elided, though the percentage is lower (74.3% when there is no dative). Absolutive subject (S) topics are also elided most of the time, but, again, less often than ergatives (81.3% with datives and 73.7% when they are the only argument).

In the excerpt in (4.25:2-4), we can see three examples of elided topic main clauses, the most common type of information structure type. They follow the existential/presentative clause in (4.25:1) (with extraposed or delayed focus).

(4.25) 93C1A01

- 1 Orduan da= baserritar bat,
then he.is farmer one
- 2 eta bere- m- eskalera batekin igotzen-da= arbol batean,
and his m- ladder one:COM he.goes.up tree one:LOC
- 3 eta artzen-ditu sagarrak,
and he.takes.them apples:DEF
- 4 eta ematen-du= - zaku batean,
and he.puts.it [sic] bag one:LOC

“Then there is a farmer, / and he goes up a tree with a ladder of his, / and he’s getting apples, / and puts it [sic] in a basket,”

Of these three clauses, one of the continuous/active topics is an S (absolute ‘subject’) and two are A’s (ergative ‘subjects’). These two types of arguments constitute the vast majority of the continuous or covert topics, as we can see in Table 4-5.

Covert topic	Information Structure	All	All%	Main	Main%
Ergative	<i>Covert_Topic_A</i>	1354	54.1%	940	52.5%
Absolute (subject)	<i>Covert_Topic_S</i>	964	38.5%	711	39.7%
Dative	<i>Covert_Topic_I</i>	104	4.2%	75	4.2%
Absolute (object)	<i>Covert_Topic_O</i>	79	3.2%	64	3.6%
TOTAL		2501	100.0%	1790	100.0%

Table 4-5: Covert topics in the Spoken Basque Corpus classified by case of topic, for all main and completive affirmative finite statements in main clauses.

As I mentioned in Chapter 3, when more than one referent seems to be active at and the arguments of an assertion are covertly expressed, it is not always clear which of the arguments is *the* topic of each assertion. In the excerpt in (4.26) we can see a typical series of actions and events in which more than one referent remains active and thus are not expressed covertly. In the text prior to this excerpt Chaplin had been the only active referent for some time and thus typically the topic. In the first assertion in this excerpt, in (4.26:34), the police (non-specific) is introduced in athetic (SV) sentence. In the following assertion in (4.26:35) Chaplin is the topic and it is overtly expressed, in part

because the predicate is non-finite and in part perhaps because it is contrastive. In the assertions in (4.26:37, 38, 41) both referents (Chaplin and the police) are active and covert, which referent is intended in each case being recoverable from the meaning of the predicates. Which entity is the topic in these assertions would seem to be primarily a matter of ‘perspective’ or empathy, that is, it depends on whose point of view is being taken, and thus the assertions are, I believe, pragmatically ambiguous, rather than vague.²⁹

(4.26) 93C2B04 (Chaplin is the ongoing topic)

- 33 Da karo!
and of.course
- 34 zea\ polizia etorten-da,
that.is police:DEF it.comes
- 35 da be- besti oi trapuz eskuen.
and oth- other that rag:DEF:INSTR hand:DEF:LOC
- 36 {LAUGH} -
- 37 Da= arrapau-itteute.
and they.have.caught.him
- 38 Da uste-dute <liderra dala>.
and they.think.it leader:DEF he.is:COMPL
- 39 Manifestapeneko= zea, liderra.
demonstration:GEN what:DEF leader:DEF
- 40 {Bai}
yes
- 41 Ta kartzela sartze-ute, noski, -
and jail:DEF:ALL they.put.him.in of.course
- 42 da kartzela=n daola,
and jail:DEF:LOC he.is:ADV
- 43 pues= bazkaltzen ai die.
well have.lunch:IMPFV busy they.are

“And of course! / uh the police comes, / and that guy, with the cloth in his hand. / {Laughs} / And they get him. / And they think he’s the leader. / The leader of the demonstration. / {Yes} / And they put him in jail, of course, / and while he’s in jail, / well they’re eating lunch. ...”

In (4.27:60) below we see an example of a covert dative topic in a clause with an absolutive (subject) argument and a dative argument. The overt absolutive argument is

postverbal, as often happens to absolutes which are not topics in speech (though not so much in writing). Again, this inversion signals verb or, rather in this case, polarity (emphatic) focus. (As we will see in Chapter 5, it is quite common in Basque for unaccusative predicates to have dative topics, in addition to assertion internal (non-topic) absolutes.)

(4.27) 93C2A02

- 56 Bueno, ba juten-da bide batetik,
well well he.goes road one:ABL
57 eta= gurutzatzen-da neska batekin. ez?
and he.crosses (runs into) girl one:COM no?
58 {Bai.}
yes
59 Bizikletaz.
bicycle:INSTR
60 Ta erortzen-zaio txanoa,
and it.falls.to.him hood:DEF

“Well, he goes along this path, / and he meets this girl, right? / {Yes.} / On a bicycle. / And his hat falls off.”

A similar type of complement-verb inversion can be seen in (4.27:57) (*he comes across a girl*), here between the verb and a commitative argument. Here, however, it is not clear whether the assertion is meant to be emphatic (with polarity focus) or whether the complement is an extraposed focus, although the latter is more likely. (Notice that *the girl* in this example is harder to accommodate than *the hat* in the previous one.) The intonational cues are ambiguous.

In other words, the order VX is ambiguous between focus extraposition and verb focus and only intonational cues may disambiguate the two structures, something which isn't always as easy in practice as we might expect. Thus such assertions may be pragmatically ambiguous, although the type of ambiguity involved here is not so great as

to be dysfunctional. This may be seen as a problem with the pervasiveness with which focus extraposition is used in some varieties of Basque. On the other hand, a similar ambiguity may result in VO languages, such as Spanish between postverbal focus and verb/polarity focus.

In (4.28:38) it is also the referent of a covert dative argument that is the undisputed continuous topic (*they give him food*). Here, however, the predicate is transitive, but the referent of the ergative argument is non-specific (*they*, i.e. *the police*, or *the guards*).

(4.28) 93C1B06

36 Eta orduan ba ^klaro
and then well of.course
37 ba bera oso ondo dago kartzelan. ez?:
well he very well he.is jail:DEF:LOC no
38 <LIST **Ematen-diote janaria,**
they.give.it.to.him food:DEF
39 dauka logela bat >, -
he.has.it bedroom one
40 eta dena, ez?
and all no

“And then well of course / well he’s very happy in jail, right?: / **They give him food,** / he has his own bedroom, / and everything, right?”

Notice that in this sentence, as well as in the following one, the overt absolutive (object) argument (*food*) is also postverbal, which leaves the verb again in focus position. These do not seem to be emphatic assertions, however, with VO structure, but rather extraposed focus assertions, with V Q structure. In standard Basque these assertions would have QV order, but, like I said, in spoken Basque, focus extraposition is quite common, especially for some speakers (cf. Chapter 6).

Another interesting fact about these two VO assertions is that they have a special ‘listing’ intonation with a rather pronounced final rise, much like the equivalent English sentences would. The speaker signifies by this intonation pattern that there is a long list of improvements in Chaplin’s situation since he rescued the guards, of which the speaker is providing two salient examples.

4.2.4 Overt topics in the Spoken Basque Corpus

Sentence-body (asserted) clauses with an overt topics are not as numerous in the spoken Basque corpus as those with a covert one. In Table 4-6 we can see the breakdown of all the topics in finite statements by case. Here too, we can see that the majority of the topics are ‘subjects’: ergatives (A) and intransitive absolutes (S).

Overt topic	Information Structure	All	All%	Main	Main%
Ergative	<i>Overt_Topic_A</i>	157	38.7%	150	38.8%
Absolute (subject)	<i>Overt_Topic_S</i>	225	55.4%	216	55.8%
Absolute (object)	<i>Overt_Topic_O</i>	18	4.4%	16	4.1%
Dative	<i>Overt_Topic_I</i>	6	1.5%	5	1.3%
TOTAL		406	100.0%	387	100.0%

Table 4-6: Covert topics in the Spoken Basque Corpus classified by case, for all affirmative finite statements (*All*) and all affirmative finite statements in main clauses (*Main*).

As we can see, there are more overt absolute ‘subject’ (S) topics than ‘ergative’ (A) topics. This is the reverse of the situation we found when the topics were covert.

In the sentences in (4.29:31&32) below we can see typical examples of overt topics, both of them ergative and both of them followed by some hesitations and pauses. Chaplin and his mean cell mate are both quite accessible at this point (they had both been

topics in the preceding paragraph), but a new referent is introduced in the preceding statement in (4.29:30), and they also need to be distinguished from each other.

(4.29) 93C1A02

28 Orduan jateko ordura doaz,

then eat:NOMIN:GEN hour:DEF:ALL they.are.going

29 ondoan dira, -

next:DEF:LOC they.are

30 ta ondoan dago xiki bat.

and next he.is small one

31 Orduan bo= **indartsuak** e= - e= ogi bat du,

then well strong:DEF:ERG uh uh bread one he.has.it

32 eta= **gizonak** e- e nai du ogi puska bat,

and man:DEF:ERG uh- uh desire he.has.it bread piece one

“Then they go to lunch time [sic], / they’re next to each other, / and next there is a little guy. / Then the strong guy has some bread, / and the guy [Chaplin] wants a piece of it.”

In a minority of cases, overt topics are objects, just as in the case of covert topics. This typically happens when the referent of the ergative argument is less topical than that of the object or when it is the focus of the assertion. In (4.30:168) we see an example in which the object is the overt topic and the referent of the ergative argument is an understood, non-specific referent, namely *they*, i.e. *the police*.

(4.30) 93C1A10

168 =Orduan andrea kartzelara eramaten-dute. -

then woman:DEF jail:DEF:ALL they.carry.her

169 Ta gero gizonak berriz beste tontakeri bat egin-du,

and then man:DEF:ERG again another silly.thing one he.has.done.it

“Then they take the woman to jail. / And then the man, he did another silly thing again.”

Notice that the object must be outside the rheme since there is another element in preverbal, assertion-initial position, the allative phrase *kartzelara* “to the jail”, and there can only be one assertion element before the verb. Topicalized objects are also found in

languages like English, particularly if they are contrastive topics, but English has an alternative construction to use in these cases, namely the passive construction, which is often used to promote objects to the topic role. This alternative does not exist in Basque, as we have seen. Thus topicalization is not necessarily felt to make the topic contrastive, though it could, with the right intonation.

Of course, in such a situation the object's referent need not be a topic and in fact it could have remained inside the assertion, though not in focus position, in Basque just as in English. Thus the sentence in (4.30:168) could have been rendered as in (4.31).

- (4.31) Orduan kartzelara eramaten-dute andrea.
then woman:DEF jail:DEF:ALL they.carry.her
“Then they take the woman to jail.” (~ “Then the woman was taken to jail”)

In this case the object, with an equally accessible referent, is in assertion-tail position. Fronting is more likely if the object is contrastive or if it is less predictable (i.e. if there is referential interference from other referents). Postposing, on the other hand, is more likely if the assertion is emphatic (i.e. if the allative focus is very salient/focal).

The conditions under which the referent of an argument which is not the default topic will become the topic when the would-be topic is not the topic for some reason, seem to be somewhat variable, both cross-linguistically and for any one language, within certain parameters. It seems, for instance, that contrastive elements are more likely to be topicalized than are non-contrastive ones.

Notice in the sentence following the one I have been discussing, the one in (4.30:169), there is another overt topic, in this case a contrastive ergative topic. The

adverbial *berriz* “again” here is a topicalizing conjunction, or connector, and not a complement of the verb. I will discuss this type of connector below.

A second type of situation in which an absolutive object is the topic, besides when the ergative is non-specific or the new-focus, is found when the ergative is a contrastive focus, such as in example (4.32:361) below.

- (4.32) 93C2B01
360 Bañe= -
but
361 ogiye Txarlotek zeuken.
bread:DEF Chaplin:ERG he.had.it
“But - / Chaplin had the bread.” (Lit. “The bread, Chaplin had it.”)

In this example the topic (the referent of the absolutive argument) is not contrastive, although it could have been in a different context, but the referent of the ergative is indeed contrastive, though it wouldn’t necessarily have to be.

In spoken Spanish object topicalization is also relatively common since the Spanish promotional passive construction is not used in speech (and the passive is not used with stative verbs such as *have* in the example above, anyway). As I mentioned earlier, when the object is the topic an ‘agreement’ object clitic is always present next to the verb.³⁰ Unlike in Basque, however, the normal position for the focus subject would be postverbal, as in (4.33a), though in an emphatic assertion it could also be rheme-initial and preverbal, as in (4.33b), at least in some dialects of Spanish. (As in the case of light subject pronouns in English, such object clitics in Spanish do not count as taking up a ‘position’.)

- (4.33) a. El pan lo tenía Charlot.
 the bread it he.had Chaplin
 b. El pan Charlot lo tenía. (*very focal focus*)
 the bread Chaplin it he.had
 “Chaplin had the bread” (~ “It was Chaplin who had the bread”)

So far we have only seen major or core arguments that are topics. Occasionally, though quite rarely, we find topicalized arguments which are not major arguments of the verb, such as the one in (4.34:118). The topicalized nominal here is in the commitative case, and it is indeed an argument of the verb (it is part of its valence), and doesn't seem to be dislocated.

- (4.34) 93C2B01
 117 ta an ibili batea ta bestea,
 and there go:PFV one:DEF and other:DEF
 118 ta azkenin ogi puskiekin bea geau-da.
 and end:DEF:LOC bread piece:DEF:COM he has.remained/has.been.left
 “and both of them at it, / and finally he ends up with the bread.”

Examples such as this one are rare, however, and this is the only such example in the whole spoken corpus. However, it is exceptional cases such as this one that reveal the pragmatic nature of the operation involved, despite the high correlation between the pragmatic function topic and certain grammatical roles.

4.2.5 Dislocated topics in the Basque corpus

The topics we have seen so far were all inside the assertion's intonation unit and not dislocated, even if they are accented. In the Spoken Basque Corpus there are 141 cases of topics which are clearly dislocated in sentence-head intonation units.³¹ These arguments are overwhelmingly case marked (when they are not absolutes with zero case marking). These ‘dislocated’ topics are almost always highly accessible, discourse old

referents, but which are somewhat unexpected in the particular context since they are not very important or haven't been mentioned for some time. Sometimes these verbalizations contain descriptions of the referent which situate the referent before asserting something about it, such as in (4.35) below.

(4.35) 93C1A05

114 eben= gero, -
uh then

115 be- bestea,
other:DEF

116 tontoa dena,
dumb he.is:REL:DEF

117 etortzen-da,
he.comes

“well, then, / the other guy, / the dumb one, / he comes (back).”

Thus we find that topics are often dislocated when followed by a parenthetical of some kind, including parenthetical assertions.

Occasionally we find that different setting phrases, especially if they are somehow interconnected, they can be grouped or ‘compressed’ into a complex intonation unit, what we might call a mega-intonation unit, such as in (4.36:40), where we find three such phrases, all describing the same referent, forming what seems to be a single, complex intonation unit with three accentual peaks.

- (4.36) 93C1A16
 38 Gero,
 later
 39 momentuan, or, -
 moment:DEF:LOC there
 40 ebe beste batek, - e= bere ondoan zegoena, txikiak,
 uh other one:ERG uh his side:DEF:LOC he.was:REL:DEF, small:DEF:ERG
 41 artzen-du gatz potea, ...
 he.takes.it salt pot:DEF
 “Then, / right afterwards, there, / uh another guy, the one who was next to him,
 the little one, / he takes the salt-shaker, ...”

Sometimes the ‘dislocated’ topic intonation unit is followed by a significant pause, something which suggests that the topic in such sentences is dislocated, not because of any pragmatic properties it might have, but simply to give the speaker time to verbalize the predicate.

Occasionally a dislocated topic is unidentifiable. This happens in four cases in the spoken corpus, which I reproduce here below. In one of them, the one in (4.37:45), the unidentifiable referent is itself grounded by the mere fact that it is sitting down next to the active referent. Notice the long pause after the sentence-head topic.

- (4.37) 93C2B08
 43 Eta= ..
 and
 44 baskatzen ari zen bitartean,
 have.lunch:IMPFV busy he.was:REL meanwhile:LOC
 45 beste alboko batek, ...
 other side:DEF:GEN one:ERG
 46 e= - droga itxurako zerbait botatzen-dio.
 uh drug semblance:DEF:GEN something he.throw.something.to.him
 “And / while he was having lunch, / another guy on the other side (of him), / uh
 he dumps something that looks like drugs.”

As I argued in Chapter 3, such dislocated topics are somewhat equivalent to an existential (or other presentativethetic clause) in which the referent in question is the focus of thethetic assertion.

In (4.38:99) a completely new referent is introduced as a sentence-head topic. Notice here too the pause following such an intonation unit. This is the type of context in which we may find a new referent coded with the demonstrative adjective *this* in English (cf. *and this girl, she steals a loaf of bread*).

(4.38) 93C2A07

96 Orduan, ..
then

97 e=
uh

98 ya kaliao dao,
already street:DEF:LOC he.is

99 eta= neska batek, ..
and girl one:ERG

100 lapurtzen-du ogi bat. ...
she.steals.it bread one

101 Ta neskak artzen-du ogia,
and girl:DEF:ERG she.takes.it bread:DEF

102 ta korrika asten-da, ..
and run:ADV she.begins

“Then, / uh / he’s back on the street, / and this girl, / she steals a loaf of bread. / And the girl takes the bread, / and (she) starts running.”

In this case, unlike in the previous one, the unidentifiable referent in question is not grounded to the discourse context in any way.

In the remaining two examples, also with fully unidentifiable *and* ungrounded referents, there is no significant pause after the sentence-head nominal. In both examples the referent in question appears suddenly on the scene, in the middle of a sequence of events, as it were. I suspect that this may have something to do with the fact that this

rather unusual strategy is used to bring these referents into the discourse. The referent introduced in this way in (4.39:296) is a rather minor or insignificant part of the story.

(4.39) 93C2A09

294 Txarlot eramaten-dute, -
Chaplin they.take.him.away

295 baiño/
but

296 e= beste neska batek,
uh other girl one:ERG

297 e= es[θ]ena guztiya ikusiya zeukan,
scene whole:DEF see:PFV:DEF she.had.it

298 eta ordun ikusiya zun no- nola tx- neska izandu-zen ba= - gaizkillea. ez?
and then see:PFV:DEF she.had.it ho- how Ch- girl:DEF she.had.been well
wrongdoer:DEF no?

“They take Chaplin away, / but! / uh another girl, / she’d seen the whole scene, / and so she’d seen how it was the girl who was guilty, right?”

Finally, the last example of an unidentifiable dislocated topic, seen here in (4.40), is indeed an important one in the story. This case is also interesting because the referent is verbalized in three separate sentence-head intonation units, with different descriptive phrases, each one making the referent more specific.

(4.40) 93C3A12

- 104 Ba kalera ateratzen-da, -
well street:DEF:ALL he.comes.out
105 te kalean dagoen- dagoela, ..
and street:DEF:LOC he.is:SUBORD- he.is:ADV
106 ba bueno\ e=
well well uh
107 emakume batek,
woman one:ERG
108 neska gazte batek,
girl young one:ERG
109 e= neska pobre batek, ..
uh girl poor one:ERG
110 ba= la- ogi bat lapurtze-u. ..
well st- bread one she.steals.it

“Well he comes out (of the jail), / and while he’s out on the street, / well uh / a woman, / a young girl, / uh a poor girl, / she steals a loaf of bread.”

Givón has claimed that the referents of clause-internal ‘topicalized’ nominals are consistently more accessible than those of external, or dislocated, topics, which have typically been mentioned very recently. I do not doubt that the accessibility of the topic referent is somehow correlated with whether a topic is dislocated or not, but it seems to me that the function of this dichotomy (dislocating vs. not dislocating an overt topic) is not the coding of the degree of accessibility of a referent. Thus we find a much variation between the extremes of unidentifiability and activeness of a referent and their degree of intonational separation from the rheme/predicate. And I suspect that other factors may be involved, such as rhetorical factors which go beyond the sentence.

In particular it seems for example that topic dislocation is associated with the beginning of new thematic paragraphs or topic chains, the dislocation of a nominal serving to set up the thematic ground for that paragraph, which is the topic for the paragraph initial assertion. Otherwise, referents destined to become topics are typically

brought about as foci. In other words, it may be that, besides reflecting cognitive production constraints and facilitating processing, the main *purpose* of dislocating a topic is to achieve a certain rhetorical effect and not to reflect any objective properties of the status of the referent in question in the minds of the speaker or hearer

Fronting without dislocation (Givón's 'Y-movement'), on the other hand, seems to be about switching among accessible referents inside a thematic paragraph. Occasionally, however, events which interrupt a paragraph in which the relevant referent is not accessible enough are also likely to display dislocated topics if there is no event connecting the new event and topic to the preceding ones, which allows the new topic to be the focus of an assertion in which the old topic is still the topic.

We also must keep in mind in this discussion that we are not talking here about a dichotomy between dislocation and no dislocation, but about a continuum from full dislocation, with a following pause and perhaps even topic repetition, and full integration of the topic into the assertion as a pronominal (or 'agreement') or 'lightly accented' nominal. This fits in with the claim that this.

4.2.6 Split external and internal topics

I argued in Chapter 3 that the notion of topic can be seen as having two related facets to it, each one emphasized by different investigators in the past: an outward looking facet, the discourse linking role, and an inward looking one, the predicand (aboutness) role. That is, the topic is 'what the predicate/comment is *about*', as well as what the state-of-affairs is connected to or pegged onto in the referential world of discourse (the discourse model).

I suggested then that in the *prototypical* case these two roles are both fulfilled by the same referent. However, as we have seen, sometimes a point can be made that an assertion has more than one topic, either in special construction, such as the so-called Chinese or double topic construction, or when two arguments of the verb are topicalized, or when one is topicalized and the other is covertly expressed. Sentences seemingly having two topics are found in the Basque corpus, though apparently never more than two.

In Chapter 3 I suggested that in two-topic constructions typically it seems that the ‘external’ one or the covert one, is the one which links the assertion to the discourse, whereas the other one, the ‘internal’ one, can be seen as that referent which the assertion is primarily about. In the examples of this type that we find in the Basque corpus, the external topic is almost always covert and the internal one is overt, as in (4.41:196) below.

- (4.41) 93C2B01
 193 Urduen,
 then
 194 dana indartu bezela,
 all strengthened as
 195 o bildur guziyek aldein-deye. -
 or fear all:ERG they.have.taken.leave.to.him
 196 Ta tipo aundi orri esan-diyo ezetz,
 and guy big that:DAT he.has.said.it.to.him that.no
 “Then, / he gets all strong, / or he’s not afraid anymore. / And he says to that big guy, ‘hell no’.” (cf. Spanish *Y al tipo grande ese le dice que no*)

In this excerpt, the ongoing topic is Chaplin, which in the clause in (4.41:196) is the referent of the covertly expressed ergative argument. But the assertion also has a

topicalized dative argument, one whose referent, “the big guy”, is quite accessible but apparently not sufficiently so.

The same phenomenon can be observed in (4.42:86) below, which has a covert ergative linking topic. In addition there is what seems to be a second topic, a ‘local topic’ in the absolutive (object) role.

(4.42) 93C2A01

84 bueno berak jotzen-die biri atea- atearakin aurpegian,
well he he.hits.them two:DEF:DAT door:DEF- door:DEF:COM face:DEF:LOC
85 eta bueno
and well
86 denak lurrean uzten-ditu.
all:PL ground:DEF:LOC he.leaves.them

“Well, he hits both of them with the door on the face, / and, well, / all of them he knocks them down.”

Most of the double topic examples are of this kind. It is very rare that two topics are overtly expressed, other than perhaps when the assertion is being repeated for clarity, as in (4.43).

(4.43) 93C2B05

53 Eta= .. <VeryFast gorruak ies itte-jotzak >
and .. hat:DEF:ERG escape do:IMPFV-it.does.it.to.him
54 {Nola?}
how? (what?)
55 Gorruak mutillai - ies-ittejotzak.
hat:DEF:ERG boy:DEF:DAT escape it.has.made.it.to.him
“And ... his hat flies off. / What? / The boy’s hat it flies off.”

It is interesting that here the hat is the ‘external’ topic and the boy the ‘internal’ one, whereas we might have expected the order to be the reverse. As I said in Chapter 3, however, the order of setting elements among each other, whether they are ‘topics’ or not, has to do with relative scope and not with any other factors such as topicality.

4.2.7 Verb topics

When we speak of topics we typically mean referents or entities, and more precisely the entities which fill one of the major roles or arguments of a clause's predicate. Exceptionally, however, other types of 'ideas', besides entities, can be used as settings for the assertion. Basically, almost any element of a sentence, for instance, can become a setting, coming before the assertion. In Basque this option extends to the verbal idea itself, a phenomenon which involves an accessible verbal idea which is typically contrastive. This phenomenon, which is relatively rare in English, seems to be much less marked in languages like Basque and Spanish. In (4.44A) we can see an English example of the type of sentence that I have in mind. In this context the verbal idea (*read*) is very active, as well as contrastive (*read* vs. *skim*).

(4.44) Q: Did you read the proposal?

A: Well, read, I didn't really read it, but I did skim it.

The VP-preposing construction in English, which is a highly specialized construction, seems to be similar to this type of verb fronting (topicalization), e.g. *He said he would read the proposal and read the proposal he did*. Notice that with VP-preposing the 'predicate idea' in question must also be active.³² This seems to be related to the fact that non-referential ideas of this sort do not last long in consciousness, at least not as ideas which can serve for grounding propositions.

Verb topicalization is often found in Basque when the action expressed by the verb always contrasts with another action (cf. Osa 1990:181), producing a construction "with a very distinct Basque flavor" (Altube 1929:32). Below in (4.45) and (4.46) we can

see two typical examples of this construction, taken from Altube's classic work on Basque syntax (Altube 1929:33).

- (4.45) *Ekarri neuk egin/ekarri neban.*
bring:PFV I:EMPH:ERG do/make:PFV/bring:PFV I.had.it
“As for doing (it), I did it.”

- (4.46) *Irakurri gitxi egiten/irakurten dabe euskaldunak.*
read:PFV little do:IMPFV/read:IMPFV they.have.it Basques:DEF:ERG
“As for reading, Basques don't do very much of it.”

In (4.45) the ergative argument is in focus position, since it is contrastive or very salient (extrinsically emphatic). In (4.46), on the other hand, the focus is the adverbial *gitxi* “little” and thus the (accessible) ergative argument is in postverbal position.

Note that the ‘topicalized’ verb is always in the perfective participle form, the unmarked form of the verb in most dialects (cf. Trask 1995), and inside the assertion proper there is either the same verb in finite form or, in some dialects, the pro-verb *egin* “do, make”, taking its place.

Osa mentions the interesting fact that Basque allows the topicalized verb to be an antitopic, i.e. to be post-rhematic, something which a language such as Spanish, which also allows verbs to be topicalized quite freely, does not allow (Osa 1990:176-8). He gives the examples in (4.47a&b) with a fronted and a backed verb topic respectively (Osa 1990:178, exs. 97 and 98).

- (4.47) a *Etorri egunero dator.*
come:PFV every.day he.is.coming
b *Egunero dator etorri.*
every.day he.is.coming / come:PFV
“(As for) coming, he comes every day.”

The possibility of a verb topic participating in the antitopic construction seems to me to reflect the relatively high degree of grammaticalization of this construction in Basque, even though the construction itself is quite rare, especially with a backed topic.³³ Thus the following example from the Written Basque Corpus is the only example of verb topicalization in either the written or spoken corpora.

(4.48) *Behi euskaldun baten memoriak*, p. 164

- a. Zu lau mendi horien artean bizi-izan-zara beti,
you four mendi those:GEN between:DEF:LOC you.have.lived always
- b. eta jakin ez dakizu,
and know not you.know.it
- c. baina munduan gauza asko daude.
but world:DEF:LOC thing many they.are

“You have lived among those four mountains all your life, / and **you don’t know this**, / but in the world there are lots of (other) things.”

This example is interesting because here the verb is not at all active, or even accessible, but must be accommodated. (Notice that this clause performs a concessive function with respect to the following assertion.)

As I mentioned in Chapter 2, Basque predicates are often complex in that they consist of a low content verb and a nominal. When such verbs are topicalized, the whole verbal complex is fronted. We can see two such verb topics in the following ‘lapidary’ statements (from Osa 1990:182, ex. 121).

- (4.49) Amets egin / sarri, borroka egin / beti.
dream make:PFV / often/frequently figh/struggle make:PFV / always
“Dream often, struggle always”

These two topic-predicate assertions are also special in other ways. Notice that the rhemes consist of adverbial predicates and no verb, and that the verb topics are new, though as often happens in such lapidary statements, they are easy to accommodate.

4.3 INVERTED TOPICS, ANTITOPICS, RIGHT-DISLOCATION AND AFTERTHOUGHTS

4.3.1 Inverted topics or 'antitopics'

In addition to the universal position for overt topics in pre-predicate (clause-initial, or sentence-head) position, some have argued that under some circumstances, in some topic-initial languages, and not just in verb-initial languages, topics can also be placed after the verb, constituting what Chafe terms 'antitopics'. Antitopics are clause-internal, postverbal constituents, i.e. nominals which are located in the same intonation unit as the rest of the clause (and unaccented), and not dislocated (and thus accented). From a pragmatic perspective, they are accessible or at least they can be easily accommodated in context as if they were accessible.³⁴ Antitopics, thus, cannot be contrastive topics, for example. From a syntactic point of view, antitopics are co-indexed in the clause by either a covert pronoun, as in English and French, or by agreement, as in Basque or Spanish.

As Lambrecht and others have noticed, the antitopic construction has been recognized for a long time, though it seems that there is much terminological confusion. The classical term for this construction is 'epexegetis', and more common names are 'extraposition' (Jespersen 1933/1964), and 'right dislocation' (cf. Lambrecht 1994:202), even though here there is no second intonation unit involved.³⁵ In addition there seems to have been some confusion between this type of constituent and another type of constituent which does come after the clause proper in its own intonation unit, namely 'afterthoughts', cf., e.g., Vennemann 1974, Hyman 1975, Givón 1976, etc. (cf. Lambrecht 1981, 1994:203, 354fn55).

Antitopics are found in Basque, but they are much less common than (clause-initial) topics and they are especially infrequent in spoken narrative. Givón 1983b also finds that this construction is about 10 times less frequent than left-dislocation (cf. also Givón 1990:761). When they occur they are always associated in Basque (and in other languages) with a very salient, or focal, (assertion-initial) focus constituent, such as are found in emphatic and contrastive assertions. As I mentioned in Chapter 3, in the Spoken Basque Corpus about half of the antitopic constructions are found in copula clauses with an identificational predicate.

It seems to me that the close association between antitopics (postverbal topics) and emphatic assertions is a universal fact which has not been sufficiently appreciated in discussions of this construction type. Such assertions are disruptive and thus typically quite rare in continuous narrative. In some languages, or in some contexts within a language, such a construction may become more prevalent as a way of coding certain rhetorical meanings, in motivated extensions of its original meaning perhaps, with the resulting concomitant loss of emphasis associated with the original, more marked construction.

Topic backing is part of a more general *setting backing* strategy which results in the focus constituent being in sentence initial position in this type of continuity breaking or out-of-the blue type of emphatic assertion. Setting adverbials, for instance, are postposed just like topics are, as we can see in (4.50).

(4.50) I saw your mother when I went to the store today.

The position of the topic subject *I* in English is, of course, quite fixed, but a similar sentence in Spanish could have the subject pronoun in postverbal position, as we can see in (4.51).

(4.51) Vi a tu madre hoy yo cuando fui a la tienda.
I.saw to your mother today I when I.went to the store

This pronoun, of course, is completely redundant (and optional) here, and certainly it is not contrastive. Its main purpose seems to be to help code (along with the marked intonation) that this is an emphatic assertion.

As we saw in Chapter 3, it is very important to distinguish clearly between the inversion of subjects which are not topics, but foci, inthetic sentences, in many (VO) languages with postverbal main focus position, such as Spanish, and inverted, or postverbal, antitopics, found in languages with either preverbal or postverbal main focus position. This is so even though some of the contexts in which they occur are similar, namely contexts in which there is no topic continuity, such as out of the blue statements, although, overall, such contexts too are more likely to have topic-comment clauses than eitherthetic or antitopic clauses.

Givón (1990) notices the two types of inverted sentences in Spanish (thetic and ‘right-dislocated’) and claims that the two constructions may be neutralized. His examples, which are reproduced in (4.52) below, involve the verb *morirse* ‘(to) die’, which is often realized withthetic information structure (the underlining marking the main accent was not in the original). Givón mentions that the comma which he writes before the antitopic in (4.52a) may be but “a perceived ‘pause’”, one which “may in fact be only a tonal or *intonational* cue”, something with which I would agree, since the

'right-dislocated' subject is in the same intonation unit as the rest of the clause (Givón 1990:762).

(4.52) a. **R-dislocation:**

... luego se murió, Juan, y ...
then REFLX he.died John and
"... then he died, John, and ..."

b. **VS word-order:**

... luego se murió Juan y ...
then REFLX he.died John and
"... then John died and ..."

What Givón seems to fail to notice is the very different types of information structure involved here. He fails to note the accentual, and thus also intonational, differences which exist between these two sentences, other than the "intonational cue" before the antitopic. What Givón does not mention is that the subject in (4.52b) is not a topic at all, but rather the focus of the assertion.

On the other hand, as I mentioned in Chapter 3, there is a context in which the two constructions may indeed be neutralized, and that is when thethetic assertion is emphatic the polarity is the focus. Thus if thethetic assertion in (4.52b) was emphatic, the accent would go on the finite verb, with the outcome being very similar, accentually and in terms of word order, to the right-dislocation (antitopic) construction in (4.52a).

It remains to be seen, however, whether there remain intonational differences can differentiate the two constructions. I believe that differences do exist, even if at times they are hard to detect. Informationally too the two sentences are somewhat different since thethetic sentence, whether emphatic or not, informs us that something happened, whereas the antitopic construction informs us that something happened *to Juan*, i.e. *Juan* is the (inverted) topic of the assertion. Other differences remain between the two

constructions. Thus, for instance, it seems thatthetic assertions cannot accommodate a second focal complement since such a complement would vie for the focus role with the subject, whereas antitopic assertions may have, and typically have, other focal complements, which are the (very focal) focus of the assertion.

4.3.2 Antitopics and right-dislocation

As we have seen, the ‘antitopic construction’ which Chafe’s (1976) described for some Native American languages is based on the same principle to the English ‘right dislocation’ construction (cf., e.g., Givón 1983b, Ziv 1994), and the French ‘right-detachment construction’ (cf. Lambrecht 1994:202), among others. Of course, although the principle, or the motivation, for the different constructions in different languages is the same, the actual uses and particular idiosyncrasies of the relevant constructions which embody this principle in different languages is not the always the same. In fact the uses and markedness of the construction may vary from one variety to another within the same language.

According to Givón right-dislocated elements in spoken English have highly accessible referents, with a referential distance of 1, just like for zero anaphora and slightly less than for anaphoric pronouns (Givón 1983b). Similarly, Lambrecht, who gives the sentence in (4.53) below as an example of ‘antitopic’ in English, argues that antitopics “cannot indicate a new topic or a topic shift” (Lambrecht 1994:204), a more modest requirement than Givón’s. (As with Givón’s Spanish examples above, the comma in Lambrecht’s example doesn’t signify the end of an intonation unit and the beginning of an other, or, even less so, a real pause, but rather seems to be the standard way of

signifying in writing the intonational shift associated with the preceding focus constituent.)

(4.53) He's a nice GUY, your brother (Lambrecht 1994:203, ex. 4.59)

It seems to me, however, that accessibility is not a particular requirement for antitopics or dislocated subjects. It seems to me, rather, that the ideas expressed by antitopics must be quite accessible, just like the ideas expressed by any other rheme-internal non-focus constituents must be very accessible. Most importantly, however, I do not believe that postposing of the topic is a way of coding, or a direct reflection of, the accessibility of its referent, as Givón seems to suggest. The purpose of topic inversion is simply to allow the focus constituent to come early and, indirectly, to code that the focus is very focal, something which typically happens when the assertion itself is itself inordinately salient.

4.3.3 Antitopics and afterthoughts

One of the types of elements that antitopics have been confused with are afterthoughts (cf., e.g., Hyman 1975, Harris 1985). Ziv (1994), who uses the label 'right-dislocation' for the antitopic construction, has analyzed in some detail the structural, intonational and cognitive differences between antitopics and afterthoughts. What Ziv means by afterthoughts coincides with what I described earlier as unplanned corrections and additions in sentence-tail intonation units. The contrasted characteristics he found are summarized in Table 4-7.

Right dislocation (antitopics)	Afterthoughts
<ul style="list-style-type: none"> • Single intonational contour with preceding clause, with no preceding pause • Unstressed • Clause final <ul style="list-style-type: none"> • The NP is necessarily coreferential with some NP which precedes it in the clause • Textually or situationally evoked, easily recoverable entity. <ul style="list-style-type: none"> • “Discourse organizers”; “part of carefully planned speech” (p. 641) 	<ul style="list-style-type: none"> • Separate intonation unit, with distinct pause preceding • Stressed • They may be final or may be added as a parenthetical in other positions in the sentence • They “may contain corrections of reference ... where no coreference holds” (p. 639) • Corrective discourse function (“the error concerns either the identity or the relative ease of retrievability of the discourse referent in question”, p. 640) • “Incidental feature of careless performance error” (p. 641)

Table 4-7: Differences between right dislocation (‘antitopics’) and afterthoughts, according to Ziv 1994.

It is clear from this list that antitopics are not afterthoughts, nor is it likely that the former derive from the latter (cf. Harris 1985).³⁶ Thus whereas antitopics are planned constituents which are incorporated into the clausal intonation unit, afterthoughts are unplanned repair mechanisms, which follow the repaired element, typically after the end of the clause, though not necessarily so. They are thus a type of element which is found in sentence-tail intonation units, or else in parenthetical intonation units.

There are certain aspects of these lists which may be open to challenge, however. Thus I believe, for instance, that afterthoughts need not be preceded by a “distinct pause”, and that its separation from the preceding elements is primarily intonational. Sometimes the correction comes so rapidly that a pause may be missing.

The requirement that antitopics be clause final also doesn’t seem to hold either, at least not in languages in which word order is less rigid than in English and in which these constructions are more grammaticalized. In these languages, which include Basque and

Spanish, we find that antitopics need not be clause final, but merely need to be unaccented postverbal elements, that is elements of the assertion tail. In other words, although in some languages antitopics may be ‘tacked on’ to the end of the assertion, in other languages such elements may be much more closely integrated into the clause (assertion tail).

Thus we see that right-dislocations are not, intonationally speaking, the converse of left-dislocations, since the latter involve a separate intonation unit. Also, the two constructions obviously perform very different discourse functions. Because of their different informational and rhetorical functions dislocated topics and inverted topics (antitopics) are not in any way interchangeable. Also, because of the different positions they occupy in the sentence their accessibility requirements are quite different, although that is never a determinant of the topic’s position.

The phenomena of antitopics and afterthoughts are of particular relevance to word order and word order change, for, as Stockwell (1977) has reminded us, there has been a view, going back to McKnight 1897, and revived more recently in Hyman 1975, according to which “a principal motivation in the rise of postverbal complementation was by way of ‘afterthoughts’” (Stockwell 1977:298), a label which has sometimes been used to include antitopics, at least in English.

In the original formulation of this hypothesis the notion of ‘afterthoughts’ was used to account for movement out of the ‘brace’ formed by the finite and non-finite verbs in Germanic languages. Hyman (1975) sought a similar explanation in order to account for the diachronic shift from verb-final order to verb-medial order in Bantu languages. It is not clear, however, whether what these and other investigators had in mind were actual

(accented) afterthoughts or (unaccented) postposed topics *and other setting constituents* associated (originally) with emphatic constructions.

Despite Hyman's claims, it doesn't seem to me that either afterthoughts or antitopics are a mechanism by which languages change word order type, although antitopics may perhaps become so unmarked in some languages so as to make postverbal position the 'unmarked' position for overt subjects. As I suggested in Chapter 3, I believe that the major motivation for word order change, in particular for the change from OV to VO, has to do with change in the primary position of the focus constituent.

4.3.4 Uses of the antitopic construction

The antitopic or right-dislocation construction is a very widespread construction cross-linguistically, found even in quite strictly verb-final languages, such as Turkish, at least under some circumstances (cf. Erguvanlı 1984). The actual instantiation of this strategy into constructions may vary from language to language, but the general motivation seems to be the same iconic and universal one that I have described above. The antitopic may be assertion final in less grammaticalized versions of the construction, or it may enjoy more freedom of placement in assertion-tail area of the clause, particularly when the focus is preverbal. The conditions for this 'backing strategy' are two: First the postposed element must be either quite accessible or easy to accommodate into the assertion, and secondly the assertion must have a focus element which is significantly more salient than average, something which calls for its early placement in the clause.

As I said, languages differ as to when they allow such postposing and how they exploit it rhetorically. In many languages, for instance, such postposing is either common (Basque, Spanish) or required (English) in content questions constructions, assertions which by definition have salient focus elements. The question in (4.54) is an example of an antitopic in French, from Lambrecht (1986a).

(4.54) a OÙ est-ce qu'il est, ce lycée? 'where is this lycée' (p. 309, ex. 8.1b)

Another use of this construction is reported by Givón, according to whom in Ute right-dislocated pronouns "indicate the *end of a thematic paragraph* ... they signal to the hearer the **cataphoric discontinuity** of the referent—and also of theme" (Givón 1990:761, cf. Givón 1983a&b&c).

If we look at Lambrecht's (1986a) examples of the antitopic construction in French, or Chafe's (1976:53) original examples from Seneca, or examples of the English right dislocation construction, as in (4.53) above, we immediately notice that these are not regular assertions, but rather they are always exclamatory, emphatic assertions or one type or another, which tend to occur in isolation and not in the middle of a paragraph, for instance, since they are rather disruptive to the flow of discourse. In (4.55) below we see four examples of the antitopics in affirmative statements in French from Lambrecht 1986a.

- (4.55) a. Ça n'a pas de goût, ce poulet. 'this chicken has no taste' (p. 312, ex. 8.2H:a)
b. Mérite des baffes, ce petit con 'he deserves a slap in the face, this little jerk' (p. 312, ex. 8.3a)
c. Sont en plastique maintenant, les tables de bistrot? 'are made of plastic nowadays, bistrot tables?' (p. 313, ex. 8.3b)
d. Bizarre, ce truc 'strange, that thing' (p. 313, ex. 8.4a)

Other English constructions, such as the English *it*-cleft construction, also reflect this same inverted order strategy, seemingly also motivated by the existence of a very salient focus which ‘wants’ to be coded early in the clause (e.g. *It was a banana that I ate*). The same thing is true of ‘inverted’ *wh*-clefts (e.g. *A banana’s what I ate*).

Another construction which seems to reflect the same strategy of placing a highly salient focus close to the beginning of the sentence is so-called ‘subject extraposition’, as exemplified in (4.56) below, which is an exclamation and which in spoken English the ‘subject pronoun’ and the copula could be omitted, something which I view as a reflection of the iconic urgency of expressing the focus (*amazing*).

(4.56) (It’s) amazing what a difference that makes!

This doesn’t mean that all examples of ‘subject extraposition’ involve the inversion of setting constituents. In some contexts, and with some predicates, subject extraposition is related to the strategy of focus extraposition. Extraposed foci, unlike postposed topics and other settings, are accented and may be in their own intonation unit following the assertion proper (although it may also be more closely integrated into the assertion). We can see an example of such an extraposed focus in (4.57).

(4.57) It is clear (to me) (,) that this makes a difference.

In this construction both the subject and the predicate are focal. Thus, unless the sentential subject expresses a very accessible proposition it does not qualify as a proper ground (cf. *That this makes a difference is clear to me*). The compromise way of dealing with such assertions is to extrapose the sentential subject, allowing both the subject and

the predicate to be accented in a type of construction which seems to be composed of two assertions: (1) *This makes a difference* and (2) *That this makes a difference is clear*.

As I mentioned earlier, I believe that the phenomenon of subject-auxiliary inversion in English, as well as subject-verb inversion in other languages such as Spanish and Basque, is also a reflex of the same general strategy of placing a salient focus first in the clause in assertions with very focal foci and inverting the setting elements. In English subject-auxiliary inversion in questions is obligatory, whereas in other languages, such as Basque or Spanish, it is simply a possible, and common, option. In (4.58) we can see several examples which show us the different possibilities for subject positioning in a content question in Spanish. In (4.58a) the subject is clause final, in (4.58b) it is postverbal. In the remaining two an argument is ‘topicalized’, the subject in (4.58c) and the object in (4.58d) (notice the clitic object pronoun).²³

- (4.58) a. ¿Cuándo le vio a María Juan? Q V O S
 b. ¿Cuándo le vio Juan a María? Q V S O
 c. Juan ¿cuándo le vio a María? S Q V O
 d. A María ¿cuándo le vio Juan? O Q V S
 “When did Juan see María?”

In (4.58c) the subject is clearly the topic and in (4.58d) the object is the topic. As for the questions in (4.58b) and (4.58c) it is not possible to tell out of context which one is antitopic if any. The relative order of the subject and object does not seem to help us in this respect. Of course, it could be that neither is the (anti)topic. This is a common phenomenon associated with emphatic assertions and antitopics, namely that the notion of topic becomes less relevant.

Note finally that these different possible positions for the subject in content questions in Spanish are exactly the same ones that are possible in declarative assertions (statements) when there is a salient focus element in assertion-initial position, as we can see in (4.59).

- | | | |
|--------|-------------------------------------|---------|
| (4.59) | a. <u>Ayer</u> le vio a María Juan. | F V O S |
| | b. <u>Ayer</u> le vio Juan a María. | F V S O |
| | c. <u>Juan ayer</u> le vio a María. | S F V O |
| | d. A <u>María ayer</u> le vio Juan. | O F V S |
- “Juan saw María yesterday.”

In other words, the pragmatic principles which motivate the different possible orderings of constituents in content questions are the same as those which motivate them in all assertions with a preverbal (highly focal) focus constituent.

4.3.5 The antitopic construction in the Spoken Basque Corpus

As we saw in Table 4-1, there are at most 45 statements in the spoken Basque corpus which contain antitopics, although with some of them there is a question as to whether they are cases of antitopics or not, namely those with copula verbs and an identificational focus. We can see the breakdown of the types of statements which I classified as antitopic inversions in Table 4-8 (the whole list of sentences can be seen in Appendix 4.1).

Antitopic type	No.	Example
Identificational referent focus with copula	23	<i>She's the thief</i>
Identificational locative focus with copula	4	<i>There's the thief</i>
Predicate focus with copula verb	1	<i>He's a thief, that guy!</i>
Ergative subject inversion	8	<i>They call him, the boys</i>
Absolutive (subject) inversion	9	<i>He really got strong, that Chaplin</i>
TOTAL	45	

Table 4-8: Types and number of statements with inverted subjects in the Spoken Basque Corpus which seem to be instances of antitopics.³⁷

As we can see the largest group contains those which contain an identificational focus in a copula construction. We can see an example in (4.60).

- (4.60) 93C1B06
 200b “ni ni izan naiz la- lapurra,”
 I I be:PFV I.am th- thief:DEF
 “Me, I was the thief”

There is no doubt that the pronoun *ni* “I” is the focus of the assertion, but is the postverbal nominal an antitopic or is it simply a predicate that has been ‘bumped’ into the postverbal (assertion-tail) area by a subject focus? The latter might choice might seem to make more sense. On the other hand, with the very similar sentences in which the focus is not a nominal, but a locative, as in (4.61), we may want to say that the postposed nominal is the subject, since there is no other candidate for that role.

- (4.61) 93C3A09
 5 eta= an dau bea.
 and there he.is he
 “and there he is”

On the other hand, there is one example of a copula verb with two nominals in which the focus is clearly not identificational, but predicational, and here there is no doubt that the postverbal nominal is an antitopic, as we can see in the example in (4.62).

- (4.62) 93C2B06
 214 “lapurra da oi!”
 thief:DEF he.is that(one)
 “That guy’s a thief!” (~ “He’s a thief, that guy”)

The large number of identificational focus constructions is due to the existence of two scenes in the Chaplin film which call for such an identification, a contrastive and a non-contrastive one.³⁸ It is also interesting that 18 out of the 23 referential identifications are in completive clauses.

The remaining antitopic clauses contain inverted subjects: 8 ergative subjects and 9 absolutive ones. Six of them have redundant epithet subjects, that is, phrases which describe in some way an already active or very accessible referent, such as *the guy*, or *that Chaplin guy*, in English. These epithet nominals were *tipoa* “the guy” (3 occurrences), *gizona* “the man” (1 occurrence), and *Txaplin onek* “this Chaplin (guy)” (2 occurrences). We can see an example in (4.63:106) below.

- (4.63) 93C3A10
 104 Eta azkenean,
 and end:DEF:LOC
 105 ez dakit - droga zen edo beste zerbait zen,
 not I.know.it drug it.was:WHETHER or other something it.was:WHETHER
 106 baino seko mareauta gelditzen-da **tipoa**. ez?
 but suddenly/completely dizzy he.remains/becomes guy:DEF no?
 “And finally, / I don’t know if it was a drug or what it was, / but **the guy** ends up totally dizzy. You know.”

The example in (4.64:38) also seems to have an epithet like antitopic, namely the clause final absolutive phrase *iruak* “the three (of them)”. The referent in question is active at this time, having just been reintroduced in the immediately preceding context.

(4.64) 93C1B03

38 eben pasatzen-dira aurre-aurretik iruak,
uh they.go.by front-front:ABL three:DEF
“uh they pass right in front (of him) *the three (of them)*”

This example is also interesting because unlike in the majority of the other examples the focus is not preverbal, but rather extraposed. This is indeed common in Spanish, but it seems odd in Basque (cf. Spanish *Pasan, justo por delante los tres*). There is only one other example such as this in the corpus (93C1B06:60). (The speakers in both cases are children whose first language is not Basque.)

In (4.65:124) we see another example of inversion and a highly topical referent in the absolutive role. Here the allative phrase in assertion-initial position is a somewhat contrastive focus.³⁹

(4.65) 93C2A07

120 Orduan okiña juten-da poliziana,
then baker:DEF he.goes policeman:DEF:ALL
121 ta esaten dio: X ‘Aizu!’
and he.says.it.to.him X? Listen!
122 neska izan-dala’,
girl:DEF she.has.been:COMPL
123 ta bueno
and well
124 Neskana juten-da **polizia**,
girl:DEF:ALL he.goes policeman:DEF
125 eta Txarlot libre uzten-dute. ...
and Chaplin free they.leave.him

“Then the baker goes to the policeman, / and he says ‘Hey! / it was the girl!’, / and well, / The policeman goes (to where) the girl (is), / and they free Chaplin.”

In Spanish too this sentence would be rendered with an antitopic, although the focus in this case would be postverbal, cf. *Va donde la chica el policía*, unless the focus was very salient, either contrastive or surprising, cf. *Donde la chica va el policía!*.

In (4.66:333) we see a clause with an inverted ergative subject phrase: *The movie continues a bit longer*.⁴⁰ Here too there is an assertion-initial focus (*pixkat geixio* “a bit longer”). The subject refers to the movie that the speaker is relating, which thus can be very easily accommodated.

(4.66) 93C2A09

333 Ta geo pixkat geixio kar- jarraitzen-du **pelikulak**,
and later bit:one more ? it.continues movie:DEF:ERG

334 baño moztu in-da.
but cut:PFV do:PFV-it.is

“And then the movie goes on a little bit longer, / but it got cut off.”

The final example I will present is the one in (4.67:25) below. The referent of the clause-final ergative nominal, *mutikoek* “the boys”, is very accessible. They had been active up until the previous clause.

(4.67) 93C3A08

24 e- a jartze-a berrize bizikletan, -
uh ? he.gets.on again bicycle:DEF:LOC

25 eta - deittu-itte-xobe **mutikoek**,
and call:PFV-do:IMPFV-they.do.it.to.him boys:DEF:ERG

26 e- e- esaten,
uh uh say:IMPFV

27 nola txapela laga-dun. ..
how hat:DEF he.has.left.it:SUBORD

“he gets back on the bicycle, / and **the boys** call to him, / saying, / how he’s left his hat behind.”

The verb is the focus here, as evidenced by the verb focus *egin* construction already mentioned. But notice that here it is not clear that the inverted ergative is an antitopic at all. Actually we could view the referent of the dative argument, the boy, as the (covert) topic, which would entail that the ergative is not an antitopic, but rather a rheme-internal (non-setting) complement which is less focal than the verb and thus is placed in assertion-

tail position. In other words, the sentence is pragmatically ambiguous. (Here too, the Spanish version of this sentence would also have an inverted subject, cf. *y le llaman los chicos*).

As we can see, there are several complicating factors which make the issue of inverted 'subjects' a complex area to study. It is clear that in all these cases the subjects' referents are not foci, as in trulythetic sentences, but non-focus complements whose inversion is motivated by the presence of a very focal focus, typically in rheme-initial (and clausal) position. On the other hand, it is not always clear to what extent the referents of these inverted subjects have properties of topics, or to what extent the properties of topics are relevant in this context of emphatic assertions. In Chapter 5 we will look at other examples inverted topics. I will now turn to the written Basque corpus, and next to other written corpora where examples of subject inversion are somewhat more numerous than in the spoken corpus.

4.3.6 Antitopics in narrative written style: the Written Basque Corpus

In writing, the placement of a topic-like element after the verb is also clearly correlated with the presence of a salient focus element in assertion-initial position, and referents which are quite accessible, or else which can be easily accommodated. The higher the degree of salience of a focus, the more likely an accessible topic will be to invert, and the more likely it will be to be interpreted as a contrastive topic if it doesn't. But here too we find a fairly large number of cases in which it is not clear that the inverted subject is an antitopic or an inverted non-topic argument.

As we can see in Table 4-9, in the Written Basque Corpus there are 121 cases of ‘subject inversion’, that is cases in which the prime topic candidate (ergative or absolutive) is in postverbal position. (All these sentences are listed in Appendices 4.2 and 4.3.)

Type	No.	Notes
Post-quote	48	Other codes = <i>Post_Quote</i>
Comment-topic (copula)	36	Predicate Arguments = <i>Copula</i>
Comment-topic (non-copula)	19	Information Structure = 9 <i>Invert_Topic_A</i> , 10 <i>Invert_Topic_S</i> ; 2 are 3 rd person desideratives
Absolutive (object) topic	3	Other codes = <i>Object_Topic</i> ; Information Structure = <i>Covert_Topic_O</i>
Dative topic	12	Other codes = <i>Dative_Topic</i> ; Information Structure = 11 <i>Covert_Topic_I</i> , 1 <i>Overt_Topic_I</i>
Thetic-like	3	Information Structure = <i>No_Topic??</i>
TOTAL	121	(62 main clauses; 48 post quote clauses, 6 dependent clauses)

Table 4-9: Affirmative, main and dependent clauses with ‘subject’ (prime topic candidate) inversion in the Written Basque Corpus.⁴¹

The single most common type of ‘subject inversion’ in the written corpus, 2/5 of all, is cases in which a quote is in focus position. In written narrative, much like in English narrative, but unlike in spoken narrative, the verb of saying overwhelmingly follows the quote, and in this construction when the ‘subject’ (the source of the quote) is overt, it is always postverbal, as in (4.68).

(4.68) *Behi euskaldun baten memoriak*, p. 14

‘Sinestea libre al da?’ galdetu-nion **nik** Bidani hari

believe:NOMIN:DEF free QN it.is ask:PFV-I.had.it.to.her I:ERG Bidani that:DAT

“ ‘Is one allowed to think freely?’ I asked Bidani.”

In spoken Basque quotes or direct speech complements such as this would almost always be extraposed and even dislocated into intonation units of their own, following the main clause (in sentence tails).

This example is interesting for a number of reasons. For one, the referent of the ergative argument is not only very topical, its overt mention is unnecessary since the preceding verb already codes that the referent of the ergative argument is the speaker. Overt mention of first and second subject pronouns is typically associated with their referents being contrastive topics. In this case the pronoun could conceivably be seen as contrastive since this question is a reaction to something that Bidani, the writer's interlocutor, had previously said to her. Finally, the order of the ergative and dative arguments following the verb could be reversed without any clear pragmatic changes in the assertion obtaining.

The next most common situation in which subjects are inverted in the written corpus have copula verbs, such as those in examples (4.69), (4.70) and (4.71) below. All of these assertions are emphatic, or exclamatory, non-continuous, disruptive assertions, and they all have very salient focus constituents in rheme-initial position.

(4.69) *Behi euskaldun baten memoriak*, p. 27
nire-nireak ziren bere ahoko ile haiek.
my/mine-my/mine:DEF:PL they.were his mouth:DEF:GEN hair those
“Those hairs in his mouth were my very own” (“They were my very own, those hairs in his mouth”)

(4.70) *Behi euskaldun baten memoriak*, p. 16
holakoxea zara zu!
thus:GEN:EMPH:DEF you.are you
“That's what you're like.” (“You're just like that”)

(4.71) *Behi euskaldun baten memoriak*, p. 18
Ni nintzen behi hori.
 I I.was cow that
 “That cow was me.” (“It was me, that cow”)

It is not surprising that the antitopic construction is so prevalent with copula verbs, since first of all copula constructions are used to identify referents which have certain properties (identificational focus) or to match certain properties with a referent (predicational focus). Here, it seems the focus in the first two assertions could be either predicational or identificational, depending on the context, whereas in (4.71) it would have to be identificational. Notice also that when the verb is a rather empty copula the focus element is that much more salient since it doesn't have to compete with the verb for the status of being the most informative element of the assertion.

As we can see in Table 4-10 copula clauses with an inverted topic such as the ones we have just seen constitute a significant percentage, 15.9%, of all finite statement copula clauses in the Written Basque Corpus, more than twice as common as topicless existential (thetic) clauses. (This list doesn't include some of the asserted, non-main clauses included in the previous table.)

Information structure	No.		Information Structure Code
Topic Comment	61	30.3%	<i>Overt_Topic*</i>
(TopicCOVERT) Comment	90	44.8%	<i>Covert_Topic*</i>
Comment Topic	32	15.9%	<i>Invert_Topic*</i>
Comment (i.e. thetic, existential)	18	9.0%	<i>Non_Topic*</i>
TOTAL	201	100.0%	

Table 4-10: Main and dependent finite copula clauses in the Written Basque Corpus.⁴²

The antitopics in these copula statements aren't always active or recently mentioned referents, but they are always at least easy to accommodate from the context.

Thus in (4.72) the subject's referent "the mountains" is certainly accessible in the story, an ever-present part of the scenery, although it hadn't been mentioned recently.

- (4.72) *Behi euskaldun baten memoriak*, p. 24
otsoz beteta daudela mendiak.
wolf:INSTR fill:PFV:ADV they.are:COMPL mountains:DEF
"that the mountains are full of wolves"

A similar thing can be said about the antitopic in (4.73), "thinking".⁴³

- (4.73) *Behi euskaldun baten memoriak*, p. 26
"Hori da pentsatzea,
that it.is think:NOMIN:DEF
"Good thinking!" (lit. "That's thinking")

In (4.74) below the antitopic, i.e. the proposition *crossing on foot or by bicycle*, is not discourse given, though in this context it is quite easy to accommodate: the writer is discussing ways of crossing to the other side of the bay.

- (4.74) *Kuba triste dago*, p. 19-20
Debekatua dago oinez edo bizikletaz zeharkatzea
forbid:PFV:DEF it.is foot:INSTR or bicycle:INSTR cross:NOMIN:DEF
"It's forbidden to cross on foot or by bicycle"

The antitopic in (4.75), *the free elections*, mentioned by one of the characters in the book, hasn't been mentioned before in the text, and thus the idea may be new to the reader, but it can be surmised that it is accessible or easy to accommodate for the intended hearer, especially since it is definite.

- (4.75) *Kuba triste dago*, p. 22
'<Duela gutxi> izan-dira hauteskunde libreak,'
it.has.it:COMPL little they.have.been election free:PL:DEF
"Free elections were just a little while ago."; "Free elections took place recently." (Cf. "Just a little while ago you had free elections.")

Antitopics, or antitopic-like inverted subjects, are also found with regular (non-copula) predicates, but those cases are proportionally less frequent. As we saw in Table 4-9, only at most 19 such sentences can be found in the Written Basque Corpus (see Appendix 4.3). We can see one such as in example (4.76) below.

(4.76) *Behi euskaldun baten memoriak*, p. 14
 holaxe hitzegiten-du nire barruko ahotsak,
 thus:EMPH it.speaks.it my inside:GEN voice:DEF:ERG
 “Thus speaks my inner voice”; “That’s how my inner voice speaks”

(4.77) *Behi euskaldun baten memoriak*, p. 28
 Barre egin-dugu denok.
 laughter make:PFV-we.have.it all:PL:ERG
 “We all laughed.”

Two of the examples that I have classified as antitopic constructions are actually not statement assertions, but desiderative assertions (cf. English *Let...*, *May...*) (cf. Appendix 4.3:8-9). Desiderative assertions, just like imperatives, as we shall see in Chapter 7, are typically emphatic, with the focus on the verb, and thus it is not surprising that their subjects are inverted. In this clause type it is the verb or the polarity and not some other complement which is the assertion-initial focus receiving the assertion-initial accent, as in (4.78).

(4.78) *Behi euskaldun baten memoriak*, p. 14
 “Ikas-dezala **munduak** behiaren haunditasuna!”
 learn:ROOT-it.may.have.it:COMPL world:DEF:ERG cow:DEF:GEN greatness:DEF
 “ ‘May the world learn (about) the greatness of cows.’”

As we can see in this sentence, not all antitopic nominals are in clause-final position. In five of the nineteen tokens (four of the seventeen non-desiderative cases)

there is an additional postverbal element and in all these cases this other element is final and the presumed antitopic is postverbal, as we can see in (4.79) below.

- (4.79) *Behi euskaldun baten memoriak*, p. 26
halaxe esan-zion otso batek besteari:
thus:EMPH say:PFV-he.had.it.to.him wolf one:ERG other:DEF:DAT
“Thus spoke one wolf to the other: ...”

In two of these five cases, the clause-final, post-antitopic element doesn't have an idea that is truly discourse given, as in this same sentence in (4.79) for instance, although it can certainly be easily accommodated.

In addition to these seemingly antitopic construction, there are an additional 15 sentences which also have postverbal subjects, and a salient focus, but here another argument of the verb, an absolutive object or a dative, seems to be filling the topic role in those sentences. In (4.80) we can see an example of one such statement.

- (4.80) *Kuba triste dago*, p. 32
Harritu egin-nau honek,
surprise/amaze:PFV make/do:PFV-he.has.me this:ERG
“He really amazed me, this guy.”

The referent of the inverted ergative argument is active here, and so is the object's referent (the speaker). The verb is highly focal, as marked by the *egin* verb focus construction. Thus this sentence exemplifies the neutralization that exists between what might seem to be inverted topics (antitopics) and inverted arguments which are not topics. Both types of arguments, of course must be quite accessible.

The following example in (4.81) is similar to the preceding one. Here the topic is the referent of the covert dative argument.

- (4.81) *Kuba triste dago*, p. 19
 Asko gustatzen-omen-zaio turistekin jardutea;
 much please:IMPFV-QUOTATIVE-it.is.to.him tourists:DEF:COM
 do/be.busy:NOMIN:DEF
 “He really likes to hang around tourists” (~ “It really pleases him to hang around tourists.”)

The postposed absolute expresses an idea which hasn’t literally been mentioned but which is quite easy to accommodate in this context.

Finally, there are three sentences in the written corpus which contain inverted subjects, but which it is not clear whether they are topicless (thetic) sentences with verb focus or antitopics. In the one in (4.82), it is fairly clear that this is not a subject focus construction.

- (4.82) *Behi euskaldun baten memoriak*, p. 30
 Heldu-da ordua,
 arrive:PFV-it.is hour:DEF
 “The hour has come.”

The accent in this sentence must be on the verb and not on the postverbal subject. This can only be due to the fact that the assertion is emphatic, i.e. it is like an exclamation (cf. Chapter 7). The same thing is true of the next example in (4.83c).

- (4.83) *Behi euskaldun baten memoriak*, p. 20
- a. ... behia dago.
 ... cow:DEF it.is
 - b. Behia, eta ez, esate baterako, xator bakar bat ere.
 cow:DEF and not say:NOMIN one:ALL:GEN mole single one also
 - c. Udazkenaren epeltasun hartan ibili ziren xatorrak han eta hemen, ...
 fall:DEF:GEN warmth that:LOC go:PFV they.were moles:DEF there and here
- “(a) The cow is (here). (b) The cow, and not, for example, a single mole. (c) In the warmth of the fall there were indeed moles here and there, ...” (= “Moles did appear here and there.”)

It seems to me that these sentences are emphatic, which is why the finite verb is the focus. The inverted ‘subjects’, on the other hand, do not seem to be antitopics. I will argue in Chapter 7 that these are perhaps best seen as emphaticthetic (topicless) sentences. The fact that the ideas expressed by the inverted subjects are quite accessible (*the moles*) or stereotypical (*the hour*), in other words, low in focality, warrants their not being the focus of thethetic assertion and allows the finite verb to fill that role instead.

To summarize, in written Basque inverted subjects are always unaccented, and topical (accessible), unless, of course, they are extraposed. Whether they are topics or not may be a moot question since inverted subjects always have the same referential properties, regardless of the predicate type and of whether there another argument’s referent is presumably filling the topic role. Let us look next at similar cases of subject inversion in journalistic writing, where they tend to be more common than in narrative (at least in some styles).

4.3.7 Antitopics in journalistic written style

As we have just seen, the Written Basque Corpus displays a higher percentage of antitopic constructions, or inverted non-focus subjects, than the Spoken Basque Corpus. (Focus extraposing, including subject focus extraposing, on the other hand, is much more common in speech than in writing, as we will see in Chapter 6.) One might have expected this to be the case for a number of reasons. First of all, writing lacks the phonetic and intonational resources of speech to mark an assertion as emphatic and thus inversion might be expected to be used for this purpose in writing. Secondly, it might seem to be the case that accessible topics are more likely to be coded overtly in writing

than in speech, which would contribute to the number of candidates for inversion. Finally, the written corpus, unlike the spoken narrative corpus, contains a fair amount of conversation, or dialogue, and emphatic assertions seem to be more common in conversation than in narrative.

I have also found that inverted subjects such as the ones we have seen, can be quite common also in some newspaper articles and other similar written genres, at least in some journalists' styles. In fact the construction is so common, relative speaking, in some cases, that it seems that in those contexts it loses some of its markedness (emphasis).

To illustrate how the antitopic construction can be exploited in some contexts, I will give some examples from a short newspaper article which has a particularly large number of such antitopic constructions, 12 of them to be exact out of 31 declarative assertions, or almost 2/5 of the total number of statements. The article, which is reproduced in Appendix 4.4, is about the public *presentation* of the recent first-time *publication* by the *Iztueta Foundation* of a 19th century *manuscript* written by *Iztueta*. All of these referents are active or very accessible ever since they are mentioned, or suggested, by the article's title.

In (4.84) we can see the first two sentences of the article's second paragraph, right after the publication of the book is announced (the focus accent is underlined, the postverbal topic is in bold).

- (4.84) *Diario Vasco*, 9/27/96, Felix Ibargutxi
- a. Atzo egin-zen liburuaren **aurkezpena**, Donostiako KMn.
 yesterday do:PFV-it.was book:DEF:GEN presentation:DEF, Donostia:GEN
 KM:LOC
- b. Hain zuzen ere, hantxe gordetzen-da **eskuizkribua**
 so correct also there:EMPH keep:IMPFV-it.is hand:writing:DEF
 “The presentation of the book took place yesterday, in Donostia’s K[oldo]
 M[itxelena Culture Center]. Right there is where the manuscript is kept.”

Notice that the inverted subjects in these two statements refer to the presentation and the manuscript respectively. In (4.85a&b) we see two further statements displaying inverted subjects, immediately following the ones just mentioned.

- (4.85) *Diario Vasco*, 9/27/96
- a. baina hala eta guztiz ere balio handia du **hiztegiak**.
 but thus and completely also value great:DEF it.has dictionary:DEF:ERG
- b. Gainera, beti da ona **argitara eman gabe zegoen zerbait argitaratzea**
 additionally always it.is good:DEF light:ALL give:PFV without it.was:REL
 something publish:NOMIN:DEF
 “but despite all that the dictionary of great value. / Besides, it’s always good to
 publish something which had hitherto been unpublished.”

Notice that most of these antitopics are different formulations, or verbalizations, of the same two very active referents: *the dictionary* and *the publishing of the dictionary*. In the following example in (4.86) the antitopic is the *author of the dictionary* (Iztueta), the other very active referent that this article revolves around.

- (4.86) *Diario Vasco*, 9/27/96
- Mujikak komentatu-zuenez, inguru-mingurutan galtzen-da batzutan **Iztueta**, ...
 M-ERG he.commented.it:ADV roundabout/circumlocution:LOC he.gets.lost
 sometimes Iztueta
 “As Mujika commented, Iztueta sometimes gets lost beating around the bush.”

The very first sentence of this article, which we can see here in (4.87), is also quite interesting. This OVA statement announces the publication of this old *manuscript* (focus) by the *Iztueta Foundation* (antitopic).

- (4.87) *Diario Vasco*, 9/27/96, beginning of 1st paragraph, after the title
Juan Ignazio Iztuetaren *Vocabulario de vascuence* argitaratu-du **Iztueta
Fundazioak**.
JII dictionary of Basque publish:PFV-it.has.it Iztueta Foundation:DEF:ERG
“The Iztueta Foundation has published Juan Ignazio Iztueta’s *Vocabulario de
vascuence*.”

This inversion highlights the rhematically salient element, which is the (preverbal) object. On the other hand, the ‘antitopic’ (*the Iztueta Foundation*) was presumably an unidentifiable entity at this point, although it seems to be easy to accommodate in this context, or else that the writer is assuming familiarity with the Foundation. It does seem that the postposing of the topic here serves the purpose of keeping the reader’s attention on the focus (the book) without the distracting presence of an assertion-initial topic (the publisher).

On the other hand, notice that although the Foundation, the inverted subject’s referent, is the agent here, it is not very likely that it is the pragmatic topic of this assertion. In other words, it doesn’t seem that this assertion is at all *about* the Iztueta Foundation, but about the publishing of the manuscript by the foundation. In other words, it seems to me that this construction is being used here in an extended sense of the original (emphatic) one in order to diminish the centrality of the topic and to emphasize the whole event.

The English equivalent of this particular type of statement with subject inversion would be perhaps a passive construction with subject focus, such as in example (4.88), with subject focus.

(4.88) A dictionary has been published by the Ford Foundation.

A difference might seem that such focus subjects are somewhat odd when they are as long and complex as the actual focus constituent in (4.87). Of course, this Basque example itself is perhaps not much less odd sounding either.

It seems to me that the repeated use of the antitopic construction in this article, in cases where it would not be expected in normal spoken or written narrative for instance, is a specialized use of this construction in a certain genre or style within a genre. The ‘overuse’ of this normally marked (emphatic) construction, does seem to carry with it a lessening in the markedness associated with the subject inversion (antitopic) construction in normal contexts.

4.4 TOPICALIZATION IN BASQUE

4.4.1 Introduction

As I mentioned earlier in this chapter, Basque is what Li and Thompson (1976a) call a topic-prominent language. To begin with, Basque doesn’t have any grammatical coding, either morphological or positional, for subject arguments. Furthermore, Basque doesn’t have grammatical or role changing constructions which are used in order to make a non-default argument the topic.

In addition, Basque has certain properties which highlight the notion of topic and which are missing from less topic-prominent languages. Thus, for instance, Basque makes extensive use of ‘topicalizers’, that is, particles or adverbials which separate the topic and other settings from the assertion proper. More interestingly, Basque topicalizes elements which we do not normally expect to see topicalized. I am referring not just to assertion tail (non-focus) accessible complements, which as we already saw are ‘optionally’ topicalized in many languages, but also to elements which in other languages would be the focus of the assertion, such as additive foci (*also*) and a type of exclusive focus (*only*).

In this section I will first look at Basque topicalizers, and in particular the additive topicalizer *ere* “also” (the exclusive topicalizer *besterik* will be discussed in Chapter 7 when we look at negatives, since the exclusive particle *only* is rendered in Basque most typically as *other not*). Finally I will look at some interesting existentials in northern dialects of Basque in which a rather inaccessible referent is coded as a topic and not as a focus as in other dialects and other languages.

4.4.2 Discourse connectors and topicalizers

Every language has discourse particles, or connectors, which help to situate an assertion with respect to others in the discourse. The most common type are coordinating conjunctions, which join assertions of equal status (often temporally sequenced). Others code a dependence relationship (subordinating conjunctions) and join dependent clauses to main clauses, be they asserted (in sentence tails) or not (in sentence heads). In addition, there are some conjunctions which code a semi-dependence, or paratactic,

relations of one assertion to another assertion (cf. Mann and Thompson 1986a, 1986b, 1987, 1988, 1989, 1992 for types of relatedness among sentences, whether overtly marked or not). These are particles such as *however*, *on the other hand*, *at least*, *therefore*, *so*, *besides*, and the like.

These particles typically go at the beginning of the added clause, i.e. between the clauses, in an iconic way. This is the typical position for conjunctions in Basque as well, though a few of them are clause-final, such as especial uses of the (normally clause-initial) conjunctions *eta* “and”, and *baina* “but”, with a causative and concessive meaning, respectively, or they go before the verb, such as causative *bait-* and conditional *ba-*. Clauses with final and preverbal connectors seem to be more tightly integrated to the main clause, that is, they are found in what I call sentence-tails.

Sometimes these ‘trailing’ assertions involve a setting change, very often a topic change, which is sometimes a change of a contrastive nature. In these cases the particles are often placed *after* the new setting or topic and before the assertion, as in (4.89a), though not necessarily, cf. (4.89b).

- (4.89) a. Mary came. John however didn't come.
b. Mary came. However John didn't come.

These settings which come before the connector may be dislocated, but they need not be. We can see that in situations such as the one in (4.89a) the connectors act as ‘topicalizers’ or topic or setting ‘shifters’ (cf. Euskaltzaindia 1987b:495; Osa 1990:249).

The Basque Language Academy has published a very detailed classification and description of clausal connectors in Basque (Euskaltzaindia 1990). They distinguish five different types of such paratactic, quasi-dependent sentences and connector types, as we

can see in Table 4-11. These connectors join clauses together, or rather they connect one clause to another with a certain dependency function added, such as addition, contrast, concession, or consequence.

Paratactic clauses	Sample junctors
Additives (<i>emendiozkoak</i>)	<i>eta</i> “and”, <i>ere</i> “also”, <i>gainera</i> “besides”, <i>bestalde</i> “on the other hand”, <i>behintzat</i> “at least, at any rate”
Alternatives (<i>hautakariak</i>)	<i>edo</i> “or”, <i>bestela</i> “otherwise, if not”, <i>osterantzean</i> “otherwise, or else”
Adversatives (<i>aurkaritzakoak</i>)	<i>ordea</i> “but, however”, <i>berriz</i> “again; however”, <i>ostera</i> “however, but”, <i>aldiz</i> “but, on the other hand, however”, <i>bitartean</i> “meanwhile”, <i>dena den</i> “anyway, in any case”, etc.
Consecutives (<i>ondoriozkoak</i>)	<i>beraz</i> “therefore, so”, <i>bada (ba)</i> “so, then; well”, <i>hortaz</i> “therefore, consequently”, <i>orduan</i> “then”
Causatives (<i>kausazkoak</i>)	<i>zeren</i> “because; for”, <i>zergatik</i> “for” (lit. “why”), <i>alabaina</i> “in fact”, <i>bada</i> “so”, <i>izan ere</i> “in fact”

Table 4-11: Classification of the most common paratactic clause types and connectors in Basque by function; from Euskaltzaindia 1990. Some are topic-shifters as well.

Basque seems to make a significant amount of use of topic (setting) shifters. Take for instance the rather common connector *behintzat* “at least, as for”, which the Academy classifies as an ‘additive’.⁴⁴ This connector always follows the setting element, as we can see in (4.90), in either an affirmative context (4.90A’, from Euskaltzaindia 1990:66), or a negative one in (4.90A’’).

(4.90) Q: Etorri da norbait?

come:PFV it.is someone

A’: Koldo eta Jone **behintzat** etorri-dira

Koldo and Jone at.least come:PFV-they.are

A’’: Koldo eta Jone **behintzat** ez-dira etorri

Koldo and Jone at.least not-they.are come:PFV

Q: “Has anybody come?” / A’: “Koldo and Jone at least have come. (I don’t know about any others).” / A’’: “Koldo and Jone at least haven’t come.”

Other connectors apply to the whole assertion, without involving the topic, although the topic may indeed change, e.g. the additive connector *gainera* “besides, in addition”. Another additive connector, *ere* “also”, on the other hand, is always a topic shifter, as we will see in the following section.

4.4.3 The topicalizing, additive particle ERE ‘also’

4.4.3.1 Introduction

Ere “also, too” is an additive particle or connector, similar to English *too*, but which always follows a setting or topic phrase, and immediately precedes the assertion proper (cf. Rebuschi 1984/1982, cited in Oihartzabal 1984:360). We can see an example of this particle in (4.91b) (from Euskaltzaindia 1990:35).

- (4.91) A: Miren eta Maddalen afaltzera joan-ziren.
M and M have.dinner:NOMIN:ALL go:PFV-they.were
B: Peru ere haiekin joan-zen.
P also they:COM go:PFV-he.was
A: “Miren and Maddalen went to have dinner.” B: “Peru too went with them”
(= “Peru went with them too”).

This connector is used with any kind of setting constituent, not just topics, and seems to be highly grammaticalized. In other words, the settings need not be dislocated if they wouldn’t be dislocated under normal circumstances, i.e. if the additive connector wasn’t there.

What is interesting about this particle and this construction is that this is the only way in Basque to contrast additive elements of this sort. In other words, unlike in other languages where an additive element may be either the topic or the focus of the assertion,

in Basque it must always be the topic. Thus, in English for example we find different possibilities for adding a new referent to a very accessible, open proposition, by means of the additive particles *also* and *too*, some of which are exemplified in (4.92) below.⁴⁵ (I am not dealing here with the situation in which a whole new assertion, and thus proposition, is added to the discourse, e.g. *Also, John came*, for which a different connector would be used in Basque).

- (4.92) a. Also John came. (subject, focus)
 b. John too came. (subject, focus)
 c. John too came by. (subject, topic)
 d. John ate beans too. (object, focus)
 e. John also ate beans. (object, focus)
 f. ? Beans too John ate. (object, topic)
 g. ?* Beans John also ate. (object, topic)

As we can see, the preferred way to place an additive element in English is as in focus position. In Basque, on the other hand, in each of these situations the new, somewhat contrastive element would have to be a topic or setting element and could never be the focus of the assertion. That means that the additive particle can only be a topicalizer and never a focalizer, as in other languages.

In (4.93) below we see an example of an additive assertion of this sort (from Euskaltzaindia 1990:38). We can see that the content question in (4.93A) (*Who is going hiking?*) receives the reply in (4.93B) in which the focus corresponds to the ‘new’ element (*Antton is going*). Notice that this element must be the focus of the assertion and could not be the topic, in which case the synthetic, finite verb would have to receive the *ba-* particle.

- (4.93) A: Nor doa gaur mendira?
 who s/he.is.going today mountain:DEF:ALL
 “Who is going hiking today?”
- B: Antton doa (*Antton badoa.)
 Antton he.is.going (*Antton EMPH:he.is.going)
 “Antton is going.” (*Antton = Focus*)
- C: (Eta) Amaia ere badoa. (*Amaia ere doa.)
 and Amaia also EMPH:she.is.going (* Amaia also she.is.going)
 “(And) Amaia too is going” / “Amaia is going too” (*Amaia = topic*)

However, notice that if we want to add another element to the answer in (4.93B), such as is done in (4.93C) (*Amaia is going too, Also Amaia is going*), the new element must be the topic, regardless of the fact that it is the ‘new’ part of the proposition and that the predicate is the ‘old’ part. Notice that here the *ba-* particle on the finite verb is required since the synthetic verb is in focus position (i.e. the polarity is the focus by default).

In (4.94) we can see another example of an additive topic, one which involves a generic referent which is supposedly easy to accommodate in this context (from Odriozola and Zabala 1992:48).

- (4.94) Anelidoak ezezik, moluskuak ere, ikasiko-ditugu
 annelids:DEF not.only mollusks:DEF also, learn/study:FUT-we.have.them
 “Not only annelids, we will also study mollusks.” (“We will study not only annelids, but mollusks as well.”)

The comma following the additive topic was in the original source for this sentence. I do not think, however, that a full intonational break is needed, such as the one following the initial phrase (*Anélidoak ezezik*). In other words the topic in the *ere* construction need not be dislocated and seems to be quite integrated into the clause.

4.4.3.2 *ERE* and unusual topics

The fact that the topicalizing *ere* construction is the only means of expressing additive elements to an assertion is responsible for some ‘unusual’ topics which we do not often expect to see topicalized, such as a non-finite clausal complement, for instance, as in the example in (4.95) below (from Osa 1990:263).

- (4.95) Esaten bakarrik ez, entzuten ere badaki harek.
say:IMPFV only not, listen:IMPFV also EMPH:he.knows.it that:ERG
“He doesn’t just know how to speak, he knows how to listen too.”

Notice that although the additive, contrasted element *listening* is ‘new’, the fact that it is in the same semantic set as the active idea *speaking* makes it somehow accessible enough and worthy of being topicalized.

When the focus is the verb itself, verb topicalization is also one possibility, as we can see in the second clause in (4.96). Notice that the verbal idea it contrasts with (*hitze*gin “speak”) is also the topic of the first clause. The verb here is repeated.

- (4.96) Hitzegin bakarrik ez-du egiten, entzun ere entzuten-du.
speak:PFV only not-she.has.it do:IMPFV listen also listen:IMPFV-she.has.it
“She doesn’t just speak, she also listens” (= “She listens too.”)

With verbs, however, there is another possibility, namely verb focusing, as in (4.97), and in this case the particle *ere* is not a topicalizer.

- (4.97) Ez-du bakarrik hitzegiten, entzuten-du ere.
not-she.has.it only speak:IMPFV listen:IMPFV-she.has.it also
“She doesn’t just speak, she also listens.”

Notice that the verbal idea in the first clause that the additive verbal idea contrasts with is also be focused. (In speech *ere* reduces to *e* or *be* in some dialects, thus the second clause would be commonly rendered in Gipuzkoan as *entzuteue* “she also listens.”)

The example in (4.98) (Euskaltzaindia 1990:41, ex. 57) is also unusual since the topicalized element is a complement clause, and it quite clearly provides new information which could not in any way have been accessible to the hearer.

- (4.98) Lehendakariak Madrilerako joateko asmoa duela ere irakurri-dugu
president:DEF:ERG Madrid:ALL go:NOMIN:GEN intention:DEF he.has.it:COMPL
also read:PFV-we.have.it
“We have also read that the (Basque) president intends to go to Madrid.” (I.e. in addition to reading other things.)

In other words, this kind of topicalized element is reminiscent of the non-identifiable dislocated topics which we saw above. There is one way, however, in which the proposition expressed by the topicalized completive clause is accessible (easy to accommodate), namely that it belongs to a set of things which have been read, some of which the hearer is already familiar with (i.e. are discourse given).

The additive particle when combined with the negation operator produce interesting combinations which often have alternative means of expression in other languages. Thus *also X not* is equivalent to *not X either*. In (4.99) we can see an example of this type of assertion in Basque, which also must make use of the particle *ere*.

- (4.99) Liburu hori ere ez-dut irakurri.
book that also not-I.have.it read:PFV
“I haven’t read that book either.”
OR “I haven’t even read that book.” (= “I haven’t read even that book.”)

In Basque the combination *also not X* can also be used to express the meaning *not even X* in the right context, as we can see in (4.100).

- (4.100) Liburu bat ere ez-dut irakurri.
book one also not-I.have.it read:PFV
“I haven’t even read one book.”

This is the kind of additive element which most languages would probably not code as a topic/setting element, given its low topicality. But in Basque, as long as the sentence is in a context in which the idea of reading books is accessible to the hearer, topicalization of the additive element is a possible strategy. In this case, however, this is not the only strategy and at least in some contexts, the additive phrase can be assertion internal, as in (4.101), although it is not clear whether it is the focus or not (cf., e.g., Euskaltzaindia 1993:64).

- (4.101) Ez-dut liburu bat ere irakurri.
not-I.have.it book one also read:PFV
“I haven’t even read one book.”

From this construction with *bat* “one” + *ere* “also” in a negative context derives the standard negative expression equivalent of English *none*, *no one*, *nothing*, etc. In other words this expressions is not composed of a negative operator having scope over an assertion’s fragment (focus) placed to its right, as in English and other languages, cf. *not[one]*. Rather it is composed of a topic fragment (one) and the negative word used as a pro-assertion (not), as the focus fragment of an active open proposition:

- (4.102) English: not (even) one.
Basque: one also not.

In (4.103) we see the Basque expression used in context, with optional elision of the remainder of the predicate.

- (4.103) Q: *Libururik irakurri-duzu?*
 book:PART read:PFV-you.have.it
 A: *Bat ere ez(-dut irakurri).*
 one also not(-I.have.it read:PFV)
 Q: “Have you read any books?”; A: “None”, “Not one” (lit. “One also not”)

Actually, the combination of *bat* “one” and *ere* “also” has become lexicalized and can now also be used as a focus in a negative assertion, such as in (4.104).

- (4.104) *Ez-du batere irakurri.*
 not-she.has.it one:also(anything) read:PFV
 “She didn’t read anything.”

This *batere* cannot be used as a nominal modifier (*batere libururik* “any book(s)”), though it is used as an adjectival modifier in negative contexts (*batere erraza* “easy at all”).

Finally, let us look at one additional interesting use of the particle *ere*, namely with conditional (*if*) clauses in order to form concessive (*even if* and *even though*) clauses, which are prototypical setting clauses. In (4.105) we can see an example of a sentence-head concessive, which is ambiguous between a realis sense and an irrealis (counterfactual) sense (from Zubiri and Zubiri 1995:702, ex 6.928).⁴⁶

- (4.105) *Sinesten ez-baduzu ere, egia da.*
 believe:IMPFV not-if:you.have.it also truth:DEF it.is
 “Although/Even if you don’t believe it, it’s true.”

This is particularly interesting in light of the claims which have been made in the literature about conditionals being topics, promoted by the fact that in some languages conditionals receive the same marking as topics do (cf. Haiman 1978a, 1986; Thompson and Longacre 1985; Jacobsen 1992).

It is important to note that additive phrases, just like all other settings and topics, can also be postposed, that is, they can be placed after the focused element in the assertion tail. This, as usual, requires that the postposed element be quite accessible and that the assertion be emphatic. In (4.106) below we can see an example of a sentence with a postposed additive topic (from Euskaltzaindia 1990:40, ex. 50).

- (4.106) Iruñera joango-da gaur Miren ere.
Iruñea:ALL go:FUT-she.is today Miren too
“Miren too will go to Iruñea today.” (= “Miren ere gaur Iruñera joango da.”)

In other words, *Miren* here is an antitopic.⁴⁷ The example in (4.107) contains a backed concessive, which, again, is distinguished from a sentence tail constituent or clause in that it is not accented.

- (4.107) Konpondu-dut, zirudien baino zailagoa izan-bada ere.
fix:PFV-I.have.it it.seemed:REL than difficult:MORE:DEF be:PFV-if:it.is also
“I fixed it, though it was more difficult than it seemed.”

In this section I have described in general lines the usage of the additive particle *ere*. The purpose of my exposition has been to show that this particle is a ‘topicalizer’ and that therefore some unusual, or unexpected, elements end up being topicalized. In practice matters are a bit more complex. To begin with speakers of north-eastern dialects of Basque seem to use this particle differently from the majority of Basque speakers, who speak western (‘non-north-eastern’) dialects. In north-eastern dialects it seems that this particle is used as, or can be used as, a focalizer (cf. Oihartzabal 1984, Euskaltzaindia 1990:41).

In western dialects too this possibility seems to be creeping in through the use of the particle *ere* in combination with the ‘particle’ *baita* (< bai + eta) ‘yes and’. This use

of *ere* as a focalizer is apparent in an alternative of the ‘regular’ construction in (4.108a) below, in which the assertion proper is elided, and only the additive topic is overtly expressed (cf. “Jane too”). In the alternative to such an additive fragment assertion, shown here in (4.108b), the particle *ere* is preposed and attached to *baita* (cf. “also Jane”).

- (4.108) a. Peru gaur etorriko-da, **baita** Jone **ere**.
 Peru today come:FUT-he.is yes.and Jone also.
 b. Peru gaur etorriko-da, **baita ere** Jone.
 Peru today come:FUT-he.is yes.and also Jone
 “Peru will come today, and Jone too;” “... and also Jone.”

In this construction, the copulative particle *baita* has ‘attracted’ the additive particle *ere*, leaving the ‘topic’ behind. It seems that the *ere* particle in this case is being reinterpreted as a focalizer by Basque speakers, and that imperfect speakers are even spreading the new use to new contexts (cf. Euskaltzaindia 1990:46).⁴⁸

4.4.4 Low topicality topics in existentials

As I mentioned in the previous section, Oihartzabal (1984) has reported that the use of *ere* as a strictly topicalizing particle is not extensive to north-eastern Basque dialects. In these dialects *ere* is used as a focalizer, or so it seems from his examples, such as the ones in (4.109) and (4.110) (from Oihartzabal 1984:361).

- (4.109) Berriz *ere* diot.
 again also I.say.it
 “I say it again.”
 (4.110) (Gezurti eta) gaixto *ere* da.
 liar and bad/mischievous also he.is
 “(He’s a liar and) he’s bad too.”

The fact that there is no polarity particle *ba* on these synthetic verbs, shows that the element preceding the finite verb, and not the verb or the polarity, is the focus.

Oihartzabal, however, in the same article describes a different set of constructions in these dialects, which display a type of topicalization of low-topicality elements which is not typically found in the western/southern dialects. Actually the elements which are topicalized in these dialects would be foci in non-north-eastern dialects, and not merely as assertion-tail extra-rhematic elements.

The topicalization that Oihartzabal describes involves existential sentences and possessive-like existential sentences. Existential sentences, of course, can have eitherthetic or topic-comment structure, typically depending on whether the referent in question is identifiable or not, as in *There is a rule* vs. *The rule exists*. What we find in these dialects of Basque is that existentials are routinely expressed with a topic-comment information structure configuration, as in (4.111) (from Oihartzabal 1984:365, cf. also Euskaltzaindia 1987b:497).

(4.111) *Topic-comment existential; north-eastern dialects*

Ogia bada.

bread:DEF EMPH:it.is

“There is bread.” (Lit. “Bread there is”)

In this sentence, the fact that the finite verb has the emphatic particle *ba* shows that the preceding nominal is the topic and not the focus of the assertion.

The reason why these dialects topicalize the absolutive argument in these clauses seems to be related to the fact that, if they didn't, the resulting clause would be ambiguous. As Oihartzabal notices, when the absolutive argument is the focus instead of

the topic, as in (4.112), the clause means *it is bread*, with the absolutive acting as a predicate nominal, and not *there is bread*.

(4.112) *Predicate nominal absolutive; covert absolutive referent*

Ogia da.
bread:DEF it.is
“It is bread.”

In other words, an absolutive nominal in focus position with a copula verb must be interpreted as a predicate nominal and not as the subject focus of an existential assertion. In the non-north-eastern dialects, in which both absolutives would be the focus this is not a problem because they use a second copula verb, *egon* “be” to express the existential meaning, as in (4.113) below. In this assertion the absolutive is indeed in focus position.

(4.113) *Thetic (topicless) existential; non-north-eastern dialects*

Ogia dago.
bread:DEF it.is
“There is bread.”

There are some contexts in which in non-north-eastern dialects too we find existentials with topic-comment information structure of the type found in north-eastern dialects. That is, we find that subjects with referents which are low in topicality and which would normally be the focus of the assertion are instead topicalize, as long as they are easy to accommodate in the context in which they are used. Thus, for instance, it is common for taverns in the Basque Country to have signs such as the one in (4.114) indicating the availability of some particular food or drink.

(4.114) *Salda badago.*
broth/soup:DEF EMPH:it.is
“There is soup.”

Given that finding soup (*salda*) in such an establishment is not surprising, that is enough for the referent of the assertion's only argument to be considered topical enough to be the topic rather than the focus, as it would be in, say, English or Spanish.

However, if the subject is not the focus, being the topic is not the only possible alternative option. As we can see in (4.115) below, it can also be postposed, as an antitopic, if the assertion is emphatic, thus producing an assertion with VS order which, superficially is identical to the order found inthetic existential assertions in VO languages, in which the subject is the focus.

(4.115) *Badago salda.*
EMPH:it.is broth/soup:DEF
"There is soup."

In Chapter 5 we will see that this latter option is quite common in spoken Basque nowadays, being used even when the referent in question would not seem to be quite accessible enough and thus should have been the focus (as it would have been in standard written Basque). Inversion cannot apply, however, if the idea involved is very focal and not easy to accommodate. One possible motivation for the wide use of emphatic comment-topic structures or emphaticthetic sentences for existential and presentative assertions is the fact that these sentences, but not regularthetic (subject focus) sentences, have the same order in Basque and Spanish.

This dichotomy between a construction in which the absolutive is the topic and one in which it is the focus also has a transitive analogue as well in north-eastern dialects. The dichotomy here involves what we could call a transitive 'existential', a possessive construction, and a transitive copula construction in which a 'benefactive' argument is

coded in the ergative role. In other words, these sentences are very much like the existential and copula sentences we saw above, but they add an ergative argument which codes the an entity which bears a certain relationship to the referent of the absolutive argument. In (4.116) below we can see the transitive equivalent of the copula construction with a predicate nominal in focus position (cf. Oihartzabal 1984:364).

- (4.116) *Predicate nominal absolutive; covert ergative and absolutive arguments*
Lagunak ditut.
friends:DEF I.have.them
“They’re my friends (friends of mine).”

The sentence in (4.116) doesn’t mean *I have friends*, as we might have expected, but rather *They are my friends*, *They are friends to me*, or even more literally, *I have them (as) friends*. In other words, the nominal coded in the absolutive role is actually not the object of the verb, but a predicate nominal about that absolutive argument, which is elided. The actual possessive meaning (*I have friends*), in which the existence of a referent is asserted (with respect to another entity) would be expressed by the sentence in (4.117), in which the referential absolutive argument is topicalized, even though it doesn’t seem to be at all topical.

- (4.117) *Object topic; covert ergative*
Lagunak baditut
friends:DEF EMPH:I.have.them
“I have friends.” (Lit. “Friends I have.”)

Such pairs of sentences seem much less mysterious when they involve a different type of non-verbal predicate, such as a locative, instead of a predicate nominal, although the two types are completely parallel. Thus the sentence in (4.118) below, for instance, is analogous to the one in (4.116), but the predicate phrase is a locative.

(4.118) *Predicate locative; covert ergative and absolutive arguments*

Etxean ditut.

house:DEF:LOC I.have.them

“They’re at home.” (Lit. “I have them at home.”)

In (4.119), on the other hand, we can see the version transitive existential version of this sentence in which the locative is not a predicate but a locative complement which here is topicalized.

(4.119) *Setting locative; covert ergative and absolutive arguments*

Etxean baditut.

house:DEF:LOC EMPH:I.have.them

“There’s (I have) some at home.” (Lit. “At home I have them/some.”)

The topicalized locative in this sentence could also be placed, unaccented, in the assertion tail as an anti-setting with very much the same meaning, cf. *Baditut etxean* “I do have them/some at home”.

Oihartzabal provides one further interesting set of examples displaying the difference in the use of this ‘transitive existence’ (possession) construction in the north-eastern and the non-north-eastern dialects. The assertions in question consists of answers to the question *What do you have in your pocket?* (analogous to existential *What is there in your pocket?*), which we can see in (4.120).

(4.120) *Zer duzu sakelan?*

what you.have.it pocket:DEF:LOC

“What do you have in your pocket?” (~ “What is there in your pocket?”)

Because of the nature of the question, we would expect the nominal which corresponds to the question word to be the focus, just as in (4.121).

(4.121) *Non-north-eastern dialects*
Dirua dut.
money:DEF I.have.it
“I have money.”

Oihartzabal argues, however, that that is what we find in the non-north-eastern dialects. In the north-eastern dialects the most natural way to answer this question would be as in (4.122), with the absolutive argument outside the assertion in topic position

(4.122) *North-eastern dialects*
Sosa badut.
money:DEF EMPH:I.have.it
“I have money.” (Lit. “Money I have it.”)

One may argue that even here the topicalization is motivated, since the referent of the object nominal (*money*) can be accommodated in this context, for money is the sort of thing that one carries around in pockets. Still, the parallelism between the content question and answer construction in this case is broken in the north-eastern dialects.

The fact that in some dialects topicalization of such barely topical elements is possible seems to me to be, along with the topicalization of relatively inaccessible additive elements, an extreme case of the exploitation of the topic-comment configuration, something which perhaps to be expected in a topic-prominent language like Basque.⁴⁹ In other words, in Basque the topicality (accessibility) requirement for an element to be topicalized is obviously relaxed in some contexts. Thus we can see that the choice of whether to topicalize some element of an assertion or not may differ somewhat from language to language, and even from dialect to dialect within any one language, depending on factors having to do with the information structure system and the system of constructions of the language (variety).

NOTES TO CHAPTER 4

1. About the notion 'psychological subject' see von der Gabelentz 1869, 1960/1881, 1972/1901:369 (cf. Schlobinski and Schütze-Coburn 1992:89).
2. The notion of topic prominent vs. subject prominent languages is related to the notion of grammatical vs. pragmatic word order discussed in Thompson 1978, and it is also related to the distinction between Type A and Type B languages discussed by Perlmutter (1971:Ch. 5), which, of course, as is now recognized, is not an absolute dichotomy (cf., e.g., Haiman 1974b, 1991).
3. Bossong (1984) has argued that in Basque the notion absolutive, and not the notion subject, is relevant for matters of continuity and argument elision. Abaitua and Trask's (1987) response makes it clear that in terms of topic continuity Basque has the same preferences as other languages (except perhaps the proverbial Dyirbal).
4. Voice changing constructions are more common in some languages than in others. German, for instance, makes less use of constructions which promote elements to subject (cf. Hawkins 1986). Spanish, on the other hand, even lacks a promotional passive construction (there is one such construction, but its use is almost completely restricted to formal writing). Also, Spanish has less freedom than English as to which roles may become subject.
5. As Lambrecht states, "languages which require a subject NP are a small minority among the languages of the world. For example, Gilligan (1987) demonstrates that 'Pro-Drop languages' represent the vast majority of the world's languages" (Lambrecht 1994:191). Furthermore, he argues, "the occurrence of lexical subject NPs is indeed a relative anomaly across languages (compared to the occurrence of 'detached' NPs) is confirmed by the fact that there exist languages in which lexical NPs never function as subject arguments (Van Valin 1985, Jelinek 1984, Mithun 1986). On the other hand, languages which prohibit NP detachment do not seem to exist" (Lambrecht 1994:191). However, he goes on to add, "even in 'non-Pro-Drop' languages canonical sentences with lexical topic NPs are rare, hence 'statistically anomalous,' in spontaneous discourse" (Lambrecht 1994:191).
6. Oyharçabal (1985a) assumes that this 'subject' must be part of the rheme: "*If jendea is somehow the topic, I do not think that it is outside the rheme (from the point of view of information, it is a new element for instance)*" (Oyharçabal 1985a:108; my translation, J.A.) [*"jendea gisa batez mintzagai baldin bada, ez dut uste irazkinetik kanpo dagoen (informazioari dagokionez, elementu berria da adibidez)*"].
7. The other example mentioned in Euskaltzaindia 1987a in conjunction with these ones are exemplifying topic-postposing is that of an inverted thematic ground, i.e. an 'anti-topic'.

8. Sasse says: “Events are not prototypical predicates, ... but may become predicates in the presence of an independent entity with which they are linked. This entity is normally the most autonomous—that is, event-independent—member of their typical sets of participants. The prototypical predicate, on the other hand, is the state- or quality-descriptive predicate, because it is the only one which really ascribes a ‘property’ to an entity” (Sasse 1987:565).

9. Many languages (though not English) allow the use of the negative word as a predicate in addition to it being used as a ‘statement’ (e.g. English *no*) or an operator (e.g. English *not*). As an operator, the negative word *ez* in Basque comes before the negated material. The sentence in (4.14:2) could have alternatively been presented with the negative as an operator (*Bueno, ez manifa* “Well, not a demonstration”). See also, e.g., 93C3A12:117.

10. Basque linguist Altube proposed a certain type of apostrophe-like, semi-commas to be used in written Basque to express a minimal intonation break found between overt topics and other grounds, on the one hand, and the rheme, on the other, within the main- or body-intonation unit (Altube 1956; cf. Villasante 1980:248). Villasante (1980:249) argues that Altube’s proposed punctuation method is made necessary by the proposals he makes with respect to word order.

11. Givón 1990 mentions that the same type of function for right dislocation in Japanese, a rather strictly verb-final language. His source is Fujii 1989.

12. Aissen calls left-dislocated topics, the ‘external topic construction’, which “is similar to what is usually termed LEFT-DISLOCATION in the literature” (Aissen 1992:47-48). Although different Mayan languages have clause initial topics, “Tzotzil and Jakaltek topics are less integrated into basic clause structure, being essentially prefixed to what is otherwise a fully well-formed clause. The connection to the following clause is pragmatic, not syntactic: the clause must be ‘about’ the topic. By contrast, Tz’utujil topics appear to be much more tightly connected to the clause that follows. Two specific differences are that Tzotzil and Jakaltek topics do not occur in embedded clauses, while those of Tz’utujil do. Another is that third-person pronouns do not generally function as topics in Tzotzil and Jakaltek, but they do so freely in Tz’utujil. These differences suggest two sorts of entities.” (Aissen 1992:44).

13. Givón calls clause-internal topicalization “Y-movement”. About the referents of Y-moved phrases he finds that “[t]heir anaphoric predictability is also underscored by their characteristic *low referential distance*. Typically, their discourse antecedent is found within the preceding 2-3 clauses. Further the *referential interference* measure of Y-moved referents is usually high.” (Givón 1990:753).

14. Mitxelena has the following to say about this: “I would say, rather, that the pause [after the topic] ... is just a virtual one: it may be there, but it doesn’t have to. Actually, any element of the sentence, starting with the subject or the object, may be followed by a pause, *as long as it is not the focus*. The same thing happens with the comma, which may

be written or not, though preferably not. The distinction in this language [Basque] between topic and what de Rijk calls 'left dislocation', seems to me to be more imaginary than real" (Michelena 1987:470; my translation, J.A.; italics in the original version) ["Yo diría más bien que la pausa [después del tópico] ... es puramente virtual: puede aparecer, pero no tiene que hacerlo por necesidad. En realidad, cualquier elemento de la frase, empezando por el sujeto o el objeto, puede ir seguido de pausa, *con tal de que no sea el foco*. Lo mismo pasa con la coma, que se pone o no se pone, antes no que sí. La distinción para esta lengua del tópico y de lo que de Rijk llama «left dislocation», se me figura más ilusoria que real"].

15. Gundel proposes a 'topic familiarity principle' according to which "the topic must be chosen from information which is in some sense already given or shared by speaker and addressee" (Gundel 1985:86-87). Lambrecht (1986b, 1988b) discusses a "topic acceptability scale", according to which active elements are most acceptable (or least marked) as topics and unidentifiable ('brand new') referents are least acceptable.

16. For Reinhart (1982/1981) the minimal condition for something having topic status is that it can be identified, even if it is new, perhaps by itself being ground by an identifiable referent (e.g. *A daughter of a friend of mine, she got her BA in two years*, cf. Gundel 1985:89, ex. 13).

17. Both examples have sentence grounds which cataphorically refer to the subjects in question: According to Lambrecht, "[t]his kind of cataphoric reference to pragmatically non-accessible items is a rhetorical convention, which is based on the rule of accommodation for pragmatic presuppositions" (Lambrecht 1994:197). He also argues that this is a well known phenomenon: "This common phenomenon whereby a writer introduces a referent via a linguistic expression or grammatical construction which normally requires the presupposition that the referent is already introduced is discussed by Clark & Haviland (1977) under the name of 'addition'" (Lambrecht 1994:197).

18. Notice, however, that a context may be found for this sentence in which the verb *sat* can be understood as the focus of the assertion, namely one in which the verb is a contrastive focus, but in that situation it is more likely that the verb will be decomposed into a finite and a verbal part (e.g. *was sitting*), with the verbal part receiving the focus accent.

19. Under *Other* I have included two groups. The first one (60 ea.; 57 in main clauses) consists of main clauses with completive complements, and which seemed to have a modality/evidentiality function and not to have a topic. There were 60 of these clauses, or 1.6% of all, 57 of which were main clauses, or 2% of all main clauses. The Information Structure code for these clauses is *Modality/Evidentiality*.

The second group contains unclear cases. There were 168 of these clauses (4.6%), of which 133 were in main clauses (4.8%). For 60 of these there is doubt as to whether they are referentially grounded or not; 46 have a non-specific subject; 51 have narrow focus;

42 are main clauses which seem to be marking evidentiality of modality of a lower complement clause, which is really the main clause; 8 have non-subject links; 8 are either ungrounded or grounded with antitopic; 6 arethetic or presentationals; 3 are grounded or presentationals; 4 are grounded or narrow focus.

20. In this and following tables I refer to main and dependent declarative assertions (statements), unless otherwise specified. Speech Act code = *Body_Statement**.

²¹ Information Structure = **Topic _A/S/O/I**. The first column contains the Predicate Arguments code. In parentheses are the numbers of these topics which are covert topics.

22. Ergatives (A): 273/307 = 88.9%; 1070/1220 = 87.7%. Intransitive absolutes (S): 13/16 = 81.3%; 741/1006 = 73.7%. Transitive absolutes (O): 0/1 = 0%; 84/113 = 74.3%; Datives (I): 47/54 = 87%; 55/58 = 94.8%.

23. In Standard Spanish, human accusative arguments receive the same ‘case marker’ as the dative, rather than receiving no case marking like inanimate accusative arguments. This is called the ‘personal *a*’ in traditional grammar, cf. *Vi un libro* “I saw a book” vs. *Vi a Juan* “I saw Juan”. In some dialects of Spanish, such as the one spoken in the Basque Country, this is taken one step further and the ‘agreement’ clitic pronoun for human accusatives has the same form (*le(s)*) as the dative, and different from the inanimate accusative clitic pronouns (*la/o(s)*). This phenomenon goes by the name of *leísmo*. In the Basque corpus these clauses have been coded by the value *Dative_Human_Obj* in the *Other codes* code.

24. According to Givón, “the subject is consistently more topical than the direct object, and the direct object more topical than the indirect object” (Givón 1990:901). The following comments by Chafe on this topic are relevant: “it is true that most subjects are of either primary or secondary importance. These two properties are effects of the role of subjects as the grammaticized expression of starting points. Givón’s model would be more congruent with the present one if topic were equated with starting point, referential accessibility with activation cost, and thematic importance with what I am calling referential importance. It would be necessary, however, to recognize the separate status of all three, being alert to the various ways in which they interact as well as the functional reasons for such interactions” (Chafe 1994a:184).

25. For similar examples in the Spoken Basque Corpus see 93C2A03:156, 93C3A07:157.

26. The 5 with the absolute topic and inverted ergative are: 93C1A10:42, 93C1A16:6, 93C1B06:208, 93C2A03:298, and 93C2A07:145.

27. The 10 with the topic absolutes are (93C+): 1A02:17, 1A02:72, 1A10:148, 2A01:116b, 2A03:255a, 2B04:160a, 2B08:132, 3A07:12, 3A07:157, and 3A10:193.

28. The use of *le* instead of *la* in this context is normal in the dialect of Spanish spoken in the Basque Country.

29. The last two clauses of this excerpt are also interesting. The setting clause in (4.26:42) is about Chaplin, the only argument, which is covertly expressed. The covert topic of the following and final assertion (*they're eating lunch*), i.e. the non-specific *they*, includes Chaplin *and* the people that supposedly are in jail with him.

30. With a vague generic or non-referential objects there is no agreement marker in Spanish, e.g. *Pan no tiene nadie* "Bread nobody has", or with assertion-initial focus subject, *Pan nadie tiene*; *Dos panes no tiene nadie* "Two loaves of bread nobody has", or *Dos panes nadie tiene*.

31. *Unit type* code = *S_Head_Topic*. In addition there are 22 more topics without the proper intonational closure which seem to be interrupted clauses; *Unit type* code = *S_Head_Topic_Qn*.

32. VP-preposing is a minor English construction in which a highly topical, indeed active, 'verb-phrase', i.e. a non-finite clause with an externally 'controlled' topic, is fronted. Cf. Ward 1990.

33. These verbal topics may be accompanied by topicalizers (see below), such as the additive topicalizer *ere* "also", e.g. *Etorri ere / etorri da* "Come, he has done that too" (Osa 1990:181).

34. Lambrecht states that "high accessibility of the referent is a general condition for appropriate use of the antitopic construction across languages ... In contrast with left-detachment, the lexical (or independent-pronominal) topic expression in right-detaches position cannot indicate a new topic or a topic shift. ... The crosslinguistic tendency for right-detachment constructions to be used in discourse contexts in which the topic referent is already highly salient, and for left-detachment constructions to be reserved for topic-announcing or topic-shifting contexts" (Lambrecht 1994:203-4).

35. Lambrecht informs us that "[t]his construction has been variously referred to in the literature as 'epexegesis' (a term from classical grammar), 'inverted word order' (a translation of the term *devoir cümle* used by Turkish scholars; see Erguvanlı 1984), 'extraposition' (Jespersen 1964/1933:154ff.), and 'right dislocation.' The detached constituent itself has been referred to as 'de-focused NP,' 'afterthought NP,' 'post-predicate constituent,' 'tail' (Dik 1980[1981], Vallduví [1990]), and 'antitopic' (Chafe 1976)" (Lambrecht 1994:202-3).

36. Harris 1985, for instance, argues that "there exists a verb-initial order, occurring less often but still in significant quantities, which no doubt began life *as a pragmatically controlled 'afterthought' mechanism (broadly defined)*, and which is progressively grammaticalising in French" (Harris 1985:244, my italics, J.A.). Lambrecht also argues that antitopics must be distinguished from afterthoughts, cf. Lambrecht 1994:203, 354fn55.

³⁷ Information structure = *Invert_Topic**.

38. The two situations are: (1) Chaplin as the leader of the demonstration, and (2) The girl as the one who stole the bread. The first identification is not contrastive. Rather, Chaplin is chosen as the leader out of all the people in the demonstration. In the second case the identification is clearly contrastive, since the girl contrasts with Chaplin, the one the police initially arrests for this crime.

39. The girl that the policeman goes to contrasts with Chaplin, whom the policeman had captured first, thinking that he was the thief.

40. This particular predicate, *jarraitu* “continue; follow”, is one of a handful of verbs with an ergative argument but no absolutive argument, though it may take a dative argument.

⁴¹ Unit type = *Fin_Cl_**; Other Codes = *Subj_Inversion*; Polarity = *Aff*.

⁴² Unit Type = *Fin_Cl_**; Predicate Arguments = *Copula*; Speech Act = **Statement**; Polarity = *Aff*.

43. Other codes = **Accommodated_Topic**.

44. It is not clear to me that *behintzat* belongs with the ‘additives’ connectors, as classified by the Academy. Perhaps ‘restrictive’ would be a better classification. Sometimes substituted with *gutxienez* “at least”, undoubtedly under the influence of Spanish *al menos* or *por lo menos* “at least”, which has two functions of connector and adverb/intensifier (cf. Euskaltzaindia 1990:65-6).

45. Spanish is like English in that both topic and focus possibilities seem to be available. Spanish, though, only has one such adverbial, namely *también*. It’s position seems to be restricted to pre- and post-rhematic position.

46. There is also another (less common) finite concessive for with *ere*, namely the one formed with the *-(la)rik* setting adverbial clause (*-(la)rik ere*). There are also non-finite concessives with *ere*, one formed with a non-finite ‘conditional’, *-ez gero ere*, others formed with an adverbial (participial) clauses (*-ta ere* and *-rik ere*, which are synonymous), as well as with other non-finite clauses: *gabe ere* “although without -ing”, *-t(z)ekotan ere*, and *-agatik ere*. In addition there are two other concessive construction families, one with *arren* preceded by a perfective non-finite clause or a finite clause with the ending *-n*, and the other also with the same two types of clauses, but *preceded* by the subordinator *nahiz eta*.

47. The same source also has the sentence *Miren joango da gaur ere Iruñera* (p. 40, ex. 51). Here *gaur* “today” is the postposed setting, which accompanies *Iruñera* “to Iruña” in the assertion tail. *Miren*, on the other hand, must be the focus. In other words this sentence means something like: “It is Miren who will go to to Iruña again today.”

48. Euskaltzaindia has noted both of these facts, namely that the nature of the construction is changed and that its use is spreading. Thus they say that by placing *ere*

before the 'topic' "it seems that the we are breaking a basic rule for the use of *ere*. Although this construction does appear in classical writers, there is no doubt that it has been in the 20th century that it has seen its greatest expansion" (Euskaltzaindia 1990:46; my translation, J.A.) ["Hau egitean, jakina, badirudi *ere*-ren erabileraren funtsezko arau bat hausten dugula, esan den bezala, lokailu honek beti behar baitu aurretik sintagma bat edo osagai bat. Itzuli hau idazle klasikoetan *ere* agertzen den arren, ez da dudarik batere XX. mendean ezagutu duela hedaturarik handiena."].

49. Besides this use of topicalization in existentials and possessive existentials, Oihartzabal 1984 also mentions the extension of this very same type of topicalization to the verb *jakin* "know" (which also has synthetic forms) (Oihartzabal 1984:365-67). This use of topicalization seems to be available to topical referents (such as *nik* "I", non-contrastive), easily accommodated ones (cf. *euskara badakit* "I speak Basque"), or even non-specific ones (e.g. *Amitz liburu baditu* "she has lots of books", but *liburu guti du* "she has few books"). Euskaltzaindia (1987b:496) mentions another example of the type of meaning change that obtains sometimes depending on whether something is topicalized or not.

Chapter 5

Topicless clauses, object topics, and antitopics

5.1 THEMATICALLY GROUNDED AND UNGROUNDED SENTENCES

5.1.1 Introduction

Topicless (thematically ungrounded) assertions, as defined in Chapter 3, are exceptional in natural language, and thus are much less common than grounded ones. They are the exception which proves the rule. It seems to be a trait of human communication that, barring exceptional circumstances, propositions must be thematically grounded. That is, they must be predications and have a topic-comment information structure configuration. The thematic ground (topic) is typically active or very accessible, predictable, and covertly coded in a majority of clauses, at least in narrative. When the topic is less accessible, it is coded more ‘overtly’, by an overt nominal which is either ‘dislocated’ or, if the referent is accessible enough, at the beginning of the clause in the same intonation unit.

However, there are contexts in which the topic-comment articulation is either not possible or is for some reason overridden. These are cases in which either there is no available referent to be made the topic, or the available referent is not topical enough, or in which the whole event is presented (asserted) as a whole and not as a predication. It seems that all languages have special ways of marking these so-called ‘thetic’ sentences

(cf., e.g., Kuroda 1972, Sasse 1987),¹ ‘neutral descriptions’ (Kuno 1972), ‘news sentences’ (Schmerling 1976), ‘event-reporting sentences’ (Lambrecht 1987, etc.), and so on (cf. Lambrecht 1994:138).

In this section I will describe the characteristics ofthetic assertions and the way they are coded cross-linguistically and in Basque. In the following sections I will go through Sasse’s functional typology ofthetic sentences, which is based on a studies of clauses with subject inversion in SVO languages with flexible subjects. We will see that, although Basque is not an SVO language ‘subject-inversion’ is a common enough phenomenon in some cases. Many cases of ‘subject inversion’, both in SVO languages and in Basque, seem to be actually cases of topic inversion in emphatic sentences and not strictlythetic sentences at all.

5.1.2 Theticity: coding strategies and explanations

As I argued in Chapter 3,thetic assertions are those in which the subject is not the topic, but the focus of the assertion, and there is no other argument to fill the topic role.* In these assertions, the subject thus must be placed in a focus position. As we can see in the following examples,thetic sentences are indistinguishable from ‘narrow focus’ constructions in which the subject is the focus. In (5.1) we can seen an example from English and in (5.2) one from Spanish.

- (5.1) a. Who came? My mother came.
 b. What happened? My mother came.
-

* When I say that the subject is the topic or the focus, I mean, of course, that *the idea expressed by the subject*, and not the nominal itself, is the topic or the focus. When I want to refer to the nominal in question I will call it the subject constituent, the subject expression, or the topic/focus expression.

- (5.2) a. ¿Quién vino? Vino mi madre. (Emphatic: ¡Mi madre vino!)
 b. ¿Qué pasó? (Que) vino mi madre.

The accentual prominence of the focus in thethetic sentences may not be as great as in the other narrow focus construction given the lesser relative salience of the focus in the former case, but there is little doubt that the subject is treated as the focus of the assertion.

In most languages the focus subject of athetic sentence is placed in the same position as other complements when they are focused: preverbal position in OV languages such as Basque, and postverbal position in VO languages, such as Spanish. English is special in this regard in that a focus subject is placed in a position which is not normally used by non-subject foci, unless they are very salient.

Languages may also code topic and non-topic subjects differently morphologically, as is the case in Japanese, with a non-topic *ga* marker or a topic *wa* marker; cf. e.g. Kuroda 1972. Other languages may use special complex construction to indicate that the subject is not the focus, as in spoken French, with a cleft-like focus construction (cf. Lambrecht 1986, 1994).

As we saw in Chapter 3, languages may employ more than one construction or mechanism to markthetic sentences, or different types ofthetic sentences. Thus English, for instance, has a specialthetic existential construction, existentials being perhaps the most common type ofthetic assertion, in which the ‘subject’ is inverted with the verb.

The fact that the subject is not the topic, however, doesn’t always mean that a clause is topicless (thematically ungrounded), since a non-subject argument could be the topic, something which happens in a minority of cases, as we saw in Chapter 4. We also

find that a thetic clause may be grounded, not by a thematic ground, but rather by a circumstantial ground, such as a locative (cf. below).

It has been noted for a long time that with a certain class of predicates, the syntactic characteristics which I have attributed to topiclessness are not exceptional, but rather typical, if not the norm. Thus, for instance, it has been noted that in languages which allow subject inversion, inverting is the very common with certain verbs, e.g. *arrive*, *fall*, etc., and that with the same class of verbs in English typically the subject is integrated into the intonation contour of the assertion and receives the assertion's accent. We can see typical examples of this from English and Spanish in Table 5-1.

	Thetic / Subject focus		Topic-comment	
English	{ The <u>mail</u> arrived. }	<u>S</u> V	The <u>mail</u> /it { arrived. }	<u>S</u> <u>V</u>
Spanish	{ Ha llegado el <u>correo</u> }	V <u>S</u>	El <u>correo</u> /Ø { ha llegado. }	<u>S</u> <u>V</u>

Table 5-1: Typical thetic vs. topic-comment clauses in English and Spanish. Curly brackets signify the extent of the assertion proper.

The inversion and peculiar accentuation of subjects have been associated with, and attributed to, the semantic characteristics of these verbs, which are said to be 'stative'. They have also been associated with, and attributed to, the semantic characteristics ('semantic roles') of the topic-candidate (subject) of these predicates. Subjects of these predicates are typically 'non-agentive', such as 'patients', 'undergoers', or 'themes'. Some investigators have even claimed that the reason for characteristics such as inversion in these clauses are syntactic, and that they stem from the fact that these subjects are, in fact, objects underlyingly.²

Syntactic and semantic approaches to the study of 'thetic' or 'topicless' constructions have established some very interesting correlations, primarily semantic

ones. Still, I do not believe that the behavioral properties of these sentences are *caused* by any of these correlated properties. I believe the behavioral properties are caused by the fact that the subject, the topic candidate, is not the topic but the focus. The true correlations then are those between semantic role and predicate type on the one hand, and the likelihood that a ‘subject’ will not be the topic. The semantic characteristics of predicates and arguments only make it more likely that these predicates will be found in thetic constructions, but—crucially—it doesn’t guarantee it, nor does it preclude other predicates from appearing in thetic constructions.

Although thetic constructions are largely coextensive with so-called ‘unaccusative’ predicates—that is “existentials, verbs of appearance, psych-verbs, meteorological conditions, and the like” (Sasse 1995a:23)—the correlation is not perfect, and the reason is that theticity is a discourse-pragmatic phenomenon and not a semantic or syntactic one. Thus, as we have seen, prototypically thetic predicates may be found in categorial (topic-comment) constructions. This happens, for instance, when an optional adjunct is added to the assertion, as in example (5.3).

- (5.3) a. The mail arrived early.
b. El correo llegó temprano.

The reason for this difference is, of course, that in this case the optional temporal argument is an obvious candidate for the focus role and thus the subject has no option but to be the topic, especially since the focus (*early*) predicates a property *about* the subject argument (*the mail*).

It may be countered that subject inversion is possible in a language like Spanish in clauses with other complements, as we can see in (5.4).

- (5.4) Llegó temprano el correo hoy!
 it.arrived early the mail today
 “The mail arrived early today.”

In this example, however, the optional argument (*temprano* “early”) is still the focus (and a predicate) and the subject is an example of an antitopic. In other words, this is obviously an emphatic assertion. The inverted subject still conveys the entity (topic) that the predicate is about, but is placed in postverbal position to indicate that the focus is very salient. The effect produced by such an assertion is one of surprise. The antitopic, as we saw, must be accessible (given or easy to accommodate).

As we saw in Chapter 3, in a language like Basque a focus subject is not inverted at all, but goes where all focus constituents go, namely in preverbal position. Topics in emphatic assertions, on the other hand, may very well be inverted. When the subject is inverted, however, it is not always clear whether its referent is indeed the topic of the assertion or that it matters much since the distinction between an antitopic and a non-topic non-focus argument is neutralized. As we can see in Table 5-2, the basic topic-comment (1) andthetic (2) sentences in Basque are both SV sentences. Thetic sentences are identical to narrow focus sentences (3).

	Information Structure	Example	
1	Topic-comment	Kor <u>re</u> oa { <u>ir</u> itsi-da}	<u>S</u> <u>V</u>
2	Thetic	{Kor <u>re</u> oa iritsi-da.}	<u>S</u> <u>V</u>
3	Other subject focus (preverbal)	{Kor <u>re</u> oa iritsi-da.}	<u>S</u> <u>V</u>
4	Other subject focus (extraposed)	{ <u>I</u> ritsi-da, kor <u>re</u> oa.}	<u>V</u> , <u>S</u>
5	Comment-topic	{ <u>I</u> ritsi-da} korreoa.	<u>V</u> <u>S</u>
6	Thetic with verb/polarity focus	{ <u>I</u> ritsi-da korreoa.}	<u>V</u> <u>S</u>

Table 5-2: Different information structure types and their effects on Basque word order. Brackets enclose the assertion proper (verb and focus) and exclude topics, antitopics, and assertion tail elements.

However, subjects can be postverbal for one of 3 reasons: when the focus is extraposed (4), when the assertion is emphatic and the topic is an antitopic (5), or when the clause isthetic but either the verbal idea or the polarity are the focus (and thus the verb must be in focus position).

A preliminary look at the Spoken Basque Corpus definitely gives the impression that inversion is a major way of coding theticity in this language. Thus there are about 400 clauses in this corpus with predicates which are prototypically thetic: existentials (134 occurrences) and presentationals with new referents with verbs such as *etorri* “come” (92 occurrences), *agertu* “appear” (14 occurrences), *azaldu* “appear (12), and *ateratu* “come out” (9 occurrences). Interestingly almost 2/3 of these clauses (63%) display subject inversion, whereas only 1/3 (37%) have the subject before the verb. If we look at a sample of Basque written narrative, however, we see that these clauses have mostly preverbal subjects. The answer to this dilemma is, I believe, that some of the inverted subjects are actually extraposed foci (which are very rare in writing) or are actually antitopics in emphatic sentences. I will try to justify this position in the remainder of this chapter.

5.1.3 Constraints on thetic subject inversion

Although in prototypical thetic cases, such as existentials and impersonal event sentences, it is easy to see the reasons for the preference for the lack of a grounding element, in other cases, whether a sentence is thematically grounded or not seems to be to some extent a matter of speaker choice. The prototypical cases are clear enough and they tend to be marked in all languages as thetic (subject focus) in one way or another. But

languages may differ as to whether to allow, or encourage, thetic framing in other, less prototypical cases.

This is the conclusion reached by Sasse (1995a) in his summary of the studies in the volume edited by Sasse and Matras (1995). This volume contains several studies of (supposedly thetic) subject-inversion in different European languages: Latin, Italian, Spanish, Russian, Hungarian, Modern Greek, Romani, and Turkish. The conclusion that Sasse draws is that there is a language type, to which all of these languages except Turkish belong, which form thetic sentences by means of subject-verb inversion.³

According to Sasse, one major finding of these studies is that subject inversion in these languages is subject to a variety of language internal constraints. That is, subject inversion is more or less liberally applied and extended to non-prototypical cases in different languages, so that we find “polyfunctionality and ... subtle language-specific constraints governing the use of the constructions in question” (Sasse 1995a:7).⁴ In other words, he argues, “in each language there are restrictions with respect to the applicability of VS order as an explicit sign of ‘theticity’, and the number and nature of these restrictions seem to differ from language to language [3.5]... Comparably few such restrictions have been found in the Balkan languages and in Russian, and comparably many in Italian and Spanish, with Hungarian standing somewhere in between” (Sasse 1995a:13). Thus some languages would code as thetic only the prototypical, core cases, whereas other languages would have extended this information frame to other less clear-cut but still motivated cases.

It is hard to evaluate Sasse’s hypothesis from the meager data presented by different authors, which are taken mostly from written sources. It is particularly

worrisome that Sasse does not mention the normal association between thetic inversion of the subject and its role as focus of the assertion, or the other possible reasons for the inversion of a subject in these languages, namely focus extraposition and, in particular, topic inversion in emphatic assertions. As I have already mentioned, and will become more clear in this chapter, emphatic assertions share some pragmatic characteristics with thetic assertions, and in languages in which postverbal position is the main focus position for all complements, including subjects, both sentence types display subject verb inversion. As we have seen, however, the two sentence types should be kept quite separate since they have very different focus structures.

I have argued that there is a strong, overwhelming tendency in natural language for normal declarative assertions to be predications and for them to have a topic which is what the predication is about and what links the assertion to the world of discourse. I believe that the reasons why an assertion might reject this pragmatic structure in normal (non-emphatic) assertions have to do with two discourse pragmatic factors:

- Low topicality of the subject's referent, such as in the case of unidentifiable referents, which means that it is a poor candidate for the **topic role**. The semantics of the subject, i.e. its semantic role, may also be a factor in the topic-worthiness of a referent, given the strong correlation between topics and 'actors' (and foci and 'undergoers').
- The lack of a better candidate for the **focus role** in the assertion. Of course the verb itself (or the polarity of the assertion) can always be the focus, but, as we have seen, that is typically a marked choice in natural language.

Although the first of these factors has often been appealed to, in one form or another, to explain thetic sentences, the second factor is crucial as well. One way in which investigators have attempted to capture this generalization is by arguing that verbs in thetic constructions must be intransitive or that, at least in some languages, they must have only one argument, the ‘mono-argumentality’ constraint of Bolkestein 1995 (cf. also Sasse 1995a:20-21). I believe that these two hypotheses are mistaken and that the actual reason why subjects in transitive clauses (with an overt object) or intransitive clauses with an additional complement or adjunct tend to be topics is that the additional complements are more likely to be the focus (sometimes even a predicate focus) and prevent the subject from filling that role.

I believe that whether a subject’s referent will be the focus of an assertion instead of its topic is a fairly straightforward matter. In part it depends on characteristics of the subject itself such as the topicality of the subject’s referent. Thus a very accessible referent will not make a good focus, unless it is contrastive. It also depends on the topicality of the subject’s semantic role, with Actor-like roles being more ‘topical’ than Undergoer-type roles (cf. Foley and Van Valin 1984). But it also depends on other characteristics of the assertion, and of the state-of-affairs expressed by the assertion more generally (including the aspect of the predicate), characteristics which correlate with whether the assertion can be interpreted as a predication or not.

It is true, as we saw in chapters 3 and 4, that languages may differ somewhat as to whether a ‘subject’ will be the topic or the focus of an assertion, but for the most part I believe that languages tend to be fairly consistent in this matter. Thus I believe that the differences which Sasse observed among languages (or dialects, or styles) with respect to

the degree with which they use subject inversion cannot be related to a difference in requirements to make sentences *thetic*. I believe that the difference is related to the ease with which languages invert subjects which are *not* foci, i.e. the degree to which topics can be made *antitopic*s in more or less emphatic constructions.

If we find that a language makes great use of topic inversion (i.e. *antitopic*s) we can deduce that either the text we are using contains a great number of emphatic assertions, a sign of a lively style or a peculiar genre for example, or that the *antitopic* construction in this particular language (or context of language use) has lost some of its markedness is used quite liberally. I will return to this issue below.

5.1.4 Thetic functions: Typologies of *thetic* sentences

The closest thing that we have to a typology of *thetic* sentences, that is of the contexts in which they are found and the uses that they are put to, is perhaps the one found in Sasse 1987, 1995a, 1995b. Sasse's typology presents a number of important insights. This typology also has as some severe limitations stemming from (1) the lack of recognition of the importance of the focus role (in addition to the topic role); (2) the use of subject inversion as the point of departure for the typology, where in fact inversion can be caused by various factors; and (3) the nature of the data, which comes almost exclusively from written sources, and in some cases from extremely specialized contexts (genres, styles), such as newspaper headlines.⁵

In the following sections I will follow Sasse's (1987) original typology, as well as his (1995a) review of different analyses of VS structures in different languages, and his own (1995b) contrastive study of VS sentences in Modern Greek and Hungarian. This last contrastive study is particularly interesting to us since it compares Greek, a language

which is similar to Spanish in being SVO with flexible subject order in which the primary focus position is postverbal, and Hungarian, a language similar to Basque in that its primary focus position seems to be preverbal but is not strictly verb-final.

The major overarching unifying characteristic of all thetic sentences, according to Sasse, is that they “appear often (but not exclusively) after the explicit or implicit question ‘what happened?’, in exclamations, in presentative and existence-asserting functions, in sentences establishing scenery, in explanations—in short, *at any point in a text where information is not given about someone or something, but about an entire state of affairs*” (Sasse 1987:535; my italics, J.A.). The major thetic functions, or the “typical domains for thetic expressions”, as identified by Sasse 1987 are summarized in Table 5-3 (Sasse 1987:566, ex. 134).

Function	Uses
Existential statements	presence, appearance, continuation, etc., positively and negatively
Explanations	with or without preceding questions such as ‘what happened?’, ‘why did it happen?’, etc.
Surprising/unexpected events	
General statements	aphorisms, etc.
Background descriptions	local, temporal, etc., setting

Table 5-3: Functions and uses of thetic structures according to Sasse 1987.

Sasse’s more recent typology is briefly summarized in Table 5-4 (cf. Sasse 1995a, 1995b). As we can see, there have been some minor changes in the new typology.

Function	Uses
Annunciative	All-news statements
Explanative	Describe happenings
Introductive	Existentials and presentatives
Descriptive	Background statements
Sudden events	
Discontinuative	Episode opening; given subjects
Reactive/consequential	The event is a reaction; given subjects

Table 5-4: Thetic functions according to Sasse 1995a.

In the next few sections I will analyze these functions in turn and in particular the manner in which they are realized in Basque, but also in other languages. As we will see, the distinction between inverted focus subjects and inverted topic subjects will turn out to be a crucial one. On the other hand, some of the uses of emphatic assertions are similar enough to the typical uses of thetic assertions so as to cause a certain amount of confusion.

5.2 ANNOUNCEMENTS

5.2.1 Sasse's definition and examples

The first type of function for thetic clauses that Sasse discusses is what he calls the “annunciative function” or the “announcement strategy”, where “the existence of a certain previously unknown state of affairs is announced” (Sasse 1995b:167), “where the presuppositional status of both V and S is lowest” (Sasse 1995a:14), and where the subject is not “shared by the audience” (Sasse 1995a:14), i.e. is ‘new’. Lambrecht too mentions ‘out-of-the-blue’ or ‘all-news’ statements as being the prototypical thetic or ‘wide-focus’ statement.

Sasse, who as I said restricts himself to written sources, novels and newspapers primarily, sees this as an evolution or “further specialization of the general ‘existential’ character of ‘thetic’ VS utterances,” one that “has evolved into a strategy of announcements for informations [*sic*] given to an addressee ‘out of the blue’ employed in *exclamations*, ads, headlines, announcements of social events, etc.” (Sasse 1995b:167; my italics, J.A.).

It is true that we typically find thetic sentences in these contexts, since these are contexts in which the topicality (accessibility) of the subject’s referent, and thus its topic-worthiness is diminished. However, as I have already argued, the opposite is not the case. That is, by no means are all sentences found in these contexts thetic, since most of them do in fact have a topic-comment configuration (i.e. are ‘categorical’ assertions). The preferred strategy for presenting information (‘news’) in natural language, including all-news or out-of-the-blue utterances, is for there to be a grounding referent, a thematic ground (topic) to which the new information is connected, of which the new information is predicated. A clause under these circumstances of low presuppositionality will be thetic (topicless) if and only if a predicate interpretation is not possible, for any of the reasons given above, such as whether or not there is a better focus candidate than the subject’s referent.

Sasse gives the example from Modern Greek in (5.5) (= Sasse 1995b:167, no. 69) as a typical example of an annunciative (all-news) thetic sentence in natural language. It doesn’t seem very surprising in this context that the subject of such a statement should not be the topic since it doesn’t seem that the announcement is about Kostas but about a certain phone call that he made, that is, about the whole state of affairs or event.

- (5.5) telefonise o Kostas
has.called the Kostas
“Kostas called!”

Thus *Kostas* here is not the topic but the focus of the assertion. In a language like Spanish a VS assertion of this type could also be an emphatic assertion if the finite verb and not the subject was the focus, as indicated by the main accent of the assertion, cf. (5.6).

- (5.6) a. Ha llamado Kostas. “Kostas called.”
b. Ha llamado Kostas. “Kostas has/did called.”

Still, an out-of-the-blue assertion such as this one may still have a topic under some circumstances. To begin with such assertions referring to happenings are often grounded by the referent that the information is relevant to. Thus, in languages such as Basque or Spanish, a sentence which announces a phone call out of the blue, often has a *benefactive topic*, as in the Spanish example in (5.7), in which the benefactive (coded by the *te* “you” clitic pronoun) grounds the clause on the listener, which, not surprisingly comes in clause initial position.⁶

- (5.7) (Te) ha llamado Kostas
(you:DAT) has called Kostas
“Kostas called (you)”

On the other hand, if the assertion has a more likely focus candidate than the subject’s referent, such as a temporal (predicate) phrase, the subject may be the topic, as in (5.8).

- (5.8) Kostas (te) ha llamado hace un rato.
Kostas (you:DAT) he.has called it.makes a while
“Kostas called a while ago.”

There is yet another alternative in a context such as this, and that is for the assertion to be presented emphatically. In Spanish, this would be coded typically simply by inverting the subject constituent, as in (5.9). Alternatively, the focus constituent could be fronted to preverbal position to make it more emphatic, as in (5.10).

(5.9) (Te) ha llamado ahora mismo Kostas.
(you:DAT) he.has called now same Kostas
“Kostas just called (you).”

(5.10) Ahora mismo (te) ha llamado Kostas.
now same (you:DAT) he.has called Kostas
“Kostas just called (you).”

Note that these sentences could hardly be emphatic with a temporal phrase such as the one in (5.8): *hace un rato* “a while ago”. The immediacy of the call expressed by the temporal phrase *ahora mismo* “right now” in (5.9) and (5.10) make this focus more focal and thus a better candidate to be coded as an emphatic focus. In these sentences Kostas, as the antitopic, must also be relatively accessible, or at least not too unexpected, surprising, or difficult to accommodate. The same thing is not true in (5.8) above in which Kostas is a regular topic and all it needs to be is identifiable.

In Basque we find a similar situation for announcements. When the subject’s referent is relatively inaccessible and there is no better focus candidate, the subject itself may be the focus, which means that it is placed in assertion-initial position and is the accented (focus) element, cf. (5.11), with an analytic verb, and (5.12), with a synthetic verb.

(5.11) Polizia etorri-da. (SV)
police:DEF come:PFV-it.is
“The police came.”

(5.12) Polizia dator. (SV)
police:DEF it.is.coming
“The police’s coming.”

But if there is a better focus candidate, the subject readily becomes the topic, as in the case of (5.13), with a temporal adverb in the focus role, or as in (5.14) with an object fulfilling that role.

(5.13) Polizia berehala etorri-da. (SXV)
police:DEF right.away come:PFV-it.is
“The police came right away.”

(5.14) Poliziak liderrak atxilotu-ditu.
police:DEF:ERG leaders:DEF arrest:PFV-it.has.them
“The police (has) arrested the leaders.”

But, just as in Spanish, a topic-comment sentence which is emphatic, one that has an emphatic focus, such a contrastive one, or which are emphatic rheme-focus assertions, as announcements often are, can be turned into a comment-topic sentence with topic inversion, as we can see in (5.15) and (5.16) below.

(5.15) Berehala etorri-da polizia. (XVS)
right.away come:PFV-it.is police:DEF
“The police came right away.”

(5.16) Liderrak atxilotu-ditu poliziak. (OVS)
leaders:DEF arrest:PFV-it.has.them police:DEF:ERG
“The police (has) arrested the leaders.”

In either case the subject could be clause initial, and thus the topic, but it is inverted to code the high salience of the assertion. When the whole rheme is the focus, the finite verb fills the focus role and is found accented in focus position. In such assertions, the proposition and all its parts (including the topic) are quite accessible. Thus, for instance,

the assertions in (5.17a&b) could be uttered when the arrival of the police had been expected.

- (5.17) a. *Etorri-da polizia!*
come:PFV-it.is police:DEF
“The police came!”
- b. *Badator polizia!*
EMPH:it.is.coming police:DEF
“The police’s coming!”

Emphatic assertions with topic inversion are somewhat similar pragmatically tothetic assertions, as we have seen, but whereas in languages like Spanish in both cases the subject is postposed (although for different reasons), in a language like Basque, only the former have postposed subjects (unless, of course, a focus subject is extraposed). Unless, of course, the thetic assertion is emphatic and the subject is easy to accommodate, in which case the resulting VS sentence is indistinguishable from an intransitive antitopic construction. The only difference is that in the latter case the topic can be fronted, whereas in the former it cannot. In north-eastern dialects, however, as we saw in Chapter 4, subjects with low-accessibility referents in existential clauses can be the topic of the assertion.

5.2.2 All-news statements in newspaper headlines

Sasse notices that subject inversion is frequently found in these languages for specialized functions such as newspaper headlines, in cases in which inversion might not be used in speech, such as in the Greek example in (5.18) (= Sasse 1995b:167, no. 68).

- (5.18) *epistrefi sto proskinio i katastopia ton paradhosiakon methodhon.*
returns to.the scene the espionage of.the traditional methods
“Traditional methods of espionage (are) returning to the scene.”

In Hungarian, Sasse argues, this function is also performed by a subject-inversion construction, at least in newspaper headlines (those are Sasse's only examples). This is the case even in sentences without a finite verb where the predicate is non-verbal and there is no copula verb, as in the example in (5.19).

- (5.19) **Hazánkban az örmény külügyminiszter.**
homeland.our.in the Armenian foreign-minister
"Armenian foreign minister in our country."

Although these specialized 'announcements' may involve thetic sentences, one may wonder whether all such inversions are truly thetic. As I argued above, a sentence with a postverbal subject in these languages may be other things besides a thetic assertion, that is an assertion with a new-focus subject. It may also be an emphatic assertion with an inverted topic. With the right intonation, however, it could also be a sentence with a right dislocated (extraposed) focus subject.

The particular domain of newspaper headlines seems like a very likely one for emphatic assertions to be used in, given the 'lapidary' function of these assertions and the fact that the topics of these clauses tend to be fairly stereotyped and accessible (e.g. *the president, the State Department*). On the other hand, as I mentioned in Chapter 3, it is undeniable that there exists some pragmatic and rhetorical similarity between thetic clauses and emphatic assertions with topic inversion. Below I analyze newspaper headlines in Spanish and Basque, which shed some light on this topic.

5.2.3 All-news sentences in Spanish newspaper headlines

To get a better sense of the phenomenon of inversion in headlines in a subject-inverting language, I looked at headlines of the Internet editions of several Spanish

language newspapers for May 3, 1996. I found that no single strategy is followed consistently in all Spanish newspapers. In peninsular Spanish newspapers, inversion in headlines is not very common (although it may have been at one time).

Thus, in *La Vanguardia de Barcelona*,⁷ for example, we find one single case of inversion out of 60 headlines, an idiom (for a complete list of all the headlines see Appendix 5.1).⁸ We do not even find subject inversion with predicates such as *morir* “die” and *dispararse* “shoot up”, such as in the two headlines in (5.20), both of which are predicates which usually preferthetic information structure in speech.

(5.20) *La Vanguardia*, May 3, 1996

- a. **La actriz María Luisa Ponte muere a los 77 años en Aranjuez.**
“Actress María Luisa Ponte dies at 77 years of age in Aranjuez.”
- b. **El gasto en pensiones se dispara.**
“Pension expenditure shoots up.”

A full half of the headlines consist of transitive sentences; another ten are intransitive, fifteen are non-clausal, one is an imperative (a quote), and one is a VO clause with an indefinite subject. Among the non-clausal headlines, we find bipartite assertions with clear topic-comment structure, and ‘understood’ copulas, such as the ones in (5.21a&b).

(5.21) *La Vanguardia*, May 3, 1996

- a. **La condena del caso Bulger, cuestionada.**
“The sentence in the Bulger case, (is being) questioned.”
- b. **Chapertons y su original mundo de neumáticos, en el Villarroel.**
“Chapertons and his original world of tires, in the Villarroel theatre.”

El Periódico, another Spanish language newspaper from Barcelona, had five cases of subject inversion in its 43 headlines, two of which are inside quotation marks, i.e. are quotes (cf. Appendix 5.2).⁹ All the verbs involved in these inversions are prototypical subject inverting verbs found inthetic sentences with focus subjects: *empezar* “begin”,

aparecer “appear”, and *arrancar* “to start up”, *gustar* “to please”, and *resultar* “turn out”, as we can see in (5.22a-e). The example in the quote in (5.22d) is not strictlythetic, since the referent of the dative argument is clearly the topic here.

(5.22) *El Periódico*, May 3, 1996

- a. *Empieza la matanza de las ‘vacas locas’.*
“The killing of mad cows begins.”
- b. *Aparece un tercer sospechoso del crimen de Alcàsser.*
“A third suspect of the Alcàsser crime surfaces.”
- c. *Arranca el boicot contra Francia.*
“The boicot against France begins.”
- d. *Stuart BAIRD: “Me gusta que me corra la adrenalina por el cuerpo, es como una adicción.”*
Stuart Baird: “I like for adrenaline to run through my body, it is like an addiction.”
- e. *Charlie Sheen: “Resulta más fácil relacionarse con una prostituta.”*
“Charlie Sheen: “It is easier having a relationship with a prostitute.”

In addition to these five inverted clauses with inversion (all of which are intransitive), there are 23 transitive clauses (none of which has an inverted subject), 8 non-clausal headlines, 6 intransitives with preverbal (topic) subjects, and one which is a special construction in a quote (an exclamation which is really an inversion).

In the headlines for the Mexican newspaper *La Jornada*, on the other hand, we find numerous inversions, including in transitive clauses, and including inversions which would be extremely marked, if at all possible, in the spoken language (cf. Appendix 5.3).¹⁰ Of the 34 headlines (for the front and back pages of the newspaper), 10 are non-clausal, 9 have no overt subjects, 6 have overt clause-initial subjects, and 9 have subject inversion. Of the 9 clauses with subject inversion, one is a content question, 2 are intransitive clauses, those in (5.23a-b), and 6 are transitive clauses, namely those in (5.24a-f).

(5.23) *La Jornada*, May 3, 1996

- a. En un año bajaron en 3 mil millones de barriles **las reservas de crudo**.
“In one year the crude reserves went down by 3 billion barrels.”
- b. Urge **nueva base teórica para la izquierda**: Bovero.
“A new theoretical base for the left is needed urgently, says Bovero.”

(5.24) *La Jornada*, May 3, 1996

- a. Elogia **el Presidente** a organismos que han evitado cerrar empresas.
“The president praises organizations which have avoided closing businesses.”
- b. Investigará **la CIDH** denuncias de abusos contra mexicanos en la frontera.
“The CIDH will investigate accusations of abuses against Mexicans at the border.”
- c. Aprueba **el Senado de EU** proyecto antimigratorio.
“The US Senate approves the anti-immigration project.”
- d. Advierte **Clinton** que no lo sancionará si no revisan restricciones a beneficios sociales.
“Clinton warns that he will not sanction it if they don’t revise restrictions to social benefits.”
- e. Anuncia **EU** la captura de 130 presuntos narcos.
“The United States announces the capture of 130 alleged narcotraffickers.”
- f. Acuerda **asamblea indígena** proponer iniciativas sobre regiones autónomas.
“Indigenous assembly resolves to propose initiatives about autonomous regions.”

The two intransitive sentences with subject inversion in (5.23a-b) are rather typicalthetic sentences in which the subject is obviously the focus of the assertion. Not so the VSO transitive sentences in (5.24a-f). In these the focus is obviously the assertion final complements, often clausal complements. In other words, in these clauses, when they are read aloud, there are two accents, one on the verb and one on the clause-final object. The subjects are treated as (unaccented) antitopics that are easy to accommodate to readers of a newspaper (*the president, the US Senate, etc.*).¹¹

On the other hand, there seems to be a way in which these transitive sentences, all of which would sound odd in speech under normal circumstances, are like thethetic ones. As I discussed in Chapter 3, emphatic assertions seem to share withthetic sentences the

general feel of being more like event assertions than like predications. Thus it would seem that in this particular genre, in this particular variety of Spanish, the emphatic assertion construction has been adopted for the purpose of either making the assertions in the headlines more salient or for the purpose of de-emphasizing or playing down the predication aspect of the assertion.

One very interesting fact about this choice is that it is confined to this genre and somewhat arbitrary, since in the same context the Iberian Spanish newspapers seem to follow the opposite tendency of emphasizing the topic-centered predicate status of the assertions. That is, the choice of presenting an assertion as emphatic (in this context) is a rhetorical one. Furthermore, the reason for using this assertion ‘strategy’ may have something to do with the (limited) pragmatic similarity between emphatic assertions andthetic assertions of being ‘event centered’ as opposed to ‘topic centered’. Below I will argue that a very similar thing may be happening in Basque presentational sentences.

5.2.4 All-news sentences in Basque newspaper headlines

As in the case of the Mexican newspaper *La Jornada*, but unlike in the case of the two Spanish newspapers, Basque newspaper headlines also often display subject inversions which would be rather marked in actual speech.¹²

Thus we find that the headlines for the front page of the newspaper *Egunkaria* for May 3, 1996, contains a fair amount of subject inversions, including transitive subject (ergative argument) inversions, which—like in the case of *La Jornada*—would be quite marked in the spoken language (cf. Appendix 5.4). Of the 12 front-page headlines available, 3 are non-clausal and 1 has no overt subject. Of the remaining 8 headlines containing finite statements with overt subjects, 5 have uninverted (clause-initial)

subjects, and 3 have inverted subjects. Of the inverted subjects, 1 is an absolutive subject, the one in (5.25) below, and 2 of which are ergative arguments, those in (5.26), and (5.27).

The intransitive sentence in (5.25) is a complex one, displaying two coordinated clauses, the first one of which is non-finite, a common construction in Basque for event pairs (see Section 5.6.3 below). Furthermore, the combined two clauses share a subject (absolutive argument), which follows the complex clauses. In this case it seems that the subject is not an extraposed focus, but an antitopic. The sentence final temporal phrase does seem to be an anti-setting.

(5.25) *Egunkaria*, May 3, 1996

Gaur hasi eta bihar amaituko-da **Aznarren inbestidura saioa** *Espainiako Legebiltzarrean*.

today begin:PFV and tomorrow it.will.finish Aznar:GEN investiture attempt:DEF
Spain:GEN parliament:DEF:LOC

“**Aznar’s investiture attempt** will start today and finish tomorrow in the Spanish Parliament.”

The next two examples have transitive verbs and inverted ergative phrases. Both seem to be cases of antitopic constructions. In the headline in (5.26), with OVS order, the preverbal object is obviously the focus and the clause-final subject is quite clearly an antitopic phrase (a well known publishing house and thus easy to accommodate in this context). Thus, this is not athetic clause, but an ‘emphatic’ antitopic assertion, much like the headlines we saw in the Mexican newspapers. The only difference is that the focus phrase here is preverbal, whereas those in *La Jornada* were clause-final or extraposed.

(5.26) *Egunkaria*, May 3, 1996

Jorge Oteizaren arkitektura-lanak biltzen-dituen katalogo-liburua aurkeztu du
Iruñan **Pamiela argitaletxeak**

Jorge Oteiza:GEN architecture-works it.collects.them:REL catalog-book:DEF
it.has.presented.it in Iruña Pamiela publisher:DEF:ERG

“**Pamiela publishing house** has introduced a book-catalog which collects Jorge Oteiza’s architectural works”

The example in (5.27) below also has an antitopic ergative argument, but here the absolutive argument (the ‘object’) is topicalized and the focus position is filled by the adverbial predicate (cf. *leave free*, *let go free*).

(5.27) *Egunkaria*, May 3, 1996

Kartel bat dela-eta Ertzaintzak atxilotutako Roberto Sampedro, HBko
komunikazio arduraduna, aske utzi du **epaileak**.

poster/sign one because Basque.Police:DEF:ERG imprison:PFV:ADV:GEN R.S.,
HB[Basque political party]:GEN communication person.in.charge:DEF,
free:ADV he.has.left.him judge:DEF:ERG

“**The judge** freed R.S., the person in charge of communications for HB, who the Basque police had imprisoned because of a publicity sign.”

The following example in (5.28) below is one of the five headlines which have uninverted, clause-initial ‘subjects’. This sentence displays the order SXVO, where the ergative S is the topic, the X is the focus, and the O is an unaccented assertion-tail constituent with an accommodated (accessible) referent.

(5.28) *Egunkaria*, May 3, 1996

Josu Urrutikoetxeak gaur beteko-du Frantziak ezarritako espetxe zigorra.

Josu Urrutikoetxea:ERG today he.will.fill.it France:ERG establish:PFV:ADV:GEN
prison punishment:DEF

“Josu Urrutikoetxea will complete today the prison sentence established by France”

The same antitopic inversion phenomenon which we saw above can be observed in the triple headline in (5.29), from the same newspaper but from a different date. The first line, the noun phrase in (5.29a), expresses the general discourse topic for the article:

The/a plan to have public servants learn Basque. The next two headlines in (5.29b) and (5.29c) state two aspects of the position of a certain political party, coded as an inverted ergative phrase, towards said plan.

(5.29) *Egunkaria*, May 14, 1996

- a. Funtzionarioak euskalduntzeko plana
public.servants Basque.speaker.becoming:NOMIN:GEN plan:DEF
- b. Euskara erabilera-hizkuntza izatea gehiegizkotzat jo-du **PSE-EE alderdiak**
Basque:DEF use-language:DEF be:NOMIN:DEF excessive:DESTIN it.has.hit.it
PSE-EE party:DEF:ERG
- c. Zerbitzu-hizkuntza izatea nahikoa dela diote **sozialistek**
service-language:DEF be:NOMIN:DEF sufficient it.is:COMPL they.say.it
socialists:DEF:ERG

“(a) Plan to ‘Basquicize’ government workers. / (b) The PSE-EE (Socialist Party of Euskadi-Euskadiko Ezkerra party) believes it’s excessive to make Basque a working language. / (c) The socialists say it’s enough for it to be a service-language.”

The two statements in (5.29b) and (5.29c) have identical structure. Both sentences display the order OXVS, where the O, a nominalized clause, is the topic, the adverbial X is the focus, and the postverbal ergative argument is an antitopic constituent.

In the following excerpt in (5.30) from a front-page short news piece, we can see two more examples of this type of subject inversion. The first headline, the noun phrase (5.30a), sets the discourse topic and second headline, the sentence in (5.30b), states the most important event. The first and second sentences of the article proper, those in (5.30c&d), both have OVS order, in which the O is the focus and the S is inverted, i.e. is an antitopic phrase in assertion-tail position.¹³

- (5.30) *Euskaldunon Egunkaria*, Sep. 24, 1996
- a. **POLIZIAREN SAREKADA LONDRESEN**
 police:DEF:GEN raid:DEF London:LOC
 - b. **IRAKo ustezko kide bat tiroz hil-dute**
 IRA:GEN belief:INSTR:GEN member one shot:INSTR they.have.killed.him
 - c. **HAMAR TONA** lehergai aurkitu-du **Scotland Yardek** Londresen egindako
 polizi sarekada batean,
 ten ton explosive it.has.found.it Scotland Yard:ERG London:LOC
 do:PFV:ADV:GEN police raid one:LOC
 - d. eta ekimen berean IRAko ustezko kide bat hil-dute tiroz **agenteek**, ...
 and initiative same:LOC IRA:GEN belief:INSTR:GEN member one
 they.have.killed.him shot:INSTR agents:DEF:ERG
- “(a) Police raid in London / (b) They have shot dead a presumed IRA member / (c) **Scotland yard** found ten tons of explosives in a police raid carried out in London, / (d) and in that same action **the agents** shot dead a presumed IRA member.

What is interesting about these two examples is that it is quite clear here, more so perhaps than in the previous examples, that these two assertions are *not* ‘about’ the antitopics, i.e. about *Scotland Yard* and *the (Yard’s) agents*, respectively. The referents of these inverted topics, besides being quite accessible in this context despite not having been mentioned directly before (they can be ‘inferred’ from the context), seem to be quite unimportant and not what the assertion is about necessarily.

I would like to suggest that what is actually happening here is that the emphatic assertion (antitopic) construction, which in Basque is used normally when the focus is particularly salient (contrastive or emphatic), is being used, in this genre at least, to emphasize the event itself and de-emphasize the topic, perhaps to put extra emphasis on the whole assertion and on state-of-affairs and to de-emphasize the topic-centeredness associated with regular assertions. In other words, the emphatic assertion construction is being used in a way that is reminiscent ofthetic assertions. The major difference is that in the antitopic construction the postposed subject’s referent is not the focus and must be

quite accessible (or easy to accommodate), whereas in the thetic construction the subject must be the focus and typically is quite inaccessible. I am not saying that this is true about *all* antitopic assertions, but it may be that the construction has become extended to cover this new function.

5.2.5 Interruptions and sudden events

According to Sasse, assertions describing interruptions and other sudden events which break the continuity of discourse are coded by thetic assertions, i.e. clauses with inverted subjects in the relevant languages that he studies. He calls this the “interruptive function”, or the “Sudden Event strategy”, used “in cases of sudden, interruptive events”, such as “[t]he phone or the door-bell ringing” (Sasse 1995a:15). We can see an example from Spanish (5.31) (the phone ringing example).

(5.31) Suena el teléfono.
it.sounds the telephone
“The phone’s ringing.”

Such statements are typically not assertions (predications) about the phone or the door-bell, but rather they announce a stereotypical event (e.g. *The ringing of the phone*). If the event was non-stereotypical, as Bolinger for instance has argued, a topic-comment configuration would be called for, e.g. *The lamp is ringing*. This statement (typically) expresses that something is happening to the lamp and not that something is happening, and thus the subject would normally be the subject here.

Sasse argues that, unlike announcements, the interruptive function “may occur text- or paragraph-internally as well as text- or paragraph-initially” and “may occur in chains” (Sasse 1995:15) (According to Sasse, this property is shared with thetic

assertions with a descriptive function, cf. below.) Sasse mentions the Modern Greek sentence in (5.32) as an example of this type of clause paragraph medial interruptivethetic clause (Sasse 1995b:166, ex. 64).

- (5.32) dhiesxize ti sala tu ksenodhoxiu ja na fi*ji*, otan **ton stamatise mja adhinati foni**
crossed the hall of.the hotel for SUBJUNCTIVE go.away when **him stopped a**
faint voice
“he crossed the hotel hall to leave when *a faint voice stopped him*”

I believe that it is mistaken to consider that expressing interruptions and sudden events is a separate ‘function’ ofthetic assertions. Some of Sasse’s examples, such as the phone ringing example, are clearly event centered *announcements* with non-topic subjects which may occur with minimal or no context, but they may be found paragraph medially to express such events.

Other examples, such as the one in (5.32) don’t seem to me to bethetic at all. Rather, here an argument other than the subject, in this case the object, is the topic and thus the subject behaves as a focus complement. The fact that the semantic role of the subject in this case is not agentive, no doubt contributes to the fact that the default topic selection choice is bypassed.

Finally, it seems to me that interruptions and sudden events are often expressed by means of emphatic statements. Thus assertions expressing such events may sometimes display inversion not because the subject is the focus, but because it is an inverted topic. The crucial difference can only be detected by knowing where the main accent lies.

5.3 EXPLANATIONS

5.3.1 Explanations vs. announcements

The thetic construction is also said to be used when an explanation is called for. Explanations are a lot like announcements (cf. Sasse 1995b:167), but they are different from them in that these statements are not completely out of the blue since there is a certain sense or expectation on the part of the listener that something happened or that something requires an explanation. Sasse calls this the “explanatory function” of thetic utterances. According to him, the “presuppositional background” of such sentences is something like a “statement requiring further explanatory elaboration” or “a question such as ‘what is the matter?’” (Sasse 1995b:167).

It doesn't seem to me that the distinction between ‘announcements’ and ‘explanations’ is a clear-cut one, and typically one and the same sentence can be used either to make an out-of-the blue announcement or to give an explanation. Whether a listener expects or should expect an explanation many times is not something clear-cut, especially if there is no overt question involved (such as *What happened?*).

Actually what a speaker means as an explanation could very well be taken as an announcement by the listener. But, as we will see below, a language may have specialized constructions used only to provide explanations and not to make announcements. Let us look first however at the claim that explanations are (typically) thetic assertions.

5.3.2 Ungrounded explanations

According to Sasse, in Modern Greek “there is a feeling that [VS]¹⁴ is frequent or even sometimes preferred in situations such as” the one in (5.33) (Sasse 1995b:167, ex. 70).

(5.33) ti ejine? - Pethane o pateras mu [VS] / o pateras mu pethane [SV]
what happened - died the father my / the father my died
“What happened? - My father (has) died”

In other words,thetic clauses are found when something in the context requires an explanation (whether the explanation is requested or not). As we can see, however, a topic-comment configuration is also possible in this context. In fact, as was the case with announcements, the fact thatthetic assertions are used to give explanations doesn’t mean that all explanations arethetic. Nothing is farther from the truth.

The same is true in Hungarian, according to Sasse. However, he goes on to say, in this language this type of explanation is expressed with athetic subject-inversion construction only in certain cases, namely when the subject is not inherently topical,¹⁵ and it is restricted to certain semantic areas in which low-topicality subjects prevail, such as those involving “natural states, bodily conditions, sudden events, appearance, damage, or loss” (Sasse 1995b:182) and events such as “[d]oor opening, phone-ringing and similar disturbing events as well as loss, damage and disasters prevail” (Sasse 1995b:184), where the subject is not salient enough for the proposition to be seen as a predication for that referent, but rather the whole event is seen as a unitary happening.

Lambrecht (1994) also discusses an example of what could be seen as an explanation (though it could also be an announcement), one which would display subject inversion in European languages such as Italian or Spanish, and an accented assertion-

initial subject in English. The context for Lambrecht's example is a woman entering a bus, all flustered and loaded with bags. Feeling the looks of the other passengers, she utters the sentence in (5.34) as an explanation for her appearance.

(5.34) My CAR broke down.

Indeed this seems to be athetic sentence with a focus subject. As Lambrecht notes this sentence looks identical to a 'narrow focus' sentence (such as in the answer to the question: *What broke down?*). The reason for this, and this is my interpretation, not Lambrecht's, is that in both cases the subject is the focus.

Notice however that although this assertion is certainly not a predication about the car, it might be interpreted pragmatically as a predication about the speaker, since it expresses that something happened *to her*. As we will see in the next section, in languages such as Basque and Spanish this fact is formalized by making the speaker an argument of the assertion and by making it the topic in this type of explanatory assertion.

5.3.3 Object topics in happens tance sentences

I agree with Sasse that most explanations do not actually have thetic structure. Only explanations which refer to 'happenings' or non-caused events with low topicality arguments are *sometimes* thetic. As long as these states of affairs cannot easily be seen as predications about the subject, the subject will not be treated as the ground for the assertion and the subject will typically be the focus of the assertion.

But, as I mentioned, explanations often do indeed have a topic, but it is just not the subject, but another argument of the verb. That is, sometimes assertions which make announcements or give explanations, and in particular those which appear in a context in

which something is assumed to have happened to someone or something, in many languages typically do contain a topic, which is not the subject. This referent is actually the reason for (or source of) the expectation for an explanation on the part of the listener in the first place (the speaker in the case of example 5.34).

Affected referents, however, are rarely found filling subject roles. In some languages, such as Basque and Spanish, such topics are coded as optional additional arguments of the verb coded as dative (indirect object) arguments, and as is often the case with very accessible (predictable) topics, they are typically coded only covertly with dative benefactive agreement (or verb clitic). Thus Lambrecht mentions the Italian version of the sentence in (5.34), which has subject inversion and which is a perfect example of this phenomenon. The almost identical Spanish version, shown here in (5.35), is almost identical to the Italian version.¹⁶

- (5.35) Se me ha averiado el coche.
REFLEX to.me it.has broken.down the car
“My car broke down.” (Lit. “The car broke down on me.”)

English does not have an similarly highly grammaticalized way of coding a ‘benefactive’ argument at the clause level. In the English versions of such ‘happstance’ sentences what seems to be the topic in languages like Italian and Spanish is often coded as a possessive pronoun attached to the subject, as in (5.34) above.

It is not clear whether the referent encoded as a possessive in English could be seen as the topic of the assertion, since topics are typically arguments of the verb, although, pragmatically speaking, it does seem to perform the same function as the Italian and Spanish dative arguments. Interestingly, however, this possessive pronoun in

English goes at the beginning of the clause, just as the benefactive topic pronouns in Italian or Spanish.

'Benefactive grounding' is very common for what Sasse calls "bodily indispositions", e.g. (5.36), and "physical affections", e.g. (5.37), as well as other "sudden disturbing events", as in the examples in (5.38a&b) (cf. Sasse 1995b:162-3).

(5.36) Me sangra la nariz.

to.me it.bleeds the nose

"My nose is bleeding." (Cf. "I got a nosebleed.")

(5.37) Me ha picado un mosquito.

to.me it.has pricked one mosquito

"A mosquito bit me." (Cf. "I got a mosquito bite.")

(5.38) a. Se me ha caído la taza.

REFLX to.me it.has fallen the cup

"I dropped the cup." (Lit. "The cup fell down on me.")

b. Se me ha quemado el pan.

REFLX to.me it.has burned the bread

"I burned the bread." (Lit. "The bread burned on me.")

What all these sentences have in common is not that they are thematically ungrounded (thetic or topicless) but rather that the thematic ground is not the subject but the dative/benefactive. Notice that in the last two examples in (5.38) the English versions have experiencer *subjects*, something which would seem quite odd or impossible in some languages. The lack of a grammaticalized benefactive role may perhaps be what motivates the existence of subjects with such roles in English, a phenomenon which has puzzled grammarians for a long time. Notice also the tendency for the assertions in (5.36) and (5.37) to be realized in English with sentences in which the topic is the subject (cf. *I got a nosebleed, I got a mosquito bite*).

A major difference between on one hand what Sasse calls “bodily indispositions” and “physical affections”, and on the other “sudden disturbing events”, is that the former are always thematically grounded (just not by the subject’s referent), whereas the latter may or may not be grounded. Thus the two Spanish assertions in (5.38) would indeed bethetic sentences if they were missing the benefactive argument, as we can see in (5.39a&b).

- (5.39) a. Se ha caído la taza.
REFLX it.has fallen the cup
“The cup fell/dropped.”
- b. Se ha quemado el pan.
REFLX it.has burned the bread
“The bread burned.”

5.3.4 Some further examples of thetic explanations

To elaborate on the peculiarities of explanatory assertions, let us look at the examples that Lambrecht (1994) uses to demonstrate the difference between topic-comment (“predicate focus”) information structure, as in example (5.40A) below, andthetic (“wide focus”) information structure, as in example (5.41A).

- (5.40) *Topic-comment sentence* (“predicate focus”):
Q: How’s your neck?
A: My neck (it) HURTS

- (5.41) *‘Thetic’ sentence* (“wide/utterance focus”):
Q: What’s the matter? [understood: *with you*]
A: My NECK hurts.

In the topic-comment sentence the subject is the topic and the verb receives the assertion's accent. In thethetic sentence, on the other hand, the assertion-initial subject receives that accent.

Although the sentence in (5.41A) seems to be formallythetic (subject focus, no topic), from a pragmatic perspective we could say that the speaker is somehow what the assertion is about and thus a topic of sorts. In the question that this is an answer to we can also discern an 'understood' topic: *What is the matter with you?* In Italian and Spanish, as well as in Basque, the speaker in these sentences would definitely be the topic, coded as a dative argument of the verb, as we can see in the Spanish version of (5.41) in (5.42).¹⁷

(5.42) Spanish version of (5.41)

Q: ¿Qué (te) pasa?

what (to.you) it.happens

"What's the matter (with you)?"

A: Me duele el cuello.

to.me it.hurts the neck

"My neck hurts."

The sentence in (5.42A) is notthetic. The subject is not the topic, however, but rather the dative argument's referent is.

The fact that explanations often have topics, albeit non-subject ones, is not the only reason for denying a correlation betweenthetic structure and explanations (explanatory assertions). The truth is that most explanations do not havethetic structure or unlikely topics at all, a fact that Lambrecht forgets to mention. Other equally common types of answers (explanations) to the same question are regular topic-comment assertions with subject topics, in English as well as in other languages, as we can see in the examples in (5.43) and (5.44) below.

(5.43) Q: What's the matter? [understood: *with you*]

A₁: I had an accident.

A₂: I have a headache.

A₃: I fell down.

(5.44) Q: What's the matter? [understood: *with him*]

A₁: He had an accident

A₂: He has a headache

A₃: He fell down.

These sentences are indeed *about* the referent of their subjects, not only in the loose pragmatic sense mentioned earlier about the sentence in (5.41A), but more generally as a regular topic coded as an argument. These examples show that answers to these 'what's-the-matter' type questions need not bethetic. In fact, perhaps they rarely are.

5.3.5 A special type explanation construction

In addition to normal topic-comment andthetic sentences, languages often use a special construction or strategy for answering *What's the matter*-type questions. In this particular strategy/construction, an entity is first identified (in its own intonation unit) as what *the matter* is. Immediately afterwards, the speaker (typically) proceeds to state what exactly the matter is with that referent in an assertion which uses the new referent as the topic. This type of sentence is exemplified in (5.45A).

(5.45) Q: What's the matter?

A: My neck. It hurts.

This construction, which contains two intonation units, is equivalent to the sentence in (5.41) above (*My neck hurts*). Notice that this initial nominal is very different from a dislocated topic, as shown by the different intonational pattern (with a period rather than a comma intonation contour). In fact the noun phrase in the first intonation

unit in (5.45A) is equivalent to a presentative-likethetic clause should perhaps be seen as the focus of an ‘understood’ assertion, cf. *My neck's the matter. It hurts*. In other words it would be an assertion’s focus fragment, much like fragment answers to content questions. This focus fragment may not be enough of an explanation and thus it is typically followed by an assertion which elaborates on the explanation, which is in fact a predication about the referent introduced in the preceding intonation unit.

Basque and Spanish have a somewhat more specialized complex construction to express this same explanative function, which differs from the English one in that the sentence which follows the newly established entity typically displays a ‘complementizer’ (*que* in the case of Spanish and *-la* in the case of Basque). In other words, these clauses are identical to two types of dependent clauses in these languages: (1) completive complement clauses (assertions); and (2) the assertions which follow newly presented referents in presentative constructions (see below). Thus the Spanish sentence in (5.46A). is equivalent to the English one in (5.45A) above.

(5.46) Q: ¿Qué pasa?

“What’s the matter”

A: El cuello. (Que) me duele

the neck (understood: the speaker’s neck), (that) to.me it.hurts

“My neck. It hurts”

The second part of this construction, i.e. the predication, doesn’t seem to be added as an afterthought, but seems to be, at least most of the time, a fully planned addition, and the whole two-part (two intonation unit) construction seems to be quite integrated. The reason for the complementizer is that there is (or there was at one time in the development of the construction) an ‘understood’ verb here. Thus, answers to *What’s the*

matter-type questions in these languages which do not have a bipartite structure may also display the complementizer, as we can see in (5.47).

- (5.47) Q: ¿Qué (te) pasa?
“What’s the matter (with you)?”
A: ((Me pasa) que) me duele el cuello.
((to.me it.happens) that) to.me it.hurts the neck
“My neck hurts.”

We should not, however, think of the bipartite explanation construction in (5.46A) as being *derived* from one-part explanations such as this one in (5.47A), although it may be seen as a blend of two constructions. Notice that occasionally we find in actual speech that the entity that is presented in the first part of the construction doesn’t fill an argument role in the following assertion, as we can see in (5.48) below.

- (5.48) El cuello. Que tengo tortícolis.
the neck that I.have torticollis/stiff.neck
“My neck. I have a stiff neck.”

In explanations which involve more than one entity, different choices are possible for the entity which is presented first. Thus we find cases such as the one in (5.49) in which the cause of the problem (*the neighbor’s dog*) and not the affected part (*my leg*) is chosen as the main entity in the explanation (although it could be the other way around).

- (5.49) El perro del vecino. Que me ha mordido la pierna.
the dog of.the neighbor, that to.me it.has bitten the leg
“The neighbor’s dog. It bit my leg.”

Notice also that the initial noun phrase need not have an identifiable (or otherwise grounded) referent, as we can see in (5.50) below.

- (5.50) Un perro. Que me ha mordido la pierna.
one dog that to.me it.has bitten the leg
“A dog bit my leg.” (cf. “A dog bit me on the leg.”)

It is interesting that one of the examples of the ‘explanatory function’ in Sasse’s study of Hungarian, namely the example in (5.51:B), has exactly this same bipartite structure (Sasse 1995b:184, ex. 121).

- (5.51) A: Mi történt?
what happened
“What happened?”
B: barátom ... Elütötte egy autót.
the friend.my ran.over.him a car
“My friend ... he was run over by a car.”

Here a referent is established as the source of the speaker’s ‘problem’, the ‘focus’ of the answer, and *at the same time* it is established as the topic for the following sentence where the explanation is elaborated. And since this referent, which fills the object role, is the topic of the assertion, the subject of that clause (“a car”) cannot be the topic and thus it is inverted. The same thing would happen in Spanish, for instance: *Un amigo mío, que le ha atropellado un coche.*¹⁸

As I said, Basque has an equivalent construction to the Spanish one, with the predication following the noun phrase typically marked by the completive subordinator *-la*, which is also used (in most dialects) in post-presentative finite predicates, as we will see in Section 5.4.7 below. We can see an example in (5.52) below.

- (5.52) A: Emakume!
woman!
B: Bai?
yes?
A: Poltsa. Erori egin zaizula.
bag:DEF fall:PFV do:PFV it.is.to.you:COMPL
“A: Lady! B: Yes? A: Your bag. It fell down.”

Notice too that although speaker A originates this exchange, A's second construction is indeed an 'explanation' given by A to B as to why A is calling B. In this example the 'dummy' verb *egin* "do" indicates that the verb is the focus.

This bipartite construction blend is interesting in part because of its similarity to presentational constructions which are followed by a finite clause acting as a predication about the newly introduced referent (cf. English *I saw a man yesterday, who had a broken leg*). We will look at presentative assertions and post-presentational predicates in the next section.

5.4 INTRODUCTIONS

5.4.1 Existentials

Existential clauses are very often topicless (thetic) across languages. The reason for this is, of course, that the referent in existentials is typically unidentifiable, i.e. most inaccessible and thus not topical enough to be the topic. Occasionally existentials have topic-comment structure, such as those formed with the verb *exist* in English (e.g. *God doesn't exist*), which occurs when the idea in question is accessible.

It has been noted that existentials are very often used to introduce new referents into the discourse, especially unidentifiable referents, for the purpose of then saying something about them. This, however, is not the only way new referents can be introduced, as we have already seen and will see again below. Also, not all existentials introduce referents which are destined to become topics.

Because the subject of existential constructions is (typically) the focus, we find that in many languages, primarily languages which place the focus after the verb, existential sentences involve subject inversion. English has a specialized existential construction, the *there*-construction, in which the focus subject is postverbal (as in most VO languages). This in fact is the main construction which allows focus subject-verb inversion in modern English. Spanish uses the defective verb *haber*, previously meaning “have”, from Latin *habēre* “hold, handle”, in which the noun phrase with the new referent used to be an object, rather than the subject, and the focus. Basque uses either the copula verb *izan* “be” or, more generally, the stative/locative copula *egon* “be (at a place)” in existential sentences (see below).

Inversion is not always associated with ‘presentations’, however, even in languages which do allow inversion. In Hungarian, for instance, according to Sasse 1995b, the introductory function is not served by a subject-inversion construction, other than in “traditional genres such as fairy tales” (Sasse 1995b:182). In other words, in Hungarian, much like in Basque and other “OV” languages, the existential subject is in preverbal focus position. In Basque, however, inversion is also possible, such as at the beginning of fairy tales, but also quite commonly in speech, as we will see (though not in standard, written Basque).

5.4.2 Presentationals and focus

New referents need not be introduced in specialized, ‘presentative’ constructions. There are two major ways in which referents may be introduced into the discourse. The first one, as we saw in Chapter 4, is by means of a ‘dislocated’ topic nominal in an

intonation unit preceding an assertion (predication) about them. These referents are typically accessible to the hearer, although there are exceptions, as we saw.

The second way in which a referent is introduced into the discourse is as the focus of an assertion *in any grammatical role, not just as focus subjects*. Typically they are introduced in the role of objects (e.g. *I saw a man*), or as some other complement of the verb (e.g. *I ran into a man*), including the subject (e.g. *A man came*). It is examples such as this last one, in which the introductory focus is the subject of the assertion, that have attracted the greatest attention of grammarians, but it would seem that this is just one more possibility among the different possibilities of argument focus. Subject focus presentations, of course, are unusual because they're typically topicless orthetic (as long as an object is not the topic).

The semantically rather 'empty' 'existential' verb we have been discussing is not the only verb with which referents are introduced into the discourse in the subject role. Every language has other, typically less specialized, predicates which are used commonly to introduce referents as subject foci. Sasse (1995b) classifies them as 'static existentials' and 'dynamic existentials'. The difference between the two types is that the latter involve motion of the entity (*come, enter*), whereas the former don't (*be, lie, rest*). All these are predicates in which the subject's semantic role is not very topical, and thus poor topic candidates and good focus candidates, even if the entity in question is accessible. Dynamic existentials are, according to Sasse, "all verbs of 'inceptive existence' in the broadest sense, i.e. verbs of appearance, happening, occurrence, beginning, establishing, opening, inauguration, arrival, and the like" (Sasse 1995b:161).¹⁹

Very often, assertions with this kind of predicates and thetic structure (focus subjects) do not introduce new referents into the discourse, but reintroduce, or bring back referents which are accessible from the prior context ('discourse-old') or which are at least identifiable to the hearer ('hearer-old') and thus we cannot say that these predicates are specialized to introduce *new* referents. Some existential constructions, such as the *there*-construction in English, are specialized to introduce unidentifiable referents.

Sasse also mentions other verbs which are semantically related to these verbs and which also have a tendency to have focus subjects, i.e. dynamic verbs of "inceptive non-existence", "such as verbs of disappearance, ending, disrupting, closing, stopping, deferring, loss, damage, mishaps and the like" (Sasse 1995b:161). These verbs tend to have more accessible subjects and are thus more likely to be found also in non-thetic assertions. However, because their semantic characteristics, including the semantic role of their subjects, these verbs have a strong tendency to appear in thetic structures rather than being used as predicates.

After the existential construction, 'dynamic existentials' seem to be the preferred way of introducing referents into the discourse as subjects in the Spoken Basque Corpus, as we will see below, the most common such predicate being *etorri* "come". These dynamic presentationals, however, just as often as not, they reintroduce accessible ('discourse old') referents, and are indistinguishable from announcements or 'sudden events' constructions (cf. below). Very common, similar Basque verbs, which could be seen as either dynamic or static 'existentials', are *agertu* "appear" and *azaldu* "show up".

But, as I said, thetic assertions with such intransitive predicates are not the only, or even the major mechanism for introducing referents into the discourse. Whenever

possible referents are introduced in the focus role in clauses which have a topic. In such clauses the new referent is introduced by ‘connecting’ or ‘grounding’ it to an existing topical referent, e.g. *then she saw a man*, Intransitive ‘presentational’ predicates are primarily employed when such a connection is not possible. Perhaps the most common grammatical role for referents to be presented in is direct object position, also absolutes in Basque, especially with verbs such as *eduki* “have” and *ikusi* “see”. In the case of this last verb, the assertion may be ‘impersonal’, that is, the ergative role may be suppressed and understood to be generic, as in example (5.53) below.

(5.53) 93C2B03

- 2 *pelikulie asten-da mendixen-* ..
 movie:DEF it.starts mountain:DEF:LOC
- 3 *Bueno. mendixan ikusten-da/ gison bat,*
 well mountain:DEF:LOC X.sees.him man one
- 4 *=e= zuaitz baten,*
 uh tree one:LOC

“Well then, / the movie starts in the mountains. / Well in the mountain you can see a man, (~ “... a man can be seen.”, “one sees a man”) / uh on a tree”

Unlike in accusative languages, the absolute argument here keeps the same marking (absolute) whether the verb has an ergative argument or not (cf. Chapter 2).²⁰ In this example the focus constituent is rheme-final or extraposed.

5.4.3 Circumstantially grounded presentatives

As I already said, existentials, whether used for introducing major entities into the discourse or not, and other typicalthetic sentences, are typically not thematically grounded (i.e. they do not have topics).²¹ However, we very often find that these clauses are indeed grounded, just not *thematically* grounded, but rather grounded by a locative, temporal, or other circumstantial ‘referent’. In other words, very frequently we find that

thesethetic sentences are grounded by some aspect of the context, typically a place or a time to locate the new referent with respect of this most important contextual information.

This phenomenon is so common that languages may have special constructions in which a circumstantial, often a locative, is involved in such a presentation clause-internally, i.e. in the position typically reserved for overt topics, and not dislocated as a regular setting. In English there is a construction, or family of constructions, for instance, often known as ‘locative inversions’, which fits exactly this description, e.g. *In the patio sat one of the guests*. This construction in English may take only a handful of (intransitive) verbs and its function is to introduce a referent into the discourse (cf., e.g., Hetzron 1971, 1975; Bolinger 1972b, 1977a; Kuno 1971; Hooper and Thompson 1973; Green 1980; Bresnan 1994; and Birner 1994). These sentences are not common in speech, however. They are clearly associated with written narrative genres, perhaps as a remnant from an earlier stage of the language when such constructions were more common.

In these clauses the ‘locative ground’ is indeed outside the rheme, much like overt topic subjects (thematic grounds) are in canonical topic-comment sentences. It is thus no surprise that some investigators have argued that the locative (ground) in these clauses is actually *the* ‘subject’, since they look and act so much like (topic) subjects in canonical sentences. It is the fact that these assertions don’t have a topic that makes it possible for a locative to become their major grounding element and thus to display such a high degree of integration with the rest of the assertion, such as being in the same intonation unit as the rest of the clause as opposed to being ‘dislocated’.

Osa (1990) very perceptively noticed that Basque oral narratives very often begin with either one of the formulas which we can see in (5.54), with a time or place adverbial in clause initial position and a nominal introducing a referent in either preverbal or postverbal position.

- (5.54) a. Time/place-adverbial **Verb** New-Referent
b. Time/place-adverbial New-Referent **Verb**

The new referent introduced in this construction is typically an important one, typically a major protagonist. Osa provides some examples from a well-known collection of oral narratives by Basque anthropologist Barandiaran (1973), including the ones in (5.55) and (5.56) below (from Barandiaran 1973, pp. 313, and 307, respectively; cited in Osa 1990:135). (These sentences are from two different dialects, which explains some of the differences.)

- (5.55) Bein batian **aita ta ama** bizi ziran.
once one:LOC father and mother live:PFV they.were
“Once there lived a mother and father ...”

- (5.56) Elixalde baten bixi zien **andragixon bi**
church:area one:LOC live:PFV they.were woman:man two
“Near this church there lived a woman and a man (a couple)”

Osa assumes, wrongly I believe, that the adverbial in (5.56) is in focus-position since it comes before the verb. I believe that it is fairly clear that in both cases the temporal is outside the assertion proper, as a setting or ground, and that the referent of the subject nominal is indeed the focus of the assertion. The difference between the two examples is simply that in the first one the focus is in assertion-initial preverbal position whereas in the second one the focus is assertion-final (extraposed, delayed, or dislocated), not postverbal (cf. Chapter 3). In other words, in the second example the postverbal nominal

would be accented as the focus and the verb itself would be most likely bear a noticeable accent as well.

5.4.4 The form of Basque existentials and otherthetic sentences

On the surface, by looking at the Spoken Basque Corpus one may get the impression that Basque has two or three strategies for presenting referents in existential and other such ‘thetic’ sentences. In one, the nominal which codes the new referent comes before the verb, just like in (OV) languages with preverbal focus position. But in the others, the nominal comes *after* the verb, just like in (VO) languages with postverbal focus position, such as Spanish. Thus in (5.57) we can see an example of preverbal existential presentation from the Spoken Basque Corpus, whereas in (5.58) and in (5.59) we see examples of postverbal presentations.

(5.57) 93C2B01

464 **Andre= gizen beltz bat** zeon,
woman fat black one she.was
“There was a big black woman,”

(5.58) 93C1B06

167 eta dago **neska bat**, -
and she.is girl one
“and there is a girl.

(5.59) 93C1A02

1 Badago **gizon bat**,
EMPH:he.is man one
2 eta= .. kamioi bat ikusten-du.
and truck one see:IMPFV-he.has.it
“There is a man, / and he sees a truck.”

The variation is primarily found in speech, rather than in writing. In the Written Basque Corpus, for instance, where the existential construction is quite rare, they all have SV structure, i.e. there are no examples of VS existentials. In the Spoken Basque Corpus on the other hand, where the construction is quite common, in a majority of cases, 86/125 or 68.8% of the time, the absolutive nominal follows the verb.²²

It should be clear from what we saw in Chapter 3 that this is due to the fact that, in addition to the most basicthetic structure in which the focus subject is preverbal, Basque has two additional alternatives which are used primarily in the spoken language, both of which result in the nominal coming after the verb.

The first strategy, exemplified here in (5.58) is focus delay or extraposition. As we can see, the synthetic finite verb in this case does not bear the emphatic particle *ba*, thus signifying that neither the verb nor the polarity is the focus of the assertion, but rather the extraposed nominal is the focus. On the other hand, the normal reasons for using the focus delay construction do not seem to apply in this case. It seems that the use of the focus extraposition construction in this particular context is to some extent ‘institutionalized’ or grammaticalized, perhaps under the influence of the surrounding VO languages.

The other strategy which results in the presented nominal being postverbal is exemplified in (5.59) and seems more surprising. As we can see in this example, the verb bears the emphatic particle *ba*, which indicates that the whole rheme, presented emphatically (the ‘polarity’) is the focus of the assertion, and the subject is inverted. This state of affairs, as we saw in Chapter 3, can result from either an emphatic assertion with topic postposing, or from an emphaticthetic sentence which never had a topic to begin

with, to use a derivational metaphor. In these sentences, unlike in un-emphaticthetic (subject focus) sentences, the subject is not the focus of the assertion, even if the referent in question is new. On the other hand, it doesn't seem that it could be the topic either, given its low topicality. But notice also that to be unaccented they must be easy to accommodate. The fact that these subjects typically have stereotypical referents may help us understand why they can be 'presented' in such a non-focus position.

On the other hand, that this might be an antitopic construction (or that it might have been at one time), is suggested by the existence of topic-comment existential assertions with low topicality topics, as we saw in Chapter 4. In other words, sentences such as the one in (5.59) could be seen as the antitopic equivalent of such topic-comment existential sentences with 'new' topics. In other words, the existential assertions in (5.59) would be the emphatic version of *Gizon bat badago* "A man, there is", "there is (indeed) a man". Such sentences are used primarily in north-eastern dialects of Basque, but to some extent in all dialects as well.

The relative high frequency of such sentences with subject inversion in speech might be due to the similarity complete parallelism of these Basque sentences with Romance ones, or at least Spanish ones.

Emphaticthetic sentences may have relatively 'new' subjects in other languages as well. Such assertions are typically not found in presentative contexts, however, but rather when an assertion contrasts with another one, such as in the exchange in (5.60B).

- (5.60) A: There's nobody in there.
B: No, no, there is a man. I've seen him.

Notice, however, that in this context the subject's referent (*a man*), or rather the *idea* about such a referent, although unidentifiable (and thus indefinite) must be somehow

accessible from the context. In the context for this sentence, the idea that *someone* might be found must have been brought up (actually, the assertion in 5.60A might be enough to evoke the presence of *a man*). Notice also that, even in this context, this type of sentence becomes less acceptable the more specific the new referent is (unless it had been mentioned before), e.g. #?No, no, there is a lawyer. I've seen him.

We find the same two possibilities (preverbal and 'postverbal' focus) with other common 'presentative' predicates as well, such as *etorri* "come" and *azaldu* or *agertu*, both meaning "appear", all high-frequency predicates in the corpus. The main difference is that with these verbs, all of which are conjugated analytically, there is no formal indication that the verb/rheme is the focus (as we saw the *ba* particle is only used with synthetic verbs), and thus intonational cues are crucial for differentiating a verb/polarity focus sentence and an extraposed focus subject sentence. On the other hand, intonational cues are not always as clear as we might want them to be in speech and there is reason to believe that in this particular context the different strategies may be neutralized and reanalyzed as postverbal focusing, at least for some speakers.

There are 85 declarative sentences with *etorri* "come" in the Spoken Basque Corpus and, of these, 39, or 45.3%, display 'subject inversion' (cf. Appendix 5.6 for a full list).²³ The vast majority of these inversions are clear examples of focus delay/extraposition. Some, however, are examples of the antitopic construction, and some also display topic-comment structure (i.e. the subjects are topics). In (5.61) below we see an example of a preverbal subject-focus presentation, i.e. a presentative with SV order, and in (5.62) an example of a postverbal extraposed subject-focus presentation, i.e. a presentative with VS order.

(5.61) 93C3A07

114 eta **manifestazix**o bat zetorren.
and demonstration one it.was.coming
“and there was a demonstration coming.”

(5.62) 93C3A12

40 Alako batean ba= **etortzen**-dia, **polizi** batzuk edo, ..
thus:GEN one:LOC well come:IMPFV-they.are police some:PL or
“Then all of a sudden some policemen come or something, ...”

The intonational break before the postverbal subject in this latter sentence leaves no doubt that the nominal is indeed the focus, just like in the preceding one. In the following example in (5.63) the focus not only not only clause final with a preceding minor intonation break, but is clearly in a separate intonation unit as well.

(5.63) 93C2B01

20 zer ta or **etortze**-zako,
what and there come:IMPFV-it.is.to.him
21 bere atzetik **manifestazi**yuba. –
his behind:ABL demonstration:DEF
“right then and there comes, / from behind him a demonstration.”

In this example the reason for dislocating the focus subject seems to be to produce a suspense effect in the presentation. In (5.64), on the other hand, the reason for placing the focus subject in final position seems to be the speaker’s hesitation as to how to verbalize this referent.

(5.64) 93C2B06

26 Etortzen-da=- **etortzen**-dia= ^zea - madero bat,
come:IMPFV-he.is come:IMPFV-they.are what – policeman one
“(Then) comes – come – uh a policeman.”

Although there are plenty of clear cases of assertion-final focus subjects, sometimes in rapid speech an extraposed focus may be hard to tell apart from a postverbal focus. It is

quite clear, however, that postverbal position is not a focus position in Basque, the way that it is in VO languages (except perhaps for some imperfect speakers in some contexts).

No *azaldu* or *etorri*thetic sentences were found in the written corpus for comparison. A computer search in the additional three-novel corpus reveals only a handful ofthetic sentences with *agertu* and *azaldu* “appear”, all of which have the nominal in preverbal position. We can see two examples with *agertu* in (5.65) and (5.66).

(5.65) *Hil ala bizi*
behi eskeleto bat agertu-zen.
cow skeleton one appear:PFV-it.was
“a cow’s skeleton appeared”

(5.66) *Otto Pette*
Mugarre kapitaina agertu-zen,
Mugarre captain:DEF appear:PFV-he.was
“Captain Mugarre appeared”

Thus it isn’t just a fact about narrative thatthetic sentences in general, and presentativethetic clauses in particular, have extraposed focus nominals, but it is primarily a fact about *spoken* narrative, since focus extraposition is much more frequent in speech (in writing the focus extraposition construction is typically used when the focus constituent is long and complex, cf. Chapter 6).

5.4.5 The delayed focus strategy in presentations

The preference for focus extraposition in speech may be in great part due to the constraints of on-line speech production, as evidenced by the fact that many times extraposed foci in speech are preceded by pauses. This would be a motivation for occasionally using a construction which is typically used with complex foci. The great

prevalence of this construction in Basque speech might need additional motivations, however. One possibility is that Romance influence is motivating speakers to use a focus construction which looks like a Romance focus construction, namely clause final or extraposed focus. A further motivation for the use of this construction in presentative situations might be rhetorical, namely that it is ideally suited for introducing referents into the discourse, at least in some contexts (cf. Hetzron 1975).

In Chapter 3 I mentioned that focus extraposition is sometimes used when the focus serves an *identificational function*, that is, for identifying, or making more specific a referent whose existence is already known. This is similar to the use of the *wh*-cleft construction with focus extraposition, e.g. *That's not what I ate. What I ate was, a banana*. This seems to produce a 'suspense effect' on the presentation. However, an identificational focus assertion is nothing more than a presentative assertion in which the existence of a vaguely defined entity is known beforehand. We can see an example of this in (5.67) below.

- (5.67) a. There was somebody in that room.
b. There was, an old woman.

The assertion (5.67a) is a regular existential, in which a new referent is introduced, but the focus phrase is extremely non-specific (*somebody*). The second existential assertion makes the referent much more specific, more identifiable, and more accessible. And, interestingly, the second existential in this type of context often displays an extraposed focus intonation pattern, with an intonation break before the focus. In English the difference is simply intonational, but in a language such as Basque, there is a much more noticeable difference in the order of the focus element, which comes after the verb

(actually, clause finally or extraposed) rather than before the verb. We can see this in the Basque version of (5.67) in (5.68) .

- (5.68) a. Norbait zegoen gela horretan.
 someone she.was room that:LOC
 b. Zegoen, .. atso bat
 she.was .. old.woman one
 a. “There was someone in that room.” b. “There was an old woman.”

In a context such as this the extraposition creates the ‘suspense effect’. In this discourse the speaker pre-announces the existence of a certain individual first, which is then made more specific, or ‘identified’, with an extraposed focus existential construction.

We may speculate that this identificational focus construction has been appropriated, or extended, to code normal presentative assertions, perhaps under the pressure to use presentative assertions with ‘postverbal’ focus constituents. In the example we have just seen the expectation that there is a referent which needs to be identified is given by the prior context. Sometimes, however, the expectation about the existence of an entity can be inferred by the hearer without it being made explicit by the speaker. In fact, given enough of a pause between the verb and the focus, the verb itself can be said to create the expectation that there is an entity waiting to be identified.

5.4.6 Subject inversion in elicited sentences

In elicited Basque sentences, much like in written Basque, subject inversion in presentational-like sentences is much less frequent than in the spoken narratives from the Spoken Basque Corpus. However, with some predicates which are often associated withthetic sentences, such as *azaldu* “appear”, I found that subject inversion was commonly produced in elicited sentences, such as the example in (5.69), even by speakers who

consistently place focal subjects in preverbal position (the Spanish sentence used to elicit this sentence was: *Ha aparecido mi libro*).

- (5.69) *Azaldu*-da nere liburua
appear:PFV-it.is my book:DEF
“My book showed up.”

The reason for this inversion is not surprising, however. It is clear that this is an emphatic assertion with polarity focus. Notice that the subject’s referent (*my book*) is identifiable, as must be the idea that this entity had been lost, and the subject could have also been a topic in this context. Thus the subject can only be an antitopic, or else athetic sentence with rheme focus, either one of which explains the position of the very focal focus verb in focus position and the inversion of the subject. In either case the verb, and not the subject receives the focus accent, cf. *Azaldu da nire liburua*. Inversion would not be possible, on the other hand, if the subject’s referent was new, and thus a stronger focus candidate, cf. *Liburu bat azaldu-da* (SV) “A book showed up”.

If fact, elicitation of all types of sentences which in Spanish would typically display subject inversion consistently produced sentences in which the subject was preverbal in Basque. There were a few exceptions to this, the vast majority of them contain subjects which are not foci, but antitopics (for a complete list of 182 sentences, see Appendix 5.5). Thus, besides the example in (5.69), there were several cases with subject inversion in which the sentences were obviously emphatic, such as exclamations, cf. (5.70) and (5.71).

- (5.70) *Badaki nagusiak!*
EMPH:she.knows.it boss:DEF:ERG
“The boss knows it already!”
(Original Spanish version: “¡Ya lo sabe el jefe!”)

- (5.71) Izango-direla arazoak!
 be:FUT-they.are:COMPL problems
 “Will there be problems!”
 (Original: “¡Que si habrá problemas!”)

Subject inversion was also found in desiderative assertions which were clearly emphatic assertions, such as (5.72) below.

- (5.72) Ordain-dezala nagusiak!
 pay:ROOT-he.may.have.it:COMPL boss:DEF:ERG
 “Let the boss pay!”
 (Original: “¡Que pague el jefe!”)

Some antitopics were motivated by the fact that the context required that the topic (and the whole proposition) be very accessible, as is the case in the example in (5.73) below.²⁴

- (5.73) Arrazoi horregatik itzuli-zen bere etxera Miren.
 reason that:by return:PFV-she.was her house:DEF:ALL Miren
 “That’s why Mary went back home.”
 (Original: “Por esa razón volvió a su casa Miren.”)

The only possible (imagined) context here thus makes the focus constituent significantly more salient than average, which, as we saw, is a condition which favors emphatic information structure and topic inversion.

A subject would also be postposed if the topic and focus roles were filled up by referents not in the subject role and as long as the subject’s referent met the topicality (accessibility) condition for being in the assertion tail, as in example (5.74).

- (5.74) Mireni (T) beldur (F) ematen-zion **espero ez zuen bisita harek**.
 Miren:DAT fear give:IMPFV-it.did.it.to.her expectation not she.had.it:REL visit
 that:ERG
 “Miren was scared about the unexpected visit.”
 (Original: “A Miren le asustaba la inesperada visita.”)

Here the overt dative argument's referent is the topic, the non-referential bare noun *beldur* "fear" is the focus, and the postverbal subject is simply a given complement in the assertion tail.

Another context in which the elicitation method produced sentences with subject inversion was one in which verbs of saying followed quotes, such as in example (5.75) below. Here the quote is obviously the focus of the assertion.

- (5.75) "Ez-dugu amore emango" esan zuen **Gobernuko Presidenteak**.
not-we.have.it yielding give:FUT say:PFV he.had.it government:GEN
president:DEF:ERG
"We will not give up," said the President of the Government."
(Original: "No cederemos," dijo el Presidente del Gobierno.)"

This type of subject inversion is very common in writing in many languages, and it is regularly used in the written Basque corpus. It is never used, however, in the Spoken Basque Corpus, where extraposed quotes are always used.

In addition, there are three cases of subject inversion which are different from these antitopic sentences. The first one, seen here in (5.76), is the Basque version of the example we saw in (5.18) above (on pg. 568), from Sasse (1995b:167, no. 68).

- (5.76) Berriz ere itzuli-dira **ohizko espioitza metodoak**.
again also return:PFV-they.are habit:INSTR:GEN espionage methods:DEF
"Traditional espionage methods have returned to the scene."
(Original "Vuelven a la escena los tradicionales métodos de espionaje.")

The accentual and intonation cues of this sentence make it clear that we are dealing with an extraposed focus. The subject-focus extraposition construction might have been preferred due to the length and complexity of the focus phrase, which, as I said, is one of the main correlates of focus extraposition, perhaps the major one in written Basque. Needless to say, the focus could also be placed in preverbal position in this sentence

(SV). Notice that the English translation of this sentence that I have given above would most naturally have a topic-comment configuration with the phrase *to the scene* as the focus constituent.

Finally, there are two elicited sentences, which existential sentences in this subject's elicited corpus in which the subject is inverted, but in which the finite verb is coded with the emphatic particle *ba* which indicates that the polarity, and not the subject, is the focus. Let us look at these two sentences in turn.

The referent of the inverted subject of the assertion in (5.77) is not identifiable to the hearer, but notice that it must be easy to accommodate. The sentence states that a person who doesn't know a particular information *does* exist. Thus, the idea that some such person exists (or doesn't exist) must have been brought up or suggested in the imagined context in which the subject placed such a sentence. In other words, the idea of such a person (not any particular referent) must be accessible, and what is asserted here is its existence.

(5.77) **Bada nagusia datorrela ez dakienik**

EMPH:it.is boss:DEF he.is.coming:COMPL not he.knows.it:REL:PART

"Someone doesn't know that the boss is coming." (Lit. "There is someone who doesn't know that the boss is coming")

(Original: "Alguien no sabe que va a venir el jefe")

Needless to say, the subject of this sentence could also come before the verb, but in that case it would not be the focus, but the topic. Notice that the structure of this sentence is quite different from the one used to elicit it.

Finally, the sentence in (5.78) is much harder to explain. This existential sentence is an example of the stereotypical way of starting traditional tales (cf. *Once upon a time there was a ...*). This sentence has a postverbal subject with an unidentifiable referent,

the major protagonist of the story. Thus we would have expected it to be the focus of the assertion. Notice, however, that the verb bears the particle *ba*, indicating that the polarity is the focus.

- (5.78) Behin batean bazen **oso alferra zen errege bat**.
once one:LOC EMPH:he.was very lazy he.was:REL king one
“Once upon a time there was a very lazy king.”
(Original: “Erase una vez un rey muy perezoso.”)

The problem is that the idea expressed by the subject would seem to be most inaccessible and thus it shouldn't qualify as an antitopic (or as a plain assertion-tail complement either). In fact, the subject would seem to be the most focal element in the assertion and thus should be the focus.

On one hand, this sentence might not seem very unusual in light of the existential sentences that we saw in Chapter 4 with rather inaccessible topics. This assertion type is quite generalized, however, to all dialects and not just the north-eastern ones. In addition, it is obvious that this type of sentence performs a very special function, namely starting traditional tales, tales which have stereotypical, unsurprising, and even 'familiar' characters. Thus this use of the emphatic (polarity focus), antitopic construction could be quite conventionalized and motivated. One major problem with this explanation, however, is that the 'antitopic' in these sentences cannot be topicalized, as antitopics usually can.

But it could also be that the inverted subject is not an antitopic at all. It could be that the subject is a regular complement of a topicless assertion which would have been the focus if it hadn't been for the fact that the whole assertion is emphatic. As I argued in Chapter 3, when a whole assertion is emphatic (i.e. more salient than 'average' assertions), this added salience, which is given to a whole assertion, manifests itself in the

bearer of the polarity, that is, in the finite verb. In other words, it is not that the subject is not focal, it's just that the polarity is more focal and thus is 'chosen' as the focus of the assertion. I believe this to be the most likely explanation for this sentence type.

One thing that remains unexplained (unmotivated) is why an emphatic assertion would be used in this context. It is interesting in this regard to notice that similar sentences in Spanish display similar focus structure, with the focus accent falling on the finite verb. In this language the typical construction for initiating traditional tales also has an accented existential verb, an inverted formulaic temporal phrase, and a clause-final subject, as in example (5.79).

- (5.79) Erase una vez **un rey muy perezoso**.
it.was:REFLX one time one king very lazy
"Once upon a time there was a very lazy king."

Sentences like the one we've been discussing with a non-focus, unidentifiable subject are not only found in traditional tale beginnings, however. Occasionally we find such a sentence in the Spoken Basque Corpus, such as the one in example (5.80).

- (5.80) 93C3A09
26 Ta gero badator **mutiko bat**, bixikletan, -
and then/later EMPH:he.is.coming boy one, bicycle:DEF:LOC
"And then there's a boy coming, on his bicycle,"

I would also interpret this sentence as an emphaticthetic sentence with verb focus. Here too the subject could not be topicalized (at least in the dialect that the speaker of this sentence uses). I would classify this type of sentence as a very salient assertion, used in surprising, emphatic or contrastive contexts.

5.4.7 Post-presentational predicates

It has often been noted that presentational sentences are often followed by predications about that newly introduced referent (cf., e.g., Fox 1985, Lakoff 1987, Lambrecht 1988b, Givón 1990).²⁵ This phenomenon is not restricted to presentational sentences such as existentials in which the focus is the subject, but is found whenever a referent is presented in the focus role, whether identifiable or not.

I would like to argue that these post-presentational predicates (PPP's) are a type of sentence-tail constituent, much like secondary predicates, which predicate something of the main clause's focus referent. These optional elements are typically in their own intonation units and they have been described as 'extraposed' modifiers of the focus nominal in some theories of grammar.²⁶

PPP's may be (1) non-verbal, (2) verbal and non-finite, or (3) verbal and finite. The non-verbal and non-finite verbal predicates found here are indistinguishable from those predicates found in copula clauses and those used as secondary predicates in other constructions, as we can see in (5.81) and (5.82), respectively.

- (5.81) a. There is a man (,) in the backyard (,) reading. (cf. John is [in the backyard])
b. I saw a man (,) in the backyard (,) reading.

- (5.82) a. There is a book there (,) holding the door. (cf. The book is [holding the door])
b. I found a book there (,) holding the door.

In languages like Basque and Spanish we find that this predicate in some cases may switch places with the focus nominal whose referent 'governs' the predicate. This happens often with simple non-finite verbal predicates. Actually what seems to be going

on is that the main verb and the predicate verb attract each other. We can see an example from the spoken corpus in (5.83:8) below.

(5.83) 93C3A10

8 eta an ikusten-du ba **pasatzen** e zea bat. **kamioneta** bat.
and there he.sees.it well pass.by:IMPFV uh what:DEF one small.truck one
“and there he sees a pickup truck go by.” (cf. Spanish: “vio pasar una camioneta”)

The third type of PPP consist of finite clauses. The topic for this type of predicate clause is, like in the examples we have just seen, the main clause’s focus. In some languages, these finite clauses are indistinguishable from finite predicates found in declarative assertions, as shown by Lambrecht 1988b for a dialect of English, and exemplified by the sentence in (5.84), the title of Lambrecht’s paper.

(5.84) There was a farmer had a dog.

In most languages perhaps the form of these finite ‘amalgamated predicates’, as Lambrecht’s (1988b) calls them, are special constructions which may look like other constructions in the language but which have specific properties, deriving from their use as asserted predicates. In standard English these finite predicates have the form of relative clauses, as we can see in (5.85).

(5.85) There was a farmer (,) who had a dog.

However, unlike relative clauses, these clauses are unmistakably assertions (predications) and not modifiers. In other languages finite secondary predicates may yet take different shapes.²⁷

I would like to emphasize that these finite predicates are not different from other secondary predicates. What all these complex sentences share is that the focus of the first

clause is the topic of the second clause. Also, the focus of the first clause need not be a new, unidentifiable referent, as we can see in (5.86).

(5.86) Then I saw my mother (,) who was leaving the house right then.

In Basque, finite PPP's may take one of two possible shapes, depending primarily on the dialect. The first type looks just like a headless relative clause, which are nothing but relative clauses followed by definite nominal declensions. In (5.87) we see the schema for such sentences with definite absolutive coding (where n = relative marker, and a = absolutive, definite, singular).

(5.87) a. [... V]- n - a

The second possible shape that these finite amalgamated predicates may take is shared with two other Basque constructions: setting adverbial clauses with the ending *-la* (also *-larik*), and completive complement (*that*) clauses. This type of clause is schematized in (5.88).²⁸

(5.88) [... V]- la

Like all asserted predications, however, these are not strictly verb-final, as we will see. This is obviously true of the second type of finite PPP, which is obviously derived from a completive clause, and these clauses, just like main clauses, are never strictly verb final. As to the first type, which is obviously derived from a (parenthetical) headless relative clause, which *is* indeed typically strictly verb final, it may be verb final in writing, but in speech we find that complements are typically placed after the verb.

5.4.8 Basque -NA finite predicate clauses

One type of ‘post-presentative’ finite predicate clause in Basque has the shape of headless relative clauses, as we can see in example (5.89) below. As we can see this ‘headless relative clause’ is not verb final.

(5.89) 93C2A09

271 zegon ebai= e kamion bat,
it.was also uh truck one

272 e= ogiyak .. uzten ai zena momentu artan denda artan, -
uh loaves leave:IMPFV doing it.was:REL:DEF moment that:LOC store
that:LOC

“there was also a truck, / which was dropping off bread at that store at that moment.”

In this example the finite complement is not in the same intonation unit as the presentation, but it is in the same sentence, in a sentence-tail tail intonation unit. Sometimes, however, the two are more closely integrated, as in (5.90), where the two seem to be in a macro-intonation-unit.

(5.90) 93C1B03

17 Badago beste neska bat <pasatzen-dena bere aurretik>, -
EMPH:it.is another girl one she.passes.by:REL:DEF his front:ABL

“There is another girl, who goes by in front of him”

The degree of separation vs. integration is not always completely obvious, since the pause between the two units may be quite small. I have calculated, however, that in the Spoken Basque Corpus PPP’s which are in separate units are twice as numerous as those which are more integrated into the main clause: 33 vs. 16.

Notice that these clauses are functionally different from true headless relatives, since the latter select from a number of alternatives and are not asserted (cf. *the one who...*), i.e. are not predicational. It is obvious, however, that that is the source of this

type of PPP. One and the same sentence could conceivably be ambiguous between the two meanings.

5.4.9 Basque -LA finite predicate clauses

The second type of ‘post-presentational’ finite predicate clause is identical to finite completive clauses, which are also typically asserted, despite being formally dependent on another verb. This second type is twice as frequent as the former in the Spoken Basque Corpus. We can see an example in (5.91) below, with the second clause in a separate intonation unit. (Other codes = *Presentative_Predicate_FCl.*)

(5.91) 93C2A01

2 bueno ba ateatzen-da kale erdian - **kamioi bat**,
well well it.comes.out street middle:DEF:LOC truck one

3 daukala atzean egur bat,
it.has.it:COMPL behind:DEF:LOC wood/log one

“well then this truck comes out in the middle of the street, / which has a log in the back.”

In (5.92) we see another such sentence, this time with the finite predicate clause in the same intonation unit.²⁹

(5.92) 93C1B08

154 ateratzerakoan,
leave:NOMIN:ALL:GEN:DEF:LOC

155 jotzen-da **neska batekin** <nai-diola lapurtu>
he.hits(intransitive) girl one:COM she.wants.it.to.him:COMPL rob:PFV

“And when he’s about to leave, / he runs into a girl who wants to rob him”

The completive clause is obviously the historical source for this type of finite PPP clause, just as the relative clause is the historical source for its equivalent in English. (As I mentioned earlier, both these *-la* clauses are also similar in form to the adverbial setting (absolute) clauses in *-la(rik)*, from which the completive clause most likely derives.)³⁰

As to the question of how the finite PPP clause might have evolved from the completive clause, I would like to conjecture that this happened, probably not too long ago by reanalysis from constructions with verbs such as *ikusi* “see”, verbs which take both completive and nominal objects. In the Spoken Basque Corpus we can see some examples of such ‘missing link’ structures. In the example in (5.93:21), where the object is preverbal and the finite predicate is postverbal, it is clear that the *-la* clause is an added predicate about the focus referent (cf. *He sees a girl, who is coming*), and not a complement clause (cf. *He sees that a girl is coming*).

(5.93) 93C2A10

20 Ta biden dijola bizikletan, -

and road:DEF:LOC he.is.going:COMPL bicycle:DEF:LOC

21 ba beste neska bat ikuste-u <datorela baita bizikletan>, ..

well another girl one he.sees.her she.is.coming:COMPL also
bicycle:DEF:LOC

22 ta ari be- beira gelditze-a, -

and her:DAT loo- looking he.remains

“And while he’s going down the road on his bicycle, / he sees another girl also coming on a bicycle, / and he starts looking at her”

In the example in (5.94:4) below, however, where the object is postverbal, it is not clear what kind of structure we are dealing with: a completive clause, or a nominal object followed by a finite clause predicate.

(5.94) 93C3A11

3 =Pelikula da= nekazari bati buruzkoa. ez?

movie:DEF it.is farmer one:DAT about no?

4 Ordun ikusten-da nekazari bat ari-dala <madariak biltzen>

then he.is.seen farmer one he.is.busy:COMPL pears pick:IMPFV

“The movie is about a farmer, right? / Then you can see (1) a farmer who is picking pears / (2) that a farmer is picking pears”

Presumably the intonation and accentual properties of the sentence would allow us to distinguish the two possibilities, but in practice this isn’t always an easy task.

A different kind of structural ambiguity is found with the ‘auxiliary’ verb *ari* “be busy, engaged”. This verb is used as an auxiliary verb to form continuous imperfective forms of verbs which lack a synthetic conjugation. Much like the copula verb in Spanish or English, this verb takes an imperfective, non-finite clause as its complement. (The upper verb *ari* is always intransitive and does not code the arguments of the lower verb). However, the verb may also be used by itself and is sometimes used to introduce new referents. Here too we find sometimes a certain structural ambiguity which depends on whether the non-finite complement clause is interpreted as a non-finite PPP clause or as a complement of *ari*. We can see an example of this phenomenon in (5.95) below.

(5.95) 93C2A10

1 Ba= gizon bat ai da <udareak biltzen>,
well man one doing he.is pears pick:IMPFV

2 arbola baten eskillaratan iyota. ...
tree one:LOC ladder:LOC go.up:PFV:ADV

“Well a man is (busy) picking pears, / up on a ladder in a tree”

In other words, is the non-finite clause (*udareak biltzen* “picking pears”) a complement of *ari* or a post-presentative predicate? Sometimes intonation makes it clear which one was intended, but other times it is not so obvious.³¹

5.4.10 Finite predicate clauses without subordinators

Sometimes we find in the spoken corpus that regular, ‘main’ finite clauses (without any dependency markers) follow a presentative, much like in the sentences mentioned in Lambrecht 1988b. There are three examples of this type in the Spoken Basque Corpus, all of which come from the Hendaia corpus, where a northern dialect is

spoken (although most of the speakers in this group spoke a mixture of dialects). We can see two of these sentences here in (5.96) and (5.97).

(5.96) 93C1A01

- 11 Eta= gero badago= .. e= - badago= aur bat,
and later EMPH:it.is uh EMPH:it.is child one
12 pasatzen-da bere bizikletarekin,
he.passes.by his bicycle:DEF:COM

“And then there is a child, / he goes by on his bicycle”

(5.97) 93C1A04

- 1 Badago gizon bat,
EMPH:it.is man one
2 artzen-ditu= - udareak, ..
he.takes.them pears

“There is this guy, / he’s picking pears”

In the case of third example, the sentence in (5.98), the predicate clause is actually conjoined to the presentative clause which precedes it. Thus whereas the previous two clauses are similar to those that Lambrecht studied, this one is not.

(5.98) 93C1A02

- 20 orduan kartzelan, badago indartsu bat. -
then jail:DEF:LOC, EMPH:it.is strong one
21 Eta josten ari da.
and sew:IMPFV doing he.is

“then in the jail, there is a strong guy. / And he’s sewing.”

In the example in (5.99) below the predicate clause is even more independent, or less integrated with the presentative, since it displays an overt (demonstrative) pronoun referring to the focus of the presentative clause.

(5.99) 93C2B06

72 Txarloten ondoan, zegoen tipo txiki bat. -

Chaplin:GEN side:DEF:LOC he.was guy small one

73 Eta arek zeukan e= papelina bat. o zerbaite. ez? ..

and he:ERG he.had.it uh little.paper one. or something. no?

“Next to Chaplin there was a little guy. / And he had a little paper or something, right?”

This type of sentences reinforces my intuition that there is a much tighter relationship between a (‘presentative’) main clause and a finite PPP clause than between a main clause and a separate (conjoined) main clause..

5.4.11 More on the position of a presentational focus phrase

Inversion of subjects in presentative (focus subject) and presentative-like (antitopic subject) is quite common in spoken Basque, as we have seen. I have attributed this to the fact to the appropriation of the (identificational) extraposition focus construction and the antitopic construction for this function. I suggested that the ultimate motivation for this might be Romance influence. In this section I would like to suggest another possible, or perhaps additional, reason for why there might be a tendency for new referent (presentative) focus phrase to be in postverbal rather than in rheme-initial, preverbal position. The reason I would like to suggest is the attraction of a post-presentative predicate of the type we have seen.

As I have argued, post-presentative predicates are rather tightly integrated with the preceding clause. Given that such predicates are quite common in narrative, we may wonder whether there might be some iconic pressure for the focus phrase and the PPP to be in close proximity. This hypothesis is schematized in (5.100) below.

(5.100) [PRESENTATION ... NP_{NEW} ⇒ ...] ← [PREDICATION ...]

This iconic pressure or motivation might provide some (added) incentive for the speaker to look for (permissible) ways to postpone the focus phrase.

This possible iconic motivation for extraposing focus phrases with new referents receives some possible confirmation from an interesting observation made by Osa (1990) about word order in the writings of Basque 19th century author and translator Lardizabal. Although Lardizabal is known for a high degree of consistency in placing the verb in clause-final position (and thus the focus in preverbal position, cf. Chapter 2), Osa notices that one of the few places where he places complements after the verb is when they are followed by a “relative clause in apposition”, such as in the case we can see in (5.101) (Lardizabal 1855/1957:100, Osa 1990:224, ex. 239).

(5.101) *Testamentu Berriko Kondaira edo Historia*

a. ... Jesus-ek ipiñi zuen **beste parabola**,

Jesus:ERG place/put:PFV he.had.it (an)other parable

b. Eliz-Ama Santak ain sarrero Meza Santan irakurtzen digun hau.

Church-Mother holy:DEF:ERG so often mass holy:DEF:LOC read:IMPFV

it.has.it.to.us:REL this

“Jesus told them another parable, this one that the holy mother Church reads to us so often in the holy mass.”

The appositional phrase in this case is a parenthetical relative clause, headed by a demonstrative pronoun, and not strictly a predicate, but the similarity in function is clear.

Yet one argument against this hypothesis might be the fact noticed earlier that there seems to be a semantic ‘attraction’ between main verbs and PPP’s. As we saw above, this attraction has formal consequences in Basque and Spanish under certain circumstances, causing a main verb and a (non-finite) PPP to be placed next to each other.

Yet another possible motivating factor for using the focus extraposition construction with presentative foci in an OV language with a single major focus position might be to differentiate these presentative foci from other types of foci, such as contrastive ones. As I said, OV languages typically have only one major (non-extraposed) focus position, whereas VO languages use postverbal position for average salience foci and preverbal position for very salient foci. The exploitation of the secondary (extraposed) position to code this distinction might result from Romance influence in what might be a case of grammatical convergence.

5.4.12 Post-presentative predications and ‘extraposition’

As I said earlier, the finite version of the post-presentative predicate in English has the same form as a relative clause and has sometimes been analyzed as an actual relative clause which has been (synchronically) ‘extraposed’. The same thing is true for Spanish and many VO languages.

This type of derivational analysis is not possible for finite PPP clauses in Basque, and it doesn’t seem to me that it is the correct synchronic analysis for the English construction either. What may seem plausible from a purely structural point of view, becomes much less appealing when we realize how different the functions performed by the two types of clauses are. A post-presentative finite predicate clause does not function as a relative clause, i.e. a non-asserted clause used for identificational purposes, but rather as an asserted predication, as has been noted for instance by Ziv (cf. e.g. Ziv and Cole 1974, Ziv 1975).³² A synchronic parallel seems more plausible for finite PPP clauses and *parenthetical* relative clauses, which are indeed (parenthetical) assertions.

Having denied that these finite PPP clauses in English are really relative clauses out of place, I would like to suggest that perhaps a diachronic relationship *in the opposite direction* might be possible under some circumstances. I am referring to the possibility that (post-nominal) relative clauses might ever possibly be derived from finite PPP clauses by a process of reanalysis. While there are no real indications that this process is taking place in modern Basque, sometimes finite PPP clauses have an uncanny similarity to actual relative clauses, such as the one in (5.102) (cf. Other codes = *Pseudo_RelCl*).

(5.102) 93C1A05

24 ematen dute norbaitekin ebe <ez duela maite>, ..
give:IMPFV they.have.him someone:COM uh not he.has.it:COMPL love
“they put him with someone who doesn’t love him.”

In other words, I am suggesting that this type of PPP clause might be in some languages the source of postverbal relative clauses, whenever the ‘need’ for such clauses arises. Post-nominal, parenthetical headless relative clauses might be another possible source.

The phenomenon of ‘extraposition’, which is not restricted to ‘relative clauses’ but in general to phrases of different types, all of which are capable of functioning as predicates, has attracted much attention in recent years due to the unusual or ‘unexpected’ characteristics of such ‘extraposed’ elements, such as the impossibility of ‘extracting’ elements from them (cf. Ross 1967).³³ It seems to me that these formal accounts fail to understand this phenomenon because they fail to see the functional differences between true embedding inside the asserted clause proper, from which ‘extraction’ is possible under some limited circumstances, and elements which are outside the clause in setting or tail units which are peripheral to the main clause, from which extraction is never possible. It is fairly clear that extraction from setting (sentence-head) or post-assertion (sentence-tail) clauses is never possible in any language and that only under some limited

circumstances are such long-distance dependencies possible, such as when the ‘gap’ is inside a (non-extraposed) focus constituent.

5.4.13 Presentations in a picture descriptions

Alongside the main *Pear* and *Chaplin* story collection project, I conducted an additional study involving descriptions of still pictures. In this experiment, subjects were shown a number of pictures from a illustrated children’s books depicting a variety of situations. The subjects were asked to provide a brief description of the situation depicted in the pictures.

The descriptions often begin with the ‘presentation’ of an entity which is central to the situation (and thus topical) followed by a predication about that referent, as the one we see in (5.103).

(5.103) 93-DEU-01-A-03

18 Eta orain ba\
and now well

19 gizon bat dao,
man one he.is

20 bere laneko bazkaria txakurrari ematen.

his work:DEF:GEN lunch:DEF dog:DEF:DAT give:IMPFV

“And now uh, / there is a man, / giving his bag lunch to the/a dog.”

A variation on this pattern can be seen in (5.104), in which the location (a locative referent) is introduced and then something is said about that location, namely the objects it contains (also a presentational).

(5.104) 93-DEU-01-A-03

- 1 Bueno logela bat da,
well bedroom one it.is
- 2 eta bertan oe bat eta poto pila bat daude.
and same:LOC bed one and can pile one they.are

“Well it/there is a bedroom, / and (right) there there is a bed and a bunch of cans.”

In other words, the most common strategy employed to make a description, at least in this rather artificial context, is for an entity to be chosen in order to ground the situation on that entity. The entity then becomes the thematic (or circumstantial) ground for the predication that follows.

Often, however, the new referent, the topic-to-be, is presented without a verb, as in example (5.105).

(5.105) 93-DEU-01-A-05

- 17 Aurra batzu.
child several
- 18 txakur batekin jolasean.
dog one:COM play:DEF:LOC

“Some kids, / playing with a dog.”

This looks at first sight as a ‘dislocated’ topic, an unidentifiable entity of the type which we saw are relatively rare in this context. On the other hand, we could also see it as a presentation in which the presentative (copula) verb is omitted or elided. The fact that the subsequent predicate is not a regular sentence but a non-finite PPP argues for the latter interpretation, as does the intonational contour of the first unit. Notice that if the predicate was a finite clause it would have to have the subordinator associated with finite PPP’s (*-la* or *-na*).

The two parts of the sentence: the new topic and the predication may be ‘collapsed’ into a single bipartite intonation unit, as in the example in (5.106).

- (5.106) 93-DEU-01-A-05
 22 Neska bat pianoa jotzen
 girl one piano:DEF play:IMPFV
 “A girl (_s) playing the piano.”

It is obvious here that there are two sub-units in this intonation unit, as evidenced by the two accents.

The sentence in (5.107) is equivalent to the one in (5.106), except that here the PPP is a finite clause, with the *-la* ending.

- (5.107) 93-IKASBIDE-01-A-04
 Ba\ mutiko batek kendu-diola tapa zakarrontzi bati.
 well boy one:ERG take.away:PFV-he.has.it.to.it:COMPL cover:DEF trash-can
 one:DAT
 “Well, a boy who has taken the lid off a trash-can”; or, alternatively, “(I see) that a boy has taken the lid off a trash-can”

However, notice that this sentence could also be seen as a completive complement of an ‘understood’ evidential verb, such as *ikusi* “see”. As I mentioned above in Section 5.4.9, completive clauses such as this may have evolved into the finite PPP clause construction in (some dialects of) Basque.

Finally, I will mention yet another construction which speakers used in this study to describe the situations depicted in the pictures, namely a noun phrase modified by a (true and preverbal) relative clause, as in (5.108).

- (5.108) 93-DEUSTU-01-A-05
 31 Eta eskurik gabe txirrindan doan aur bat.
 and hand:PART without bicycle:LOC he.is.going:REL child one
 “And a child who is riding a bicycle with no hands.”

This is interesting, of course, because the secondary predicates found in languages like Spanish and English, and even in some dialects of Basque, have the form, though not the function, of relative clauses. This example could represent a calque from Spanish, but it

could also represent an alternative descriptive ‘strategy’, in which a description is made by modifying a generic entity rather than by predicating something of it. I believe this could provide some insight into the relationship between modificational and predicative relative clauses in some languages.

5.5 DESCRIPTIONS

5.5.1 Background descriptions

Thetic utterances are also used, according to Sasse, to provide background descriptions, for “describing background scenery ... made up of static or dynamic existential statements referring to weather, landscape, etc.” (Sasse 1995b:168) such as in the Greek example in (5.109) (Sasse 1995b:168, ex. 71).

- (5.109) *stin prokimea tu nisiu ixē singendrothi olo to plithos ton*
at.the pier of.the island had gathered all the crowd of.the summer-guests
“the entire crowd of the summer-guests had gathered at the pier of the island”

This sentence is reminiscent of the locative inversion clauses which we saw earlier in this chapter (also see below). Thus, as with other so-called thetic sentences in which the subject is not the topic, we could say that this is a *locationally grounded* statement, as opposed to a *thematically grounded* one.

According to Sasse, just like all other thetic clauses, clauses such as the one in (5.109) do not

continue statements about a topical subject commented on in the preceding sentences. Rather, they disrupt topic continuity and either denote sudden events or indicate a shift from the main story-line to a background description (scene-setting). In both of them, the entities constituting the

referents of the subjects of the construction usually carry low textual presuppositionality. (Sasse 1995a:15)

Descriptive ‘thetic’ clauses, he goes on to explain, “usually (but not always) involve existential verbs (often in a disguised form denoting a certain prototypical property of the subject such [as the] *shining* of the sun) and describe environmental conditions presented as a background to the main story line” (Sasse 1995a:15). The classical example of this type of thetic clause would be *The SUN was shining*, which in Spanish for example would have an inverted subject (unless the subject was the topic and the assertion was *about* the sun), cf. *Brillaba el SOL*. One may wonder, however, whether this clause is thetic because it is a background description or because it is a statements about a stereotypical atmospheric state-of-affairs and not about a particular referent? One clear difference between these sentences and the event-centered and referent-centered thetic assertions that we saw before is the fact that they typically display imperfective aspect, as we can see in the Spanish example, since they do not code events but states.³⁴

According to Sasse, these statements (like statements about interruptions and sudden events) can occur both “text- or paragraph-internally as well as text- or paragraph-initially” and “may occur in chains, i.e. there can be several interruptive events simultaneously or several states constituting a complex scene-description” (Sasse 1995a:15). In other words, they do not occur in isolated contexts, like announcements or explanations do.

Hungarian too supposedly makes extensive use of ‘thetic’ (subject inversion) assertions for this purpose of giving background descriptions, such as the one in rather poetic example in (5.110) (= Sasse 1995b:185, ex. 123).

(5.110) meztelenül meredeztek égnek a fák ágai
naked towered.above toward.the.sky the trees branches.their
“Naked towered the branches of the trees towards the sky”

According to Sasse, however, unlike all thethetic sentences which we have seen so far, which usually contained little more than the verb and the (focus) subject, these statements displaying subject inversion are somewhat different since they “are usually of a more elaborate type than the interruptive or explanative ones. They resemble the patterns found in the headlines in that they usually have focal or even topical material in front of the verb” (Sasse 1995b:184). I think these characteristics suggest that these Hungarian clauses are notthetic (subject-focus) assertions, but a, perhaps specialized, type of emphatic assertion with inverted, but non-focus subjects.

I believe that the Spanish analogue of the Hungarian sentence in (5.110) could be illustrative. Normally, there is no doubt that such a sentence would have a preverbal subject (cf. *Las ramas de los árboles se levantaban desnudas hacia el cielo*). However, an assertion with an inverted subject is also possible, as we can see in (5.111), although it is quite marked and sounds rather literary or poetic-like.

(5.111) Desnudas se levantaban hacia el cielo las ramas de los árboles.
naked REFLEX they.were.rising towards the sky the branches of the trees
“The branches of the trees towered towards the sky naked.”

In this example the predicate adjective *desnudas* is in preverbal focus position and it is obviously a specialized *emphatic assertion*, one that is conventionalized and associated with literary and rather poetic language. Notice that the subject’s referent would have to be accessible, as is always the case with non-focus inverted subjects. (In a rather different context the preposed adjectival phrase could indeed be a setting, in which case the subject would be an extraposed focus. Such a sentence would only make sense as the

answer to a question such as *What towered towards the sky naked?* Unfortunately, Sasse does not provide information about the intonational and accentual properties of these sentences or the contexts in which they were found.)

Sasse also notes that there is a “particularly frequent pattern” found in the (written) Hungarian data of descriptive sentences with subject inversion, namely the one schematized here in (5.112).

(5.112) Locative Manner **Verb** Subject,

This is reminiscent of the presentational sentences that we saw earlier which had a locative ground and a focus subject. On the other hand, it seems quite clear that in these sentences the Manner phrase is indeed the focus of the assertion and that the subject is neither the topic nor the focus but an additional complement, which is why it is postverbal.

The fact that these clauses don't have topics would be in line with the purported function of this clause type, namely to establish some background for the discourse, as opposed to predicating something about an entity. The sentence in (5.113) is Sasse's example for this particular pattern (Sasse 1995b:185, ex. 126).

(5.113) a belső csarnokban nyugtalanul fészkelődtek a várakozók.
the inner hall.in ancily [*sic*] shuffled the waiting:PL
“The people waiting scooted back and forth ancily.”

It seems clear that the manner adverb in this type of sentence is an assertion initial focus, which would mean that the subject cannot be the focus. Nor can it be the topic since these clauses lack a thematic ground (topic) and have instead a locative ground.

Sasse mentions, in passing, that sometimes subject inversion seems to be triggered by the presence of setting adverbials. He argues that, although the data is inconclusive, there is some suggestive evidence that some inversions are triggered by fronted setting material (Sasse 1995a:12-13).³⁵ This observation lends support to my hypothesis that these clauses are grounded by a locative or other setting adverbial, as opposed to by the referent of one of the arguments of the verb. It isn't that subject inversion is *triggered* by the setting material, for there is nothing preventing the coexistence of a thematic ground and a (dislocated) setting locative. However, in these sentences the setting locative is really a locative ground, one that is much more closely integrated with the assertion (pragmatically and intonationally). And, because the ground for this type of sentence is locative, there is no thematic ground. This type of sentence is different from the presentational sentences with circumstantial grounds that we saw earlier in that here the subject is not the focus of the assertion.

5.5.2 Thetic sentences and the setting function

The function of providing background descriptions is often associated with non-asserted, setting clauses, rather than with asserted ones. As I mentioned in Chapter 3, (dependent) setting clauses in languages may be topic-linked with the main, asserted clause, but this is not a requirement, nor does it seem to be a strong preference. Setting clauses very often display subject inversion in Spanish, although often it seems like an optional alternative in a way that is very different from the alternation as found in asserted clauses.³⁶

In addition to setting clauses which are clearly marked as such by subordinators and other coding devices, it seems that sometimes setting like functions are performed by

unmarked 'main'-like clauses. In (5.114) below we can see an example of what I have in mind, which is a reconstructed excerpt overheard on Maine Public Radio on October 16, 1996.

- (5.114) a. When you look at the polls ...
b. I was looking at the polls yesterday,
c. and frankly, (it looks like) you don't have a chance.

The clause in (5.114a) is clearly a (sentence-head) setting clause with a temporal (or conditional) function. The clause in (5.114b), on the other hand, although it is basically a rewording or paraphrase of the previous clause, looks like a main clause. Functionally, however, we could perhaps say that this clause has a background or setting-like function with respect to the actual assertion, i.e. the clause in (5.114c).

We might say that these are 'dependent' clauses in disguise. In fact, there is a type of construction in English (and in Spanish) in which a setting clause and an asserted clause are involved, but the formal dependency markers seem to be reversed. We can see an example of this in (5.115)

- (5.115) I was peacefully sitting there, when this man came and sat next to me.

Here the setting/background sentence is coded as a main clause, whereas the truly asserted clause has a temporal subordinator, which is somewhat equivalent to a coordinating conjunction, cf. *and then*.

It turns out that clauses such as this one in (5.114b), which look like a main clause, but which actually function as setting clauses providing background information for a following assertion, often display subject inversion in Spanish, such as we can see

in example (5.116) (the inverted subject *yo* could go either after the verb, after the adverb, or after the direct object).

(5.116) Estaba mirando *yo* ayer las encuestas, y es que no tienes ninguna posibilidad.
was looking I yesterday the polls, and it.is that not you.have any chance
“I was looking at the polls yesterday, and you really don’t have a chance.”

Now the question is, what is the motivation for subject inversion in this case, which seems highly preferable to non-inversion (even if the topic of the main clause was the same as that of the ‘background’ clause). The subject is certainly not the focus, since it must be unaccented. On the other hand, it doesn’t seem to be the topic either, since this doesn’t seem to be a statement about the subject’s referent.

In fact this doesn’t seem to be a statement (an asserted clause) at all, but rather a setting or background sentence for the actual assertion in the following clause. The inversion of the (highly topical) subject in such a ‘setting’-like main clause might actually be *a way of coding the fact that this is not an asserted clause but a setting clause*. Notice that the subject pronoun in this example is superfluous for referential purposes, since the verb already codes that the subject is the speaker. But the overt subject is obviously not contrastive either, as overt first person pronoun subjects often are.

Even if we accept that this clause, despite looking like an asserted clause, acts as a non-asserted setting clause, one may wonder as to *why* it is coded this way (with subject inversion), in other words we may wonder as to the origin of the construction. It may be that this was originally a true antitopic (emphatic) construction, since it displays the same unaccented inverted subject (with a topical referent). On the other hand, subject inversion in this case may simply be related to the inversion that is commonly found in

setting clauses, almost as to indicate that they are not assertions *about* the subject. Notice the similarity between these setting clauses and imperfective ‘absolute’ sentences in Spanish, which also display (obligatory) subject inversion, e.g. (5.117).

- (5.117) Estando mirando yo ayer las encuestas, ...
being looking I yesterday the polls
“While I was looking at the polls yesterday, ...”

In this construction, the inverted subject *yo* could go after any of the other phrases, including the copula *estando* “being”, but never before the verb. Although the meaning of this (rather literary) construction is quite different from the other (much more common) one, notice that subject inversion seems to be one of the main ways of coding this construction, given that it doesn’t contain any subordinator. This supports the conjecture made earlier that inversion in the finite setting clause in (5.116) is what actually codes this sentence as a setting sentence and thus it is neither a sign of it being athetic or an emphatic assertion.

5.5.3 Inversion in background Basque clauses

Basque setting constructions are, by and large, strictly verb-final. We can attribute this to the fact that since these clauses are not asserted, word order is not used to code pragmatic roles. That is, setting clauses do not have topics of foci, though they may certain have contrastive and linking elements, as we saw in Chapter 3. Thus ordering alternations are not used to code pragmatic functions in these clauses since there are no pragmatic functions (roles) to code. (An exception to the verb-final pattern is, for example, the *nola* “how; since” construction, which seems to be borrowed from Spanish, as a substitute for strictly verb-final *-la(rik)* clauses.)

In this context, it is interesting to notice that there are in the Basque corpus setting-like main clauses, which just like the Spanish sentence in example (5.116) above, display subject inversion. We can see an example of this in (5.118) below, taken from the spoken corpus. The clause in (5.118:36) is a regular main clause with a covert topic, but the following clause in (5.118:37) has an overt postverbal subject with a topical referent (*denak* “all”) with the same referent as the topic of the preceding clause.

(5.118) 93C2A01

35 Eta=

And

36 bueno eseritzen dira maian,

well sit:IMPFV they.are table:DEF:LOC

37 eta daude **denak** or maian ya jateko prest,

and they.are **all:PL** there table:DEF:LOC already eat:NOMIN:GEN ready

38 eta dago makurtuta.

and he.is bend:PFV:ADV

39 Eta= emat-

and giv-

40 bueno botatzen diote bazkaria- bazkaria platerera,

well dump:IMPFV they.have.it.to.him lunch:DEF lunch:DEF plate:DEF:ALL

41 eta bea ez-da konturatzen. ez?

and he not-he.is realize:IMPFV no?

“And .. / well they sit at the table, / and they’re all sitting at the table all ready to eat, / and he’s bending down. / And - / well they put some food on his plate, / but he doesn’t realize, right?”

The clause in (5.118:37) seems to be a background statement, much like in the Spanish case we saw earlier. Like in the Spanish sentence, the subject here is inverted (notice that in English the quantifier *all* could also be inverted, or ‘floated’, to a position behind the verb, cf. *They were all sitting at the table ready to eat, and ...*). The next clause in (5.118:38) also seems to be a background clause, though it has a covert topic, the story’s protagonist (*Chaplin*). The ‘real’ assertions that these setting-like main clauses seem to be building up to are, of course, those in (5.118:40&41).

One very interesting thing about the clause in (5.118:37) is that it has a synthetic verb, and no preverbal focus and still it doesn't bear the polarity particle *ba*. As we saw already seen this type of behavior is only expected in non-asserted Basque clauses, such as relative clauses and setting adverbial clauses. As we will see below, lack of a *ba* particle on synthetic verbs is generalized in this type of paragraph initial clause. Of course, this could also perhaps indicate that these are verb-focus assertion. After all, the verb does receive the main accent of these clauses. However, such assertions with synthetic verb focus but without the particle *ba* are not typically found elsewhere in Basque.

The 'main' clause with subject inversion in (5.119:32) below also seems to be associated to this same setting-like function, cf. *The truck was (already) leaving, and the flag falls off*. The referent of the inverted subject here too is active. In Spanish too this context would also typically call for subject inversion, cf. *(Ya) se iba el camión ese, y se le cae la bandera* "the truck was (already) leaving and its flag falls off". (Notice in both English and Spanish the optional presence of the two equivalent temporal adverbial *already* and *ya*, which may help code this function in these languages.)

(5.119) 93C2A09

31 Eta zij- e=
and go- uh=

32 zijuan kamiona- - kamion oi,
it.was.going truck:DEF - truck that

33 eta erortzen-zaio bandera.
and fall:IMPFV-it.is.to.it flag:DEF

"And it wa- / that truck was leaving, / and (cf. when) the flag fell down."

Notice here too the lack of the *BA* particle on the bare, clause-initial synthetic verb. It is possible that this type of inversion, unlike the other types which we have seen, could be

attributable to Spanish influence. In other words, it could be that this setting-like main clause might have been borrowed. But inversion here could also be seen as being internally motivated by the fact that the referents of these inverted subjects, like the referents of inverted subjects in emphatic assertions, are not ‘true topics’, or at least are less topic-like than they would be in regular assertions.

The written corpus does not contain any examples of inversions of this type. One of the novels from which part of the written corpus was built does contain one, and only one, such clause, namely the one in (5.120a) below.

(5.120) *Behi euskaldun baten memoriak*, p. 171-2

a. Egun batez, ari-zen **Haundia Xarmegarria zure berririk** kantatzen,
 day one:INSTR busy he.was big:DEF XZB (song title) sing:IMPFV

b. eta hosto bat erori-zitzaion balkoiaren ondoko intxaurrari.
 and leaf one fall:PFV-it.had.it.to.it balcony:DEF:GEN side:DEF:GEN
 walnut.tree:DEF:DAT

c. Kolpetik ixildu-zen.
 blow:DEF:ABL shut.up:PFV-he.was

“One day, Big Guy was singing *Xarmegarria zure berririk*, / and (when) a leaf fell off from the walnut tree next to the balcony. / He immediately stopped singing.”

Notice here too the this clause performs the same type of setting function, the kind that may be followed by an asserted *when* clause in English and Spanish, cf. *He was singing that song, when a leaf fell from the tree.*

5.6 EPISODE OPENINGS

5.6.1 Episode opening sentences

Sasse argues that there are two functions of ‘inverted clauses’ which are unlike all other functions in that the subject is ‘given’ (i.e. *discourse* old).³⁷ One is what he calls the ‘discontinuative’ or ‘episode opening’ function and the other is the ‘reactive/consequential’ function (see below).

I do not believe that we are dealing with two distinct functions here. New paragraphs (‘episodes’) of the type that Sasse describes often express a series of actions which constitute a reaction to some prior event. In addition, some of Sasse’s examples of ‘reactive’ clauses seem to be exclamations or, in other words, emphatic assertions (e.g. Sasse 1995b:169, ex. 74). Thus I do not believe there is really such a thing as a reactive function for clauses with subject inversion, although there does seem to be an episode opening function (which includes ‘reactive episode’s). I will discuss these sentences in this section. (Emphatic assertions will be discussed in greater detail in Chapter 7.) All these clauses, I will argue, are more like antitopic assertion constructions, or perhaps even topicless constructions of a different (setting-like) sort, and not thetic (subject focus) assertions at all.

What Sasse calls discontinuative clauses (assertions) have a general “episode opening function”, that is, they introduce a new episode, a new paragraph, in a narrative, and thus establish a new topic chain. He even argues that inversion is what signals topic discontinuity when he says that “it appears to be a function of VS clauses in some of our languages to *signalize* this type of topic discontinuity” (Sasse 1995a:16; my italics).

(Compare this with the claim I mentioned in earlier chapters that in verb-initial languages topic-fronting is what codes discontinuity in episode opening clauses.)

In some ways, the ‘discontinuative clauses’ which Sasse mentions seem similar to the setting-like main clauses which I discussed in the previous section. In addition, Sasse’s description of the referents of inverted subjects in these constructions is indeed very reminiscent of the subjects in the setting main clauses. He notes that the ‘given’ subject is “usually a main character of the story, even the speech participants themselves in the form of explicit personal pronouns,” although “[t]hese pronouns, however, are not contrastive in the usual sense and do not normally bear strong accent” (Sasse 1995a:16). In the excerpt in (5.121), from a written Greek source, we can see one of Sasse’s examples of the discontinuative function (from Sasse 1995b:168)..

- (5.121) (*Background*: The old couple had a quarrel in the evening. Then they went to bed.)
telos pandon sikothike tin alli mera to proi **o anthropos** ke pije ...
end of.all got.up the other day the morning **the man** and went to
‘anyway, **the man** got up the next morning and went to ...’

Notice that the man’s getting up seems to be a preparatory or *preliminary action* for the one that follows, and thus more akin in function to a setting clause (a background assertion) than to a main asserted clause.

On the other hand, these examples are also unlike the ones in the previous section. These clauses indeed seem to be asserted and cannot easily be replaced by setting clauses. In addition, they seem to be emphatic in nature. It is not easy to tell, however, from Sasse’s few examples, which rather devoid of context and intonational information.

Sasse claims that this use of subject inversion has been proposed for languages such as Russian, Modern Greek, and Romani. He claims that its existence has been

denied for Italian, Spanish, and Hungarian, although it “has been reported for Latin” (Sasse 1995a:16).³⁸ This statement would seem to me to be inaccurate as far as Spanish is concerned, which, as we will see, it does make use of inversion in clauses which initiate an episode and which typically describe a reactive situation. We should note, however, that in Spanish the use of inversion is not obligatory in these contexts. Rather, it is present only when the assertion is seemingly emphatic, something which is associated with more vivid styles of narrative. They seem to be examples of very focal verb-focus with a concomitant subject (antitopic) inversion. (In other words, that is its motivation, since the use of this clause type for this function is somewhat conventionalized and thus not strictly predictable.)

Sasse presents some further evidence that these are not truethetic (subject focus) assertions when he notes that in the case of these episode opening utterances “there are no preferences whatsoever with regard to the semantic type, the aktionsart or the argument structure of their verbs. The verbs occurring in these clauses are not restricted to existentials even in the broadest sense, and transitive action verbs prevail” (Sasse 1995b:169). This is because in this construction the subject’s referent is not the focus of the assertion, as is the case in truethetic assertions, but rather an antitopic or a non-topic, non-focus rhematic element.

5.6.2 Spanish examples and conventionalized formulae

As I said in the previous section, there is a particular type of inverted clause in Spanish which seems to fit the description of (at least some of) Sasse’s episode opening clauses, clauses which also very often, though not always, express a *reaction* to some previously expressed state of affairs, or at least express some kind of *unexpected* and

surprising action. The excerpt in (5.122b) contains a made-up example of a discontinuative sentence with VS order from Spanish. Actually, the clause in (5.122a) also displays subject inversion, but this is clearly a descriptive setting-like main clause of the type I described in the previous section.

- (5.122) a. Estaba hablando yo con Juan tranquilamente,
I.was speaking I with Juan calmly
b. y entonces saca Juan la carta del bolsillo,
and then he.takes.out Juan the letter from.the pocket
c. y me la enseña.
and to.me it he.shows
“(a) I was just talking with Juan, / (b) and he takes the letter out of his pocket, / (c) and he shows it to me.”

Notice the use of the ‘historical present’ in this sentence in (5.122b) which, although it is not required, is very common in this type of sentence. (The historical present adds *vividness* to the narrative.) Notice that the inverted subject’s referent is also very accessible. It seems to me that this is an emphatic verb-focus assertion depicting a surprising and unexpected event. This new, surprising event breaks up the continuity of the discourse and often initiates a new chain of events.

Inversion in this type of context is not obligatory in any sense. Very often what we find in this context is that the subject is actually dislocated, yet another strategy often used in paragraph beginnings, cf. (5.123).

- (5.123) ... y entonces Juan, saca la carta del bolsillo, y ...
... and then Juan, he.takes.out the letter from.the pocket and
“... and then Juan, he takes the letter from his pocket, and ...”

This sentence too, by dislocating the subject, indicates discontinuity. It doesn’t seem that the effect produced by dislocation in this sentence in (5.123) is very different from the

one produced by inversion in (5.122b). Emphasis and surprise can also be expressed by means of the sentence with the dislocated subject by intonational cues, for instance.

This way of beginning (typically reactive) paragraphs has been ‘institutionalized’, or grammaticalized, in Spanish into a somewhat idiomatic constructions involving either the verb *coger* “take, grab” or the verb *ir* “go”. They may be used literally, but often they are used ‘metaphorically’. In other words nothing is necessarily *taken*, and no movement does necessarily take place. In (5.124) and (5.125) below we can see two examples of this very common construction (at least in northern peninsular Spanish).

(5.124) (Entonces) coge Juan, se levanta, va hacia la ventana, y ...
then he.takes Juan REFLX he.gets.up, he.goes towards the window, and
“Then Juan goes (lit. takes), gets up, goes towards the window and ...”

(5.125) (Entonces) va Juan y pega un grito, que se oye hasta en la calle.
then he.goes Juan and he.hits a scream that REFLX it.hears even in the street
“Then Juan goes and screams so loud that it could be heard from the street.”

In both examples the verb involved in the first clause does not have its literal meaning, but a metaphorical, extended one (although this type of sentence can be used with these verbs with the literal sense as well). The crucial point here is that these clauses begin a series of surprising or unexpected events. As we will see in the next section this construction is also found in spoken Basque, including some of the same idiomatic versions.

5.6.3 Surprising episode openings in Basque

Because of the relative productivity of the antitopic (emphatic assertion) construction in Basque, and given my interpretation of Sasse’s discontinuative/reactive function as an example of antitopic construction (or a topicless construction with a non-

focus subject), we might have expected this rhetorical, surprising episode beginning function to be one of those that this construction is used for in this language. And indeed, we do find a not insignificant number of examples in the spoken corpus of such paragraph initial clauses. There are 35 fairly clear-cut examples of this construction in the Spoken Basque Corpus with the verbs *joan* “go”, *hartu* “take”, or *hasi* “begin”, and most of them with inverted subjects (25/35) though some are bare verbs (6/35): 2 AV, 2 SV, 6 V, 7 VA, 2 VAO, 2 VO, 12 VS, 1 VSXtzen, 1 VXS (Other codes = *Hartu-Joan_Cn*). We can see an example with *joan* “go” in (5.126) below (a discontinuative but not reactive example). It is not clear here whether the sense of the verb *joan* is literal or metaphorical.

- (5.126) 93C1B08
 104 Ta dijoa **polizia**,
 and he.is.going **policeman:DEF**
 105 ta artzen-diote gizon orri,
 and they.take.it.to.him man that:DAT
 “And then the police goes, / and they take that man,”

Notice that here too the synthetic verb does not bear the *ba* particle, thus showing that this is not an assertion with polarity focus. This indicates that perhaps the sentence is seen as an assertion with verb-focus, though this would be rather unusual with a synthetic verb, or else it is a non-asserted, setting-like clause. Note that these are rather different from emphatic asserted clauses with inverted subjects, such as the one in (4.127:32) below. Truly asserted clauses with synthetic verbs always bear the polarity marker.

(4.127) 93C2B03

30 artzen-dau otzara oso bat,
take:IMPFV-he.has.it basket whole one

31 <@ bisikletan gaien jarri, @>
bicycle:DEF:GEN top:DEF:LOC put:PFV

32 ta badoia mutiku oi. ..
and EMPH:he.is.going boy that

“he takes a whole basket, / he puts it on top of the bicycle, / and (then) that boy leaves.”

In (5.128:204) we can see another example of this construction, this time with *hartu* “take”. This is an idiomatic example with a non-literal sense of the verb *hartu*.

(5.128) 93C2B06

203 eta gizona - dendara sartzen-dan bezela, -
and man:DEF store:DEF:ALL he.enters:REL as

204 ba al- artzen-du **neskak**,
well ?- she.takes.it **girl:DEF:ERG**

205 eta- ogi bat lapurtzen-du,
and bread one she.steals.it

206 ta korrika alde- al- aldegiten-du. -
and run:ADV leave- lea- he.takes.leave

“and since the man goes inside the store, / well the girl goes, / and she steals a loaf of bread, / and she takes off running.”

As I said earlier, although the use of the antitopic construction for this function is motivated, there doesn't seem to be anything obligatory or necessary about its use to express this ‘rhetorical’ function. It is also clear that some speakers are much more likely to use this particular rhetorical device than others. Thus some speakers use it several times whereas most don't use it at all. For this and other reasons I am inclined to believe that this construction is probably a convergence type calque from Spanish, although to some extent it is motivated by similarities with native Basque constructions.

We have already run into this type of indeterminacy when we discussed emphatic assertion inversion in journalistic styles of writing, and headlines in particular. It seems

that the antitopic construction, which is primarily associated with emphatic assertions and very focal foci, can be put to other, somewhat related uses in the language which are not strictly predictable but which are nonetheless motivated.

I should note that Basque has another ‘native’ construction which is used in very much the same context to express a preliminary and sudden event such as in a topic chain. This construction involves a non-finite clause with the verb in the perfective participle form. We can see an example of this type of sentence in (4.127:31). This construction, as all non-finite clauses more generally is strictly verb final. We can see an example with *hartu* “take” in (4.129:146) below.

(4.129) 93C3A12

145 Ordun/ ba\ polizia ba ikustean ori,
then well policeman:DEF well see:NOMIN:DEF:LOC that

146 ba poliziak artu,
well policeman:DEF:ERG take:PFV

147 =ta= ..
and

148 atxilotzen-du,
arrest:IMPFV-he.has.him

149 eta eramaten-du. -
and take.away:IMPFV-he.has.him

“Then well when the policeman sees that, / well *the policeman goes/takes (him)*, / and .. / he arrests him, / and he takes him away.”

This construction however, was much less frequent in the spoken corpus than the finite version, with or without inversion, and most speakers never use it at all.

Besides the verbs *joan* “go” and *hartu* “take”, one other verb which is found relatively often in episode beginnings with subject inversion is *hasi* “begin”, which is often used as an aspectual verb. This is not surprising since this verb indicates the beginning of an event or an action and can give the impression of suddenness, a trait

which is often associated with emphatic statements, and also with episode openings of the type we have seen above.

In the Spoken Basque Corpus there are 206 clauses with the verb *hasi* “begin”, a very common verb, as we can see.³⁹ Of these, 88 had a non-finite imperfective complement clause (in *-tzen*), typically following the verb *hasi*. Among these we find 8 cases in which there is an overt, inverted subject, a significant number, relatively speaking. For five of these the verb has a sentential complement. In the vast majority of these cases, in particular the ones with a sentential complement, we find that the event or action that is beginning is followed by another event which contrasts with the first one, interrupts it, or is somehow a reaction to it. In other words, the beginning event is in some way a setting for the ‘big’ event that follows. In (5.130), below we can see one of these examples in which the incipient event is interrupted.

(5.130) 93C2A01 (Ikasbide)

31 Eta= jo- asi bear zen **gizon aundia** jotzen,
and hi- begin need he.was man big:DEF hit:IMPFV

32 eta= deitzen-dute bazkaltzeko ordua dela.
and call:IMPFV-they.have.it eat.lunch:NOMIN:GEN hour:DEF it.is:COMPL

“And the big guy was going to begin hitting (him), / when they call that it’s time to go to lunch.”

In (5.131) we see a second example. Here the event that has just begun (*everybody begins to leave*) contrasts with another event (*Chaplin turning around in circles*).

(5.131) 93C1B08

119 Ta asten-dira,
and begin:IMPFV-they.are
120 asten-dira **denak** ibiltzen,
begin:IMPFV-they.are all:PL go:IMPFV
121 ta bera bueltak ematen.
and he turns give:IMPFV

“And they start, / they all start to leave, / and he was going around in circles.”

Finally, in (5.132) we see another typical example. Here too the sentence expressing an action which is about to begin is followed one expressing a contrastive, and perhaps also interruptive, reaction on the part of Chaplin.

(5.132) 93C2B06

135 Derrepente Txarlotekin egiten-dute topo. ez? -
suddenly Chaplin:COM do:IMPFV-they.have.it encounter no?

136 Eta asten z- zaio **tipoa**= - zerbait esaten,
and begin:IMPFV he.is.to.him guy:DEF something say:IMPFV

137 Txarlotek kontestatu egiten-dio,
Chaplin:ERG answer:PFV do:IMPFV-he.has.it.to.him

138 eta- asi- asi-zaio tiroka. -
and begin:PFV begin:PFV-he.is.to.him shoot:ADV

“Suddenly they run into Chaplin, right? / And the guy begins to say something to him, / Chaplin answers back, / and he begins shooting at him”

In all these examples with *hasi* we can see similarities with both the setting-like main clauses we saw earlier and the episode beginning clauses in we have seen in this section.

5.7 CONCLUSIONS

In this chapter I have explored in some depth the claim that some asserted clauses do not have topics and how this affects the form of such assertions in different languages. The major structural phenomenon associated with these sentences in many languages is subject inversion.

As we have seen, not all clauses which display subject inversion in different languages are of the same type. The fact that only a subset of such clauses in languages such as Spanish do so too in languages like Basque shows that there is more than one type of inversion involved.

In one case, the subject is the focus of the assertion. The assertion then may or may not have another topic, chosen from among the other arguments of the verb (or an added benefactive one). This is the reason why these assertions have the same formal properties as assertions in which the subject is a contrastive focus or an 'answer' focus. In other words, in all these cases the subject's referent fills the focus role, even though their function is somewhat different.

In languages in which foci are normally preverbal, such as Basque, that is where subjects go in subject-focus assertions as well. In languages in which foci are normally postverbal, such as Spanish, that is also where subjects are found in subject-focus assertions. In languages such as English in which focus subjects (but typically not other foci) are placed in rheme-initial position, that is also where such subjects are found as well. Other languages, on the other hand, such as spoken French, may have more specialized means of coding these focus subjects, such as cleft (focus) constructions.

Thus in many (non-rigid VO) languages, but not in OV-type languages, subject inversion is a natural result of the subject being the focus of the assertion. But in those (and other) languages subject inversion may also have a different motivation. Subject inversion is also found in cases in which the subject is somehow an unusual type of topic in emphatic assertions with a very salient focus element, or assertions which are overall very salient (such as exclamations).

The motivation for subject inversion in these cases is presumably related to the urgency of coding the focus constituent early in the clause, even in clause-initial position. Inversion in emphatic assertions is rarely required, however, and a similar effect can be achieved sometimes by left-dislocating a topic-subject. It does seem, however, that these

subject inversion, or antitopic, constructions may acquire other uses as well, such as indicating that the clause is not a predication in the same sense that normal (declarative) assertions are typically predications. Main-like clauses which function more like paragraph initial settings than like true assertions may adopt this form in some languages, for instance. This would be a motivated, but in no way predictable, extension of the main use of the antitopic construction.

Even in languages such as Spanish in which both antitopic assertions and subject-focus assertions result in subject inversion, the characteristics of these inverted subjects are very different from each other. In the case of subject-focus assertions, the subject, as the focus, must be accented and must be in focus position. Their referents are typically quite inaccessible and the assertion cannot be understood as a predication or 'comment' about them. Also, because they fill the focus role, the presence of other rhematic complements is highly curtailed.

In the case of emphatic assertions, on the other hand, the subjects must be unaccented, just like all other rhematic elements which are not focal, and their referents must be very accessible or easy to accommodate from the context, as is the case with all non-topic, non-focus rhematic complements in any assertion.

As to the pragmatic role of antitopics, it seems that the postposed topics of the most basic emphatic assertions can typically be seen as the referent that the predication is *about*, i.e. as topics. However, given the nature and functions of these emphatic assertions, the notion of predication is much less central to them than it is to regular (declarative) assertions, and it seems that a holistic (topicless, non-predicational or ungrounded) interpretation of the state affairs is favored in many cases.

A third source for postverbal subjects seems to come from emphatic assertions in which the subject is not the topic to begin with and in fact is a strong focus candidate. When the subject is focal and is not the topic, in normal declarative assertions we expect the subject to be the focus (as we just saw). But when a whole assertion, and not just the focus, is very salient and the verb or the polarity may 'outrank' the subject in focality and may be made the focus. In an OV language with consistent rheme-initial focus position this means that the finite verb must be in that position and the subject must be postposed.

The consequences of these phenomena for Basque word order are several. In subject assertions we expect subjects to come in preverbal, focus position and to be accented, just as any other focus constituent. In emphatic assertions, on the other hand we expect the subject (when not left-dislocated) to be in postverbal position and to be unaccented. This is indeed what we find most of the time, particularly in standard, written Basque. In some varieties of Basque, however, we find complications which muddle the picture somewhat.

To begin with we find that under some circumstances, which are particularly common in speech, a focus may be extraposed to clause-final or post-clausal position. Since most clauses have only at most one complement (the focus), in most assertions with focus extraposition the focus comes actually right 'after' the verb, which makes it seem that it is postverbal much like in Spanish and VO languages in general. In Basque, however, we do find an intonational break between the verb and the focus, which is not found between a verb and a postverbal (non-extraposed) focus in VO languages (although focus extraposition is also found in those languages).

It is conceivable that extraposed foci might come to be reanalyzed as postverbal foci, especially given their relative frequency in spoken Basque. This is most likely to occur perhaps among non-fluent speakers who are more fluent in a VO language. Basque dominant speakers may also eventually be affected by this misanalysis.

A second complication comes from assertions which look like emphatic assertions with polarity focus and a postverbal subject, but which have functions typically associated withthetic assertions, and in particular with presentative assertions. Of course, although these subjects look like antitopics, they may also be non-topics: focal complements which invert with the verb because the verb is more focal in these emphatic assertions. However, the increased use of this emphatic construction for the purpose of introducing referents into the discourse might be motivated by the example of neighboring languages which introduce such referents in postverbal (focus) position. In other words, this would be another example of convergence.

NOTES TO CHAPTER 5

1. Cf. also Sasse 1995a,b, Lambrecht 1986, 1994:138ff. As we saw in Chapter 3, the name *thetic* and the distinction between *thetic* and ‘*categorical*’ (topic comment) structure, goes back to Brentano, and his student Marty, who “made a distinction between two types of human JUDGMENT. ... Brentano and Marty claimed that sentences can express two distinct types of judgment. The CATEGORICAL judgment, which is expressed in the traditional SUBJECT-PREDICATE sentence type ... In contrast, the THETIC judgment involves only the recognition or rejection of some judgment material, without predicating this judgment of some independently recognized subject” (Lambrecht 1994:139).
2. That, of course, requires a belief in synchronic transmutation which, as I explained in Chapter 1, I do not subscribe to. The ‘unaccusative hypothesis’, as this explanation is called, dates back to Perlmutter 1978. For a look at unaccusative verbs in Basque, see Levin 1983, 1989. Cf. also, e.g., Van Valin 1987 and Bresnan 1994.
3. Turkish uses VX inversion “for quite different purposes than in all the other languages ... and seems to have nothing to do with “theticity” as has been traditionally understood. Inverted elements are typically topical” (Sasse 1995a:9). It would seem that what unites most of the *thetic*-inverting languages, as opposed to Turkish, is that, with the possible exception of Hungarian, they are VO languages. But Turkish is a much more rigidly OV language than Hungarian is. As we will see later, Basque is a lot more like Hungarian than like Turkish.
4. Sasse argues that “All studies agree that VS order in the languages examined cannot be explained in terms of one single straightforward criterion, but, rather, is sensitive to a number of heterogeneous syntactic, semantic and discourse-pragmatic factors. Unless strictly grammaticalized, the impact of these factors may vary according to the **text type** due to the varying presuppositional conditions found in different genres. In particular, *the difference between spoken and written language plays a significant role*” (Sasse 1995a:9, my italics)
5. Sasse did attempt to initiate the study of spoken data, but the “spoken narrative material ... was not taken into further consideration since its transcription was stopped due to the time-consuming nature of the work” (Sasse 1995b:143).
6. Note that either with or without the benefactive topic, a sentence such as this would be odd in an announcement context if Kostas was treated as the topic, cf. #*Kostas ha llamado*. This sentence would be possible if the addressee had been expecting a call from Kostas and thus Kostas was more accessible.
7. *La Vanguardia*, Internet edition: <http://vangu.es.es>.
8. The exception is *Más vale prever...* “Lit. It’s worth more to foresee ...”. The whole proverb is *Más vale prever que lamentar*, “Better safe than sorry”.
9. *El Periódico*, Internet edition: <http://www.elperiodico.es>.

10. *La Jornada*, Internet edition: <http://serpiente.dgsca.unam.mx/jornada/>; or <http://www.nuclecu.unam.mx/~jornada/>.

11. Furthermore, newspaper headlines are not randomly found in a newspaper, but rather they are found in sections about particular topics (e.g. Politics, Sports), within which certain referents are even more accessible.

12. *Euskaldunon Egunkaria*, “the Basques’ newspaper” (or simply *Egunkaria* “the newspaper”) is the only newspaper published fully in Basque. Other newspapers that are published in the Basque Country occasionally have articles in Basque. The front page headlines and news summaries for *Egunkaria* are posted in the unofficial *Egunkaria* home page in the Internet in the following address: <http://www.kender.es/~jazanza/egun.htm>.

13. The rest of the paragraph of this front-page summary article has five more sentences (two of them quite complex), all of which have topic-comment structure, four with overt topics and two with ‘covert’ ones.

14. Sasse’s quote actually says “SV”, but this is obviously a typographical error.

15. Thus, according to Sasse,thetic structure for events in Hungarian is “restricted to fixed expressions involving a well-defined set of semantic areas with a limited amount of productivity. Like in M[odern]G[reek], with animate, especially human subjects, the S is typically topicalized (SV)” (Sasse 1995b:184).

16. The verb *averiarse* “break down”, like many Spanish intransitive verbs which typically express happenings and not actions, is the anti-causative version of the transitive (causative) verb *averiar* “break down” (which is rarely used). Expressing happenings by means of reflexive anti-causatives is so common in Spanish that some basically intransitive verbs, such as *morir* “die” and *caer* “fall” for instance, adopt (presumably by analogy) the reflexive pronoun even though it is not needed when used in the ‘happening’ or ‘event explanatory’ function, as opposed to the predicative function.

17. The addition of a ‘benefactive’ topic to athetic construction with a ‘predicate’ which typically appears in withthetic structure is common enough in many languages. The phenomenon sometimes goes by the name of ‘possessor ascension’ in the Relational Grammar tradition, from a somewhat anglocentric perspective.

18. Hungarian, like spoken Spanish, doesn’t have a promoting passive construction. It seems that subject inversion constructions are sometimes used in contexts in which the passive is often called for in a language like English (cf. Sasse 1995b:185-6). Sasse mentions athetic “OVS strategy” for Hungarian. The example he provides seems to have a topic-comment structure, it’s just that the topic in this case is the object and not the subject (Sasse 1995b:185, ex. 127).

19. Sasse mentions the following static existentials for Modern Greek: *kiriarxi* ‘reigns, prevails’; *jennate* ‘springs up, rises’; *jinete* ‘happens, takes place’ (Sasse 1995b:158). He distinguishes between “event-central” and “entity-central”thetic sentences (Sasse 1987).

The former report events, whereas the latter present entities. Lambrecht (1994:143-4) makes a similar distinction. The differences do not seem to me to be clear-cut, and I don't know of any formal differences between the 'two types' of constructions. It seems to me rather that thetic announcements can be made use of for bringing referents into the discourse in order to be talked about later (i.e. to become topics), perhaps under the influence or analogy of the (topicless) existential construction, but that doesn't mean that they constitute a these 'presentational' sentences constitute a separate or different type of thetic construction.

20. Actually, human objects of *ikusi* are often coded as datives, at least on the verb, much like in some dialects of Spanish, so in the detransitivized version the auxiliary would have (empty) absolutive and dative agreement, cf. 93C2A02:30.

21. In Chapter 4 we saw that Basque allows topicalizing of generic, mildly accessible phrases in existential constructions, but these are a minority of all existentials.

22. Information structure = *Like Non_Topic_Present**; Word order = *Like *V*S**; Predicate argument = *Copula*; Polarity = $\langle \rangle$ *neg*.

23. Information structure = *Non_Topic_Present_etorri*; Word order = *Like *V*S**; Predicate argument = *Copula*; Polarity = $\langle \rangle$ *neg*. There are no equivalent thetic sentences with *etorri* "come" in the Written Basque Corpus.

24. Whether the subject will invert in a sentence such as this is independent of its grammatical role. In this example it is an absolutive, but it could be ergative as well, e.g. *Arrazoi horregatik hil zuen txakurra nire aitak*, *Arrazoi horregatik hil zuen nire aitak txakurra* "That's why my father killed the dog."

25. According to Givón, "[t]he existential verb 'be' or 'have' has no lexical-semantic contents; it is *semantically bleached*" (Givón 1990:745). He argues that the real predicate is not the existential verb but another one, not necessarily a finite one.

26. Givón 1990, following Fox 1985, argues that these added predicates come under the same intonation unit as the presentational, though they are clearly also often found in their own sentence-tail intonation units, especially if the predicate is complex. According to Givón, "R. Fox (1985) has found that consistently, at the level of 99% in written English, some other predicate is always present with the indefinite-subject clauses. It comes *under the same intonation contour*... It is most commonly either a REL-clause, an adjective, a LOC-phrase, a noun complement, a genitive or associative phrase. And it most commonly *follows* the indefinite subject" (Givón 1990:746). Givón adds that, "[a]s noted more recently by Fox and Thompson (to appear), the extra predications that tend to appear with EPCs are used to establish the *relevance* or *connectivity* of the newly introduced referent to the particular junction of the discourse where it is introduced" (Givón 1990:746).

27. As Lambrecht very perceptibly notices, "[t]he relative clause in this construction, even though it is grammatically marked as dependent, exhibits a number of syntactic,

semantic, and pragmatic features which characterize it as a special type which could perhaps be characterized as “dependent main clause (see the formal analyses in Lambrecht 1988b for English and in Section 7.3 of Lambrecht [1986] for French)” (Lambrecht 1994:180).

28. Other codes = *Presentative+Predicate_FCI*; or Presentative_Predicate_FCI*.

29. Conjunction = -la_PPred.

30. The -la(rik) construction differs from the other two -la constructions in that (1) it is always non-asserted; (2) it is strictly verb-final, and (3) it has a variant -larik. This adverbial clause may have been the historical source for the completive clause. As is well known completive clauses often derive historically from adverbial clauses, which is why complementizers are often derived from adverbial pronouns, cf. *how* and *como* completive clauses in English and Spanish, respectively.

31. For a clear-cut example of a post-presentative predicate, cf. 93C2B01:333-335.

32. That these finite PPP's are actually asserted clauses has been denied by Huck and Na 1990, for instance.

33. More recently, Rochemont and Culicover (1990) have attempted to explain the formal characteristics of these ‘relative clauses’ in English (along with other ‘non major transformation types’) without appealing to movement.

34. A non-background description version of this last sentence, however, could also bethetic if the subject is not the topic, e.g. *Aquel día, brilló el sol* “That day the sun shone”, a global statement or announcement, again, not about the sun, but about the weather in general.

35. In his summary of subject inversion constructions in different European languages, Sasse concludes that “[i]t is difficult to say whether or not a specific type of preverbal material differing from the preceding in being “topical” rather than focal also triggers inversion in some languages, namely, adverbials of ‘setting’, i.e. expressions indicating time, place, and circumstance and setting a frame for the following predication. In Modern Greek, VS is almost obligatory or at least strongly preferred ... Russian and Romani seem to be similar ... The only language which does not have VS with “setting constituent fronting” at all is Latin. ... For Italian, it seems to depend on the text type ... For Spanish, we lack conclusive data...” (Sasse 1995a:12-13).

36. In spoken French, on the other hand, these setting clauses seem to have canonical SVO order, unlike main asserted clauses, according to Lambrecht 1994.

37. Truethetic clauses may have subjects with ‘hearer old’, or identifiable referents, but they do not have referents that are ‘discourse old’. Sasse mentions for example that “both the interruptive and the descriptive type commonly involve ‘identifiable’ subjects as well” (Sasse 1995a:16).

38. Hungarian “does not seem to have the **discontinuative** and **reactive** functions reported for M[odern]G[reek] ... Paragraphs usually begin with SV sentences ... It seems that VS clauses with a reactive function still occurred in sources of the 15th and 16th century [*sic*]”, but this “strategy seems to have been gradually abandoned” (Sasse 1995b:186).

39. Other Codes = **hasi**; Word Order = **V*S** Or **V*A**.

Chapter 6

Focality, focus, and word order

6.1 FOCALITY AND FOCUS

6.1.1 Introduction

In his monograph on Basque word order, Osa mentions how difficult it is sometimes to determine the focus of the sentence (Osa 1990:285).¹ Villasante too observes that one problem with Altube's word order rules for Basque is that they assume that all clauses have a focus, something which he doesn't think is true (cf. Chapter 2).²

The confusion in this matter is quite understandable. Altube's primary examples of focus, or *galdegaia* "the question element", were those in which the focus element was extremely salient, or focal, such as content questions and answers to such questions, or cases in which the focus was contrastive (and the rest of the proposition was already active or very accessible). However, the existence of a focus constituent in average statements, in which all or most of the elements of the clause are relatively 'new', is certainly not as obvious as in those marked cases.

I believe, however, as I argued in Chapter 3, that an argument can clearly be made for the existence of a pragmatic role focus even in all-news sentences, which have been said to be all-focus, or wide-focus assertions (cf. Lambrecht 1994). However, just as the

pragmatic role topic may have different realizations depending on its own pragmatic characteristics and those of the assertion it is in, so the formal characteristics of the focus may vary according to its pragmatic status and context.

The argument in favor of the existence of a focus *role* is supported by certain facts about assertions. The first one is that, as numerous investigators have argued, simple assertions, i.e. those expressed in a single intonation unit, typically a finite clause, always seem to have one, and only one, single element that is most 'newsworthy' or informative (in the context of the whole assertion, of course, since simple ideas do not provide information). It is true that sometimes we find complex assertions in which there seem to be a higher concentration of 'new' elements, but, as I argued in Chapter 3, additional informative elements are always added in intonation, and pragmatic, units which follow the main asserted clause, in what I have called sentence tails.

Under some special circumstances, those complex, or extended, assertions may seem to form a single overarching and complex intonational contour. The fact remains, however, that the vast majority of asserted clauses are simple assertions under a single intonation contour and with a single most informative element, bearing the main accent of the assertion. This element is the focus constituent and the idea that it expresses fills the focus role of the assertion.

It might be argued that rather than positing a focus role, it might make more sense to speak of 'focality', or 'rhematic salience', or degree of 'newsworthiness' or other such gradient parameter for the different elements of the assertion, much as some prefer to use only the gradient notion of 'topicality', without positing a topic role. I believe, however, that this notion would not be sufficient to explain or represent the intonational and

syntagmatic (formal) facts about asserted clauses. I believe that we need to posit a focus role *in addition to* the gradient notion of focality (newsworthiness, rhematic salience), just as we need to posit the notion of topic role in addition to the gradient notion of topicality (accessibility).

In addition to the well established fact that most assertions have a single, clear-cut most informative element, there are formal characteristics which are correlated with the focus role. The focus of the assertion is that element of the asserted clause (in a simple intonation unit) which receives the accent for that clause's intonation unit, the single, most prominent accent. It is true that this single accent may be more or less prominent depending on the salience of the focus with respect to the ideas expressed by the rest of the assertion, but in all cases one and only one element receives this main accent (unless, of course, the focus is extraposed into a separate or semi-separate intonation unit, in which case a clause may have more than one accented element, cf. below).

In addition there are languages, such as Basque, in which the focus role is intimately tied to a certain position in the sentence, which must always be filled. In this sense it is obvious that in every assertion a single element stands out by virtue of being placed in that position. Matters are complicated by the fact that, as we saw, some languages make use of more than one position, depending on a variety of factors, such as the salience of the focus (Spanish) or the grammatical role of the focus (English).

When more than one element in the assertion proper, that is the assertion minus any topic or setting element and additions to the assertion, is new or informative, we find that languages have defaults that determine which element will be the focus. I believe

that these defaults are not arbitrary, but rather reflect which elements are more informatively salient from a cognitive viewpoint, in other words, their focality.

The notion of focality refers to the 'noteworthiness' or 'informativeness' that the different elements of the rheme have in the context of the assertion. The focality of an idea depends in part on its accessibility, but also on other characteristics, as we saw in Chapter 3. Since simple assertions in natural language typically consist of a verb and a single (rhematic) complement (argument, adjunct, or modifier), the notion of focality usually comes down to a choice between the verb and the complement.

The most focal element, the most important, newsworthy, unexpected, informative, surprising element (in the context of the asserted proposition), will be the focus element. Additional elements, if they exist and are not topicalized (i.e. 'fronted' outside the rheme or assertion proper), must still be topical (accessible) or easy to accommodate in context, and are always unaccented. Additions or extensions to the assertion must be added in separate intonation units after the assertion proper.

Active or very accessible elements are not very focal, unless they are contrastive or otherwise unexpected or emphatic. Verbs are very focal when the ideas they express are semantically complex. Copula verbs or other quite stereotypical verbs with low semantic content, on the other hand, are low in focality. When the polarity of the proposition is contrastive or emphatic and all the other elements of the assertion are low in focality, then the polarity is the focus element and the finite verb acts as the focus constituent for the purposes of accentuation and focus position. When a verb and a complement are both quite focal, such as when they are both new and unpredictable, we

find that in declarative assertions the complement tends to be treated as the focus, especially if the verb expresses a stereotypical or basic level idea.

As we can see, the semantic and pragmatic properties of an idea which determine its focality depend in great part on its objective, or intrinsic, characteristics. In addition, an element may be more focal if the speaker wants to emphasize it for some reason, thus adding to its focality. In other words, the default focus preferences may be overridden by the speaker if he or she wants to place the emphasis on a particular idea.

In the following sections I will expand the discussion on this subject from Chapter 3, concentrating on the different types of salience, the types of elements of the assertion which may be salient (whether embedded or not),³ and some of the major coding mechanisms, in particular as they affect the word order of Basque asserted clauses.

6.1.2 Focality differences and default focus

As I said, although one of the elements of the rheme is sometimes clearly, and objectively, more salient than others, such as in answers to content questions, this is not always the case. In out-of-the-blue statements, whether they have a topic-comment configuration or not (and declarative assertions most often do), the different parts of the rheme may seem to be equally new, or equally unexpected. Or the differences might not be obvious or easy to ascertain, as Osa and Villasante noted. In these cases, as I said, the first choice seems to be the verb's complement, and this choice is always reflected accentually, as well as syntagmatically, i.e. by word order, in Basque and other languages. Occasionally both the verb and a complement are very focal. In such cases we often find that both are accented, but that the complement is extraposed to another intonation unit or

even be placed at the end of the clausal intonation unit preceded by a minor intonation break.

Thus, for example, in rhemes which contain a non-verbal predicate and a semantically 'light' verb, such as a copula or stereotypical (basic-level) verb (e.g., *I am here*, *I feel sick*, or *I finished the chapter*), the verb's complement is rather uncontroversially the focus, and the single accented element. In Basque this intrinsically focal focus element would be placed, most naturally, in rheme-initial and preverbal (focus) position. When both the verb and its rhematic complement are equally new and unexpected, as in, e.g., *I saw your father* or *I read a book*, the complement is typically the focus in non-emphatic declarative assertions, at least as long as the verb is a basic level, stereotypical, and rather unsurprising verb. Thus in Basque such a complement would be placed in focus position.

Thus we see that in Basque, under normal circumstances, a complement is the focus of the assertion. The combination of this fact and the fact that preverbal position is so strongly associated with focus complements may be the motivation behind the fact that so many common 'intransitive' predicates in Basque are coded as verb + complement units, where the complement is a non-referential object, e.g. *negar + egin* "cry" lit. "cry(n.) make/do". In these cases, of course, the non-referential nominal is typically the focus (and thus preverbal), unless some other complement is more focal. One might even speculate that the predominance of the analytic conjugation in Basque is a way of exploiting another type of 'predicate decomposition', one in which the verbal part of the verb is separated from the tense/polarity bearing element, in order to prevent the finite

verb from being in focus position (which always denotes polarity focus), in cases where there are no focal complements to fill the focus role.

6.1.3 Default focus and basic word order

Given that most studies of Basque word order have used decontextualized sentences in which all the parts of the rheme are typically equally new, it is not surprising that the conclusion is often reached that the basic word order in Basque is (S)OV. However, in naturally occurring connected language, the majority of sentences are not of this decontextualized type. In connected speech, the two or more elements of a rheme are generally not equally salient and a verb may be more focal than the (referent of the) complement for a variety of reasons, such as the high accessibility of the referent. This then results in the verb being the focus and thus being in focus position.⁴

As we can see, the ordering of elements inside the assertion depends on the relative focality of the ideas expressed by the verb and its complements, which determines which one will be the focus and placed in focus position. This means that often the focality of a complement depends on the estimated accessibility of its referent-idea (accessibility being one component of focality), as determined by the speaker, and thus is open to interpretation. As we will see below, variation in word order preferences among speakers may be traced to this difference in interpretation.

Finally, another focus related factor is responsible for variation in word order among Basque speakers, namely the relative frequency with which a speaker makes use of the strategy of focus extraposition. As we will see in this chapter, focus extraposition

can be quite frequent in speech, especially for some speakers, although this strategy is much less common in writing.

6.2 TYPES OF FOCALITY AND FOCUS

6.2.1 Contrastiveness and focus

There is a correlation between the focus role and the idea of contrastiveness. According to perhaps most investigators, contrastive focus is a subtype of focus of assertion. In other words, some foci are contrastive (cf., e.g., Dik 1989:282; Givón 1990:Ch. 16).⁵ For Givón, contrastiveness is one type “low informational predictability,” or “counterexpectancy” (cf., e.g., Givón 1990:Ch. 16, 1995:320-321, *fn*16).⁶ And because of this counterexpectancy, he argues, there is an strong crosslinguistic tendency to place these elements at the beginning of the clause.

I believe that contrast and other types of ‘counterexpectancy’ are factors which determine the degree of focality of a rhematic element. As we saw in Chapter 3, not all contrastive elements are foci, however. That is because the notion of *focus role* is applicable only in asserted clauses, whereas the notion of contrastiveness, and perhaps also that of focality, is also relevant in clauses which are not asserted.

The correlation between contrastive elements and clause initial position mentioned by Givón and others is easy to explain. First of all, it seems that perhaps all languages make use of rheme-initial position for foci, at least for the most focal ones. Since contrastive foci are very focal, it is not surprising that they are often placed in rheme-initial position. However, rheme-initial position is not clause initial position. The

discrepancy is much lessened in practice by two factors: topics in most languages are more often than not covertly expressed by morphemes incorporated into the verb, such as ‘coreferential pronouns’, ‘agreement’, or even unaccented pronouns, as in English. In addition, when topics are covert, they can be dislocated or inverted (antitopics), thus allowing very focal foci, such as contrastive ones, to be in clause-initial position.

We must be careful when equating focus with contrast. First of all, as I said, we should keep in mind that not every contrastive element is a focus. Elements in non-asserted clauses may (occasionally) be contrastive, but that doesn’t mean that they are foci. A topic may also be contrastive, as we saw in Chapter 4. Contrastive topics are not different from other topics, other than the fact that they are always covert and typically accented, since they are not continuous and predictable.

Another problem has to do with the exact definition of contrast. As Bolinger and others have argued, contrast is not a discrete category (cf. Bolinger 1961b, Givón 1990). In a wide sense everything contrasts with everything else in the universe (of discourse), and thus in that sense every narrow focus is contrastive. So if someone says *I went home*, in a sense *home* contrasts with every other place that person could have gone to. But there is also another, more common, sense of the term contrastive in which “the alternatives are narrowed down”, in which there is a small, more or less clear-cut set of alternatives (Bolinger 1961b:61).⁷ In reality contrast in natural language is usually something in between these two extremes, with more or less obvious alternatives being involved. In other words, there is a whole range between ‘ignorance’ (lack of expectation) and ‘contrary belief’, going through ‘contrary expectation’, in Givón’s terms.⁸

Although the two extremes are not the only possibilities that we find (sometimes, for instance, the alternatives are *implicitly* reduced by the context), in practice they serve as useful points of reference in the study of focality. Thus we can say that some foci are contrastive, whereas others are simply ‘new’ or ‘informational’ (non-contrastive) foci (cf., e.g., Dik 1989:282). A ‘new’ focus may be hearer new, discourse new, or discourse old but unexpected. However, as we saw, not all ‘new’ ideas are foci of assertions, since oftentimes discourse-new and even hearer-new referents are treated as easy to accommodate from the context. In other words, they are treated as if they were accessible because they are easy to situate in the world of discourse or are non-important (props) (cf. Chafe 1994a).

Contrastive foci are typically more focal (salient) than non-contrastive, *non-emphatic* new foci (see next section). Thus we find that assertions with a contrastive focus often display formal characteristics associated with high degrees of focality, such as rheme-initial foci, and even topic inversion, as Givón has noted.

The Basque linguist Altube (1929) too distinguished between two types of foci: *opositivos* (contrastive) and *absolutivos* (absolute), which seem to correspond to (narrow sense) contrastive and non-contrastive or ‘new’ foci. His examples, however, do not always make this clear. Some of the examples he gives of ‘*opposite*’ focus do not seem necessarily contrastive, though they could be (*Who broke that? He did*), though he does say that the contrast here may be ‘implicit’ rather than ‘explicit’. According to Altube, ‘*absolute*’ foci are those which do not suggest (“at least spontaneously”) other alternatives. His examples, however, could receive either a contrastive or a non-contrastive interpretation (e.g. *Everything that was said is true*) (Altube 1929:37). Some

of his examples of *absolute* focus seem to me to be actually examples of contrastive polar focus (e.g. *Did he go home (or not)?*).

6.2.2 Intrinsic vs. extrinsic focality

The major source of focality for the ideas expressed in an assertion derives from their inherent pragmatic properties, such as accessibility, unexpectedness, inherent contrastiveness, and the like. This type of focality is an *intrinsic* or context-dependent one, which forces itself upon the speaker by the context. Focality, as well as focus choice, may also be a matter of the speaker's choice, however. This I will call *extrinsic* focality.

Some students of focality and focus have emphasized or dealt primarily with intrinsic focality. Others, however, such as Bolinger, have stressed the importance of the speaker's choice (or will) in these matters. I believe that the major factors involved in imparting focality on an element are intrinsic and thus determined by the context, such as accessibility, contrastiveness, and the like. The speaker, however, sometimes does have a choice, such as, for instance, when two elements are (intrinsically) equally good candidates for this role. In addition, the speaker may add salience, or emphasis, to a particular focus to indicate that he or she considers it particularly unexpected, surprising, or otherwise 'noteworthy'. In other words, the context typically determines which element is the most focal and thus will become the focus, but a speaker can specify the degree of focality of a focus by coding it with formal devices which are associated with high degrees of focality, such as marked accentuation or marked word order.

When a speaker imbues a focus element with a high degree of salience this typically translates into additional salience for the whole assertion. Or the speaker may imbue the whole assertion with additional ‘force’, such as is the case with exclamations. These are the assertions that I have been calling **emphatic**. Extrinsic focality is a matter of speaker choice, and as such it may be used and exploited for rhetorical purposes, as in the case of the specialized uses of the Basque antitopic construction discussed in Chapter 5.

In addition, an emphatic assertion may also be one in which the focus is *intrinsically* very focal, more focal than in ordinary assertions, such as in contradictions and other contrastive assertions. As I said, a contrastive focus is typically more focal than a non-contrastive one, and this often has formal consequences, such as the placement of the focus constituent in clause initial position and/or the inversion of the topic.

6.2.3 Assertion types and focus types

Continuous narrative sentences, perhaps the most common type of sentence in discourse, do not typically have very focal foci. In certain types of sentences, however, the presence of a salient and noticeable focus element is the norm, a fact that is often reflected in the constructions used to express such assertions. I am referring to constructions such as content questions, polarity questions, negative statements (sometimes), and emphatic statements, such as exclamations. We can see examples of these assertion types in (6.1).

- (6.1) a. Who came?
b. John did come
c. John didn't come
d. Did John come?
e. Didn't John come?
f. John came.

The focus of a **content question**, such as the one in (6.1a), is (typically) the idea expressed by the content 'word' itself. This focus is very salient in the context of the question (interrogative assertion). This is so even if the other elements of the question are themselves highly focal (which often they're not). This is because the question construction carries 'instructions' for the hearer to accommodate the open, underlying proposition, to take it for granted, or to treat it as accessible, which is why it is sometimes said to be 'presupposed'. It is probably the high intrinsic degree of focal salience associated with question words that accounts for the fact that content question constructions very often place the content word in assertion-initial position, often with topic inversion as well, much like emphatic declarative assertions, even in languages like English in which clause-initial position for foci is quite marked. This construction is very much like the Basque content question construction, with the exception that subject inversion is more 'optional' in Basque.

Affirmative polarity assertions, such as the one in (6.1b), are similar to content questions in that for them to be felicitous, the underlying proposition (minus the polarity) must be in some way accessible or easy to accommodate. In the case of this example, someone might have asked or questioned whether John did actually come, for instance. Notice that in the English construction the polarity, which is instantiated by the finite

verb, is 'separated' from the verb and placed in rheme-initial position. A similar thing happens in Basque, as we will see.

Similarly, when a **negative sentence** is uttered, (6.1a), the underlying (affirmative) proposition is often (though not always) accessible (cf. Givón 1975, 1978; Horn 1989; cf. Chapter 7). This means that, in these cases, the negation (the negative polarity) stands out as the most informative part of the sentence, i.e. is very focal. Thus for the negative assertion in (6.1b) (*John didn't come*) to be uttered felicitously, there is typically a prior expectation that John might come (i.e. the idea-proposition that John was going to come was 'accessible', though not necessarily presupposed to be true). Otherwise, the listener would be entirely justified in challenging the negative statement (by saying, for instance, *Was he supposed to come?*). Not surprisingly, here too we find that the polarity bearing finite auxiliary verb, with the negative 'operator' cliticized to it, is placed in rheme-initial position. The same thing is found in Basque, as we will see in the next chapter.

6.2.4 Optional elements and focalizing particles

Optional rhematic complements, and in particular some types of modifiers, are also very often intrinsically salient, to the extent that they may also often force the accommodation and even presupposition of the rest of the sentence (cf., e.g., Givón 1990:711-12). Manner modifiers, for instance, are typically very focal and tend to be the default focus. Thus in the sentence *The teacher kicked me on purpose*, the adverbial *on purpose* would typically be the focus and the rest of the proposition, the fact that the teacher kicked the speaker, must be either accessible or else meant to be accommodated by

the hearer. Other modifiers, such as some common temporal adverbials, on the other hand, are easier to accommodate and thus are not as focal. Thus in a sentence such as *My mother came here yesterday*, the adverbial *yesterday* is not as likely to be the focus, but rather unaccented (and thus accessible). But it could be the focus, if accented, and in that case the fact that the speaker's mother came at all must be accessible.

In other words, these optional arguments and modifiers, when they are the focus, are quite salient with respect to the rest of the elements of the assertion. This is especially true with negative statements (e.g. *Joe didn't kill the goat deliberately*) or questions (e.g. *Did Joe kill the goat deliberately?*) (cf. Givón 1990:713, 1984a:Ch. 9).⁹ Adverbs such as these are almost always foci. Other adverbs, such as *honestly* and *suddenly* in English, may either act as setting adverbs (non-rhematic) and as salient parts of the rheme (cf. Jakobson 1975:25).

There is a special kind of modifiers or 'operators', which Givón calls 'contrastive quantifiers', which cause the idea they modify to become the focus of the assertion (cf. Givón 1990:715). These are words such as *even*, *all*, *every*, *other*, *first*, *only*, *self*, *really* and *just*. Jakobson (1975) also discusses this class of words and calls them **focalizers**. He distinguishes between exclusive focalizers (e.g. *only*), particularizers (e.g. *especially*), and additives (e.g. *also*) (cf. Jakobson 1975:24). König (1991) calls them 'focus particles', and recognizes two major types: *inclusive* or *additive*, and *exclusive* particles (cf. also König 1993, Kryk-Kastovsky 1993). Nevalainen (1991) studies three such elements, which he calls "exclusive focusing adverbs": *but*, *only*, and *just* (cf. also Nevalainen 1987). Myhill and Xing (1993), also alluding to these 'focus particles',

discuss the following major types of emphatic/contrastive ‘focus functions’: *even, only, negation, really, all, finally, again, also, rather, and neither.*

It is very interesting that according to all these investigators one of the major types of so called focalizers is the ‘inclusive’ or ‘additive’ type (cf. *also*), which is expressed as an additive *focus* in so many languages. As we saw in Chapter 4, this type of additive element in Basque is expressed not in the focus role but in the topic role, and the additive particle in question, *ere* “also”, is always a topicalizer rather than a focalizer.

This, however, is not the only type of contrastive function which is expressed as a topic (as opposed to as a focus) in Basque. As we will see in next chapter, another major type of contrastive ‘function’, namely the exclusive function (cf. *only*) is also typically expressed in the topic role, or, more generally, as a setting element, even if the referent in question is not highly topical. This is done, as we will see, by means of a peculiar negative construction in which the ‘focalizer’ *only* is expressed as *not other* (e.g. *besterik ez*) and the contrasted element is the topic.

6.2.5 Summary of focality and focus types

In this section we have seen some of the ways in which foci may differ from each other according to the source and type of their focality. The 3 major parameters along which focus may vary are degree of salience/focality, contrast vs. non-contrast, and intrinsic vs. extrinsic salience, cf. Table 6-1.

DEGREE	Minor salience	←————→	Very salient
CONTRASTIVENESS	Narrow contrast	←————→	Wide contrast (‘new’)
INTRINSICNESS	Intrinsic salience	←————→	Extrinsic salience

Table 6-1: Parameters along which rhematic salience may vary.

As we have seen, foci are always more newsworthy, or noteworthy, than the other elements of the assertion (proper). However, some foci are more focal than others, depending on their own focality, as well as the focality of the other elements of the assertion.

In addition, a speaker may choose to make a focused element even more salient by emphasizing it. This option is always reflected in the degree of intensity of the focus accent, but also often in other formal characteristics of the focus constituent. These include the use of specific constructions, such as cleft focus constructions, or, when the language has alternative positions for placing focus constituents, the placing of the focus element in a particularly salient position.

Another distinction we have seen is the one between contrastive and non-contrastive foci. Because of the limited choices involved when a focus is contrastive, it seems that typically contrastive foci are more salient (all other things being equal) than non-contrastive ones and thus are more likely to display coding characteristics associated with high focality.

These parameters do not necessarily represent discrete oppositions, but are rather continua along which variation can be found. Also, as we have seen, the three parameters are not fully independent of each other. Thus, for instance, extrinsic salience always adds to the degree of salience a focus already has, and contrastive foci are typically more salient than non-contrastive ones.

In addition to these basic parameters along which foci may differ, there are other semantic and pragmatic characteristics of focus elements which differentiate them from each other, such as the function that the focus element performs in the assertion. We

have already seen examples of additive and exclusive functions, for instance. More commonly, the focus (whether contrastive or not) may be a **referential** element inside the predication, such as an object, or it may be a **predicational** element, such as when the focus is the verb or a (non-referential) non-verbal predicate.

Additionally, as we saw in Chapter 3, a focus (whether contrastive or not) may perform an identificational function, that is, it may identify an entity or idea that the hearer already has in his or her mind in vague form before the time of the assertion, such as in replies to content questions (although identifications may also come out of the blue, unexpectedly). In other words, identificational foci are found in contexts in which the hearer already knows of the existence of a referent but doesn't have enough information to make an identification.

Identificational foci differ as to how the need for an identification comes about. The existence of the referent that needs to be identified may be brought up either by the speaker (e.g. *Who stole the bread?*), by the context, or it may even be set up by the speaker in the very same sentence, as happens often happens with extraposed foci, e.g. *And then I saw=, ... my brother.* In a sentence such as this the part preceding the focus sets up the expectation that there is a referent that is going to be identified.

6.3 THE FORMAL EXPRESSION OF FOCUS AND FOCALITY

6.3.1 Introduction

In every language the focus of the assertion is coded accentually. The degree of intensity of the accent ('accentual salience') is correlated with the focus' degree of focality. In addition, as we have already seen, there is a second mechanism for coding foci and different degrees of focality, namely the positioning of the focus in different positions in the clause.

In Chapter 3 I discussed the positions which are used for focus constituents crosslinguistically, which are summarized here in Table 6-2.

	Focus position	Focality
1	Rheme-initial and preverbal (after topic/settings)	Average/High
2	Rheme-second and postverbal	Average
3	Clause-initial (with postposed topic/settings)	High
4	Clause final or dislocated (to different degrees)	High/Average

Table 6-2: Focus positions used crosslinguistically and their relative markedness.

Some languages, the so-called OV languages and/or languages with a high degree of 'free' word order, probably come close to having a single preferred position for all focus elements, regardless of focus or assertion type: rheme-initial and preverbal (1). A high degree of focality in a focus is then coded solely by phonological means, or perhaps by the backing of setting elements as well, in any (3).

A number of linguists have argued that in languages with extreme freedom of word order, such as many Native American languages, there is a tendency to place a 'most informative' or 'newsworthy' element first, presumably to the exclusion of other

positions and regardless of the type of salience. These are languages where word order is controlled to a very high degree by pragmatic principles which have not become rigidified or semi-rigidified, i.e. grammaticalized, into constructions.¹⁰

All languages seem to allow rheme-initial position for at least some types of (very focal) foci. Some languages, however, such as typically VO languages, allow more than one position for focus elements. Occasionally the position depends on the grammatical properties of the focus element, as is the case in English and perhaps other rigid SVO languages. Non-rigid SVO languages, such as Spanish, make use of postverbal position (rheme-second) (2) for all unmarked (non-emphatic) foci, and assertion initial position, with or without setting postposing, for marked (highly salient) ones (3) (so-called 'emotive' ordering in the Prague School approach, cf. Panhuis 1982, Hopper 1985). Even rigid SVO languages may allow more than one position for focus constituents, particularly with non-arguments, such as adverbs, and may also have specialized minor constructions the purpose of which is to place a focus constituent in a position more in accord with its pragmatic properties, rather than its grammatical properties, such as subject-focus inversion constructions (such as the *there* construction in English) and cleft constructions.

In addition, many languages, whether VO or OV, seem to allow the extraposition or dislocation of a focus constituent to a following intonation unit for a variety of reasons. So-called VO languages seem to use this strategy more than so-called OV languages, and particularly than rigid OV languages, which use it only under limited circumstances, if at all. The extraposed focus is not necessarily completely dislocated. More grammaticalized versions of this construction allow a closer proximity between the main

clause and the focus so that the focus may even seem to be clause final, although it always seems to retain some type of intonational break with the rest of the clause.

6.3.2 Languages with preverbal focus as the main focus strategy

As I said, there are many languages in which the focus constituent is placed exclusively, or almost exclusively, in preverbal, rheme-initial position. This has been said to be true of many languages, such as Georgian, Armenian, Turkish, and Hungarian (cf., e.g., Comrie 1984; Horvath 1981, Kim 1988). Languages which have been classified as having very free word order and ‘hard to classify’, such as many Native American languages, have also been said to place the most ‘newsworthy’ element in initial position (cf., e.g., Mithun 1987/1992).

Comrie (1984) argues that Hungarian is a language, where “the preverbal rule is very strict” (Comrie 1984:3). This, of course, is in stark contrast with Sasse’s (1995b) assessment that Hungarian displays subject inversion inthetic assertions, sincethetic assertions are those in which the subject is the focus. In fact, Hungarian seems to be very similar to Basque in terms of word order characteristics and focusing strategies.

Comrie has argued that Eastern Armenian is another language which typically places its foci in preverbal position, but that it allows them to be postverbal under some circumstances. In Armenian, he argues, the “preverbal rule” for focus is “considerably less” strict than in Hungarian (Comrie 1984:3). Thus in Armenian, unlike in Hungarian, “the answer to a question with an interrogative pronoun” could possibly be postverbal (*ibid.*). This is exactly what we find in Basque as well, but, as I have argued, it is not that these objects are postverbal, but rather extraposed.

Actually, Comrie's description of Armenian is interesting to us since this language's focus marking preferences and word order characteristics are surprisingly similar to those of spoken Basque. Spoken Basque is like Armenian, for instance, in that (S)VO-type orders are about as common as (S)OV ones. Comrie would like to say that Armenian is an OV or verb-final language, i.e. that this order is *basic*, but he recognizes that VO order seem to be equally unmarked. (Notice that Comrie does not mention pragmatic relations in this context of word order.)

The normal order is for the subject to precede the predicate, although constituents of the predicate may be preposed, in particular to indicate topicalization. Within the predicate, there is a tendency for the verb to occur clause-finally, although further investigation will have to be required in order to determine definitively whether Armenian is basically a subject-object-verb or a subject-verb-object language. Both of the following orders are quite normal. (Comrie 1984:4).

This is yet one more example of the futility of trying to determine a basic word order for so many languages in term of its grammatical relations rather than in terms of its pragmatic relations.

In Armenian too, just like in Basque, the copula needs to have some constituent come before it, it cannot stand rheme-initially: "the preverbal position before the verb 'be' in present and imperfect indicative must always be filled, i.e., no sentence with the present or imperfect of 'be' can lack a syntactic focus" (Comrie 1984:7). This suggests to Comrie the hypothesis that "all sentences have a syntactic focus" element (i.e. a formal element in rheme initial position, other than the bare verb). What Comrie calls a syntactic focus is what I call a focus, and what he calls a pragmatic focus is what I call a very focal focus. Thus he agrees with my assessment that every assertion must have a

focus, and that when the verb is a copula, the non-verbal predicate is typically the focus (except for in those marked situations in which the polarity is the focus, of course). He also mentions that the verb may be the (syntactic) focus when he says that “[i]f no noun phrase or adverbial is specifically focused, then the syntactic focus is, by default, on the finite verb” (Comrie 1984:7).

In Armenian, when the verb is periphrastic and it has a complement which is the focus, the non-finite “lexical verb” must move out of the way to make room for the ‘syntactic focus’, which must go before the finite verb-auxiliary (Comrie 1984:7-8). In Basque, on the other hand, the finite auxiliary verb and the “lexical verb” are fused into a single phonological word and thus remain as a unit. That is, the whole verb together (the main participle plus the finite auxiliary) acts as a unit for the purpose of determining preverbal position. This is only true in affirmative statements, however, and not in negative statements, as we will see in Chapter 7.

The position of the negative particle in Armenian is also extremely similar to the position of the negative particle in Basque: rheme-initial, preverbal position, what Comrie calls “apparent focus position” (Comrie 1984:9). Comrie argues that the negative particle itself is the actual pragmatic focus, at least in some instances, though perhaps not always, despite the fact that this particle cannot bear stress of its own and is always cliticized to the next word, just like in Basque.

This, of course, is the same idea that Altube (1920, 1929) had for Basque (cf. Chapter 2). I will argue in the next chapter, however, that it is not the negative word that is the ‘focus’ (either in Basque or Armenian), but rather the *negative polarity*. Since the polarity is instantiated in the finite verb, it is the finite verb (together with the cliticized)

negative operator/particle, that is the ‘syntactic focus’ of the assertion. In fact this is only true in cases in which a negation contradicts the affirmative proposition, which is the case quite often, as I mentioned earlier in this chapter. This, of course, explains why the finite verb (together with the negative operator) are found in rheme-initial position.

Finally, I should mention Comrie’s “most intriguing” observation with respect to Armenian, namely that a “nonspecific object ... appears in syntactic focus position ..., even though [it] is not necessarily interpreted as pragmatic focus” (Comrie 1984:16). As we will see further below, in Basque too, non-referential nominals and nominals with non-specific referents in general are the objects which most strongly resist postverbal positioning. This, however, is not so intriguing when we realize that these nominals are the default focus in these cases, the most focal element in the predicate, and, furthermore, that their low topicality is what typically prevents them from being placed after the verb. In Basque, however, in most cases if the idea expressed by the non-referential nominal is ‘active’ (it has just been mentioned), another element, may be the focus and then the object must be postverbal.

6.3.3 Preverbal focus position in Basque

Basque, like perhaps all other languages which allow the verb to be clause final, has a very strong tendency to place focus constituents in preverbal, rheme-initial position. This is the primary position for focus elements, especially in standard, written Basque. In the next section I will discuss the alternative focus position, namely focus extraposition, which can be quite common in some varieties of Basque.

‘Preverbal position’, however, can mean different things. We just saw an example of this when we looked at the difference between Basque and Armenian periphrastic verbs. In Basque, with finite, synthetic verbs, preverbal position is always the position before the finite verb, e.g. (6.2).

- (6.2) *Aitak liburu bat dauka.*
 father:ERG book one he.has.it
 “Father has a book.”

When the verb is synthetic and there is no other element in the rheme, then the polarity, or the verbal idea itself, must be the focus, and this is coded by means of the (‘emphatic’) affirmative particle *ba-* before the finite verb, which typically indicates that the polarity is the focus. Sometimes it seems that this particle can also indicate that the verbal idea itself is the focus, although with these typically light, imperfective verbs verb focus is quite rare. To express verb focus typically the periphrastic construction, with *ari* “busy, engaged”, must be used instead. This is the same construction that the vast majority of verbs must use to code continuous imperfective aspect. In (6.3a&b) we can see two examples, with a transitive and an intransitive verb respectively.

- (6.3) a. *Aitak badauka.*
 father:DEF:ERG EMPH:he.has.it
 “Father does have one/it” (polarity focus)
 (?“Father has one/it” (verb focus))
- b. *Aita badator.*
 father:DEF EMPH:he.is.coming
 “Father is coming” (polarity focus)
 or “Father is coming” (verb-focus)

With analytic verbs, in affirmative sentences the two parts of the verb form a unit, even though they are written as separate words in standard orthography. The focus is simply placed before the verbal complex, as in (6.4), with the object in focus position.

- (6.4) Aitak [RHEME liburu-bat [v ikusi-du]]
father:DEF:ERG book one see:PFV-he.has.it
“Father has seen a book”

One difference between synthetic and analytic verbs is that, with analytic verbs, when the verb is the only element in the rheme or when the verb or the polarity is the focus, there is no special marking involved, such as the particle *ba*, as we can see in (6.5a&b) below.

- (6.5) a. Aitak [RHEME ikusi-du]
father:DEF:ERG see:PFV-he.has.it
“Father has seen it”
b. Aita [RHEME etorri-da]
father:DEF come:PFV-he.is
“Father has come”/ “Father came”

It is easy to see that at a time before the two parts of the verb fused, the lexical verb was placed in focus position (before the finite auxiliary) in these constructions.

In asserted negative sentences, however, as we will see in more detail in the next chapter, the two parts of the verb do not form a single word and they may not even be contiguous. The non-finite verb must obligatorily ‘move’ to postverbal position when the negative particle precedes the finite element of the verb, the auxiliary, as we can see in (6.6a&b) below.

- (6.6) a. Aitak [RHEME ez-du liburu bat ikusi]
 father:DEF:ERG not-he.has.it book one see:PFV
 “Father hasn’t seen a book”
- b. Aita [RHEME ez-da etorri]
 father:DEF not-he.is come:PFV
 “Father hasn’t come”, “Father didn’t come”

6.3.4 Emphatic assertions and emphatic focus

As I mentioned above, emphatic assertions are those which have a focus which is significantly more focal than the other elements of the rheme and more salient than the average focus. This situation may arise in a number of contexts. As we saw, contrastive foci are quite focal. Also, when the open proposition which obtains from excluding the focus element is active, the focus is also quite focal, relatively speaking. A speaker may also choose to place additional emphasis on a focus to make it more salient than in average assertions. In content questions the focus is also typically of this kind for reasons I mentioned earlier.

An emphatic focus (whether contrastive or not) in an emphatic assertion has a very strong preference for being in rheme-initial preverbal position in Basque, as well as in many VO languages, such as Spanish, for instance. Even in English, with its rigid word order, there are constructions meeting these characteristics in which the focus is placed in initial position, such as content questions and exclamations.

Preverbal positioning of emphatic foci may be accompanied by the backing of the topic to assertion-tail position, as an antitopic. This is the same position where complements which are neither settings nor the assertion focus are placed, and, like them, these postposed topics must be easy to accommodate. This is, of course, the ‘inverted

order' mentioned by Prague School scholars and the same type of phenomenon discussed by Chafe and Lambrecht as antitopics. In addition to the topic, other setting elements are also typically inverted in these constructions with very focal foci. As we saw earlier, an alternative to inversion is left-dislocation, which also allows the focus constituent to come early (or first) in the clause (though not in the sentence).

Topic inversion is perhaps particularly important in OV languages, where pre-verbal position is the only or at least the major position for focus constituents of all types, regardless of its focality level. In other words, in these languages topic inversion is one of the major cues for emphasis, besides the intensity of the focus accent. But this strategy is also found commonly in VO languages, such as Greek (cf., e.g., Sasse 1995a:11). (Note also for instance the redundant, inverted pronoun in emphatic imperatives in English: *Don't you worry!*, *Don't you do that!*.) Topic inversion is not the only way in which a focus constituent may be highlighted. The same result of placing the focus in clause- and intonation-unit initial position may be achieved by placing the overt topic in its own intonation unit ('left dislocated'), as noticed for instance by Odriozola and Zabala (1992:37-38).

Emphatic constructions of this type are disruptive of continuity in narrative and thus are not very common in this genre. They are most commonly found in conversational turns, especially in vivid conversation styles, and they are used to display surprise, unexpectedness, or 'emotivity'. Constructions displaying emphatic focus and comment-topic order sometimes have other special characteristics such as copula verb deletion, as in the following Basque example in (6.7) (mentioned in Villasante 1980:43; the intonational contour was added).

(6.7) Edérra eguraldià!
beautiful:DEF weather:DEF
“Beautiful weather!” (“Beautiful, the weather”)

The second element in this example, *eguraldia* “the weather”, would seem to be the topic (the sentence *is* about the weather). Notice that the ‘inverted’ topic (*the weather*) is not accented, even though it is presumably discourse-new, since it can be easily accommodated from the context. In this sentence the rheme’s accent is on the focus non-verbal predicate (*ederra* “beautiful”) and the intonational contour gradually descends from there.

This type of postposing, while relatively marked (statistically speaking), is perhaps more common, in Basque than in VO languages. In Spanish, however, with its flexible subject order, an identical construction to the one in (6.7) is possible as well, e.g. *¡Bonito el día!* “Beautiful day!”). In English, with its rigid word order, inverted order is less common, but also found in some specialized (‘minor’) constructions.

As we already saw in the preceding chapters, the frequency of topic inversion in Basque depends on the genre and the style. In narrative it is quite rare, outside certain specialized uses, such as identificational focus constructions with copula verbs, of the type exemplified in English by *Cháplin was the thief!*, *Hére's your book!* and in some specialized contexts, such as action chain beginnings. Overall, copula verbs are the most common type of verb found in examples of this construction, perhaps due to the inherent high salience of non-verbal predicates. In conversation, as I said, it is more common, since exclamatory, non-continuous sentences are found there more often.

In Chapter 5 we saw that these clauses are quite common in certain Basque written styles, such as newspaper articles and commentary, and especially headlines. The

following excerpt in (6.8), from the first paragraph of a newspaper commentary piece, for instance, contains two such antitopic constructions.

(6.8) *Euskaldunon Egunkaria*, Tuesday, Jan. 2, 1996, by Iñaki Heras Saizarbitoria

- a. 1996: beste geltoki bat diru batasunerantz
1996: another stop one money unity:DIRECT
- b. Tren handi bat omen da **Europa**.
train big one supposedly it.is Europe
- c. Gero eta bagoi gehiago dituen.
later and wagon more it.has.them:REL:DEF
- d. Sei, hamar, hamabi, hamabost....
six, ten, twelve, fifteen
- e. Lokomotore indartsua du **trenak**,
locomotive strong:DEF it.has.it train:DEF:ERG
- f. Alemania,
Germany
- g. eta gidari trebea,
and guide/driver/engineer capable/experienced:DEF
- h. Bundesbank delakoa. ...
Bundesbank it.is:COMPL:GEN:DEF

“1996: ANOTHER STOP ON THE ROAD TOWARDS MONETARY UNION. / **Europe is like a big train**. / One which has more and more wagons. / Six, ten, twelve, fifteen ... / **The train has a strong engine**, / Germany, / and a good driver, / the so-called Bundesbank.”

These sentences also exemplify the fact that this construction is most prevalent with a copula, as in (6.8b), and other ‘light verbs’, such as the possessive verb (‘transitive copula’), as in (6.8e). These verbs seem to be easy to accommodate in most contexts and thus are not very focal.

6.4 FOCUS EXTRAPOSITION AND BASQUE WORD ORDER

6.4.1 Focus extraposition or delay

In addition to the preverbal focusing strategy, Basque, like many other languages, particularly VO languages but perhaps not so rigidly verb-final languages, allows for the extraposition of the focus under (typically) exceptional circumstances. This is the strategy which we saw in Chapter 5 was used sometimes in presentations in speech, though typically not in writing. Rigid word order languages can achieve the same extraposing effect by using bi-clausal right-focus-cleft constructions (*wh*-clefts in English), just like the focus fronting strategy can be achieved by left-focus-cleft constructions (*it*-clefts in English).

Languages may differ in the purposes to which this extraposition strategy is used for and how often it is used. As we have seen, in some languages this strategy may be used only, for example, when the focus constituent is very long and complex, especially if the default position for the focus is preverbal or, even if it is postverbal, there are other non-focus complements which are shorter and 'simpler'.

In Basque focus right-clefting is quite common with long and complex foci. Thus we find that of the 244 finite completive (*-la*) complement clauses in the Spoken Basque Corpus, the vast majority are extraposed.¹¹ These clauses are typically the focus of the assertion involving the main verb as well as (dependent) assertions in their own right. Of these 244 clauses, only 3 precede the main verb, two of them because the dependent clauses they are in are themselves strictly verb final (non-asserted).¹² The remaining 241 completive clauses are all postverbal. The majority are extraposed, although a few are to

the right of the verb but are not foci and thus not extraposed. In 133 of the extraposed completive clauses the dependent clause is clearly in a separate intonation unit from the main clause, as in (6.9) below.

(6.9) 93C1B06

152 eta eskutitz orretan ba jartzen-du,
and letter that:LOC well put/place:IMPFV-it.has.it
153 ba Txarlot joan-daitekeela ba kalera.
well Chaplin go:PFV-he.may:COMPL well street:DEF:ALL
“and in that letter it says, / well that Chaplin can leave the jail.”

In the remaining 108 cases, however, they seemed to be integrated into the main intonation unit, even though an intonational break still can be detected. We can see an example in (6.10:90). (Notice also the antitopic in (6.10:87) and the extraposed focus in the dependent assertion in (6.10:90).)

(6.10) 93C2A06

87 e= jeisten da= e nekazaria,
uh come.down:IMPFV he.is uh farmer:DEF
88 zugaitzetik, ...
tree:DEF:ABL
89 eta= ori!
and that
90 konturatzen-da <falta zaiola udare saski bat> -
realize:IMPFV-he.is miss/lack it.is.to.him:COMPL pear basket one
“uh he comes down uh the farmer, / from the tree / and of course / he realizes that he’s missing a basket of pears.”

In writing too, a majority of completive clauses, 26 out of 44 in the written corpus, are postverbal, though the tendency is not as overwhelming here. We can see an example in (6.11). Although the written source for this sentence contains no orthographic markers indicating it, it is clear that the completive clause must be preceded by a rising ‘comma’ intonation.

- (6.11) *Kuba triste dago*, pg. 23
 a. kontua da
 story:DEF it.is
 b. <zer-zerka nagoela denbora guztian>,
 what-what:ADV I.am:COMPL time all:DEF:LOC
 “the result is that the whole time I’m asking *What? What?*”

Extrapolation is also quite common with non-finite complement clauses, which are also overwhelmingly the focus of the predicate, as in the case of (6.12:289). (Here the / mark codes the rising intonation between the clause proper and the extraposed focus.)

- (6.12) 93C2B01
 288 =ta ikusten-zan an,
 and see:IMPFV-he.was there
 289 ai zan / <periodiku letzen>.
 busy/doing he.was newspaper read:IMPFV
 “and he could be seen there, / he was reading the paper.”

The majority of these complements are postverbal in this construction with *ari* “busy”, which as I said is the only means of coding continuous imperfective aspect for all verbs which lack synthetic forms.¹³ We can see that if this construction were to become more grammaticalized, the end result could be an auxiliary (*ari*) that precedes lexical verbs.¹⁴

With less complex complements focus right-clefting is quite rare in writing, though we do find examples of it, as we will see below. It does seem, however, that in speech, and in particular the speech of some speakers (or groups of speakers) this focusing strategy is becoming increasingly common, to the extent that for some speakers the intonational markedness of this construction, in other words, the degree of clefting, is somewhat reduced.

In other words, one could say that in some cases there seems to be a reintegration, or coming together of the two parts of the rheme, perhaps resulting from reanalysis of the

construction by speakers who may be under the influence of the Romance postverbal focus ‘strategy’. This reanalysis is greatly facilitated by the fact that in most of these sentences there are no additional complements intervening between the verb and the clefted focus, the only thing that separates the two of them being the intonational break.

6.4.2 Focus-extraposition in previous studies of Basque syntax

Several authors have noticed that in Basque there is an alternative to preverbal, rheme-initial focus position in which the focus follows the verb. Thus, for instance, Villasante (1980) mentions a sentence by classical Basque author Axular in which a non-verbal predicate (which are typically foci) is postverbal (Villasante 1980:12; the sentence is from Axular 1977/1643:3, § 2).

- (6.13) *Liburutto haur da emazurtza.*
book:DIMIN this it.is orphan:DEF
“This book is an orphan.”

Basque linguist Mitxelena in his running commentary to Villasante’s book mentions that although this rule goes against Altube’s focus rule, it is perfectly normal to him as a topic-comment sentence, i.e. as a sentence in which the subject is the topic, meaning simply *This book is an orphan* or, in a paraphrase, *What happens to this book is that it is an orphan* (Villasante 1980:12).¹⁵ What neither one of these authors mentions is that the focus non-verbal predicate is not just postverbal, but extraposed, preceded by an intonational break.

This type of sentence was noticed by Altube too (1929:8ff), who disapproved of it vehemently. He complained, for instance, that in old Basque catechisms, consisting of

questions and answers about Catholic doctrine, in the answers the rhematically salient element, which sometimes is rather long, was always placed after the copula, as in (6.14). This particular example has a relatively short non-verbal predicate (Altube 1929:8; my accent marks, J.A.).

- (6.14) Q: Nor da Jaungoikoa?
 who s/he.is God:DEF
 A: Da zuzentasuna bera
 s/he.is righteousness:DEF itself:DEF
 “Q: Who is God? A: S/he is righteousness itself.”

Altube viewed this construction as an obvious Spanish calque. Other Basque linguists later on, such as Goenaga (1980:201), Osa (1990:115), and even the Basque Language Academy’s grammar (Euskaltzaindia 1987a:35), have come to the defense of this construction, which they see as perfectly normal, authentic, and rather common in Basque, particularly in spoken Basque. The Basque Language Academy, for instance, says:

It must be said that the first and foremost position for the focus is immediately preverbal position. However, sometimes—and this happens specially in spoken Basque—the *intonation may break this order*, the focus being placed to the right of the verb then. (Euskaltzaindia 1987a:35, my translation, my italics, J.A.)¹⁶

Notice the comment about the intonational break, for that is crucial here. The Academy’s grammar doesn’t mention it further, but Osa (1990) does make it clear that in this construction there is an intonational break right before the focus constituent. Notice the following example in (6.15) from Osa’s monographic study of Basque word order (Osa 1990:105).

- (6.15) Q: Nork ekarri du hori?
 who:ERG bring:PFV s/he.has.it that
 A: Hori ekarri-du // aitak
 that bring:PFV-he.has.it father:DEF:ERG
 “Q: Who brought that? A: Father brought it.”

I have added the double slash before the rhematically salient element here, even though Osa’s example does not have, for that is the marker that Osa uses elsewhere to show this intonational break in this type of construction (cf., e.g., Osa 1990:114), as did Goenaga (1980:201) before him.

In the spoken versions of examples (6.13), (6.14) and (6.15), there is a noticeable intonational break immediately before the focus constituent. This intonational break, with rising intonation immediately before the focus may be more or less noticeable, but is always there. This intonational break exists because, at least originally, the purpose of this construction was to place the focus constituent in its own intonation unit, following the rest of the clause. In addition, to the extent that there are two intonational units (or subunits) involved, an accent can be detected on the verb, which is where the intonational rise begins, especially when there are not other unaccented constituents following the verb and before the focus constituent.

It is true, as Altube’s critics claim, that this construction is not necessarily a Romance borrowing and that it has a long history in the Basque language and cannot simply be legislated against. On the other hand, it is doubtful that this construction would have acquired the high degree of usage it has in the spoken language if it hadn’t been for the fact that focus in Romance languages is typically postverbal *or* extraposed. Altube was certainly on to something when he noticed the fact that this construction was

becoming less and less marked in Basque, especially in the context in which he criticized it, i.e. in written Basque, where such a construction doesn't formally display the phonological cues that are associated with it in speech. (In present day Basque, however, this construction is associated primarily with speech, as most investigators have noted.)

Furthermore, I believe that the commonality of this bipartite construction may be contributing to its reanalysis, at least by some speakers, that is, 'reintegration' of the two parts. Eventually, as I speculated earlier, this may lead to the development of a postverbal focus strategy equivalent to that found in so-called VO languages.

6.4.3 The complexity of focus phrases and the markedness of extraposition

As I mentioned earlier, focus extraposition is particularly unmarked and motivated when the focus phrase is long and complex (motivated perhaps by processing considerations, as we saw in Chapter 1). Villasante's defense of the focus postposing construction in written Basque is interesting in this respect. Villasante mentions that, because of complaints by Altube and others about this very type of 'postverbal' placement of the focus, a new version of the catechism was prepared (in 1921), in which the answers all have preverbal focus, such as the one in (6.16) (cf. Villasante 1980:249).¹⁷ (The order here is Focus-Verb-Subject. Note that not all of Villasante's examples have such complex focus constituents.)

- (6.16) Q: Zer dira deabruak?
 what they.are devils:DEF
 A: Jaungoikoaren aurrez-aurre altxa ziralako, inpernura amilduak ezan ziran
 gogo utsak **dira** deabruak
 God:DEF:GEN front:INSTR-front raise they.were:BECAUSE hell:DEF:ALL
 throw.out:PFV:DEF:PL have:PFV they.were:REL spirit empty/imperfect:DEF:PL
 they.are devils:DEF
 “Q: What are devils? A: (lit.) “Devils are imperfect spirits who were thrown out into hell because the rose up against God.” (but XVS)

Villasante finds that the solution is not always very successful, since “such constructions, having such long phrases, produce a sensation of anxiety and of waiting for something” (*ibid.*). Furthermore, “keeping so strictly to this [XV] order, especially when the predicate is long—as often happens in written language—produces the impression of something very unusual and violent” (*ibid.*). He believes that “the most normal thing in such situations is to place the [non-verbal] predicate after the verb”, or, at the very least, to place the verb closer to the front of the clause” (*ibid.*).¹⁸

Villasante’s argument is very interesting. As unusual as such long nominal predicates are in actual speech (and there are plenty of unusual things with the focus NP in this clause), extraposition is quite clearly *unmarked* in such cases in Basque, whether it is spoken or written. Villasante’s description seems accurate: the listener/reader ‘wants’ a verb much sooner than it is available in (6.16E) and something seems amiss when it doesn’t come. This type of ‘reordering’ associated with ‘heavy’ or ‘complex’ elements seems to have a clear source in processing constraints, as we saw in Chapter 1.¹⁹

On the other hand, the ‘construction’, or ‘strategy’, is not restricted to long and complex elements such as nominals with relative clauses and complement clauses, and the fact of the matter is that often it is used with much simpler phrases without this

processing justification, perhaps just to indicate emphasis, as we saw in examples (6.13) and (6.14), or to produce certain rhetorical effects (such as suspense). Clearly, the question of how complex a phrase must be to 'require' dislocation is a relative one and one that is bound to change through time if there are additional (internal or external) motivations for using the construction. We will see below just how prevalent this construction is in selected samples from the written and spoken corpora.

Once extraposition becomes established as a common possible strategy for focus constituents and stops being a strategy used with unusually complex ones, the question of specialization naturally arises. Once a second (common) position for focus constituents of any kind becomes available, it will most likely become linked with certain pragmatic functions or with pragmatic statuses of the ideas involved. In general, however, we can say that both positions are typically used with very focal foci, since they are both very salient 'positions'. Perhaps more focal foci, such as contrastive ones, prefer clause initial position whereas less focal ('new') ones prefer clause-final (extraposed) position. On the other hand, we find that extraposed position, if clearly marked intonationally, may also have quite focal foci, even contrastive ones, as we can see in example (6.17:86)

(6.17) 93C1B08

- 83 Ta jateko tokian,
and eat:NOMIN:GEN place:DEF:LOC
84 ba\ dago ogia ta orrela. ez?
well it.is bread:DEF and thus. no?
85 Baino berak ez-du ezer,
but he:ERG not-he.has.it anything
86 ogia du, bere ondokoak.
bread:DEF he.has.it his side:GEN:DEF:ERG
87 Artu-diolako.
take:PFV-he.has.it.from.him:BECAUSE

“And in the dining place, / there is bread and stuff. right? / But he doesn’t have anything. / The guy next to him has the bread. / Cause he took it from him.”

In this case the object is the topic and the contrastive ‘subject’ focus is clearly extraposed. On the other hand, this focus constituent could also have been placed in preverbal position, with or without topic inversion, as we can see in (6.18a&b), under very much the same (intrinsic) pragmatic conditions.

- (6.18) a. Bere ondokoak du ogia.
his side:GEN:DEF:ERG he.has.it bread:DEF
b. Ogia bere ondokoak du.
bread:DEF his side:GEN:DEF:ERG he.has.it

Exactly which option is preferred in what contexts is not always easy to determine. The difference doesn’t seem to depend at all on differences in the *pragmatic status* of the referents involved but, rather, on rhetorical strategies of presenting the information. The same thing is true of Spanish, which can also use these three different orders (OV, S, OSV, and SVO) in the same context. Spanish, in addition, also has a further possible possibility, namely postverbal focus (OVS). In Spanish, this strategy would probably not be used with such a ‘marked’ or focal (contrastive) focus, however, since the postverbal

strategy is used preferably for average salience new foci in continuous clauses, i.e. those which do not disrupt the continuity of flow of discourse.²⁰

6.4.4 Extraposed focus with focus placeholders

I should mention here that the cleft-focus construction is not the only available strategy for postposing complex focus constituents. In fact, there is another similar strategy which can be used in just such situations, which consists in using a ‘placeholder’ in preverbal focus position when the actual focus constituent is placed after the asserted clause proper in a separate intonation unit. This strategy, which is also used in English and other languages, is typically used when the focus is a list of elements, as we can see in the following example (from Odriozola and Zabala 1992:49, ex. 53).

- (6.19) Lanean kontsideratutako gonbinaketak, **hurrengoak** dira: uda/goi-maila,
uda/behe-maila, negua/goi-maila eta negua/behe maila
work:DEF:LOC considered:ADV:GEN combinations:DEF, following:DEF:PL
they.are: ...
“The combinations considered in this work are **the following**: summer/high-level, summer/low-level, winter/high-level and winter/low level.”

Besides *hurrengoa(k)* “the following”, other common focus placeholders in Basque include: *hauxe* “this(emphatic)” *ondokoa* “the next/following”, and, the most common one in speech, *zera* lit. “the what”.

In this construction, like in the cleft-focus construction, the focus constituent is placed outside the asserted clause and its intonation unit, but it is also linked to an actual constituent which is inside the clause proper. This strategy is much more common in writing than in speech, where it is fairly rare. In speech the placeholder *zera* is used typically when the speaker cannot think of the appropriate wording for the focus phrase.

6.5 VERB FOCUS AND WORD ORDER

6.5.1 Deviations from default focus

As I mentioned earlier, in all-news assertions where no element of the assertion is accessible from the context, in the default case a complement, say an absolutive object, and not the verb, is chosen as the focus and thus receives the assertion's accent and is placed in rheme-initial position, thus resulting in OV-type word order. We can see an example in (6.20).

- (6.20) Anaiak sagar bat jan-du.
brother:DEF:ERG apple one eat:PFV-he.has.it
“My brother ate an apple”

With focus extraposition, this assertion could also be rendered as the one in (6.21).

- (6.21) Anaiak jan-du / sagar bat.
brother:DEF:ERG eat:PFV-he.has.it apple one
“My brother ate an apple.”

In either case, the ergative's referent is the topic. Both of these assertions could be answers to either one of the questions in (6.22), although different speakers might display different preferences depending in each context.

- (6.22) a. What happened?
b. What did your brother do?
c. What did your brother eat?

There are some contexts in which the object would *not* be the focus. Thus, for instance, if there was an optional, focal adverbial *and* the existence of apples was accessible from the context, the adverb would be the focus since it would be more focal, as in (6.23).

- (6.23) Aṇaiak azkar-azkar jan-zuen sagar bat.
 brother:DEF:ERG fast-fast eat:PFV-he.had.it apple one
 “My brother very quickly ate an apple.”

The polarity of the verbal idea could also be the focus, if they were the most focal elements. Such a sentence would look like the one in (6.24), and it could be an answer to either *What did your brother do to the apples?* (verb focus) or *Did your brother eat an apple?* (polarity focus). (As I said, the finite verb is a single constituent in affirmative clauses combining the verbal idea and the polarity.) In other words, the idea of *an apple* would have to be accessible or easy to accommodate from the context.

- (6.24) Aṇaiak jan-zuen sagar bat.
 brother:DEF:ERG eat:PFV-he.had.it apple one
 “My brother ate an apple.” / “My brother did eat the apple.”

The polarity here would be the focus, for instance, if it had been suggested that the speaker’s brother didn’t eat any apples, cf. *My brother did eat an apple*. In this case the idea (if not any referent in particular) *an apple* would be active in the context. This sentence could also be equivalent to the English sentence *My brother already ate an apple*, or *My brother ate an apple already*, if the speaker had been asked to give his brother, say, a pear (the accessible idea of a pear makes the idea of an apple easy to accommodate). The accessibility of the idea of a pear in either case, and the fact that the particular entity this phrase refers to is not important, justifies its diminished focality and the placement of the object that codes it in the assertion-tail area.

If the verbal idea would the focus in (6.24) if that idea was more focal than the idea expressed by the object (or the polarity). In such a case, the object’s idea would have to be quite accessible or easy to accommodate. If the object’s idea is active and quite

specific (and thus definite, cf. *sagarra* “the apple”), this makes the verb automatically more focal than the object, unless, of course, the object is contrastive. (In fact, if an active object is in focus position, it is typically interpreted as contrastive.) In such cases the object, if active, would most likely be elided, i.e. coded only on the verb. But if it is not active, but somewhat less accessible and predictable, it could be overtly coded in postverbal position (unaccented). In some Basque dialects, when the verb is the information focus (not contrastive or emphatic focus), a special construction may be used, which makes use of the ‘dummy’ verb *egin* “do; make”, as in (6.25).

- (6.25) *Anaiak jan egin-zuen (sagarra).*
 brother:DEF:ERG eat:PFV make/do:PFV-he.had.it (apple:DEF)
 “My brother ate it (the apple).”

This construction allows the non-finite verb (representing the verbal idea) to be placed before the verb-complex, i.e. more ‘obviously’ in focus position. We can see another example of this *egin* construction in (6.26) below, taken from the written corpus. With or without *egin* this is an example of verb focus. The context tells us that Mo, a cow, the protagonist, is up in the mountains, it has started to snow, and she can’t find her way home.

- (6.26) *Behi euskaldun baten memoriak*, p. 22
 eta elurrak ezabatu **egin**-ditu bide-bidatzurrak,
 and snow:DEF:ERG erase:PFV make/do:PFV-it.has.them path-shortcuts:DEF
 “and the snow has erased all of the mountain paths”

Here, both *the snow* (the ergative argument and topic), and the need for finding a way home are very accessible ideas. The specific referent “the mountain paths”, the object’s

referent, here has not been mentioned before, but it is perfectly easy to accommodate in this context. The most focal element is obviously the verb, which is thus the focus.

Now the major question before us is the following: How accessible does the referent of an object have to be so that it will be less focal than the verb under normal circumstances, that is, in cases in which the object's idea is not contrastive and the speaker is not adding extrinsic focality to the verb to emphasize it for some reason?

In other words, the extreme cases here are quite clear: if the object's referent is new, informative, unexpected, surprising, newsworthy (i.e. focal), it will be the focus by default; and if the object's referent is active (very accessible), the verb will be the focus of the assertion by default (as long as there is no other element that is more focal). However, what happens with those cases that fall in between, in which the object's referent is accessible to some degree, or may conceivably be accommodated by the speaker? Which one is chosen as the focus, and thus as the rheme-initial element?

I believe that this is one area which is open to interpretation and variation, and thus one in which different speakers may make different, equally well-motivated choices. Furthermore, the choice here may conceivably be affected by the resulting word order. Verb focus in some instances, and thus SVO order, may be more appealing to some speakers given the fact that that would be the mimic the order of such an assertion in Spanish, for instance (whether it had verb focus or object focus, as we saw in Chapter 3).

In the following excerpt in (6.27) we can see an example of what I have just been describing (Hendaia corpus, northern dialect).²¹ In the first assertion of this excerpt, a **girl** is introduced in the characteristic topic presentative of northern dialects (cf. 6.27:149) (cf. Chapter 4). In the next intonation unit in (6.27:150) it is described how she

steals a loaf of **bread**. The bread is in the object role and it is the (preverbal) focus of the assertion. Next we find out that the girl runs into Chaplin, the main protagonist of the story, and she falls on top of him. The next time the loaf of bread is mentioned, is in (6.27:158), where we learn that Chaplin takes the loaf from the girl. The bread is still quite accessible, though presumably not fully active, so it must be referred to with a phrase. In this case the verb is postverbal and not the focus.

(6.27) 93C1B04

- 149 Orduan e= neska bat- neska bat bazen,
 then uh girl one girl one EMPH:she.was
 150 eta= ogi bat artu-zuen, - (*OV, new object*)
 and bread/loaf one take:PFV-she.had.it
 151 lapurtu-zuen,
 steal:PFV-she.had.it
 152 eta= joan-zen,
 and go:PFV-she.was
 153 eta mutikoa etortzen-zen,
 and boy:DEF come:IMPFV he.was
 154 eta= - e neska erori-zen,
 and uh girl:DEF fall:PFV-she.was
 155 erori-zen,
 fall:PFV-she.was
 156 =eta= ..
 and
 157 bere gainera,
 his top:DEF:ALL
 158 eta= mutikoa artu-zuen ogia, (*SVO, given object*)
 and boy:DEF take:PFV-he.had.it bread:DEF

“Then uh there was a girl, / and she took **a loaf of bread**, / she stole it, / and she went away, / and the guy came, / and uh the girl fell down, / she fell down, / and ... / on top of him, / and the guy took **the bread**.”

The reason why the object is not preverbal and focus in (6.27:158) must have to do with the fact that it is so accessible, and thus less focal than the verb. On the other hand, this referent (the bread) wasn't active and it could conceivably have been the focus of the

assertion in this context, and thus preverbal, without it being interpreted as contrastive, cf. *eta mutikoa ogia artu zuen*).

In other words, either choice is well motivated: (1) the object is accessible and thus less focal than when it was new, but (2) the object is not very accessible, and certainly not active, since other things have happened since it was first mentioned. It is up to the speaker to make a choice as to how focal this idea is in the hearer's mind and code the information structure of the assertion accordingly. And it is conceivable that the influence of Romance word order could play a role in this choice, as made by different speakers. Neither choice, could be said to be wrong (ungrammatical) or odd (pragmatically inappropriate).

6.5.2 Basque 'basic word order' revisited

The Basque Language Academy, Euskaltzaindia, in their study of word order, tackles the question of basic word order in *neutral, unmarked situations* and recognizes the difficulties involved (Euskaltzaindia 1987a). By neutral or unmarked situations they seem to imply, though they do not explicitly say it, out-of-the blue or non-continuous contexts. This is presumably because those are the easiest contexts to think of in elicitation or introspection contexts, even though they aren't the most common situations in actual language use.

Euskaltzaindia presents the SOV and SVO examples in (6.28a&b) and asks which sentence is more unmarked or basic. Neither one of these examples involves extraposition or complex intonational structure. Supposedly, in both of them there is a single major

accent mark on the rheme initial element, plus a more or less noticeable accent on the on the ergative topic phrase.

- (6.28) a. Aitak / sagarra jan-du (S/OV)
father:DEF:ERG apple:DEF eat:PFV-he.has.it
“Father ate an/the apple.”
- b. Aitak / jan-du sagarra (S/VO)
father:DEF:ERG eat:PFV-he.has.it apple:DEF
“Father ate the the/?an apple”

In the words of the Academy, “choosing between these two choices is not a bit easy” (Euskaltzaindia 1987a:39).²² Still, the Academy feels that it has to choose between the two orders and, with some hedging and hesitation, they choose the SOV clause as ‘basic’, without providing any reasoning for their choice other than (a) the traditional intuition of grammarians; and (b) statistics from written texts (presumably referring to de Rijk 1969).²³ It is interesting that two pages later in this work we read that the basicness of SOV is supposedly only valid for conversational style/genre and not for narrative (Euskaltzaindia 1987a:41).

As we saw in the preceding section, which order is more ‘natural’ in this or any other case depends on the focality (assertive salience) of the two elements in the rheme *in context*. Of course, we are given no context for these assertions and thus we are left to devise our own. And it seems that contexts for both sentences are about equally easy to construct with no helping clues.

The fact that the Academy chose for its example an object which is by default definite (*sagarra* “the apple”), though in some contexts can be interpreted as unidentifiable/generic (“an apple”), of course has much to do with the fact that (6.28b), with SVO order, is about as natural out of context as (6.28b). If the object had been truly

indefinite (*sagar bat* “an/one apple”), as in (6.24) above, the SVO sentence would have been much more ‘marked’ (out of context). That is because, as we saw, for the verb or the polarity to be more focal than an unidentifiable referent, and thus the focus, they would have to be contrastive.

Notice that if the verbs in question here had been synthetic, the clause with SVO order would have been morphologically ‘marked’, since it would necessarily display the emphatic/verb-focus *ba* particle which codes the polarity as the focus, as we can see schematized in (6.29b).

- (6.29) Synthetic verbs
a. (A) [O V_f]
b. (A) [*ba*-V_f O]

When the verb is analytic, however, as I already mentioned, there is no overt morphological marker involved in either the (S)VO and (S)OV versions.

These considerations show, I believe, the dangers of speaking of a decontextualized ‘basic order’ for a language. Every ordering configuration is pegged to a set of contexts given the fact that word order codes pragmatic roles and not grammatical roles. The concern with basic word order seems to make little sense in a language which uses word order for pragmatic purposes which are dependent on and connected to the contexts in which utterances are made.

6.5.3 Unmarked VO order: Polarity focus and emphatic assertions

A number of Basque linguists have argued that in narrative, as opposed to conversation, there is a tendency to place the verb in focus (rheme-initial) position, or, in

other words, “to move all the verb’s complements after the verb” (Villasante 1980:25).²⁴ The Academy makes the same point, and specifies that this is what we find in ‘story line’, intermediate clauses: “when we’re telling something, when a succession of events are given, often, the verb is placed at the beginning of the clause (rather, at the beginning of the rheme/comment), without using the (focalizing) verb *egin*, except maybe at the end” (Euskaltzaindia 1987a:33-34, see below for further discussion).²⁵ Unfortunately, these authors provide very few examples and the same examples get repeated over and over, without there being any actual detailed studies or discussion made of this phenomenon.

An obvious reason why VO order might be more common in narrative than in decontextualized sentences is that in narrative the object’s referent is much more likely to be accessible from the context, something which will diminish the focality of the object and increase the chances that the verb will be the focus and thus precede the object. This doesn’t need to be the only reason however, as we can gather from the only example the Academy provides in their study of Basque word order which doesn’t deal with decontextualized assertions (Euskaltzaindia 1987a).

This excerpt, which the Academy presents as an example of VX order being unmarked in narrative (without any added commentary), is from Agirre’s novel *Garoa* (1966/1912). The chapter of the novel that this excerpt is taken from, is about country fairs and betting events that take place in such occasions and how addicted Basque farmers are to such betting events. The excerpt in question describes the succession of events that take place in such situations: they farmers get together here or there, they chose in favor of this or the other for this or the other reason, etc., etc. Interestingly, in all

these clauses the verb is in focus position (rheme-initial) and the complements always follow the verb, even though they are not text-given.

The explanation for this seemingly exceptional phenomenon is not so surprising, however. I believe that the reason why in these cases the verb is the most focal element in the clause and thus is in focus position is that (1) the complements, even though they are new, they are also non-referential, non-specific, not 'important', and fairly easy to accommodate; and (2) these seem to be somewhat emphatic assertions, and in the absence of a truly salient object emphatic assertions have the polarity as their focus, and thus the finite verb in focus position (much like in imperatives, for instance, as we will see in Chapter 7). In other words, these are not exactly typical continuous narrative clauses, and thus they are a poor example of what they were supposed to be an example of: unmarked (S)VX order in narrative.

I have only found a couple of examples of this type of emphatic clause-chain construction in the two novels from which the written corpus is taken, both of which, unlike the example in Agirre's novel, also display thematic continuity (they describe a sequence of events sharing a topic, and not just a list of events). The first example is the one in (6.30g&h), with a three clause chain. In the last two clauses the verb *zeharkatu* "cross" is repeated. Notice, however, that the last clause in the excerpt, the one following these three, displays XV order and not VX order as the other three.

(6.30) *Behi euskaldun baten memoriak*, p. 39

- a. Hartu-nuen dezepzioarekin
take:PFV-I.had.it:REL disappointment:DEF:COM
- b. itsutu egin-nintzen,
blind:PFV make:PFV-I.was
- c. eta ahal bezala,
and able as
- d. estrapozoka,
stumble:ADV
- e. hemen zinbuluka eroriz
here stumble:ADV fall:PFV:ADV
- f. eta han okerka neure burua altxaz,
and there turn:ADV my:EMPH head:DEF raise:ADV
- g. **aldendu egin-nintzen** halako kolpea hartutako lekutik.
separate:PFV make:PFV-I.was thus:GEN blow:DEF take:PFV:ADV:GEN
place:DEF:ABL
- h. **Zeharkatu-nituen** belazea eta errota zaharraren ataria,
cross:PFV I.had.them grassy.field:DEF and mill old:DEF:GEN gate:DEF
- i. **zeharkatu-nuen** erreka belaunetaraino bustiz,
cross:PFV-I.had.it stream:DEF knees:FIN-ALL wet:ADV
- j. eta beste aldeko baso itxira **jo-nuen**.
and other side:DEF:GEN forest dense:DEF:ALL hit:PFV-I.had.it

“(a) With that disappointment, / (b) I became blinded, / (c) and as I could, / (d) stumbling, / (e) now stumbling and falling, / (f) now raising my head looking back, / (g) I **took off** from the place where I received the blow. / (h) I **crossed** the grassy field and the gate of the old mill, / (i) I **crossed** the stream/creek getting wet up to my knees, / (j) and I went straight for dense forest on the other side.”

Notice that also in the English translation these sentences would be quite natural with verb focus, i.e. with the rheme's accent falling on the verb rather than on the complement.

Another example of this 'emphatic', action chain construction, which, as I said, is not particularly common in the written corpus, can be seen in the two first clauses in (6.31a&b). Here too the chain is also a thematic chain (with a shared topic).

(6.31) *Behi euskaldun baten memoriak*, p. 76

- a. **Makurtu-nuen** burua,
bend:PFV-I.had.it head:DEF
- b. **aurreratu-nituen** adarrak,
put.forward:PFV-I.had.them horns:DEF
- c. eta besoa **hautsi-nion** bi lekutatik.
and arm:DEF break:PFV-I.had.it.to.him two place:ABL

“(a) I lowered my head, / (b) I put forward my horns, / (c) and I broke his arm in two places.”

Notice that although the referents of the postposed objects (*the head, the horns*) are not active, they are very easy to accommodate (the speaker is, after all, a cow). Finally, here too we see that the last clause of the action chain doesn't have a rheme-initial verb, since the object's referent in this case *is* salient and focal, although the verb is focal as well. Here too, the English versions could have an emphatic focus accent on the verb rather than on the complement, as would normally be the case in a purely, non-emphatic, descriptive narrative sequence.

Other than that, clauses with rheme-initial verbs and postverbal complements are quite uncommon in this novel, and this seems to be rather typical of standard Basque, which follows a rather impeccable logic of assuming that unless there is a significant reason to make the verb or the polarity the focus of the assertion, such as a contrast or great emphasis, a complement should fill this role. After all, given (accessible) complements are for the most part not overtly expressed, and if they are, they are contrastive and thus foci.

Let us look now at three examples of assertions with polarity focus, and thus with assertion-initial verb and postverbal complements. They are from the same novel, although though not from the chapter which was included in the written corpus. The first

one is the one in (6.32) below. Here there are two emphatic clauses, which express two parallel states (*I knew X* and *I remembered Y*), rather than an action chain.

(6.32) *Behi euskaldun baten memoriak*, p. 100

- a. Zer esan nahi zuen arrozarenak?
what say:PFV desire it.had.it rice:DEF:GEN:DEF:ERG
- b. **Ezagutzen-nuen** bere iritzia,
know:IMPFV-I.had.it her opinion/point.of.view:DEF
- c. alegia <nire kabuz pentsatzen> ikasi beharra neukala,
namely my initiative:INSTR think:IMPFV learn:PFV need:DEF I.had.it:COMPL
- d. eta **gogorutzen-zitzaidan** otsoen arazo hartan berak hartutako joera
axolagabea,
and it.reminded.me wolves:DEF:GEN problem that:LOC she:ERG
take:PFV:ADV:GEN attitude carefree:DEF
- e. baina aldi hartan lagundu egin behar zidan.
but time that:LOC help:PFV do:PFV need she.had.it.to.me

“(a) What did the thing with the rice mean? / (b) I **knew** what it [my conscience] was thinking, / (c) namely that I had to learn to think for myself, / (d) and I **remembered** the carefree attitude she had adopted when we ran into the wolves, / (e) **but** at that time she just had to help me.”

In these two clauses the complements are not particularly accessible, but they are easy to accommodate. Notice that these two assertions act somewhat as *concessives* with respect to the assertion which follows in the (6.32e), cf. *I knew X* and *I remembered Y*, but ...). In English too, in these clauses the verb would most naturally be accented, thus showing that the verb, or rather the polarity part of the finite verb in this case, is the focus (cf. *I did know X*, and *I did remember Y*, but ...).

The following example in (6.33b) is from a dialogue part of the novel and not from a narrative. However, here too the polarity is emphatic and thus the focus (*I have seen that there is rice*).

(6.33) *Behi euskaldun baten memoriak*, p. 102

a. —*Ikusi al duzu hemen dagoena?*—*galdetu-nion.*

see:PFV QN you.have.it here it.is:REL:DEF ask:PFV-I.had.it.to.her

b. —*Bai, ikusi-dut* *arroza dagoela. Baina lasai.*

yes see:PFV-I.have.it rice:DEF it.is:COMPL but easy

“(a) ‘Have you seen what there is here?, I asked him. / (b) Yes, I **did** see that there is rice. But don’t worry’.”

The postverbal complement clause is accommodated since the interlocutor is obviously familiar with the information (i.e. she knew what there was in the manger). Notice that this completive clause is not accented (the focus), nor extraposed, but simply placed in the assertion-tail.

Finally, let us look at another example from the dialogue in this novel, where emphatic assertions seem to be significantly more numerous. The assertion in (6.34c) is obviously emphatic, as evidenced by the fact that the verb is in focus position even though the object’s idea is not text-given, although it can be easily accommodated in the context.

(6.34) *Behi euskaldun baten memoriak*, p. 151

a. *Goazen, bai—ametitu-nion nik—.*

let’s.go yes agree:PFV-I.had.it.to.her I:ERG

b. *Eta joan-gaitezen azkar.*

and go:PFV-we.may.be:SUBORD fast

c. *Hartuko-dugu deskantsu <salbo jarritakoan>.*

take:FUT we.have.it rest <safe put:PFV:ADV:GEN:DEF:LOC>

“‘Let’s go, yes,’ I agreed with her. / ‘And let’s go quickly. / We will **take** a rest when we’re safe.’”

In this sentence the non-referential object *deskantsu* “rest” is placed in postverbal position, even though, as I said earlier such nominals are the least likely ones to be displaced from preverbal position. It is also not given, although it can be accommodated (or inferred) from the context (taking a break is the opposite of leaving quickly).

It is interesting that in this and other examples of emphatic assertions in which the verb is ‘fronted’ in Basque, in the corresponding, analogous Spanish sentences the adverbial particle *ya* “already” would be used immediately before the verb. Much like the corresponding English adverbial *already* this particle can be used in a temporal sense in situations which invariably gives emphatic value to the polarity, e.g. *Mi hermano ya ha comido una manzana*, “My brother **already** ate an apple”, cf. example (6.24) on page 699 above.

But the use of this particle in Spanish has been extended to other non-temporal cases in which the polarity (and thus the assertion) is emphatic, and thus the finite verb is the focus of the assertion. An example of this would be the Spanish translation of sentence in (6.34c) above: *Yá tomaremos un descanso al ponernos a salvo*, Lit. “We will already take a rest when we’re safe.”²⁶ In other words, this temporal adverbial seems to have become a polarity focus marker in many contexts in Spanish. (Another polarity focus marker used in a somewhat different contrastive context is *sí* “yes”, e.g. *Sí tomaremos un descanso* “We will take a rest.”) This is why it has sometimes been noticed that there is a certain correspondence or parallelism between the Spanish particle *ya* and the Basque (emphatic) verb and polarity focus particle *ba*, which is used only with synthetic verbs.

Before ending this section I would like to mention what seems to be an example of emphatic polarity focus in a Basque proverb mentioned by Villasante (1980). The ordering characteristics of proverbs and popular sayings are quite interesting in and of themselves. Basque proverbs typically have a very noticeable bipartite, topic-comment, structure, with the comment often sometimes being a non-verbal predicate, without a

verb, and the order is always topic-comment. Villasante (1980:23) mentions the popular saying in (6.35), which has SVO order, and wonders about it and why it doesn't conform to Altube's verb-final rule.

- (6.35) Egixak erretzen-du begixa
truth:DEF:ERG burn:IMPFV-it.has.it eye:DEF
“The truth burns the eye” (cf. “The truth hurts.”)

He wonders if the reasons for the order have to do with rhyme or with giving the saying a *lapidary air* that makes it easier to remember.

It seems to me that the reason for the VO order in this example is related to the fact that the verb (*burns*) is very focal and to the fact that the generic object (*the eye*) is easy to accommodate. The alternative OV order would also be possible but in that case it seems that the object might be interpreted as contrastive (cf. *The truth burns the eye, as opposed to something else*). In addition, emphasis may also be involved in the choice of the verb as the focus.

Mitxelena, in his running commentary to Villasante's (1980) book, compares this aphorism to another one, the one in (6.36), which, as he mentions, also has a clear topic-comment structure, although this one has SOV order (Villasante 1980:23fn2).

- (6.36) Euskaldunak aski harroak izaten-dira.
Basque.people:DEF:PL rather arrogant:DEF:PL be:IMPFV-they.are
“Basque people are rather arrogant.”

Mitxelena doesn't tell us why this assertion might prefer OV order. It seems to me that here the non-verbal predicate (*rather arrogant*) is clearly more salient (focal) than the copula verb, and thus, not surprisingly, must be chosen as the focus. The only context in which the verb could be in focus position would be if the idea expressed by the object

was active and the (affirmative) polarity was contrasted with its opposite value, e.g. if someone had suggested the contrary assertion (cf. English *Basques are rather arrogant*). The focus, of course, could also be extraposed, but that construction doesn't seem to be used in proverbs (cf. *Euskaldunak izaten-dira, aski harroak*).

6.6 THE POSITION OF ABSOLUTIVE COMPLEMENTS IN THE BASQUE CORPUS

6.6.1 Rhemes with no complements or one complement

As I mentioned earlier, the most common types of main clause declarative assertions in narrative are those in which there is a verb and at most one other element in the rheme (excluding any overt topics). Rhemes consisting of a single verb in main clauses are about 15% of the total. Rhemes with a verb plus one complement in the rheme are over 50% of the total. The rest of the clauses are for the most part those which include at least one additional non-focus (unaccented) element in the assertion tail, such as postposed settings or given complements.

To begin with, the total number of affirmative main statements in which the **rheme** consists of a **single verb** in the spoken corpus is about 15% (or about 400 clauses). This includes 11.3% of all affirmative main statements which have no overt topic, i.e. clauses consisting only of a verb. It also includes about half of the 5.5% of assertions which have an overt absolute (subject) topic argument (SV clauses), and about half of the 1.7% of those which have an overt ergative argument (AV clauses) (in the other half the subject is the focus and thus part of the rheme).

The total number of clauses in the spoken corpus with two elements in the rheme, i.e. the verb plus a complement (argument or adjunct), is much larger. Of all main affirmative statements, 13.5% have only a verb and an object (OV or VO), 19.8% have a verb and a non-object complement (XV or VX). In addition, about half of the clauses with SV (5.5% of all) order and the vast majority of those with VS (6.1%) and VA (0.6%) order are two-element rhemes, since the 'subject' is not the topic but the focus. All SXV (2.1%) and AOV (0.5%) clauses have two part rhemes. Also at least half of the SVX (3.4%) and AVO (1.7%) clauses have a topic subject and thus a two-element rhemes. To this we must add clauses with a finite or non-finite (and no other) complement, with or without an overt topic, which are approximately 8% of all clauses (e.g. *XIaV* and *XIzenV* respectively).²⁷ In all we can see that the majority of clauses, over 50%, consist of a verb plus a single complement. In the following sections I will be looking in some detail at the relative order of the verb and its single (non-clausal) complement in the spoken and written corpora.

6.6.2 (A)OV vs. (A)VO order in the Written Basque Corpus

Let us look first at clauses with have an overt subject and an overt object, and no other complements. There are 24 such main affirmative sentences in the written corpus. Half of them have the order AOV and half AVO (A= ergative argument). However, a closer examination reveals that the referents of 8 of 12 of the AVO clauses are not topics, but rather foci and thus part of the rheme and rheme initial, as we can see in Table 6-3 (the sentences can be seen in Appendices 6.1 and 6.2). In these cases the objects are assertion-tail complements.²⁸

Word order	#	%	Comments
A[OV]	12	50.0%	O focus
[AV]O]	8	33.3%	A focus
A[VO]	4	16.7%	V focus: 3 <i>egin</i> , 1 <i>ba-</i>
Total	24	100.0%	

Table 6-3: AOV vs. AVO affirmative statements in the written corpus.

In these cases the postposing of the object is a natural consequence of the fact that the subject is the focus of the assertion and thus in focus position. Also the referents of the objects in these cases must be very accessible or easy to accommodate. Of these subject focus examples, 6 have obviously contrastive referents and 2 have non-contrastive ones. We see an example of this type of clause in (6.37).

- (6.37) *Behi euskaldun baten memoriak*, p. 14
 "... berak zuzenduko-ditu zure pausoak."
 she:ERG straighten:FUT she.has.them your steps:DEF
 "she will straighten your steps."

In the other 4 AVO clauses, in which the ergative referent is indeed a topic, the verbal idea or the polarity is *overtly* marked as being the focus. In 3 cases the verb is analytic and it has an *egin* new-focus marker, cf. the examples in (6.26) above. The other case, shown here in (6.38) below, has a synthetic verb and a focus *ba* marker. (Notice the *ere* "also" additive topicalizer, cf. Chapter 4.)

- (6.38) *Kuba triste dago*, p. 29
 "Nik ere **badut** bizikleta"
 I:ERG also EMPH:I.have.it bicycle:DEF
 "I have a bicycle too." (Lit. "Me too, I have a bicycle.")

Here the idea of a bicycle, and of the speaker's interlocutor having one, is text-given, though not this particular referent or token of the bicycle category. Thus placing the

object in postverbal (assertion-tail) position is justified. (Notice that the object is coded as a so-called ‘definite’ noun, even though strictly speaking the referent is unidentifiable, cf. Chapter 2.) In addition, it is clear that this sentence (from a conversation, not from narrative) is emphatic and that the polarity is the focus.

If we look now at the larger pool of OV and VO clauses *without* an overt ergative argument, we find that OV clauses are much more common than VO clauses, as we can see in Table 6-4 below. We can attribute this to the fact that in these sentences the ergative argument doesn’t ‘compete’ with the absolutive one for focus position, as was the case with AVO clauses.

Word order	#	%	Comments
[OV]	58	82.9%	O focus
[VO]	12	17.1%	V/Polarity focus: 2 <i>egin</i> , 1 <i>ba-</i>
Total	70	100.0%	

Table 6-4: OV vs. VO AVO affirmative statements in the written corpus.

Here too we find that all the VO clauses are clearly well motivated (cf. Appendix 6.3). The vast majority display emphatic verb or polarity focus: 2 with *egin*, e.g. (6.39), 1 with *ba-*, (6.40), and 9 without any formal marking. In both of these examples the ideas expressed by the object are given and the assertions seem to be emphatic. The first one has verb focus (cf. the *egin* construction) and the second one seemingly polarity focus (cf. the affirmative polarity *ba-* prefix).

- (6.39) *Kuba triste dago*, p. 33
 utzi **egin**-nuen artisau lana
 leave:PFV do:PFV-I.had.it artisan work:DEF
 “I abandoned my job as an artisan”

- (6.40) *Kuba triste dago*, p. 23
 Hemen ere badituzte neskalgunak
 here too EMPH:they.have.them girl:friends:DEF
 “They do have girlfriends here too.”

Of the 9 ‘unmarked’ sentences, two form are action chain similar to the one we saw in (6.31a&b) above. A further two have the auxiliary *behar* “must” and seem to have emphatic polarity focus, as we can seen in (6.41). Here the object has a new referent (*my daughter*), but seemingly it is easy enough to accommodate to appear in assertion tail position.

- (6.41) *Kuba triste dago*, p. 33
 Ezagutu behar duzu nire alaba,
 know/meet:PFV must you.have.it my daughter:DEF
 “You must meet my daughter, ...”

Two of the remaining five clauses, even though they look like main clauses, are actually a rare special construction which functions as setting clauses, such as the one in (6.42)

- (6.42) *Behi euskaldun baten memoriak*, p. 15
 Deituko nion Aingeru Goardako
 call:FUT I.had.it.to.her Angel Guard:GEN
 “Was/were I to call her Guardian Angel.”
 (~ “Whether I called her Guardian Angel or ...”)

This sentence is reminiscent of the setting-like main clauses that we saw in Chapter 5, which also displayed emphatic form. Notice that this sentence could be translated by means of a ‘dependent’ clause in English, or with a clause displaying subject inversion.

The rest of these cases all have unmarked verbs in focus position and they are all more or less clearly exclamatory-like and emphatic. Two of them form an action chain of the type we saw in (6.30g&h) and (6.31a&b) above. The example in (6.43) is one of the remaining clauses and is clearly emphatic.

- (6.43) *Behi euskaldun baten memoriak*, p. 18
 “Esango-dizut hitz bat,”
 say:FUT-I.have.it.to.you word one
 “I will tell you something, ...” (~ “Let me tell you something.”)

Notice that the object’s referent is new, but it is not salient enough to prevent the polarity from taking over the role of focus.

As we can see (A)OV order is predominant in the written corpus. (A)VO order is found when the ergative argument is the focus of the assertion or when the verb or the polarity are the foci, which happens only in a minority of cases. Notice that we haven’t seen any cases of focus extraposition. Let us look now at similar orders in the Spoken Basque Corpus.

6.6.3 (A)OV vs. (A)VO order in the Spoken Basque Corpus

If we look now at the (A)OV vs. (A)VO affirmative main clauses (declarative assertions) in the spoken corpus we get a rather different picture. Looking first at those with an overt ergative argument, we see that $\frac{3}{4}$ are AVO, as opposed to $\frac{1}{2}$ in the written corpus, cf. Table 6-5.

Word order	Number	Percentage	Comments
AOV	14	23.3%	12 non-ref O; 3 new-O
AVO	46	76.7%	
Total	60	100.0%	

Table 6-5: AOV vs. AVO clauses in the spoken corpus in main affirmative clauses.

Even more interesting is finding that of the 14 AOV clauses, 11 have non-referential, non-specific or abstract nouns, as in *have fear* (= be afraid), *make an escape* (= take off), *blow the whistle*, *find one’s way*, *give a turn* (= turn around), *say something-or-other*,

make a stop (= “stop”). The other 3 have new referents (all three with the numeral *bat* “one”). (The complete list of AOV clauses can be seen in Appendix 6.4.) The vast majority of clauses with referential objects, whether new or given, have AVO order. (The complete list of AVO clauses can be seen in Appendix 6.5.) Those in which the object is the focus, it is clearly extraposed. Those in which the object is not the focus, it is unaccented and in assertion-tail position. This is in stark contrast with the situation in the written corpus where objects were only postverbal if there was a good reason to make the verb or the polarity the focus, typically in contrastive situations.

If we look at transitive clauses with overt objects but covert subjects, we find very similar ratios of preverbal to postverbal objects, but very different from the situation in the written corpus, as we can see in Table 6-6. Here the percentage of VO is lesser than the percentage of AVO, again for the same reason that AVO clauses sometimes display ergative argument focus.

Word order	Number	Percentage
OV	137	36.9%
VO	234	63.1%
Total	371	100.0%

Table 6-6: OV vs. VO clauses in the spoken corpus in main affirmative clauses.

Still, here too the percentage of VO clauses with respect to OV clauses, almost $\frac{2}{3}$ vs. $\frac{1}{3}$, or 4 times greater than the ration of approximately $\frac{1}{5}$ vs. $\frac{4}{5}$ in the written corpus.

6.6.4 OV vs. VO in the *Lur* subcorpus

Now I would like to compare in some detail a sample of OV and VO clauses in the spoken corpus.²⁹ I will choose the *Lur* subcorpus for this purpose, which contains (a manageable set of) 25 OV and 25 VO clauses (more later on the different percentages in different subcorpora). In Table 6-7 we can see some of the pragmatic characteristics of the postverbal objects in the VO clauses in the *Lur* subcorpus (cf. Appendix 6.6 for a full listing).

	Given	New	TOTAL
Extraposited focus	3	8	11
Non focus	10	4	14
TOTAL	13	12	25

Table 6-7: VO clauses in the spoken *Lur* subcorpus.

As we can see, about half of these objects have new referents and half have given referents. Four of the twelve discourse-new referents can be accommodated from the context, however. Eleven of the objects seem to be extraposited focus objects, about $\frac{2}{3}$ of which are new, whereas the remaining 14 are non-focus assertion-tail complements, about $\frac{2}{3}$ of which are given (i.e. the verb/polarity is the focus). Two have a *ba* focus marker on the (synthetic) verb, 1 with a given object and one with a new object. Seven clauses display hesitation after the verb (3 with a given object and 4 with a new object).

In 11 of the 25 VO clauses, the object is an (accented) extraposited focus. In at three (and perhaps more) of these clauses the focus object is clearly in an intonation unit of its own. At least seven show signs of hesitation in the verbalization of the object. In

the example in (6.44) the intonational break and pause were noticeable enough to suggest that the object in its own intonation unit.

- (6.44) 93C3A10
152 ematen-diote,
 give:IMPFV-they.have.it.to.him
153 pixka bat e= faboreko e- tratamentu bat bezela/
 bit one uh favor:GEN uh treatment one as
 “they give him, / like special treatment or something”

The example in (6.45) also displays the intonational contour associated with focus extraposition plus the hesitation particle *ba* “well” between the two elements. (This assertion has been coded as VO even though it has a locative setting in clause-initial position.)

- (6.45) 93C3A10
36 Eta kartzelan dauka/ ba lagun bat bastante asto- astakertena.
 and jail:DEF:LOC he.has.him well companion one rather stu- stupid:DEF
 “And in jail he has well a companion who’s rather stupid.”

The following example in (6.46) shows clear signs of hesitation after the verb and before the object, which may be the ‘reason’ for its being extraposed rather than preverbal.

- (6.46) 93C3A12
48 ba= artzen-du s= droga=ren bat,
 well take:IMPFV-he.has.it s= drug=:DEF:GEN one
 “we=ll he takes ?= some kind of drug”

Two of the VO assertions are marked as having polarity focus (with *ba*), and at least one has a verb focus with an unimportant object referent. None of the verbs has the *egin* new-focus marker. Most of the VO clauses, 14/25, seem to have verb or polarity focus, whether the object is given, as in (6.47), or new and accommodated, as in (6.48) (in this last example, the cell keys are a salient part of the *jail* semantic ‘frame’).

(6.47) 93C3A09 (verb focus with given object)
74 Eta= ematen-dio txapela,
and give:IMPFV-they.have.it.to.him hat:DEF
“And= they give him his hat (back).”

(6.48) 93C3A10 (verb focus with accommodated object)
149 ta artzen-ditu giltzak,
and take:IMPFV-he.has.them keys:DEF
“and he takes the keys.”

In neither of these cases is the referent of the object so accessible and thus non-focal so as to cause the verb to be the focus by default. In other words, in either case the object could have been in focus position without being interpreted as (narrowly) contrastive.

Sometimes, however, the object’s referent is new and rather unimportant and thus easy to accommodate. These assertions have verb focus, although they can also be interpreted as mildly emphatic. That seems to be what is happening in (6.49).

(6.49) 93C3A10
136 ta artzen-du ate bat,
and take:IMPFV-he.has.it door one
“and he grabs a door, ...” (and hits the bad guys against it)

Note that this is the first clause in an action chain, a context in which emphatic clauses are often found, as we have seen.

In Table 6-8 below I have summarized the pragmatic properties and other characteristics of preverbal objects in **OV clauses in the *Lur* subcorpus** (see Appendix 6.7 for a full listing of the 25 clauses).

Object's referent	Other characteristics	No.
Given, 7 ea.	Contrastive	4
	Other	3
New, 17 ea.	Non-specific	7
	Accommodated	2
	Accommodated; Non-specific	1
	Accommodated?	3
	Non-referential	2
	Other	2
Other	O=Topic; <i>egin</i> verb-focus	1

Table 6-8: Pragmatic properties of the referents of preverbal objects (OV clauses) in the *Lur* subcorpus of the Spoken Basque Corpus (cf. Appendix 6.7).

Here 70% of the object referents are new, even though at least 3 and perhaps 6 can be accommodated from the context, and are clearly more focal than the verb and thus are the foci of the assertions, e.g. (6.50).

(6.50) 93C3A07

104 zeoze sari bat eman-dixobe. -
 what-or-what prize one give:PFV-they.have.it.to.him
 "they gave him some kind of prize."

Of the 7 given referents, 4 are clearly contrastive foci, but none of the new ones are. The object in (6.51) for instance is contrastive (they take away Chaplin, as opposed to the girl who stole the bread).

(6.51) 93C3A12

120 Txarlot eramaten-dute. - preso.
 Chaplin take.away:IMPFV-they.have.him prisoner
 "They take Chaplin. Prisoner."

As I said, a full 7 of the new referents are non-specific (6 are new), such as the one in (6.50) above, and 2 more are non-referential, as the one in (6.52).

(6.52) 93C3A09

65 =txistu itten-dioe, -

whistle make:IMPFV-they.have.it.to.him

“They whistle to him.” (Lit. “They make whistle at him”)

All these objects have a strong tendency to be preverbal (in focus position), something which Comrie (1984) was also the case for Armenian. The reason why these objects are typically not postverbal is that to be postverbal they would have to be given or easy to accommodate, something which rarely happens with non-referential ideas.

6.6.5 Some generalizations about the position of the object

We can draw some preliminary generalizations from this corpus of VO and OV clauses about object ordering preferences in speech (for this group). First of all **non-referential** phrases and those with **non-specific** referents are almost always preverbal (just like Comrie found to be the case in Armenian). I believe this is due to the fact that givenness or easy accessibility of an idea, in other words a high degree of topicality, is required for an argument to be placed in postverbal (assertion-tail) position, a pragmatic property which is missing in the ideas expressed by these phrases.

Secondly, **new referents** are somewhat more likely to be preverbal (17/29), although when they can be accommodated they may be postverbal (4 ea.). But the majority of the postverbal new referents are extraposed foci, just like the preverbal ones. Given referents are much more likely to be postverbal (13/16) and non-focus (i.e. in assertion-tail position).

Thirdly, sometimes postverbal focus objects are preceded by signs of **hesitation**, such as pauses and syllable lengthenings. This suggests that the object hasn't been fully

verbalized in the speaker's mind before starting the sentence. These objects seem to be extraposed foci. In other words, focus extraposition seems to be motivated by production delays. We could say that the speaker takes advantage of the existence of this construction in such cases. In a rigidly verb-final language in which focus extraposition is not possible (at least for simple complements), the speaker would presumably just wait to start the sentence until the focus phrase was fully verbalized. This type of postponing, of course, is not motivated in writing, since a writer doesn't produce language on line. This goes some ways towards explaining the greater preponderance of focus extraposition in speech than in writing.

Another major generalization has to do with the difference between the spoken corpus and the written corpus. Thus whereas 4/5 of clauses consisting of a verb and an object (with a covert topic) have a preverbal object in the written corpus, at most 1/2 of such clauses have a preverbal object in the spoken corpus. Although genre and style related differences between the corpora may account for some of the differences, it does seem that in postverbal positioning of objects in speech is a great deal more common.

The reasons for this difference seem to boil down to two clear tendencies which are found in speech, namely: (1) a greater propensity to use the focus extraposition; and (2) a greater propensity to make the verb/polarity the focus of the assertion when the object's referent is not very focal, that is, when it is given or new but unimportant or easy to accommodate. This latter tendency often involves what in standard Basque would be interpreted as making the assertion emphatic (giving it verb or polarity focus).

Another major finding, which was already mentioned in Chapter 2, has to do with the differences which exist among speakers and among the different subcorpora of the

Spoken Basque Corpus. In other words, the figures for the relative frequency of VO and OV order in the spoken corpus that we saw in Table 6-6, approximately 2:1 in favor of the former, are an average of fairly heterogeneous populations.

As we have seen, the *Lur* group, which is formed by very fluent adults, has a 50-50 split between VO and OV orders in affirmative statements with only those two elements in the clause. As we can see in Table 6-9, a similar ratio is found for the *Deustu* group, also made up of very fluent adults, all of them around 22 years of age and most of them from towns in which Basque is still a strong language. This group actually has a slightly higher percentage of OV clauses than of VO ones. On the other hand, the two groups made up of young children, with different degrees of fluency, though all of them attending full-time all-Basque schools since about the age of 3, shows a different picture, as we can see.

Subcorpus	OV		VO		Total	Comments
<i>Hendaia</i>	33	36.7%	57	63.3%	90	10 year olds; mixed fluency
<i>Ikasbide</i>	13	11.6%	99	88.4%	112	13 year olds; mixed fluency
<i>Deustu</i>	66	55.5%	53	44.5%	119	adults; very fluent
<i>Lur</i>	25	50.0%	25	50.0%	50	adults; very fluent
TOTAL	137	36.9%	234	63.1%	371	

Table 6-9: OV vs. VO order main affirmative clauses in the different sub-corpora of the Spoken Basque Corpus.

The *Hendaia* group (northern dialect) of 10 year-olds has an average which is very close to the mean, or 2:1 ratio in favor of VO. The *Ikasbide* group of 13 year-olds has an even higher ratio of VO clauses, namely about 9:1. This means that for this group postverbal positioning of objects is almost categorical. Remember that these are all narrative texts

about exactly the same stories, so that we cannot attribute the differences to extraneous differences (half were *Chaplin* stories and half were *Pear* stories for each group).

The reasons for the postverbal positioning of objects in the two groups of younger speakers are the same as those which account for the difference between the spoken and written corpora, namely, (1) a greater tendency to use the optional operation of focus extraposition with new foci; and (2) a greater tendency to use optional verb/polarity focus when the object can be interpreted as being less focal than the verb/polarity.

In other words, whether an object will be preverbal or postverbal in Basque depends in great part on the whether two *optional* 'rules' are used. Very focal objects, and in particular contrastive ones and non-referential ones, still have a strong preference for being preverbal foci for all speakers. Very non-focal objects, such as those with active referents, which are quite predictable and unsurprising, on the other hand, have a very strong tendency to be postverbal. But new focus objects may be optionally extraposed, and relatively accessible ones may be optionally postposed.

While these two operations are more common in speech than in writing for all speakers, they are much more common for some speakers than for others, and the ones for which they are more common are those which have a greater exposure to Romance languages, and in particular to Spanish. This correlation suggests a causal relationship, which will need to be explored in greater detail in further studies. What is significant is that, even if the differences that we observe among speakers and genres are (in part) the result of the influence of a contact language, the mechanism involved in achieving those differences consists in the higher degree of application of optional constructions and principles which already exist in the language and are not, strictly speaking, borrowed.

This type of change by which the application of native constructions is affected by constructions in a contact language which are superficially similar is the phenomenon which I mentioned in Chapter 1 which goes by the name of *convergence*, a type of veiled 'transfer' or borrowing. Although much more detailed studies still need to be made to better understand this phenomenon, the evidence seems to be rather strong in this regard.

The phenomenon of focus extraposition, for instance, is highly constrained in standard Basque, being used in well-motivated circumstances such as when a focus constituent is long and complex. In spoken Basque, however, and in particular for some speakers, focus extraposition is becoming a major way, if not the major way, of expressing new complement foci, though not, typically, contrastive or otherwise emphatic ones, or foci which are non-referential and thus closely bound semantically to the predicate. I believe that, in time, this could lead to this focus position becoming the unmarked position for this type of focus constituent.

Furthermore, the intonational split associated with the extraposition construction, which is already showing signs of being weakened, might in such a situation be totally erased, resulting in the creation of a new, unmarked position for focus constituents, namely *postverbal* position. This restructuring, reanalysis, or reintegration of the main part of the rheme (which typically consists of just a verb) and the extraposed focus would be the final stage in this process. It remains to be determined whether such a reanalysis is already underway for some speakers.

I believe that this is the main mechanism by which a so-called OV language, which typically has only one major focus position (along with possibly a minor one, 'extraposition'), would come to acquire a second major focus position (postverbal) and

thus become what is called a VO language. It seems that acquiring a second, specialized focus position, perhaps under the influence of other languages, is much easier than doing away with one (postverbal) and to analogize all focus types by placing them in a single position, the most salient of all focus positions (preverbal). (It seems that no language makes use of only postverbal position for foci, to the total exclusion of preverbal position.) This would account for the fact that there are so many attested examples of change from “OV” to “VO” order, but hardly any examples of the opposite change.

6.7 THE POSITION OF NON-ABSOLUTIVE COMPLEMENTS IN THE CORPUS

6.7.1 SXV vs. SVX clauses

In the previous section we looked at one of the most common types of transitive clauses in the language. The clauses we looked at in the previous section were composed of a verb, a covert absolute object, and an ergative argument, which could be either a covert topic, or an overt topic or a focus constituent. In this section we will look at equivalent intransitive clauses, that is clauses consisting of a verb, a non-absolute complement, such as an adverbial or postpositional phrase, and an absolute argument which may be either a covert topic, or an overt topic or a focus constituent. In other words, these clauses have either the structure (S)XV or (S)VX.

In the next two sections I will compare in some detail the relative proportions of SXV and SVX clauses in the written and spoken corpora, the properties of the complements, and the differences between these types of sentences and transitive (AOV and AVO) sentences that we saw in the preceding section. I will not discuss clauses of

this type in which the S (absolutive) argument is elided since the differences found between these clauses and those in which the S argument is overt are quite predictable.

As we can see in Table 6-10, in the written corpus the number of SXV clauses is almost twice as high as the number of SVX clauses (2:1 ratio), but the opposite is the case in the spoken corpus (2:3 ratio). (The complete list of sentences can be seen in Appendices 6.9 through 6.12.)

	SXV		SVX		Total
Written Basque Corpus	44	66.7%	22	33.3%	66
Spoken Basque Corpus	58	40.6%	85 ³⁰	59.4%	143

Table 6-10: SXV vs. SVX clauses in the spoken and written corpora in main affirmative clauses.

Thus we can see that here too there is a stronger tendency in the spoken corpus for a complement to be postverbal with respect to the verb (though less than in the case of transitive clauses). In these sentences, as we shall see, the subject could be either the topic or the focus of the assertion. But topics can also be covertly expressed. The difference between speech and writing and between transitive and intransitive clauses is quite noticeable when the subject is a covert topic, as we can see in Table 6-11 below.

	XV		VX		Total
Written Basque Corpus	121	93.1%	9	6.9%	130
Spoken Basque Corpus	234	46.0%	275	54%	509

Table 6-11: XV vs. VX clauses in the spoken and written corpora in main affirmative clauses.

The differences we detect between clauses with overt and covert subjects must be due to the fact that a significant percentage of the overt subjects are not topics but focus subjects, which causes the additional complement to be postverbal. But notice that the

ratios of preverbal to postverbal X depending on whether the subject is overt or not are much more different for the written corpus than for the spoken corpus. In other words, the differences in the position of the X whether the S argument is overt or not, are not as significant in the spoken corpus. This seems to be due to the fact that whereas in writing the major reason why an X complement is postverbal is whether the S is the focus or not, in speech there are other, more influential causes: focus extraposition and given complement postposing.

6.7.2 SXV vs. SVX sentences in the written corpus

The most common type of SXV clause in the written corpus consists of copula/existential clauses with an obviously focal non-verbal predicate (X) in preverbal position: 33 out of 44 fit this description (cf. Appendix 6.9). Of the other 11, the vast majority have copula-like, semantically ‘weak’ or stereotypical verbs, which are easy to accommodate, and the X element is also clearly the significantly more focal than the verb, and thus the focus of the assertion. These verbs are the following: transitive *izan* (transitive copula), *gertatu* “happen”, *ibili* “go, be”, *galdu* “get lost”, *eman* “seem”, *bizi* “live”, *lerratu* “slide”, *joan* “go” (with manner X complement), *ari* “be engaged” (with manner X complement), *geratu* “remain”, and *hizketatu* “speak”.

The SVX clauses are also in large part copula clauses: 11 of the 22 have a copula/existential verb (see Appendix 6.10). Most of the rest also have semantically weak or stereotypical verbs: *etorri* “come”, *jarri* “put, place, set; become”, *ibili* “go; be”, *joan* “go”, *bizi* “live”, *somatu* “suspect; suppose”, *heldu* “arrive, reach”, and *ezkondu* “get married”. In the vast majority of these SVX clauses in the written corpus the S is not the

topic, but the focus, and in the rest the verb or the polarity is obviously the focus, as we can see in Table 6-12.

SVX type	Subtype	No.
Focus S	Thetic	11
	Dative topic	6
Topic S, Focus V	<i>egin</i> verb focus	1
	<i>ba</i> verb focus	1
	Very accessible X	1
Extraposd X	Extraposd focus X	2
TOTAL		22

Table 6-12: Main, affirmative SVX statements in the Written Basque Corpus.

In 11 of the 22 sentences the S is the focus and there is no topic. Three of them are ‘existentials’ or ‘presentatives’ with new referents, e.g. (6.53). The referent of the postverbal locative (*the street*) is accessible in this context.

- (6.53) *Kuba triste dago*, p. 18 (existential, focus S)
 Jende dezente da kolean;
 people enough/quite it.is street:DEF:LOC
 “There’s quite a lot of people in the street”

In 6 of them the subject is also the focus and has a new referent but here there is a topic, namely a covert dative argument. The sentence in (6.54), introduces a new referent in subject position, but the topic is the referent of the dative argument, namely the speaker. And the postverbal complement’s idea (*on the side*) is, of course, quite accessible.

- (6.54) *Kuba triste dago*, p. 18 (existential, dative-topic, focus S)
 Hamahiru-hamalau urteko mutiko beltz bizikletadun bat jarri-zait parean.
 13-14 year:DEF:GEN boy black bicycle:having one place:PFV-he.is.to.me
 side:DEF:LOC
 “A 13-14 year-old black boy on a bicycle got next to me.”

In the remaining 5 SVX sentences, the S clearly the topic of the assertion, and it is the topic or either the verb/polarity that is the focus (3 cases) or the X is an extraposed focus (2 cases). In two of the 3 cases in which the verb/polarity is the focus, this is coded by either the *egin* construction (1 case) or the *ba* construction (1 case). Only in one of these clauses is the verb/polarity focus left morphologically unmarked, namely the sentence in (6.55), in which the idea expressed by the X is very accessible and the verb is clearly more focal. In this sentence Almita is a person who had come into the room and who had been characterized as being half asleep. Thus, it is obvious that in this sentence that concept is very accessible and thus the verb is much more focal. In fact, if the complement was to be preposed it would be interpreted as contrastive.

(6.55) *Kuba triste dago*, p. 31

Almita ere **itzarri-da** bere lozorrotik

Almita also awaken:PFV-she.is her stupor:DEF:ABL

“Almita too, she woke up from her stupor.”

As I said, two of these SVX clauses (both from the same subcorpus) are clear cases of an extraposed focus constituent, namely the ones in (6.56) and (6.57) below. In both of these cases there must be the extraposed-focus intonational break. In the first case, the extraposed complement is rather lengthy and complex conditional clause.

(6.56) *Kuba triste dago*, p. 24

Txarrena **izaten-da** larunbatean kuadrako guardia tokatzen bazaizu.

worst:DEF be:IMPFV-it.is Saturday:DEF:LOC block:DEF:GEN guard:DEF
touch:IMPFV if:it.is.to.you

“The worst thing is / when/if you get block guard duty on Saturdays.”

The motivation for the extraposition in the second case is less clear, however, though there is no doubt that this complement is the focus and that there must be an intonational break between the copula and the complement.³¹

- (6.57) *Kuba triste dago*, p. 30
sei kilometro luze duen itsasertzeko malekoia da habanarren bilgunea,
six km length it.has.it:REL coastline:DEF:GEN malecon:DEF it.is Havanans:GEN
gathering.place:DEF
“The 6 Km. long ‘malecon’ (sea-wall) along the ocean is the gathering place of
the people of Havana”

Either one of these two sentences could have been written with the focus argument before the verb without there being any difference in interpretation, other than the slight awkwardness of having a lengthy clause before the verb in the first one.

Thus we can see that there is a quite compelling and obvious reason for most of the cases in which an X complement is postverbal. The main reason is that the X is not the focus, but rather something else is, either the subject or the verb or the polarity. Only in two cases is a focus extraposed and one of these is a rather complex constituent, which as I said are prime candidates for extraposition in many languages. Let us now look at similar clauses in the spoken corpus.

6.7.3 SXV and SVX sentences in the spoken corpus

As we have seen, the ratio of SXV to SVX clauses in the spoken corpus is rather different from that in the written corpus (SXV sentences from the spoken corpus can be seen in Appendix 6.11 and SVX ones in Appendix 6.12). This is not primarily because there is a higher proportion of subject focus or of verb/polarity focus, but, rather, because of the prevalence of focus extraposition in the spoken corpus.

In Table 6-13 we can see a breakdown of the 90 SVX sentences in the spoken corpus, according to their pragmatic structure. Five of these were taken out since their classification was not obvious, since the X could be a postposed setting adverbial, an afterthought, and even in one case a (post-presentative) secondary predicate, i.e. a second rheme tacked on to the first one.

SVX type	Subtype	No.
Subject focus [<u>S</u> V X]	contrastive subject	5
	emphasis	2
	existential	6
	presentative	20
	thetic (topical subject)	5
	(subtotal)	38
Focus extraposition S [<u>V</u> , <u>X</u>]	with hesitation	13
	with 2 intonation units	3
	with extraposed quote	1
	other	25
	(subtotal)	42
Verb/Polarity focus S [<u>V</u> X]	<i>ba</i> -marked (synthetic)	3
	not marked (analytic)	2
	(subtotal)	5
SUBTOTAL		85
Other, unclear	X=setting; X=afterthought, X=2 nd predicate??	(5)
TOTAL		(90)

Table 6-13: Main, affirmative SVX statements in the Spoken Basque Corpus.

Of the remaining 85, 38, or about 45%, of the clauses have subject focus with a postposed given, accommodated complement. An even larger number of sentences, 42, almost 50%, seem to have an extraposed focus constituent. A mere 5, or about 6%, are clearly constructions in which the verb is in focus position and the verbal idea or the polarity are the focus of the assertion.

In Table 6-14 we can see a quantitative comparison of all the relevant word orders (SXV and SVX) and focus structures for the written and spoken corpora (the focus constituent is underlined). As we can see the major difference between the two corpora has to do with the degree to which focus constituents are extraposed. Still, we can see that even in the spoken corpus focus constituents are twice more likely to be preverbal than extraposed.

	Written #	Written %	Spoken #	Spoken %
<u>SVX</u>	17	25.8%	38	26.6%
<u>S</u> VX	3	4.5%	5	3.5%
SV <u>X</u>	2	3.0%	42	29.4%
(SVX)	(22)	(33.3%)	(85)	(45.9%)
<u>SXV</u>	44	66.7%	58	40.6%
TOTAL	66	100.0%	143	100.0%

Table 6-14: SXV vs. SVX clauses in the written and spoken corpora.

As we can see, SXV clauses are numerous enough in the spoken corpus, 58 or about 2/5. Like in the written corpus, the majority of verbs are semantically light or stereotypical and the more salient complement goes in typical rheme-initial position. We can see an example of an SXV clause from the spoken corpus in (6.58) below.

(6.58) 93C2B08 (SXV)

131 eta biak lurrera erortzen-dira. -
and two:DEF:PL ground:DEF:ALL fall:IMPFV-they.are
“and the two fall to the ground.”

SVX in which the X is not the focus but an ‘extra’ complement of the verb in the assertion tail are also quite numerous, about 30%. But clauses with SVX order are significantly more numerous in the spoken corpus than in the written corpus. At least 40% have subject focus (of these only 2 have a topic, namely a dative). The subject focus

SVX clauses may be new (presentative-existential), as the one in (6.59), or contrastive, as the one in (6.60). (As always, the postverbal complement's idea must be easy to accommodate.)

(6.59) 93C3A07 (Focus-S; Existential)
34 ogi zati bat zeon maixen, ..
bread piece one it.was table:DEF:LOC
“there was a piece of bread on the table”

(6.60) 93C2A05 (Focus-S; Contrast)
102 Txarlot gelditu-zen ogiarekin.
Chaplin remain:PFV-he.was bread:DEF:COM
“Chaplin ends up with the bread”

But an even larger number of these SVX clauses have a topic subject and what is clearly a clause-final ('postverbal') focus element (X), i.e. an extraposed focus, as in the examples in (6.61) and (6.62) below. These verbs are also light verbs (e.g. copulas). Such verbs are only in focus position when the polarity is the focus. Notice that the verb in (6.61) is synthetic but doesn't have the *ba-* marker, for it is not the focus.

(6.61) 93C1A16 (*Focus-Extrapolation*)
79 lapurrak zeuden polizia-ren kontra,
thieves:DEF they.were police:DEF:GEN against
“the thieves were against the police”

(6.62) 93C2B08 (*Focus-Extrapolation, Hesitation*)
2 pelikula asten-da= gizon batekin, -
movie:DEF begin:IMPFV-it.is man one:COM
“the movie begins with a man”

SVX clauses can also be cases of verb/polarity focus, as in (6.63) below. Notice that here the synthetic finite verb does bear the *ba-* polarity focus marker.

(6.63) 93C3A10 (ba-focus)
 2 Eta= Txarlot **badijoa** kolean,
 and Chaplin EMPH:he.is.going street:DEF:LOC
 “And Chaplin is walking down the street”

Notice also that the postverbal complement has a discourse-new referent, but supposedly one that is easy to accommodate (*the street*). In other cases, such as in the example in (6.64), whether the verb or the postverbal complement is the focus depends solely on whether the verb or the object is accented, since the finite verb is unmarked.

(6.64) 93C2B06
 32 Txarlot **eroitzen**-da zulo batean, ...
 Chaplin fall:IMPFV-he.is hole one:LOC
 “Chaplin falls in a hole”

Here too, the postverbal complement’s referent (*a hole*) is new, and we might have expected it to be the focus. However, perhaps because this referent is rather ‘trivial’, and/or because the assertion is emphatic, the finite verb takes over as focus, since the emphasis adds focality or salience to the polarity, as we have already seen.

With SVX clauses, in which the X may be an optional complement, i.e. an adjunct, instead of an argument, we must be careful to distinguish extraposed focus constituents from sentence-tail extensions of the assertion (‘secondary predicates’). In (6.65) below we can see one such example in which it is not clear whether the main verb, *joan* “go; leave”, is being used as an aspectual verb or as an independent main verb.

(6.65) 93C1A05
 85 eta bera joaten-da m= - bueltak ematen, -
 and he go:IMPFV-he.is um= turns:DEF give:IMPFV
 “and he leaves um going around in circles” vs. “he’s going around in circles”

The complement in (6.66) is another example of this type of indeterminacy. Here the clause final ablative phrase (*from the basket*) could be the focus of the asserted clause, or it could be a sentence-tail afterthought (see also 93C2B04:46).

(6.66) 93C1B11

65 udare guztiak erortzen-zaizkio= - zakutik.
 pear all:DEF:PL fall:IMPFV-they.are.to.him basket:DEF:ABL
 “all his pears fall out - from the basket”

There were five such unclear cases among the SVX clauses, which were not included in the count.

6.7.4 The subcorpora compared

Just as in the case of clauses with absolute objects, here too the relative ratios of (S)XV and (S)VX order are not the same for the different subcorpora, and therefore for all speakers. If we look now at clauses without an overt subject, for example, which are much more common in the corpus than those with an overt one, and in which the X is the only complement of the verb, we can see clear differences of various magnitudes among the different subcorpora, as we can see in Table 6-15.

Subcorpus	XV		VX		Total	Comments
<i>Hendaia</i>	56	55.4%	45	44.6%	101	10 year olds; mixed fluency
<i>Ikasbide</i>	38	21.7%	137	78.3%	175	13 year olds; mixed fluency
<i>Deustu</i>	113	63.1%	66	36.9%	179	adults; very fluent
<i>Lur</i>	27	50.0%	27	50.0%	54	adults; very fluent
TOTAL	234	46.0%	275	54.0%	509	

Table 6-15: XV vs. VX order main affirmative clauses in the different sub-corpora of the Spoken Basque Corpus.

For two of the subcorpora, *Ikasbide* (children) and *Lur* (adults), the ratios of XV:VX are quite similar to the ratios they had for OV:VO (cf. Table 6-9, pg. 727). As we can see, the *Ikasbide* group shows the highest degree of postverbal X's (78.3%), which is somewhat less than the proportion of postverbal O's (88.4%). This is not surprising since the proportion of X's which are foci is higher than the proportion of O's that are. The *Lur* group shows an even proportion of XV:VX, just like it did in the case of the transitive assertions.

On the other hand, the other adult group, the *Deustu* group, shows a higher proportion of XV order clauses than the *Lur* group (63.1%), which is also higher than the proportion of OV sentences (55.5%). The other children's group, the *Hendaia* group, is the one for which the comparison between transitive and intransitive clauses shows the highest difference: 55.4% XV vs. 36.7% OV. The difference is in the right direction (assuming that more X's than O's are foci) but the magnitude of the difference is greater than for the other two groups which display such a difference.

In none of the spoken corpora is the ratio of XV to VX as high as it is in the written corpora, however, where the average ratio is about 9:1, as we can see in Table 6-16. As we can see, here too there are differences between the two subcorpora in the written corpus.

Subcorpus	XV		VX		Total	Comments
<i>Behi euskaldun...</i>	41	87.2%	6	12.8%	47	all 6 with <i>egin</i> verb focus
<i>Kuba triste...</i>	81	96.4%	3	3.6%	84	1 <i>egin</i> focus; 2 unmarked V-focus
TOTAL	122	93.1%	9	6.9%	131	

Table 6-16: XV vs. VX order main affirmative clauses in the different sub-corpora of the Written Basque Corpus.

In the part of the written corpus taken from the novel *Behi euskaldun baten memoriak*, all 6 tokens of VX sentences have verb focus coded with the *egin* construction, and there are no examples of focus extraposition. In the other part of the written corpus, the one that comes from the novel *Kuba triste dago*, a mere 3 have VX order. One of these VX clauses displays *egin* verb-focus, even though the postposed element is also new, as we can see in (6.67).

- (6.67) *Kuba triste dago*, p. 28
 eta joan egin-da ia hitzik esateke.
 and go:PFV do:PFV-he.is almost word:PART say:NOMIN:WITHOUT
 “and he left almost without saying a word.”

Notice that even though I have coded this as an SVX sentence, the X complement in this case is optional and, with the right (comma) intonation, it could be interpreted as a sentence-tail addition, and might be read that way by some readers. The other two VX clauses in this written subcorpus are not formally marked as having verb or polarity focus, but it is clear that the verb is more focal than the complement, which in both cases refer to very accessible locations (locative referents), as we can see in (6.68) and (6.69) below.

- (6.68) *Kuba triste dago*, p. 17
 zuzendu-naiz mahaira.
 direct/straighten:PFV-I.am table:DEF:ALL
 “I went straight to the table.”
- (6.69) *Kuba triste dago*, p. 18
 murgildu-naiz Habana Viejan-n.
 dive/submerge:PFV-I.am Havana Old:LOC
 “I immersed myself in Old Havana.”

In either case if the complement is so accessible that if it was preverbal it would have to be interpreted as being contrastive or emphatic, since that would be the only two things that would make it more focal than the verb (and thus the focus of the assertion).

6.8 CONCLUSIONS

To summarize, we have seen in this chapter that the major determinants of word order in Basque asserted clauses are the choice of focus and the position of the focus constituent in the rheme. Two focus types, polarity focus and verb focus, are coded by placing the finite verb in focus position. In some cases the two can be distinguished formally, by means of the *ba-* construction and the *egin* construction, but the distinction is often left unmarked. The choice of focus depends on the focality of the different elements of the clause. Focality, the property also known as ‘newsworthiness’, ‘noteworthiness’, or ‘unpredictability’, depends on a number of factors, in particular the accessibility of the idea (in context), but also on other factors such as contrast and emphasis, the former being an intrinsic or context given property and the latter an extrinsic or speaker-based one.

The most basic and common focus position in Basque is rheme-initial, preverbal position. A secondary position for focus constituents is extraposed position, in which I include cases in which the focus is clause-final (and often also ‘postverbal’ if there are no other rhematic complements) but intonationally ‘clefted’. Speakers (and writers) vary as to the extent that they make use of this focusing strategy. In writing this position is almost exclusively reserved for complex constituents, such as complement clauses, as in many other languages, following a common crosslinguistic tendency which seems to be

motivated by processing considerations. (Extraposition may also be used perhaps in situations in which both the verb and a complement are very focal, thus placing both of these elements in different intonation units.) In speech, however, this somewhat optional focusing strategy is much more common, and although in some cases it may serve a rhetorical (suspense) function, in other cases its use seems to be less clearly motivated, other than perhaps by the influence of the surrounding languages which, as VO (postverbal focus) languages also use focus extraposition to a high extent, higher than it is normally expected in an OV language.

The increased use of the extraposition focusing strategy is perhaps the main mechanism by which (so-called) OV languages, in which the main focus position is preverbal, become (so-called) VO languages. Such a change would involve the reanalysis of rheme-final focus position as postverbal position, the main position for low focality foci (non-contrastive, non-emphatic) in these languages. More salient foci, such as contrastive ones, still prefer rheme-initial position, but so do such foci in perhaps most VO languages, as we have seen. The difference is one of proportion (and thus 'markedness') and therefore the change can be quite imperceptible unless the relative frequency of use is ascertained at different points in time. Such diachronic differences, however, also have parallels in synchronic variation. As we have seen, frequency analyses indicate that the tendency to extrapose foci is higher for groups of speakers for whom a Romance language may be dominant, especially for those for whom that language is Spanish.

We have also seen in this chapter the futility of trying to ascertain a basic (context independent) word order for Basque, and in particular a basic word order which is based

on grammatical as opposed to pragmatic relations or roles. Frequency studies can be quite misleading in this respect, and might lead to the conclusion that the basic word order is different for the Basque of different speakers.

We have also seen that differences may exist between speakers as to the choice of focus in particular circumstances in which the choice is not particularly obvious, such as when the verb is a stereotypical one and the idea expressed by the complement is somewhat accessible. The choice of the verb or the polarity as the focus of the assertion over a low focality complement is quite warranted in these cases, but speakers may differ as to the choice they make. This doesn't lead to miscommunication since the worst kind of possible misinterpretation would be that the assertion would be interpreted as emphatic by the hearer (with verb or polarity focus). This source of variation among speakers, however, is not of the magnitude that the differences in the use of extraposition are.

As we have seen, in declarative assertions, the default, or most common, focus is a complement of the verb, since such elements tend to be more focal than the verb or the polarity. In the next and final chapter I will consider cases in which the verb and in particular the polarity, and not a complement, are the default foci of the assertion and the consequences that this has for the ordering of the elements of such asserted clauses.

NOTES TO CHAPTER 6

1. According to Osa, “it is often very difficult to find the focus, especially when we go beyond simple expository questions and affirmative sentences” (Osa 1990:285; my translation, J.A. [“askotan oso zail gertatzen [da] fokua aurkitzea, batipat adierazpenezko galderazko eta baiezko perpaus mota soiletatik aldegiten dugunean”]). The Basque Language Academy, following a tradition started by Azkue 1891, uses the term *galdegaia* for the “element which is salient inside the predicate” (Euskaltzaindia 1987a:30; my translation, J.A.) [“irazkinaren barnean ‘hanpatua’ den elementu gisa”].

2. According to Villasante “[t]he rules for ordering elements inside the clause, as formulated by Mr. Altube, presuppose that every Basque sentence has a focus” (Villasante 1980:255; my translation, J.A.) [“Las reglas de ordenación de los diversos elementos dentro de la frase, tal como han sido formuladas por el Sr. Altube, se basan en el supuesto de que en toda frase vasca hay un elemento inquirido”]. It seems that Villasante may have been following Mitxelena on this matter. Even if Altube didn’t exactly believe this, it is true that he may have gone too far in expounding his newly found pragmatic category of ‘focus’ (*galdegaia* “question matter” in Basque). After some unfair and unfounded criticism of Altube’s analyses, Villasante adds that, “the main complaint about them [the ordering rules] is that they are based on an undemonstrated postulate, namely that in every Basque sentence there is a focus” (Villasante 1980:256; my translation, J.A.) [“El reparo mayor que se les debe hacer es que se fundan en un postulado indemostrado, a saber: que en toda frase vasca hay un elemento inquirido”].

3. Ortiz de Urbina (1989, 1993) has explored this phenomena for Basque, for instance, and found some very interesting phenomena related to this type of ‘pied-piping’.

4. Basque linguist Villasante explains the (to him) surprising frequency of VO ordering in narratives by saying something like that the verb is typically more salient in sentences in this genre. According to him, “In narrative style [i.e. genre], that is, when happenings are narrated,, it is common for the verb to come first, and this is easily explained: when happenings are told, the same enumeration of such happenings is conceived generally as what’s interesting. Now, it is the actions expressed by the verbs that constitute the nucleus of the narration and that is why there is a tendency to place the verb first in this genre. (Villasante 1980:240; my translation, J.A.) [“En el estilo narrativo, es decir, cuando se cuentan sucesos, es frecuente la anteposición del verbo, y ello se explica fácilmente: cuando se narran sucesos, la misma enumeración de los sucesos se concibe generalmente como lo interesante. Ahora bien, las acciones expresadas por los verbos son las que constituyen el núcleo de la narración y así se comprende la tendencia a anteponer el verbo en este género”]. Cf. the discussion on genres and styles in Chapter 2.

5. Dik distinguishes two types of contrastive foci: parallel and counter-prepositional. His parallel contrastive focus are what I would classify as sentences with a contrastive topic and a contrastive focus. He distinguishes four subtypes of counter-presuppositional contrastive focus: replacing (‘but’ focus), expanding (‘also’, ‘even’ focus), restricting (‘only’) and selecting (from a finite number of options).

6. Thus contrast “represents a context where the speaker assumes—or actively sets up—certain expectations on the part of the hearer, then proceeds to release information that is contrary to those expectations. More likely, contrastive constructions are also topicalizing (Givón 1990, Chapter 16)” (Givón 1995:338).

7. Bolinger suggests that “... in a broad sense every semantic peak is contrastive. Clearly in ‘let’s have a *picnic*,’ coming as a suggestion out of the blue, there is no specific contrast with ‘dinner party,’ but there is a contrast between picnicking and anything else the group might do. As the alternatives are narrowed down, we get closer to what we think of as ‘contrastive stress’” (Bolinger 1961b:87).

8. According to Givón “the notion “contrast” is—neither grammatically nor cognitively—a discrete prime. Rather, ‘contrast’ is founded on the more general cognitive dimension of **informational predictability** or its converse, **counterexpectancy**. And informational predictability is, at least in principle, a matter of degree. While grammatical constructions do not code promiscuously many points along a functional-cognitive scalar domain, they often code more than a simple binary split” (Givón 1990:700). Givón also argues that “[b]etween these two extreme points there exists, at least potentially, a vast number of intermediates, all having to do either with the speaker’s assessment of the hearer’s *degree of ignorance*, or with the speaker’s assessment of the hearer’s *strength of contrary belief*” (Givón 1990:702). In practice, in normal conversation, the norm is probably somewhere in between (Givón 1990:703).

9. Givón attributes this stronger tendency in negative clauses to the fact “negation is itself a contrary speech-act” (Givón 1990:713; cf. Givón 1984:Ch. 9). Givón adds *wh*-questions to this list, but it is obvious that his example involves a contrastive *topic* (or thematically salient referent) and not a contrastive focus. The focus in a *wh*-question is the *wh*-phrase. His example is: *Oh yeah? And what did you do?* Notice that languages with freer word order would topicalize such a contrastive topic, e.g. Spanish: *¡Ah! ¿sí? ¿Y tú qué hiciste?* or else *¡Ah! ¿sí? ¿Y qué hiciste tú?* It seems to me that in content questions the question word is always the salient part of the rheme, by definition (of the question construction.)

10. This claim extends to languages such as the following: Nandi (Creider 1977), Ojibwa (Tomlin and Rhodes 1979), Ute (Givón 1983c), Cayuga (Mithun 1987/1992, 1995), Papago (Payne 1990), Maasai and Yagua (Payne 1995). Payne (1995) has argued that in Maasai and Yagua the most newsworthy element does indeed come first, preceding the verb, but that among the postverbal elements the more topical or thematic arguments come first.

11. Conjunction = *-la_Compl.*

12. In two of these the main verb is itself a in rigidly verb-final construction, such as a causal dependent clause.

13. Of the 14 examples of this construction in main, affirmative statements in the spoken corpus, 11 have the complement after the verb and in only three of them does it precede the verb. Other codes = *Like *Ari_Tzen**; Speech act = *Like Body_Statement**.

14. The *ari* (“busy; occupied”) construction is used for expressing continuous imperfective aspect for verbs which do not have a synthetic conjugation (most verbs). This construction is not tightly grammaticalized as its English or Spanish counterparts, and thus the main ‘verb’, *ari*, retains a great deal of autonomy and the non-finite (imperfective) clause is still felt to a great degree to be a complement of *ari*. Obviously, were this construction to become more grammaticalized and were the default order to become [ari + S-tzen], then Basque would have acquired its first preverbal auxiliary verb.

In addition to this, the morphological distinction between ergative and absolutive topics would be neutralized in this construction, since *ari* always takes an absolutive topic and an auxiliary with only absolutive agreement (in other words the ‘agreement features’ of the lower verb’s complements do not ‘percolate’ up to the main verb. This would introduce a perfective-imperfective ergative split in the language of the kind that are so well known in other ergative languages.

15. If the assertion’s accent was on the subject, this sentence would mean something different, since the subject would then be the (contrastive) focus and the object would be in postverbal position from having been ‘bumped off’ preverbal position (and it would have to be extra-rhematic, i.e. accessible): *Liburutto haur da emazurtza* “This book is an orphan” = “It is this book that’s an orphan.”

16. The following is the original quote: “Esanik dago galdegaiak aditzaren ezkerreko lehen posizioan duela kokalekua. Halaz guztiz, batzutan—euskara mintzatuan bereziki gertatzen da hau—intonazioak aipatu ordena hau hauts lezake, galdegaia aditzaren eskuinean ematen delarik orduan” (Euskaltzaindia 1987a:35).

17. Notice the inverted topic. On the other hand some of the NP internal syntax is rather un-Basque-like and is perhaps a calque from Romance, such as the passive-like (relative) clause. Notice that the causal clause (*because they rose up before God*) is used here as a ‘dislocated’ setting clause.

18. Villasante’s original quotes are the following: “tal construcción, al tratarse de enunciados largos, produce en el ánimo una sensación de espera o ansiedad ... el guardar a rajatabla y a punta de lanza una tal ordenación, sobre todo cuando el predicado es largo—como sucede con frecuencia en la lengua escrita—produce una sensación de algo insólito y violento. ... Lo normal en este caso es echar el predicado después del verbo o intercalar éste dentro del grupo del predicado” (Villasante 1980:249)

19. Cf. Frazier 1979/1978, (Jerry) Fodor 1983, Frazier and (Janet) Fodor 1987, Hawkins 1990, 1994. These authors would attribute the ‘need’ for the verb to come sooner as the need for early arrival of the “mother node constructing category” (cf. Fodor 1983).

20. The Spanish sentences would be as follows: OV,S: *El pan lo tiene, el de al lado*, OSV: *El pan el de al lado lo tiene*; and SVO: *El de al lado tiene el pan*.

21. This ten year old speaker speaks a mixture of unified Basque learned in school and the local Lapurtera dialect, as described by his teacher.
22. The original quote says: “honako bi hautabide hauen arteko aukera egitea ez da biziki erraza” (Euskaltzaindia 1987a:39; my translation, J.A.).
23. About the justification for their choice (SOV) this is what they say: “In spite of [all] this, according to what the majority of grammarians tell us, it does seem that the S/OV order seems more authentic/correct/good. Statistics too, especially those done on written texts, reach the same conclusion” (Euskaltzaindia 1987a:39; my translation, J.A.) [“Halan eta guztiz ere, gramatikolari gehienkoenek diotenez, aditza emateko joera bizia dugu euskaldunek, eta euskaldun gehienen intuizioaren arauera ere, badirudi S / O A ordena hau jatorrago edo ematen dugula. Estatistikek ere, testu idatzietan egindakoek bereziki, ondorio bera ematen dute”].
24. My translation, J.A. The original says: “En la narración, pues, se tiende con frecuencia a echar los predicados detrás del verbo. Esto se observa en los escritores que son buenos narradores” (Villasante 1980:25). A number of writers have repeated this claim, which supposedly goes back to Mitxelena, cf., e.g., Villasante 1980:25; Euskaltzaindia 1987a:33, 43; Osa 1990.
25. My translation, J.A. The original says: “zerbait kontatzen ari garenean, gertaera segida bat ematen denean, sarritan, aditza perpaus hasieran (hobeki esanez, irazkinaren hastapenean) eman ohi da, *egin* erabili gabe orduan, non ez den azken aditzean.” (Euskaltzaindia 1987a:33).
26. The corresponding adverbial *already* in English can perhaps also be used metaphorically, e.g. *I already have one*, but not to the degree that the corresponding particle in Spanish can, which can be used for future situations as well, as we have just seen.
27. *Xla* stands for a finite complement clause (completive) with the subordinator *-la* suffixed to the finite verb. *Xtzen* stands for a non-finite complement, or gerundial, clause with the ‘subordinator’ *-tzen*.
28. These finite, main declarative clauses discussed from this point on have been coded thus in the database: Unit Type = *Fin_Cl_Main*; Speech Act = *Body_Statement*. This is true of all such clauses discussed from this point forward. Additional codes for this set are: Word order = AVO, AOV, VO, OV. Most of the clauses with an overt object do not have an overt ergative in the corpus: 59 OV and 16 VO. It is interesting that the proportions should be so different depending on the presence or absence of the ergative argument. This is due, no doubt, to the fact that many cases of AVO order are due to the fact that the A is focused and thus rheme initial: at least 6 and maybe 7 of the AVO examples are of this kind.
29. I will look at the OV/VO group rather than the AOV/AVO group because in the latter group it isn’t always clear whether an ergative argument is the topic or is rheme initial. It

seems clear, however, that there is a smaller percentage of rhematic subjects among AVO clauses in the spoken corpus than in the written corpus.

30. There are an additional 5 which were originally classified as SVX clauses, but which seem to be examples of other types of constructions (cf. Appendix 6.11).

31. This copula sentence could conceivably have a different reading, namely one in which the preverbal absolutive nominal is the focus and the postverbal one is an antitopic. I do not believe that this is the right interpretation in the context in which this sentence was found.

Chapter 7

Polarity, emphasis, and word order

7.1 POLARITY, FOCALITY, AND FOCUS

Polarity is a fundamental ‘feature’ of all clauses, and indeed of all asserted clauses. It is one of the most basic clausal ‘operators’ in language. The major values of this operator are affirmative and negative. Affirmative polarity is the default value for all types of clauses, whether asserted or not. In the unmarked case it is not coded in any special way. Negative polarity, on the other hand, is always marked, and negative sentences are always more complex than affirmative ones, typically having one (or more) morphemes coding this marked value of the polarity, and sometimes involving complex constructions as well.

The notion of (affirmative or negative) polarity, which is common to all clauses, must not be confused with the communicative acts of affirmation and negation of a proposition. Thus, although all clauses are affirmative or negative, not all clauses are affirmed or negated. The relevance of polarity to information structure is related to the different informational characteristics of these two types of assertions.

Asides from the differences in coding markedness, there is an important way in which negative assertions are said to be typically marked. As pointed out by Givón (1975b) and others, whereas affirmative assertions are typically put forth in the context of

'ignorance' (or at least low accessibility) of the underlying proposition, negative assertions tend to be made in contexts in which the underlying proposition is accessible, or at least easy to accommodate.¹

Thus, for example, if someone says *I am not sick* or *I didn't read the book*, typically (though not necessarily) there is some reason why the hearer might have thought or expected that the speaker might be sick or might have read a particular book. In other words, what is salient here is the falsity of the underlying predication. The consequence of this fact seems to be that negative assertions often display what I have been calling 'rheme focus', or 'polarity focus', which is why negative asserted clauses often display many of the characteristics of emphatic affirmative assertions. In English, as in perhaps most languages, the negative operator typically attaches itself to the finite verb, which stands for the assertion (the rheme), and is placed in focus position.

Similar contextual 'expectations' (i.e. accessibility of the proposition) may be behind an affirmative sentence, such as *I am sick* or *I read the book*, but these expectations seem to be more unusual, or 'marked', in affirmative assertions. In other words, they are found in those which I have called 'emphatic' (with rheme, or polarity focus). Notice that in these two examples it is the finite verb, and not the complement, behaves like the focus constituent.

In other words, it seems that in perhaps most negative sentences, just like in 'emphatic' affirmative sentences, i.e. assertions which are put forth with greater emphasis or salience than average, it is the whole assertion that is 'focused'. This is summarized in Table 7-1 below.

<i>Polarity</i>	<i>Focus</i>	<i>Example</i>	<i>Markedness</i>
Affirmative	Complement (or verb)	I read a <u>book</u> .	Unmarked
	Finite verb	I <u>did</u> read a book I <u>read</u> a book.	Marked
Negative	Complement (or verb)	I <u>didn't</u> read a <u>book</u>	Marked
	Negative finite verb	I <u>didn't</u> read a book.	Unmarked.

Table 7-1: Polarity focus vs. non-polarity focus in affirmative and negative assertions and their *relative* markedness.

In the 'marked' cases, the *negation* or *affirmation* itself is the focus, and not any of the different ideas expressed in the proposition. In these cases it seems that quite consistently it is the finite verb, standing for the whole assertion and together with the polarity operator, which acts as the focus of the assertion, as shown by the fact that in these cases it is the finite verb which is typically accented and placed in focus position. In these marked cases it is not the *polarity*, but the affirmation or negation that is focal and salient. Thus by 'polarity focus' I mean that the affirmation or the negation itself, the whole rheme, is the focus. As we will see below, this is not the only possible interpretation of the facts. Altube for instance claimed that in negative sentences it is the polarity operator itself that is the focus of the assertion.

Although the observation about the different relative markedness of polarity focus in affirmative vs. negative assertions seems to hold at some level, perhaps at least in all-news, decontextualized assertions, no actual studies seem to have been carried out which might determine how much more likely it is for a negative sentence than for an affirmative sentence to be uttered in a context in which the clause is accessible (and thus the polarity is focal). From my experience with the Spoken Basque Corpus I do not think

that the unmarkedness of negative polarity focus is a clear-cut as Givón and others have assumed it is, or at least not in all contexts, and in particularly not in narrative.

Two possible factors may cause the whole assertion to be the focus: an intrinsic one and an extrinsic one. An intrinsic factor is involved when the proposition, and all the ideas expressed in it are very accessible and the assertion's polarity (by which I mean the affirmation or negation) is inherently, or intrinsically, what is new and/or contrastive. This is found, for instance, in polarity questions and answers to such questions. An extrinsic factor is involved when the speaker makes the whole assertion emphatic, or asserts the proposition with added 'force'. This can only be done as long as the underlying proposition (and all the ideas expressed in it) are accessible or easy to accommodate.²

It seems that in English, *do*-support is used in cases of intrinsic, contrastive affirmative polarity focus, though not in cases of extrinsic polarity focus. In the future tense, on the other hand, where *do*-support cannot be used, the accented finite verb *will* may code either intrinsic or extrinsic polarity focus. In Spanish, *sí* is used to code intrinsic (contrastive) polarity focus, and *ya* is used to code extrinsic (emphatic) polarity focus, as we saw in Chapter 6.

The phenomenon of 'polarity focus' seems to be what Dik (1981) has referred to as 'polar focus', and Sasse (1995a) as 'verb focus'. We should not confuse, however, polarity focus with true verb focus, i.e. cases in which the verbal idea itself is the focus. Notice too that when the verb is the only element of the rheme, the difference between verb focus and rheme focus may be hard to differentiate. The two focus types are often coded identically, namely when the verbal idea and the polarity are coded by the same

(phonological) word, i.e. when there is only one verb, the finite verb. Very often, crosslinguistically, two elements are used to code the 'verb': a finite half and a non-finite (verbal) half, forming an analytic (periphrastic) complex verb, the purpose of which may be in part to differentiate the polarity bearing characteristics from the verbal idea itself..

In some cases, however, the two parts of an analytic verb complex act may become fused and may act as a single phonological word, at least in some contexts. This may prevent a transparent differentiation between polarity focus and verb focus. This means that, unlike in English, one cannot phonologically distinguish cases of polarity focus, e.g. *I háve seen the bòok*, and verb focus, e.g. *I have séen the bòok*. This is what has happened in Basque but, interestingly, only in affirmative assertions.

As we saw in Chapter 2 the same type of fusion has not taken place in Basque negative assertions. Here the fusion of the negative operator *ez* with the finite verb, and the fact that the finite verb often (though not always) is the focus of the assertion, seems to have prevented the fusion of the two parts of the analytic verb. This, along with the requirement that the negative operator be to the left of all the elements of the rheme over which it has scope, results in word orders in negative assertions which are rather different from those we find in affirmative assertions. In other words, the word order that obtains in Basque negative sentences, regardless of which element is the focus, is the one we would expect to find when the polarity is the focus.

7.2 NEGATION IN BASQUE

7.2.1 Introduction

As we saw in Chapter 2, negation in Basque is marked with the particle *ez* “no/not”, which is placed immediately before the finite verb, just like the affirmative particle *ba*, although this affirmative particle, unlike the negative one, is only used with synthetic verbs in most dialects. When the verb is synthetic, the particle is in immediately preverbal, rheme-initial position, much like the *ba* particle we saw in Chapter 6, as we can see in example (7.1).

- (7.1) Jon ez-dator.
Jon not-he.is.coming
“Jon is not coming.”

What is written *ez dator* as two words, is actually one phonological word: it is pronounced, and sometimes also written as *eztator* (reflecting voice assimilation).³ What syllable the accent falls on depends on phonological considerations of the resulting word, such as the number of syllables and not, for instance, on the relative focality of the ideas expressed by the different morphemes (*éza* “it isn’t”, but *eztágo* “there isn’t”).

In negative assertions, the finite verb, with the negative particle attached to it, is typically in rheme-initial position (see exception below). Any and all non-setting elements of the assertion, i.e. all but the topic and ‘topicalized’ ones, must follow the finite verb. This is true whether the verb is synthetic or periphrastic.⁴ With periphrastic verbs, the order of the finite and the non-finite verbs *must* be inverted, as in example (7.2b), corresponding to the affirmative clause in (7.2a). This is true only in finite

asserted clauses, whether they are main clauses or dependent ones (such as completive clauses).

- (7.2) a. Jon etorri-da.
Jon come:PFV-he.is
“Jon has come”, “Jon came.”
- b. Jon ez-da etorri.
Jon not-he.is come:PFV
“Jon hasn’t come,” “Jon didn’t come.”

In other words, the non-finite verb doesn’t cliticize on to the finite verb when the verb is negative, as it does in affirmative sentences. Whereas in affirmative clauses the two parts of the verb count as a unit for the purpose of determining preverbal position, acting as a single phonological word, in negative clauses the negative polarity operator particle is cliticized to the finite verb instead and the main verb must be in the post-rheme (assertion tail) area just like non-focus rhematic elements in affirmative sentences. This state of affairs iconically reflects two facts about negative sentences: (a) the negative particle has scope over the whole assertion, including the focus; and (b) in the unmarked case the focus of the assertion is the affirmation or negation, instantiated as the finite verb, that is, with the finite verb standing for the assertion.

The assertion itself, however, need not be the focus of a negative assertion. Any other element may be the focus, much like in affirmative assertions. Also, in Basque, like in all other languages, the negated finite verb must be in rheme-initial position *only if the focus is inside the scope of the negative operator*. When this is not the case, the focus must precede the negative finite verb, as in examples (7.3Q&A) below. Notice that for the question in (7.3Q) to be felicitously uttered, a corresponding, *negative* proposition must be accessible from the context (cf. *Someone hasn’t come*).

(7.3) Q: Nor ez-da etorri?
who not-s/he.is come:PFV
“Who hasn’t come?”

A: Jon ez-da etorri.
Jon not-s/he.is come:PFV
“Jon hasn’t come.” (= “It’s Jon who hasn’t come.”)

The focus constituents in these assertions are *nor* “who” and *Jon*, respectively. (We shouldn’t confuse the assertion in (7.3A) with one in which the absolutive is the topic instead of the focus, since the accentual pattern and intonation are very different in that case.) When the focus is under the scope of the negative operator, however, it is not rheme-initial as in these clauses, and as in affirmative assertions, as we can see in (7.4), in which *Jon* is a contrastive focus (with the focus accent on *Jon* and not on the finite verb).

(7.4) Ez-da Jon etorri.
not-s/he.is Jon come:PFV
“It wasn’t John who came.”

I will argue below that the fact that the focus is not rheme-initial in this case as it is in affirmative assertions is due to two factors: the iconic need to keep the negative operator to the left of the focus and the fact that the negative particle is cliticized to the finite verb, even in those, supposedly marked, cases in which the assertion itself is not the focus. Notice that the two negations in (7.3A) and (7.4) have different background ‘presuppositions’. In the former case the presupposed (and thus accessible) clause is *Someone didn’t come*, whereas in the later case it is *Someone came*.

‘Verb-auxiliary’ inversion is obligatory in negative asserted clauses in modern Basque, regardless of which element is the focus of the assertion and whether the focus precedes the finite verb or not. As several authors have pointed out, verb-auxiliary

inversion in negative assertions hasn't always been required as it is today in main asserted clauses (e.g. Michelena 1978:224; Villasante 1980:30; Euskaltzaindia 1987b:493, 1993:437).⁵ At present, however, the negation construction, with main verb inversion, is fully grammaticalized and there is some evidence that in speech, verb-auxiliary inversion is spreading to non-asserted subordinate clauses, something which is not possible in standard (written) Basque.⁶

7.2.2 The placement of operators in Basque

As we saw in Chapter 1, the fact that the negative operator in Basque comes at the beginning of the rheme that it has scope over, and not after, is not as unusual for a verb-final language as it might seem. Cross-linguistically, the negative (or emphatic affirmative) operator is very often placed *before* the finite verb and before the whole rheme over which it has scope. This is also the case in the vast majority of VO languages, but it is also the case in Basque and many other verb-final languages in which the polarity word doesn't derive from a negative auxiliary verb, as we saw in Chapter 1. This tendency for operators to be left-peripheral is rather common among 'operators' of all kinds, such as focalizers. It seems that, by preceding the elements it has scope over, an operator iconically, and anticipatorily, marks the scope over which it acts.

In Basque, in general, operators which modify noun and adjective phrases, such as 'intensifiers' and 'quantifiers ('degree adverbials'), also typically precede the modified elements, e.g. *nahiko* "rather, quite a bit", *batere* "at all", *oso* "very", *izugarri* "terribly", *guztiz* "completely", *nahiko* "enough, rather", *biziki* "quite", *erabat* "completely", *zeharo* "completely". Others, however, follow the modified element, such as *samar* "rather",

asko “much”. Some precede or follow depending on the actual word they modify, thus *aski* “rather, enough” precedes predicate adjectives but follows adverbs (e.g. *maiz aski* “often enough”).

Operators which have scope over whole rhemes (as well as possibly other types of constituents), such as intensive operators (e.g. *really* in English), are a particularly interesting and relevant type. As we will see in Section 7.5 below, these elements may be in rheme-initial position, but they can also be clause-final or, rather, ‘extraposed’, in a different intonation unit following the assertion proper, having an accent of their own (cf. *I really like him* vs. *I like him a lot*).

The negative operator is also found in ‘assertion fragments’, in which the whole of the proposition minus the focus is ‘understood’ from the context. In these cases too the operator precedes the focus in Basque, just like in English, e.g. *ez etxean* “not at home”. In these cases the focus constituent stands for the assertion and the operator is placed to its left, just like when the assertion is a full rheme.

As I mentioned above, the idea that any element of the assertion may be the focus of a negative assertion, including the whole assertion itself in what is perhaps the default case, is not the only possible interpretation of the facts. Altube (1929) and other investigators, for instance, have argued that in negative assertions it is the negative operator itself that is the focus of the assertion and that the negative morpheme, *ez* in this case, that the focus constituent, which is why it is placed in rheme-initial position.⁷ Equivalently, in affirmative assertions the emphatic morpheme *ba* would be the focus. I believe this view to be mistaken. The negative morpheme is the formal instantiation of the negative polarity *operator*, and the operator itself cannot be the focus of an assertion.

Rather, some other element must be the focus, including the whole assertion, with the finite verb standing for it and acting as the focus constituent.

One argument against Altube's position would be the fact that the negative particle is not a phonological word and cannot be accented since it is cliticized to the following verb. Txillardegui (1984), and Osa (1990) along with him, argue, however, that this phonological fact is not incompatible with the claim that the negative particle is the focus of the assertion, since we also find some demonstrative pronouns which are quite clearly foci which are also cliticized on to the verb and unaccented (Txillardegui 1984:290, Osa 1990:293).⁸

Another argument for Altube's position would seem to be the fact that in some languages it is indeed possible for the morpheme representing the polarity to receive the main accent of the clause under some circumstances. In Basque, for example, the particle *ez* can, and indeed must, be accented when it stands for the whole (negative) rheme, all the elements of which except this unpredictable operator are elided, as we can see in (7.5A). In other words, Basque doesn't need the 'support' of a finite pro-verb to stand for the rheme. The particle *ez*, however, cannot be accented when it is an operator, as in (7.5A').⁹

- | | | |
|-------------------------------|-----------------|---|
| (7.5) Q: <u>Joan</u> -zinen? | go:PFV-you.were | " <u>Did</u> you go?" |
| A: (<u>Nik</u>) <u>ez</u> . | I:ERG not | "I <u>didn't</u> " (ez = 'pro-assertion') |
| A': Ez <u>atzo</u> . | not yesterday | "Not yesterday" (ez = operator) |

In Spanish, the negative particle, or the affirmative particle for that matter, may be accented when the rheme is the focus, such as in (7.6a&b) below, but not when some other element is the focus, as we can see in (7.6c&d). These polarity particles are not

necessarily accented when the assertion is the focus, however, and the finite verb may be accented instead, as we can see in (7.6e&f).

- | | | | |
|----------|-----------------------------------|------------------------------|------------------------------------|
| (7.6) a. | Juan <u>sí</u> vino. | Juan yes he.came | “Juan <u>did</u> come.” |
| b. | Juan <u>no</u> vino. | Juan no(t) he.came | “Juan <u>didn't</u> come.” |
| c. | Juan <u>sí</u> vino <u>ayer</u> . | Juan yes he.came yesterday | “Juan did come <u>yesterday</u> .” |
| d. | Juan no vino <u>ayer</u> . | Juan no(t) he.came yesterday | “Juan <u>didn't</u> come.” |
| e. | Juan <u>sí</u> <u>vino</u> | Juan yes he.came | “Juan <u>came/did come</u> .” |
| f. | Juan no <u>vino</u> . | Juan no(t) he.came | “Juan didn't <u>come</u> .” |

As in Basque, these polarity particles are accented when they function as pro-predicates, standing for the whole assertion, e.g. (7.7a&b).

- | | | | |
|----------|------------------|------------|------------------------|
| (7.7) a. | Juan <u>sí</u> . | Juan yes | “Juan <u>did</u> .” |
| b. | Juan <u>no</u> . | Juan no(t) | “Juan <u>didn't</u> .” |

Despite the appearances, I believe that even in these cases it is not the operator itself that is the focus of the assertion. I would argue that in all these cases in which the operator is accented, the operator is standing in for the whole assertion, much like the finite verb stands for the assertion in those cases in which it is the accented element of an assertion focus clause. Notice that the polarity morphemes can also be used in these languages as polarity asserting morphemes. That is, *ez* for example can mean either “not” or “no”. That is, in Basque the negation morpheme may stand for the whole assertion when it stands for the whole negative assertion, as in (7.5E). In Spanish this may happen in the same context, but also when the rest of the rheme is not elided, since in those cases it is cliticized. The final proof that the negative morpheme is not itself the focus constituent is, I believe, that any other element of the clause may indeed be the focus, as we already saw above and will see in more detail in the next section.

7.2.3 Negative assertions in which the assertion is not the focus

As we have already seen, although the whole assertion is often the focus of a negative assertion, and the finite verb the focus constituent, this is not necessarily the case. Despite Altube's claims that the negative morpheme the focus of (all) negative assertions, it is quite obvious that elements which are in the scope of the assertion may be the focus, something which was first noticed by Lafitte (1962/1944). We saw an example of subject focus in (7.4) above, for instance. The focus may even be *outside* of the scope of negation in cases in which the negative polarity is accessible (part of the 'background presupposition'), as we saw in (7.3A). In (7.8) below we can see an example in which the absolutive object is the contrastive focus of a negative assertion.

- (7.8) Jonek ez-du liburua irakurri, egunkaria baizik
Jon:ERG not-he.has.it book:DEF read:PFV newspaper:DEF but/rather
"Jon didn't read a/the book, but the paper."

This assertion could be made felicitously in a context in which a proposition such as *Jon read a book* was quite accessible, such as after such a claim has been made. This is the same sort of contrastive element which we saw in Chapter 6, which is intrinsically focal and which is always the focus of the assertion.

Analyses of negation which view the negative word itself as being the focus of the assertion have often pondered about these sentences, starting with Lafitte, as I said. Is the negative word the focus, as its rheme-initial position seems to imply, or is the object focus, as the pragmatics suggest? Or are both foci? On the other hand, if the negative morpheme is never the focus, this problem never arises.

Although the idea that a contrastive element, such as the book in sentence (7.8), is the focus of the assertion is appealing, there are some problems which this analysis brings up which we need to explain. The problem involves the position of the focus constituent and the relation between this sentence type and the type of negative sentence in which some element other than the polarity is focused *and* fronted, in which the focus is outside the scope of negation, as in (7.3G&E) above.

In (7.9) and (7.10) below we see two sentences which constitute a minimal pair to compare the two types of focus situations in negative sentences. They differ, as we can see, in the underlying (accessible) proposition.

(7.9) *Accessible*: Jon didn't read something (= there is something Jon didn't read)

Jon_{ERG} lib_{DEF}urua ez-du irakurri.

Jon:ERG book:DEF not he.has.it read:PFV

"The book is what Jon didn't read."

(7.10) *Accessible*: Jon read something (= there is something Jon did read)

Jon_{ERG} ez-du lib_{DEF}urua irakurri.

Jon:ERG not he.has.it book:DEF read:PFV

"Jon didn't read the book."

(= "It's not the book what Jon read.")

In the case of the sentence in (7.9), what is accessible is that there is something that Jon *didn't* read. In the case of the sentence in (7.10), what is accessible is that there is something that John *did* read.

There are some north-eastern dialects in which the sentence in (7.9) could be expressed in an alternative way, namely by placing the focus element directly before the finite verb, and the negative particle would then go with the non-finite verb, as in (7.11) (cf. Lafitte 1962/1944, Laka 1994:71-72, 97fn4). This alternative would be used in cases in which the focus is highly contrastive and, thus, very focal.

- (7.11) *Jonek liburua du ez-irakurri.*
Jon:ERG book:DEF he.has.it not-read:PFV
“The book is what Jon didn’t read.”

These also seem to be the dialects in which in affirmative sentences a rheme-initial focus element can cause the non-finite verb to invert when the focus is highly contrastive, something which is not possible in standard Basque or in the dialects used by the majority of Basque speakers (which are also the dialects that are exemplified in the corpus). Perhaps the reason that this sentence is not possible in all dialects has to do with differences in the degree to which the negative particle has advanced on its way to full cliticization and the number of contexts in which cliticization is found, just like the cliticization of the finite and the non-finite verbs in these non-eastern dialects seems to be more advanced and applies to most contexts (the only exception being verbs in negative assertions).

I believe that in sentence (7.9), as well as in its alternative version in (7.11), the focus is just *the book*, as the corresponding English *it*-cleft translation makes clear. In sentence (7.10), on the other hand, what is focal is that same idea *together with the negative operator*. In other words, one could say that the focus is the ‘negative idea’ *not the book*, as it were, as, once more, the corresponding *it*-cleft in English makes clear. In other words, in (7.9) the negative operator is part of what is assumed by the speaker, but in (7.10), on the other hand, the negation is *part of* what is ‘new’ and thus salient. I believe it is more accurate, however, to say that the idea *the book* is the focus, with the negative being an operator over the whole assertion.

One ‘problem’ with the sentence in (7.10) is that if the object is the focus element, then in this construction the focus element necessarily cannot appear where we expect

foci to appear in Basque, namely in rheme-initial position, but rather in a postverbal position of sorts: truly postverbal when the verb is synthetic and verbally ‘embraciated’ when the verb is analytic.

One may wonder whether there could be an alternative for negative assertions which was more consistent with the focus positioning found in all other types of asserted clauses. The problem seems to be the requirement that the negative operator be to the left of the rheme (including the focus constituent) that it has scope over. Still, one might have expected the sentence in (7.10) to have the more iconic forms that we can see in the non-sentences in (7.12), with the focus element in rheme initial position, preceded by the negative operator, with or without non-finite verb inversion.

- (7.12) a. * Jonek [RHEME {ez liburua} du irakurri]
Jon:ERG not book:DEF he.has.it read:PFV
b. * Jonek [RHEME {ez liburua} {irakurri du}]
Jon:ERG not book:DEF read:PFV he.has.it

This solution would indeed be more iconic, or at least more consistent with the focus positioning system in Basque affirmative assertions. However, it seems that the unmarked or default situation in negative assertions, namely the one in which the finite verb acts as the focus constituent, has resulted in the phonological and lexical fusion between the negative particle and the finite verb, *even in contexts in which the finite verb is not the focus constituent*. This would be further motivated by the fact that the negative operator always has scope over the whole rheme, and not just over the focus element, and the finite verb seems to be able to stand for the whole rheme, as we have seen. Notice that the same thing is true in English for instance, since the negative particle (*not*) attaches or cliticizes itself to the finite verb, even when the finite verb is not the focus.

Thus, I believe we can say that the existence of a left-peripheral negation operator, despite being iconically motivated, brings about inconsistencies in a language with preverbal focus, inconsistencies which must be dealt with. This is, no doubt, the reason why many verb-final languages have rheme-final negative ‘particles’ (such as those derived from negative verbs), rather than rheme-initial ones. This seems to be a clear-cut case of ‘competing motivations’.

I should note, however, that there are certain specialized constructions in Basque (as well as in English), in which the negative particle can go next to the focus complement rather than next to the finite verb. Thus, in English, for instance, the negative ‘particle’ goes with a focus complement for instance when the focus element is an indefinite pronominal in rheme-initial position, such as *nothing*, *nobody*, *never*, e.g. *Nobody came*. The negative operator can also (optionally) go with a postverbal focus complement when this element is extraposed and an alternative is overtly provided in the following phrase, e.g. *I read, not a book, but a newspaper*.

Likewise in Basque, there are some constructions in which the negative operator is placed next to a focus complement, such as, for instance, some imperative assertions which lack a finite auxiliary verb. (These imperatives are nonetheless clearly asserted since they express commands and are ‘imperative assertions’.) In them the negative particle *ez* still goes in rheme-initial position but, if the focus is a complement, the negative word precedes it. The sentence in (7.13B) exemplifies this phenomenon (*Ez eni galda* “Don’t ask ME”) (from Oyharçabal 1985a:111, used to exemplify something else).¹⁰

- (7.13) A: Berant nizateke. Zer tenoretan da Pariseko trenara?
 late I.could.be what time/hour:LOC it.is Paris:GEN train:DEF
 B: **Ez eni galda**, Tolosarat johan bainiz, ez Pariserat.
 not I:DAT ask:ROOT Tolosa:ALL go:PFV because:I.am, not Paris:ALL
 “A: I might be late. What time does the train to Paris leave? B: **Don’t ask me**,
 I just came on the train to Tolosa, not the one to Paris.”

A very similar sentence, from the Spoken Basque Corpus, can be seen in (7.14), a sentence which was uttered by a listener to a story-teller who was being distracted by some children playing outside.

- (7.14) 93C1B04
 120 Aber! Peio! Ez aieri begira,
 let’s.see Peio not they:DAT look:ROOT
 “C’mon! Peio! Don’t look at them!”

These examples are rather exceptional, however, for as we will see later on, in imperative assertions, unlike in declarative ones, the focus constituent is by default the verb itself, unless a complement is significantly more focal than the verb. Thus an imperative sentence such as the one in (7.13B) would be expressed as the one in (7.15) below if the complement’s idea wasn’t contrastive (cf. *Don’t ask me* vs. *Don’t ask me*) (see Section 7.7 below for a discussion of imperative sentences in Basque).

- (7.15) Ez galdetu niri
 not ask:PFV I:DAT
 “Don’t ask me.”

(Notice that both of these sentences in (7.13B) and (7.15) happen to come from northern dialects, the same dialects which allow the negative particle *ez* to go with the non-finite verb, as opposed to the finite verb in the type of sentence in which the negation is presupposed, as in (7.11) above.)

There is also a minor construction, at least in the repertoire of some speakers, in which the negative element goes next to a focus complement and not next to the finite verb. This construction has the form [*ez X, ez Y*], where X and Y are whole rhemes (assertions) and corresponds to the English construction [*neither X, nor Y*] construction. We can see an example of this construction in (7.16) below (from Euskaltzaindia 1987b:492). Here the subjects are contrastive foci and precede the verb.

- (7.16) **Ez** berak egiten du, **ez** guri egiten uzten digu.
 not he:ERG do:IMPFV he.has.it, not we:DAT do:IMPFV leave:IMPFV he.has.it.to.us
 “Neither does he do it, nor does he let us do it.”

Notice that here the two parts of the verb do not invert as in normal negative sentences. In the example of this construction in (7.17) below, we can see that the finite verb in the second part of the construction may be elided if it is identical to the one in the preceding clause (notice that, unlike in the previous example, these two assertions share a covert topic) (from Euskaltzaindia 1987b:492).

- (7.17) **Ez** galdu da, **ez** galduko (da)
 not lose:PFV it.is not lose:FUT (it.is)
 “It hasn’t gotten lost, and it won’t get lost.”

Notice too that here the focus is not a complement, but rather the whole rheme (consisting of a single verb), but still there is no verb-auxiliary inversion.

There is also another context in which a negated focus constituent precedes the finite verb, and the two parts of the verb do not invert with each other, with the structure which I said was ungrammatical under normal circumstances: (*Topic*) *neg-Focus Verb Aux* (cf. 7.12b). This possibility is available, at least for some speakers, when there is a second operator besides the negative operator, such as in *not only* (cf. *ez soilik, ez*

bakarrik). The alternative in which the focus is inside the verbal brace is perhaps more common, however. We can see an example of the latter situation in (7.18c).

(7.18) *Argia*, May 19, 1994

a. Sugearen pozoina.

snake:DEF:GEN poison:DEF

b. Historian zehar sugeek beldurra bezala harridura ere sortu-izan-dute. ...

history:DEF:LOC through snakes:DEF:ERG fear:DEF as amazement:DEF also they.have.engendered

c. Sugeek ez-dioe **jendeari bakarrik** sortzen nazka. Edozein animalia taldek ...

snakes:DEF:ERG not-they.have.it.to.it people:DEF:DAT only cause:IMPFV aversion/disgust:DEF any animal group:ERG

“(a) The poison of snakes: / (b) Throughout history, snakes have been the source of fear as well as amazement. ... ¶ (c) Snakes do not cause aversion only to people (snakes cause aversion not only to people). Any group of animals ...”

This sentence could, at least for some speakers, be rendered with the focused element in preverbal position, as in (7.19) below.

(7.19) Sugeek [_{RHEME} <ez jendeari bakarrik> sortzen dioe nazka].

snakes:DEF:ERG not not people:DEF:DAT only cause:IMPFV they.have.it.to.them disgust:DEF

“Snakes don’t only disgust people.” (Cf. “Not only to people are snakes disgusting”, with subject inversion)

In Spanish the same contrast is found. Thus, whereas the negative particle cannot precede regular focus subjects (* *No a la gente dan asco*), this is possible when these two operators are working together (e.g. √ *No solo a la gente dan asco*).

For comparison, note that if the focus in this sentence was not inside the scope of the negation, the negative would precede the finite verb and the two parts of the verb would invert with each other, as we can see in (7.20).

- (7.20) Sugeek, jendeari bakarrik ez diote sortzen nazka.
 snakes:DEF:ERG people:DEF:DAT only not they.have.it.to.them cause:IMPFV
 disgust:DEF
 “Only to people are snakes not disgusting”

7.2.4 Can an assertion have multiple foci?

As I said earlier, Lafitte (1962/1944) noted that regular complements of a negative assertion, typically contrastive ones, could be foci.¹¹ He further entertained the thought that both the contrastive element *and* the negative operator are foci. Although I do not think that this is the right solution for the construction we have been dealing with, we may want to ask whether two-focus constructions are possible at all, whether in Basque or in any other language.

Content questions with multiple question words for instance might seem to contain multiple foci. This is what Osa argues for, against the opposing arguments of de Rijk (1978), Rebuschi (1983), and Eguzkitza (1987/1986) (cf. Osa 1990:73-74). Thus, Osa would argue that in a sentences such as the question in (7.21Q), both *nork* “who:ERG” and *nori* “who:DAT” are foci. The same thing would be true of an answer to such a question, such as the one in (7.21A).

- (7.21) Q: **Nork nori jantzi-dio txapela?**
 who:ERG who:DAT dress:PFV-s/he.has.it.to.him/her hat:DEF
 A: **Aitak semeari** (jantzi-dio).
 father:DEF:ERG son:DEF:DAT (dress:PFV-he.has.it.to.him)
 “Q: Who put the hat on whom? A: The father (put it) on the son.”

To the extent that these very special assertions seem to have two accents, one per ‘focus’, it would seem that there are indeed two foci. On the other hand, we may also view them as complex two-part foci, representing complex concepts or ideas and constituting a

single, though complex, focus constituent. Complex directionals would be another type of similarly acceptable two-part focus constituents, cf. *From where to where did you go?*

Another example of what may be a double focus construction would be, for instance, a question with a question phrase and an additional contrastive element, such as the one in (7.22).

- (7.22) **Nork irakurri-ditu bi liburu?**
who:ERG read:PFV-s/he.has.them two book
“Who has read two books?”

Does this mean that this sentence has two foci? Notice that this question would have to be uttered in a context in which the proposition that someone has read a number of books different from two is accessible (e.g. the question *Who read one book?* might have been previously asked). This might imply that the question word itself, being accessible, is not the focus here.

Notice too that in this example only the contrastive object, and not the question word, would be accented, suggesting that only that is the focus. There is one context, however, in which it seems possible for question such as the one in (7.22) to receive two accents, and supposedly contain two foci, namely if the ‘second focus’ is extraposed, i.e. in a separate intonation unit, as in (7.23).

- (7.23) Nork irakurri-ditu, bi liburu?
who:ERG read:PFV-s/he.has.them two book
“Who read, two books?”

This, of course, is a situation which is analogous to the focus extraposition construction which we saw in earlier chapters and which in some ways seems to contain two foci. As I

mentioned then, one context in which such a construction is used seems to be when the verb and another element are focal (such as when the verb is semantically complex).

Yet a different way of coding the same meaning would involve coding what in the previous examples is a contrastive focus as a contrastive topic, taking advantage of the fact that the idea it represents can be easily accommodated in this context, cf. (7.24).

- (7.24) Bi liburu nork irakurri-ditu?
two book who:ERG read:PFV-s/he.has.them
“Two books, who has read?”.

This type of topicalization is rather common and quite ‘unmarked’ in Basque, as we saw in Chapter 4 and will see again below.

The same topicalization strategy is available for elements which might otherwise be the focus of a negative assertion, such as the one we saw in (7.10) above, and repeated here in (7.25) for convenience.

- (7.25) Jonek ez-du liburua irakurri.
Jon:ERG not-he.has.it book:DEF read:PFV
“Jon didn’t read the book.”

When such an element is accessible and contrastive, it can be ‘extracted’ from the assertion and be turned into a setting/topic, a ‘contrastive topic’, as in (7.26) below.

- (7.26) Jonek liburua ez-du irakurri (egunkaria baizik).
Jon:ERG book:DEF not-he.has.it read:PFV (newspaper:DEF rather/but)
“Jon didn’t read the book (but the newspaper).”
(= “The book is not what Jon read, but (rather) the newspaper”)

The contrastive element here is not the focus of the assertion, but rather a topic. To a large extent it is the speaker’s choice whether to make a contrastive element the topic or the focus, as long as it is accessible, and different languages might display different

preferences. This is a way in which languages such as Basque may deal with the problem arising from having foci come in a different position in negative assertions from where they are expected (in affirmative assertions). I will return to this topicalizing strategy in negative assertions below.

7.2.5 Focus position in negative assertions

A number of investigators have been concerned about the issue of the exact location after the verb that postverbal focus complements are found in Basque whenever there is more than one complement inside the rheme (which is not very often in actual speech, as we shall see).

According to Lafitte (1962/1944) the postverbal focus in these cases is the one immediately following the negated auxiliary, cf. Table 7-2. According to Osa, however, this may be the preferred position in eastern dialects, but he believes that, in the west, when there is more than one complement to the right of the finite verb, the focus is (preferably) the one nearest the non-finite verb, i.e. the one farthest to the right (Osa 1990:200), cf. Table 7-2.

<i>Author</i>	<i>Focus Position</i>
Lafitte	Neg-Aux Focus ... Verb
Oyharçabal, Osa	Neg-Aux ... Focus Verb Neg-Aux ... Verb, Focus

Table 7-2: Complement focus position in negative assertions, according to different authors.

Oyharçabal (1985a) too disagrees with Lafitte's assessment that the focus is the leftmost one after the verb, even in north-eastern dialects. Like Osa, he thinks that, just

like in affirmative sentences, the most salient element is the rightmost one (right before the non-finite verb) (cf. Oyharçabal 1985a:110).¹² But he also claims that in some dialects at least the focused element may even be outside the ‘brace’, i.e. behind the non-finite verb, just as in affirmative sentences (Oyharçabal 1985a:111).¹³ It seems to me, however, that these foci do not merely *follow* the non-finite verb, but rather are *extraposed*, much like we saw happens in affirmative assertions. In other words, they are outside the rheme proper.

But restricting ourselves to foci inside the brace, it may very well be that there is no such a thing as a single focus ‘position’, as there is in the case of preverbal foci in affirmative (and some negative) assertions. Osa believes that placing the focus closest to the main (non-finite) verb is only a tendency or preference, not a requirement. Thus he notices that a very focal, optional adverbial would be the focus no matter where it was located after the verb, as we can see in example (7.27) (from Osa 1990:210, ex. 213).

- (7.27) Aitak ez-du **bere gogoz** untzia leihotik aurdiki.
 father:DEF:ERG not-he.has.it his mind:INSTR glass:DEF window:DEF:ABL
 throw:PFV
 “Father didn’t throw the glass out the window **on purpose**.”

And, as Osa also mentions, the focus could go basically anywhere in the sentence other than in rheme-initial position. That is, it can go anywhere inside the brace, as in (7.28a), after the non-finite verb, as in (7.28b), which is what I would call extraposed, or, as I mentioned earlier, it can be topicalized in clause initial position, as in (7.28c) below (cf. Osa 1990:207, ex. 200).¹⁴

- (7.28) a. Ez-dugu **ardorik** edaten, ura baizik.
 not we.have.it wine:PART drink:IMPFV water:DEF but/rather
 b. Ez-dugu edaten **ardorik**, ura baizik.
 not we.have.it drink:IMPFV wine:PART water:DEF but/rather
 c. **Ardorik**[,] ez-dugu edaten, ura baizik.
 wine:PART not-we.have.it drink:IMPFV water:DEF but/rather
 “We don’t drink wine, we drink water.” (For option c, cf. “Wine, we don’t drink, we drink water.”)

Osa does not mention that the partitive phrase in (7.28c) is outside the rheme and is the *topic* of the assertion, but there is no doubt that that exactly what it is, much like it would be in the equivalent sentence in English (cf. *Wine we don't drink*). As I said, this type of topicalization is quite common in Basque when the referent is accessible (in this case it is a generic referent, which are typically quite accessible, especially in context).

(Note that, with different intonation, the partitive in the sentence in (7.28c) could also be the (rheme-initial) focus and be outside the scope of negation. The meaning of this assertion, however, would be very different from the meaning this sentence is here, since the ‘presupposition’ would be quite different, cf. *Wine (is what) we don't drink.*)

It is clear that the fronted element in (7.28c) is indeed outside the rheme. This is even more obvious when the accessible parts of a negative rheme are ‘elided’, as it often happens in replies to polarity questions. Thus, in answering the question *Do you drink wine?*, or to *What do you drink?*, or *What will you have to drink?*, the fragment analog of the sentences in (7.28a&b), with all but the negative operator and the focus elided, would be the one (7.29a) below. This fragment analog of the sentence with topicalization in (7.28c), on the other hand, would still have the partitive phrase as the topic and the negative particle would stand for the whole predicate, as we can see in (7.29b).

- (7.29) a. *Ez ardorik* (not wine:PART) “Not wine”
 b. *Ardorik ez* (wine:PART not) “Not wine.” (Cf. “Wine I don’t.”)

The topicalization of *ardorik* “wine” is made possible by the fact that that is easy enough to accommodate even if it is not discourse-given (it is, after all, a common enough drink). Actually, the answer in (7.29b) might be more likely if wine is present in the physical context, and thus more accessible. As we will see below, such a topicalizing strategy in negative assertions is quite common in Basque, perhaps as a compensatory strategy for the inconsistencies introduced by a rheme-initial negative operator.

7.2.6 Postverbal elements in negative clauses in the spoken corpus

The issue of where exactly the postverbal focus element is located is rather moot in practice, since the number of clauses with more than one element in postverbal position, or inside the ‘brace’ between the finite and the non-finite verbs (if there is a non-finite verb), seems to be close to zero in actual speech.

Verbose sentences with three complements inside the rheme, such as the one we saw in (7.27) above, are found in grammar books and occasionally in writing, but in speech they are quite rare, as we can tell from looking at the data in the Spoken Basque Corpus, which we can see summarized in Table 7-3 below.¹⁵

Brace status	Structures	#	%	Corpora
Brace-irrelevant	$ezV_f, ezV_f V_m, ezV_f X$	108	67.5%	19H, 27I, 50D, 12L
Intact brace	$ezV_f X V_m$	25	15.6%	3H, 3I, 19D
Exbraciation	$ezV_f V_m X$	25	15.6%	10H, 8I, 4D, 3L
Partial exbraciation	$ezV_f X V_m Y$	2	1.3%	2I
TOTAL		160	100.0%	

Table 7-3: Finite negative statements in the Spoken Basque Corpus. H = *Hendaia*, I = *Ikasbide*, D = *Deustu*, L = *Lur*.

As we can see, the vast majority of negative assertions either have no postverbal complements or have at most one. Only two negative sentences have 2 rhematic complements. And quite often the verb is synthetic and thus there is brace at all. Let us look at the different groups listed in Table 7-3 in turn.

Brace-irrelevant negative sentences. As we can see in Table 7-3, in $\frac{2}{3}$ of the clauses in the spoken corpus there is no brace, nor could there be one, for either there is no non-finite verb or there are no elements to be braced.¹⁶ Of these 108 sentences, 46 have the order [R_{heme} ezV_f V_m], 12 have the order [R_{heme} ezV_f], and the rest have only a complement after the finite verb [R_{heme} ezV_f C]. In (7.30) and (7.31) below we see two examples of such sentences, the first one with a partitive object and the second one with a non-finite verb, but no other complement

(7.30) 93C2A09
 267 Eta **ez-dauka** dirurik ero.
 and not-he.has.it money:PART or
 “And he doesn’t have money, or whatever.”

(7.31) 93C1B06
 212 eta **ez-diote eraman**
 and not-they.have.(it).to.him carry:PFV
 “and they don’t take him”

Negative sentences with a brace. In the corpus, 25 negative statements have a brace with a single element inside:¹⁷ in 13 of these clauses the braced element is a regular nominal complement (1 A, 5 O, 7 other); in 5 clauses the argument is an indefinite pronoun (e.g. *ezer* “anything”); and in the other 7 it is a verbal complement of an aspectual or modal verb (1 with *behar* “need” and 6 with *nahi* “want”). In the following 4 sentences we can see examples of this type of sentence with a brace. In (7.32) we find a

predicate nominal in the brace, which is intrinsically rhematically salient, as we have seen.

- (7.32) 93C2A06
5 **ez-da** pelikula bat **izan**,
not-it.is movie one be:PFV
“it wasn’t a movie,”

In (7.33) we find a verbal complement of *nahi* “want”.

- (7.33) 93C2B06
184 kar- Txar- Txarlotek **ez-zuan** a- j- atera **nai**,
jail- Chap- Chaplin:ERG not-he.had.it go.out:PFV desire
“Chaplin didn’t want to go out,”

In (7.34) there is an allative complement inside the brace.

- (7.34) 93C2B08
86 **ez-da**= bere zeldara **sartu**,
not-het.is his cell:DEF:ALL enter:PFV
“he didn’t go into his cell.”

Finally, in (7.35) we see a non-referential object nominal in the partitive case (*kasorik* “attention”).

- (7.35) 93C2A10
8 Ta **ez-dio** kasoik **eitten** ta ezer, -
and not-he.has.it.to.him attention:PART make:IMPFV and anything
“And he doesn’t pay any attention to him or anything”

It seems that non-referential nouns, and in particular non-referential ‘indefinite pronouns’ must be inside the brace and cannot be exbraciated, as Laka (1993) for instance has noticed, as we can see in (7.36b). Focus indefinite pronouns, however, *can* be fronted to rheme-initial position, as we can see in (7.36c) (from Laka 1993, ex. 18b, fn. 14).

- (7.36) a. *ez da inor etorri*
 not s/he.is anyone come:PFV
 b. ≠ * *ez da etorri inor*
 not s/he.is come:PFV anyone
 c. = *inor ez da etorri*
 anyone not s/he.is come:PFV
 “Nobody came”

In (7.36c) intonation cues show that *inor* “anyone” is rheme-initial and focus. It certainly couldn’t be topicalized, since it does not express a topical idea. In other words, this sentence violates the rule that the negative particle *ez* must be to the left of the focus if it has scope over it, reflecting perhaps an earlier state of affairs in the language.

Partitive phrases with *generic* referents, on the other hand, may indeed be topicalized, as we already saw and as we can see in (7.37) below. The partitive noun is not rheme-initial, but pre-rhematic and thus a setting element. As I said earlier, topicalization is warranted by the fact that generic referents are typically quite topical.

- (7.37) 93C2B01
 409 *ta dirurik etzeuken.* -
 and money:PART not:he.had.it
 “and he didn’t have any money” (cf. “and money he didn’t have (any)”)

Negative sentences with exbraciation. In 25 negative statements in the spoken corpus the rhematic complement is outside the brace, i.e. is comes after the non-finite verb.¹⁸ This happens often with the clausal complements of aspectual and modal verbs such as *nahi* “want” (8 cases) and *behar* “must” (1 case). In other words, many of the exbraciated complements are sentential: 3 *-tzea* nominal clauses, 3 *-tzen* adverbial clauses, 1 embedded question, and 4 finite completive *-la* clauses. In addition, there are two clauses with partial exbraciation of elements.¹⁹

Below we can see two typical examples of these sentences displaying exbraciation. In (7.38:209) the exbraciated element is a simple allative phrase.

(7.38) 93C1B06
209 eta ez-du era- **ez-dute eramaten** kartzelara.
and not-he.has.him carr- not-they.have.him carry:IMPFV jail:DEF:ALL
“and he doesn’t tak- they don’t take him to jail”

In (7.39) the exbraciated element is a clausal complement, albeit a brief one (*jaten* “eat”).

(7.39) 93C3A10
71 =eta **ez-dio uzten** jaten,
and not-he.has.it.to.him leave:IMPFV eat:IMPFV
“and he doesn’t let him eat”

One of the exbraciated elements is even an indefinite pronoun (*ezer*), something which violates Laka’s principle about the position of these special non-referential particles.

In all these cases of exbraciation it seems to me that the complement is indeed the focus, but the focus constituent is obviously extraposed. Thus exbraciation is nothing more than the extraposition of focus constituents, parallel to the phenomenon that we saw in affirmative statements in Chapter 6 and elsewhere. This opens some interesting possibilities for analyzing other languages where exbraciation is attested, such as earlier stages in the history of English (cf. Chapter 1).

7.2.7 Some conclusions

If we compare the sentences with exbraciation and those with intact braces, we can see some interesting trends. First of all, the younger speakers are more likely to exbraciate than older (perhaps more fluent) ones: 76% of the brace-intact clauses are from adult speakers and 72% of the exbraciated examples are from the younger speakers.

Another important tendency which the data shows is that sentential complements are exbraciated more often than not, for all speakers, and regardless of the complexity of the exbraciated clause, as we saw in example (7.39) above. These are the same kind of cases in which we find focus extraposition is quite common in affirmative assertions. This reinforces the idea that all examples of exbraciation are actually examples of focus right-dislocation (extraposition), and that that is the origin of exbraciation.

Another interesting finding is that no negative statements in the corpus has a brace with more than one element inside. I think it is safe to assume that clauses with more than one rhematic element after the verb are special cases of the negative brace construction. Thus it is not surprising that speakers may have different preferences as to the exact location of additional extra-rhematic (non-focus) elements, and even the embraciated focus constituent, since the grammar doesn't seem to have come up with a uniform solution for such an uncommon situation.

I would like to conclude this section with some remarks on the effect that the Basque negation construction, with the finite verb in rheme-initial position and obligatory main verb postposing, has on the general word order of the language. To begin with, the postposing of the main (non-finite) verb in Basque sentences has the obvious effect of inverting the order of the non-finite verb and the auxiliary. Since in the majority of negative assertions in discourse there is no complement inside the brace, either because there isn't one, or because it has been exbraciated, very often the resulting order is *Neg-Aux-Verb*, an order which is strongly correlated with VO languages, as we saw in Chapter 1.

Furthermore, when the finite verb is not the focus of the assertion, the focus is in a very different position from where it would be in an affirmative assertion. If there is no exbraciation, the focus is between the two verbs, so in a sense it is postverbal (as in VO languages) and preverbal (as in OV languages) at the same time. With focus exbraciation the resulting clauses look indistinguishable from similar clauses in VO languages. If exbraciation were to become obligatory, or even more prevalent, for all speakers it would have the effect of turning negative sentences into a part of the grammar which behaves like we expect a VO language, and not an OV one, to behave, one which might naturally lead to the extraposed focus being realized as postverbal.

This, of course, would only apply to foci which are inside the scope of the negative operator, for when they are outside they are fronted to rheme-initial position, as we have seen, and as we can see in the example in (7.40).

(7.40) *Behi euskaldun baten memoriak*, p. 121

a. **Horregatik** ez-zituzten harrapatzen,

that:GEN:MATTER:ABL not-they.had.them catch:IMPFV

b. *elkarren artean* sistema bat zeukatelako

each.other:GEN between:DEF:LOC system one they.had.it:BECAUSE

“That’s why they didn’t catch them, / because they had a system (going) between the two of them.” (cf. For that reason they didn’t catch them)

But notice that such sentences with preverbal foci outside the scope of the assertion are also found in VO languages, such as English or Spanish for instance (cf., English *For that reason they didn’t catch them*, and Spanish, *Por eso no le cogieron*). However, most foci in negative sentences are indeed under the scope of negation, and thus follow the negative finite verb, and in fact there aren’t any examples such as (7.40) in the Spoken Basque Corpus.

Of course, one may argue that this order affects only negative sentences, which are but a minority of the sentences of the language. However, this type of inversion has the potential of affecting the word order of a language in more significant ways. Thus, for instance, as we have seen, in north-eastern Basque dialects non-finite verb inversion also takes place in (counter-presuppositional) assertions with heavily contrastive affirmative polarity focus, with the affirmative prefix *ba* instead of the prefix *ez* “not” before the finite verb (cf. Section 7.4 below). If such an ‘emphatic’ construction were to become less marked, i.e. if the use of this construction became more generalized, something which may have happened in other languages, the ‘basic’ word order of the language would suffer a much more drastic change.

All by itself, this the ordering found in Basque negative assertions (and in emphatic affirmative assertions in some cases) probably does not constitute a major mechanism for word order change. However, together with the other changes which seem to be going on in Basque, such as the high frequency of focus extraposition, and which have similar surface results, these constructions might conceivably be part of a ‘conspiracy’ to alter the primary position of focus constituents in Basque.

7.3 TOPICALIZATION IN NEGATIVE CLAUSES

7.3.1 Fronting in Basque negative clauses

In Chapter 4 we saw that Basque displays a tendency to ‘topicalize’, i.e. to turn into settings (including topics), elements which would be inside the rheme in other languages.²⁰ We saw, for instance, that the additive particle *ere* “also” is most of the time a ‘topicalizer’, i.e. an operator which extracts an element of the rheme and turns it into a setting element, as long as it can be accommodated (which is the minimal requirement for topic role, as we saw). Unlike similar particles in other languages, which ‘focus on’, or modify, foci inside the rheme, this particle, and others like it, turn a salient clausal element into a salient topic, instead of a salient focus.

Sometimes topicalization applies to elements which are not necessarily very topical or accessible, but which can be easily accommodated as settings or topics. We can see an example of such an unexpected topicalized phrase in (7.41) below.

- (7.41) *Behi euskaldun baten memoriak*, p. 165
Antiaju Berde eta bere laguntzaileak **behin ere** ez-dira joango hara.
Antiaju Berde and his helpers:DEF once also not-they.are go:FUT there:ALL
“AB and his helpers will never go there.” (Lit. “... once even they will not go there.”)

In this example the setting element is the temporal adverbial *behin* “once, one time”, with the additive topicalizer *ere* “also; even”. Very often the topicalized element is an argument of the verb itself, i.e. the topic, as we will see below. This is also perhaps an extreme example of topicalization of an element which is low in topicality.²¹

In the previous section we saw that in negative clauses, what could be the focus of a negative assertion, under the scope of the negative operator, could be topicalized

instead, with the negation being predicated of that referent. We saw examples of this phenomenon with nominals which had generic referents, i.e. referents which, crosslinguistically, are topical enough to be turned into settings since they can be easily accommodated. And, of course, in Chapter 4, we saw examples of additive elements which in Basque are (contrastive) topics/settings, rather than (contrastive) foci. Here we will see that this phenomenon is quite generalized and grammaticalized in Basque. In addition, it is perhaps especially common with negative sentences.

In (7.42b) below we can see another example of exactly the type of sentence that I have in mind, this one also from a novel. This is a negative existential sentence, of the type we saw in Chapter 5, with a topicalized non-referential ‘subject’ which refers to an accessible idea.

(7.42) *Behi euskaldun baten memoriak*, p. 161-2

a. —*Baina gu ez-gara behi makalak*, La Vache—*protestatu-nuen*.

but we not-we.are cow weak:DEF:PL La Vache protest:PFV-I.had.it

b. —***Makala ez den behirik ez dago!***—*hotsegin-zuen orduan zakarki*.

weak:DEF not it.is:REL cow:PART not it.is call:PFV-she.had.it.then rough:ADV

“‘But we aren’t weak cows, La Vache,’ I protested. / ‘There aren’t any cows that aren’t weak!,’ she cried then brusquely.” (Cf. “Cows that aren’t weak do not exist.”)

The partitive nominal with a generic referent is clearly in pre-rhematic topic position and the predicate (*ez dago*) is *about* that topic. Notice that there is a possible English version of this sentence in which the ‘subject’ is the topic instead of being rheme-internal as well, though this might not be as likely an option for an English speaker as it would be for a Basque speaker, or at least for some Basque speakers.

Whether a certain element of the clause is topicalized or not is to some extent the speaker’s option (as long as the referent can be accommodated). But it is possible that

languages differ as to the likelihood that the topicalizing strategy will be used, or the likelihood of using it in some contexts. I believe that it is very likely that such differences exist and that Basque is perhaps more likely to use this strategies than other languages. In other words, topicalization may be relatively unmarked in some contexts, such as perhaps in contexts in which the assertion would otherwise not have a topic and perhaps also in negative assertions. It is also possible that this strategy is more common in languages with default preverbal focus (“OV languages”).

In the example in (7.43a) below the fronted element is a non-finite complement clause in the allative case (cf. *To say more she did not dare*). What makes this phrase accessible (easy to accommodate) in this context is that one might have expected the nun Pauline Bernardette (the referent of the absolutive argument, the covert topic) to have added something to what she had already said.

(7.43) *Behi euskaldun baten memoriak*, p. 173

a. **Gehiago esatera** ez-zen ausartu,
more say:NOMIN:ALL not-she.was dare:PFV

b. baina ulertu egin-nion.
but understand:PFV do:PFV-I.had.it.to.her

“She didn’t dare *say more*, / but I understood her very well.” (Cf. “To say more, she didn’t dare.”)

Actually, if the reader had not anticipated the possibility that the nun could add something to what she had already said, this type of sentence forces the speaker to entertain such a possibility and its reasonableness. In other words, the use of topicalization may have rhetorical consequences. (This clause, of course, could have also been postverbal.)

Very often the topicalized phrase is the only setting, or topic, that the sentence has, as we saw in the existential sentence in (7.42b) above. In the example in (7.43a), the

clause already has an external topic and the predication is somehow about both topics (cf. Chapter 3). In (7.44b) below we can see another example of a negative existential sentence with a fronted element which could not have been active or even accessible, and that only with a dose of good will can be accommodated by the reader as the topic of the (non-)existential predicate.

(7.44) *Behi euskaldun baten memoriak*, p. 162

a. *Badakit laguna galtzea oso mingarria dela,*

EMPH:I.know.it friend:DEF lose:NOMIN:DEF very painful:DEF it.is:COMPL

b. *baina paseatuz sendatzen ez-den bihotzeko minik ez-dago.*

but walk:PFV:ADV cure:IMPFV not it.is:REL heart:DEF:GEN pain:PART not-it.is

“I know that losing a friend is very painful, / but there is no heartache that a nice walk can’t cure.” (Lit. “heartache that is not cured by walking doesn’t exist.”)

Here, as in (7.42b), an idea, the topic, is first set up, or announced, just to immediately afterwards have its existence denied by the speaker.

Some of these frontings sound somewhat acceptable, albeit forced, when translated literally into other languages, such as English, but most of the time they seem to be much more marked in other languages than they are in Basque. Below is yet another example with a negative existential predicate, which, as I said, is one of the most common types of predicates involved in these negative sentences with fronting. In this case the English version could also have a topic-comment structure, as we can see.

(7.45) *Behi euskaldun baten memoriak*, p. 177

- a. Memoriok irakurrita ere,
memoirs:DEF:PROX read:PFV:ADV also
- b. esplikazio hori iruditzen-zait serioena.
explanation that seem:IMPFV-it.is.to.me serious:MOST:DEF
- c. Baina, horixe, **ondo jakiterik** ez dago.
but that:EMPH well know:NOMIN:PART not it.is

“After reading these memoirs, / that explanation seems to me the most serious one. / But, of course, it’s not possible to know for sure.” (Cf. “Knowing for sure is not possible.”)

As we saw in Chapter 4, this type of ‘unexpected’ topicalization with existentials is also found with affirmative existentials, but mostly in northern dialects. Its use with negative clauses seems to be more widespread.

7.3.2 The expression of restrictive *only* operator in Basque

One particularly interesting context in which this fronting strategy is employed in Basque is in the expression of what is sometimes called the **restrictive operator** *only*, an operator which in languages such as English typically quantifies an element of the clause which is the focus of the assertion, which is why such particles are often considered to be *focalizers*.

Assertions with the restrictive operator *only* acting on the focus, as well as on the whole rheme, are often semantically equivalent to sentences in which the opposite is asserted about of the set-complement of the focus idea. Thus, for example, expressions of the type [*Only X did Y*] are synonymous with expressions such as [*Besides X others didn't do Y*]. In Basque, expressions of this latter type are perhaps a more common way of expressing this restrictive operator meaning than those of the former type, by means of the pronoun (or modifier) *beste* “other”. Basque also has an equivalent of the English

focalizing operator *only* for expressing this function, namely *soilik* (or *bakarrik*) “only”, but the topicalizing alternative seems to be more common, or at least more basic and native than the focalizing one. In fact, the focalizing alternative may have arisen as a calque from Romance, given the form of the focalizing particles *soilik* and *bakarrik*.²²

The idea expressed by a topicalized constituent in these constructions can be quite ‘new’, to the point of seemingly not being at all accessible, other than in the sense of ‘fitting into the context’ or being ‘accommodated’ (after the fact). In these cases, as we might have expected, the topicalized element is actually dislocated, into its own intonation unit, as is typically the case with settings (including topics) which are not accessible. In (7.46b) we see one such example from a novel. Since the idea here is new, it is presented in a separate intonation unit, as the comma following it shows. Then, clause internally, we can see the word *besterik*, “other”, “anything else”, “something else” outside the rheme (*beste* “other” with the partitive ending *-(r)ik*).

(7.46) *Behi euskaldun baten memoriak*, p. 123

a. Ixil-ixilik zegoen Balantzategiko haran guztia:

quiet quiet:ADV it.was Balantzategi:GEN valley whole/all:DEF

b. teilatutik erortzen-zen uraren txirraia,

roof:DEF:ABL fall:IMPFV-it.was:REL water:DEF:GEN sound:DEF

c. **besterik** ez-zen entzuten.

other:PART not-it.was hear:IMPFV

“The whole valley of Balantzategi was very quiet: / you could only hear the sound of the water that fell from the roof.” (Cf. “...the noise of the water that fell from the roof is all that could be heard.”)

Here we can see that the dislocated element acts as an clause-external setting for the negative assertion. The nominal *besterik*, expressing its set-theoretic complement (“things other than that”, “other things besides that”), is the clause-internal topic for the

assertion. This latter idea is quite accessible in this context since it is the converse of the idea expressed in the sentence-head.

In (7.47) below we can see another such example, from a different novel. Here *beste* is not used as a pronoun, but as a modifier in a noun phrase in the partitive case.

(7.47) *Otto Pette*

- a. Borrero usnaria akabatu eta handik aldegin,
hangman/executioner smell:DEF finish/kill:PFV and there:ABL take.leave:PFV
- b. **beste** pentsamendurik ez-neukan une hartan buruan.
other thought:PART not-I.had.it moment that:LOC head:DEF:LOC
“I could only think about killing that stinking hangman and getting the hell out of there.” (Lit. “To kill that stinking hangman and to get the hell out of there, / I couldn’t think about anything else at that moment.”)

As we can see this sentence can be rendered by a very similar sentence in English, even though that may not be the least marked, or common, way of expressing this meaning.

Most of the time, the idea expressed by the topicalized element is more accessible than it was in the last example, i.e. it is less unexpected or surprising. In those cases it seems that the setting nominal is much more closely integrated with the asserted clause and may even act as an internal (non-dislocated) topic. The phrase with *beste* in these cases also seems to be more closely integrated with the initial nominal, as we can see in (7.48a&b).²³

(7.48) *Behi euskaldun baten memoriak*, p. 124

- a. Baina etxeko txokoetan hautsa **besterik** ez-zuten ikusi,
but house:DEF:GEN corners:LOC dust/ash:DEF other:PART not-they.had.it
see:PFV
- b. eta basokoetan, berriz, zomorroa **besterik** ez.
and wood:DEF:GEN:PL:LOC, again/however/yes, insect:DEF other:PART not
“But around the house they only saw dust, / and around the woods, on the other hand, only bugs.” (Cf. “... other than dust we didn’t see anything;” “other than bugs, nothing.”)

It is probably too much to say that *besterik* here acts as a postposition, since the preceding nominal is a full NP (in the absolutive case), but the two seem to go together most of the time in a sort of external and internal topic blended structure. Notice, for instance, that when the contrasted element is postverbal, as it occasionally happens, the noun phrase and *besterik* remain together, as in (7.49b) and (7.50) below, acting not as the focus of the assertion, but as an antitopic.

(7.49) *Behi euskaldun baten memoriak*, p. 153

a. Mendi gailurrak aurrean eta mendi gailurrak atzean,
 mountain tops:DEF front:DEF:LOC and mountain tops:DEF behind:DEF:LOC

b. jira guztian ez-zegoen mendia **besterik**.
 turn whole:DEF:LOC not-it.was mountain:DEF other:PART

“Mountain tops in front and mountain tops behind, / all around there were only mountains.” (Cf. “There wasn’t anything but mountains.”)

(7.50) *Obabakoak*

Jadanik **ez-zaigu** abiatzea **besterik** geratzen
 already not-it.is.to.us depart:NOMIN:DEF other:PART remain:IMPFV

“Now we just have to get going.” (Cf. “Now, we don’t have anything left to do but to get going;” “Now, leaving is all that’s left to do.”)

These phrases seem to be acting as antitopics of emphatic assertions with rheme focus.

These two examples show, however, that what seemed to be two different constituents may act as a single grammatical (and semantic) unit.

This type of restrictive topicalization with *beste* is possible even when the focus is the verb. In this case the non-finite verb is topicalized and the verb *egin* “do, make” is used in its place. We can see an example of this in the second clause in (7.51). Here *entzun* “listen” is a topicalized verb.

- (7.51) Ez-du inoiz hitz-egiten, entzun **bestetik** ez-du egiten.
 not he.has.it ever word make:IMPFV listen:PFV other:PART not-he.has.it do:IMPFV
 “He never speaks, he only listens.”

Notice that the context makes the verb in the assertion quite accessible, since it belongs to the same semantic field as the verb in the preceding clause. When the ‘restricted’ element is the verb, the alternative with *bakarrik* or *soilik* “only” may be more common, with the verb in focus position, as we can see in (7.52), with *egin* verb-focus construction.

- (7.52) Ez-du inoiz hitz-egiten, entzun egiten du **bakarrik/soilik**.
 not he.has.it ever word make:IMPFV listen:PFV do:IMPFV he.has.it only
 “He never speaks, he only listens.”

Unlike in the previous example, here *entzun* “listen” is the focus of the assertion.

So far we have only seen examples of this construction in which the topicalized phrase is in the absolutive case and the *beste* phrase was in the partitive case. This indeed seems to be one of the most common situation, but others are also possible. Thus, for instance, in (7.53) below we see an example with an topic ergative phrase and an also ergative *beste* phrase (from Gurrutxaga *et al.* 1994:179).

- (7.53) Jainkoak **bestek** ez-du nik hainbat sufritu.
 God:DEF:ERG other:ERG not-s/he.has.it I:ERG as.much suffer:PFV
 “Only God has suffered as much as I have.”

And in the following sentence in (7.54) both phrases are in the locative case and are temporal phrases (*ibid.*).

- (7.54) Igandeetan **bestetan** ez-zen etortzen.
 Sundays:DEF:LOC other:LOC not-s/he.was come:IMPFV
 “She came only on Sundays.”

Besides *beste*, other restrictive operators may also be used in similar constructions. Zubiri and Zubiri 1995 mention the *baino ez* construction, *baino* being primarily used as the connector of comparative constructions (similar to English *than*, but it follows the standard of comparison and precedes the compared element, cf. Chapter 2). Below in (7.55) and (7.56) we see two examples of this construction, the former with a fronted temporal phrase (*once*) and the latter with a fronted comitative argument (*with Josu*) (from Zubiri and Zubiri 1995 (p. 200-201, exs. 2.1206, 2.1208).

(7.55) Behin **baino ez**-dut ikusi eta instant batez.
 once but not-I.have.it see:PFV and moment one:ADV
 “I have only seen it once and only for a second.”

(7.56) Josurekin **baino ez**-dut hitz-egin oraindik.
 Josu:COM but not-I.have.it word-make/do:PFV still
 “I have only spoken with Josu so far.”

There is also a very similar construction with the ‘alternative’ operator *baizik*, which primarily means “but” in the *not-X-but-Y* construction, in which it follows the alternative, cf. [*ez X, Y baizik*]. In (7.57) we can see an example of this sentence (from Zubiri and Zubiri 1995:201, ex. 2.1220).

(7.57) Hogeita bost **baizik ez**-ditut ekarri.
 twenty five but not-I.have.them bring:PFV
 “I have only brought twenty five.”

There doesn’t seem to be much information available about the dialectal distribution of these ‘restrictive constructions’ or their relative frequency with respect to the alternative focus strategy with *bakarrik/soilik* “alone; only”. As to the origin of this construction, I am inclined to think that the type of construction in which the two

nominals <NP + *other*> seem to form a single informational/semantic (and phonological) unit must be derived diachronically from the ‘double-topic construction’ we saw earlier, in which the two nominals are independent, both semantically and intonationally, as in (7.46b) and (7.47). Further research will be needed to elucidate these matters.

7.4 EMPHATIC AFFIRMATION (‘ASSERTION FOCUS’)

7.4.1 Introduction

I would like to take a closer look at the phenomenon that I have variously been referring to as emphatic assertions, affirmative assertion focus, or affirmative polarity focus. Lambrecht has argued that perhaps one should consider what he calls ‘narrow focus’ on the polarity as yet another type of information structure, “the ‘counter-assertive’ or ‘counter-presuppositional’ type proposed by Dik *et al.* ([1981]), which involves the polarity of a proposition rather than some semantic domain within it” (Lambrecht 1994:236).

As I have argued before, I do not think that ‘polarity focus’, whether affirmative or negative polarity focus, constitute a different type of information structure, or that the polarity itself is the focus, but rather that the whole assertion, the affirmation or negation is focal, simply because the whole of the rheme is accessible or easy to accommodate. In other words, in these cases the whole assertion is made with added emphasis, as in exclamations for instance. In these cases the finite verb, as the ‘head’ of the rheme acts as the focus constituent for purposes of accentuation and focus positioning. These assertions share many of the properties of assertions with a highly salient focus

constituent that we saw in previous chapters, such as the possibility of inverting the topic of the assertion.

Furthermore, ‘polarity focus’ is not a marked option under all circumstances. As we have seen, in negative polarity assertions, polarity focus may even be unmarked, or most common, at least in some contexts. In affirmative assertions, on the other hand, polarity focus seems to be more marked, as all other cases of contrastive or otherwise emphatic focus are.

As I argued earlier, a major motivation for the commonly found ‘split’ of verb forms into a finite element bearing the polarity (among other things) and an element bearing the verbal idea (also among other things) may be that of allowing verb focus and assertion focus to be differentiated from each other, as we can see in the examples in (7.58).

- (7.58) a. Bertrand was coming to the party. (affirmation focus; ‘marked’)
b. Bertrand wasn’t coming to the party. (negative polarity focus; ‘unmarked’)
c. Bertrand was coming to the party. (verb focus)
d. Bertrand was coming to the party. (complement focus, ‘unmarked’)

The union of the two ‘halves’ into a single word, however, is seemingly also very strongly motivated. These competing motivations are solved in English by allowing the decomposition of the predicate into a finite element and a verbal element for the purpose of coding polarity focus, the so-called ‘do-support’ construction.²⁴ This decomposition is obligatory in negative assertions in English, cf. (7.59c), but only used when the polarity’s value is contrastive, cf. (7.59a&b).

- (7.59) a. Albert came to the party. (verb focus or assertion focus)
b. Albert did come to the party. (affirmative polarity focus)
c. Albert didn’t come to the party. (negative polarity focus)

As with any other type of focus, the assertion may be focal (and the finite verb the focus) either because: (1) the affirmation itself is ‘new’, although the whole proposition is accessible, as in example (7.60a), (2) because of extrinsic ‘speaker’ emphasis, as in (7.60b), or (3) because the assertion is contrastive, as when the opposite (negative) assertion has been affirmed, as in (7.60c). In all cases the underlying proposition, and thus all the ideas expressed by it, must be accessible or easy to accommodate.

- (7.60) c. New: A: Did John come to the party? B: Yes, *he did (come to the party)*
a. Emphasis: What does it matter if he left early, *John did come to the party*
b. Contrast: *John did come to the party*; who said he didn’t?

As I said, ‘polarity focus’ assertions have many of the formal characteristics of other emphatic assertions in which a focus constituent is very salient. Thus for instance subject inversion is often associated with these assertions, as Sasse (1995b) for instance has noted for Modern Greek. He argues that subject inversion in ‘automatic’ in what he calls ‘polar constructions’.²⁵ In Spanish subject inversion is perhaps associated with the most emphatic (focal) types of polarity focus, e.g. (7.61), although it is never obligatory.

- (7.61) *Sí trajo Humberto el dinero.*
yes he.brought Humberto the money
“Humberto did (indeed) bring the money.”

Emphatic affirmations, as we have seen, may perform a variety of functions, and are found in a variety of contexts. Thus languages often have different constructions which, while having similar information structure, may differ as to other characteristics. Thus in (7.62) below we can see several constructions which can be used to code affirmation focus in Spanish to express a variety of rhetorical functions and displaying

differences as to the intensity of the focus. They all have the basic meaning: *Juan brought the money* (= *Juan did bring the money*).

- (7.62) a. Juan trajo el dinero. Juan he.brought the money.
b. Juan *sí* (que) trajo el dinero. Juan yes (that) he.brought the money
c. Juan claro que trajo el dinero. Juan clear that he.brought the money

In general we can say that the version in which only the finite verb is accented (which could also be an example of verb focus) is the least salient or emphatic. It is also the one least likely to display subject inversion. The other two display different polarity reinforcers: *sí* “yes” and *claro* “clear; of course”, both of which could be optionally accented instead of the verb, since they can both ‘stand for’ the affirmation. These complex constructions separate the predicate from the affirmation, thus iconically indicating that the focus is ‘outside’ the rheme itself and that none of its elements are the focus of the assertion.

7.4.2 Polarity focus with Basque synthetic verbs

As we have seen, in Basque when the verb is synthetic, the fact that the affirmation is the focus (‘polarity focus’) is coded by placing the verb in focus (rheme-initial) position instead of a complement and by placing the affirmative or emphatic prefix *ba* on the verb, as we can see in (7.63b). A complement, of course, may also be topicalized, as in (7.63c). (As we saw in Chapter 4, in north-eastern dialects this last sentence would mean “I have money”, whereas the first one would supposedly mean something like “It (= that) is money to me.”)

- (7.63) a. *Dirua* daukat. money:DEF I.have.it “I have (the) money.”
 b. *Badauk*at dirua. EMPH:I.have.it money:DEF “I do have (the) money.”
 c. *Dirua badauk*at. money:DEF EMPH:I.have.it “The money I do have it.”

This construction, as we can see, is identical to the negation construction, the only difference being that here we find the affirmative particle/operator *ba* instead of the negative one *ez*. As I said, the particle is obligatory when there is no complement to fill the focus role, and thus all rhemes with single synthetic verbs are assertion focus constructions, since the verbal idea expressed by these verbs doesn't seem to be typically the focus, given their low inherent focality. Notice too, that when the rheme is emphatic the complement need not be given (“*the money*”), but may be new and accommodated (cf. *Guess what? I have money!*).

In (7.64:76) below we can see an example of the *ba* construction from the spoken corpus.²⁶ In this example the rheme contains no other complement.

- (7.64) 93C2A08
 75 artzen-daus barriro madarixok - kapazo gañian,
 take:IMPFV-he.has.them again pears:DEF basket top:DEF:LOC
 76 eta badoia.
 and EMPH:he.is.going
 “he takes the pears again - in a basket, / and he leaves.”

Notice that here the synthetic verb, which normally would have imperfective aspect (*He's going*), when it bears the emphatic *ba* marker it must receive a perfective interpretation (*He leaves*), much like in the Spanish contrast between *ir* and *irse*.

In (7.65:4) the same verb also bears the *ba* marker. In this case, however, there is a complement which could have been the focus, namely *kaletik* “down the street.” This complement however expresses a rather stereotypical and accessible idea, as does the

verb itself, which is perhaps why the whole rheme is treated as the focus, i.e. it is an emphatic assertion.

(7.65) 93C2A07

- 1 Bueno.
well
- 2 Txarloten pelikula bat da. ..
Chaplin:GEN movie one it.is
- 3 Ta bueno.
and well
- 4 Txarlot - e badijoa kaletik, ...
Chaplin uh EMPH:he.is.going street:DEF:ABL
- 5 eta greba eguna da,
and strike day:DEF it.is

“Well. / It’s a movie about Chaplin. / And well. / Chaplin’s going down the street, / and it’s a strike day.”

In some cases it seems that the emphatic assertion performs what seems to be a ‘preparatory’ or ‘preliminary’ setting-like function for the following statements. Unlike in the previous example, in which the event expressed by the *ba*-clause and the following one seem to be quite independent from each other, in the example in (7.66:112), the imperfective assertion with *ba* seems to lead to another assertion.

(7.66) 93C3A07

- 112 Eta baze- bazebilen kalien,
and EMPH:he.was- EMPH:he.was.going street:DEF:LOC
- 113 bere bastoiakin txarangaka txarangaka,
his came:DEF:COM brass.band:ADV brass.band:ADV
- 114 eta manifestazixo bat zetorren.
and demonstration one it.was.coming

“And he was going down the street, / with his cane joyfully joyfully, / and there was a demonstration coming.” (Cf. “A demonstration was coming.”)

It is conceivable that the preference for assertion focus in these cases is related to the preference we saw in Chapter 5 for main clauses which perform this type of setting-like

function to have emphatic properties (such as subject inversion), thus indicating that these are not regular assertions with a regular focus.

The synthetic conjugation is not the only way of expressing continuous imperfective aspect for the few verbs which have a synthetic conjugation. There is a different construction which is used to code that the verbal idea is the focus. This is the same *ari*-construction which is used with verbs which do not have synthetic forms for coding continuous imperfective aspect. We can see an example in (7.67b), with the corresponding *ba*-clause in (7.67a).

- (7.67) a. Jon badator. Jon EMPH:he.is.coming “Jon is coming.”
 b. Jon etortzen ari da. Jon come:IMPFV busy/doing he.is “Jon is coming.”

These are complex constructions in which the imperfective ‘verb phrase’ (same-topic non-finite clause) is a true complement of the verb *ari*, which can be translated as “active; busy; doing”. These constructions are rarely used with these verbs, however, perhaps because the verbal idea is rarely focal enough to be the focus.

Finally, as we saw in (7.63c), in emphatic affirmations complements may be freely topicalized, since they are not foci and must be accessible. The more emphatic the assertion is, the more dislocated the topicalized elements will have to be so as to allow the rheme-initial focus constituent, the finite verb, to be in clause-initial position as well. In (7.68a) below we can see an example of a topicalized completive clause.

(7.68) *Behi euskaldun baten memoriak*, p. 113

a. —Hona jaisten-zaretela **bazekitek**,

here:ALL come.down:IMPFV-you:PL.are:COMPL EMPH:they.knew.it

b. baina zertara jaisten-zareten **ez**.

but what:ALL come.down:IMPFV-they.are:SUBORD no

“They did know that you come down here, / but what they come down for they didn’t.” (Lit. “That you come down, they already knew.”)

The sentence in (7.68a), with polarity focus, could alternatively have the completive complement in postverbal position, though it still wouldn’t be the focus. The accessibility requirement on the proposition expressed by that clause would seem quite similar in both cases.

7.4.3 Polarity focus with analytic (periphrastic) verbs in Basque

Unlike in the case of negative assertions, which have the same form with both synthetic and analytic verbs, affirmative polarity focus (emphatic affirmation) with analytic verbs is quite different from affirmative polarity focus with synthetic verbs, since in this case the two parts of the verb remain together (with a single exception in some dialects, as we will see).

The reason for the difference between the affirmative and the negative polarity construction seems to be that in most dialects the two parts of affirmative periphrastic verbs have become completely fused into phonological words which cannot be split up. In most dialects, the *BA* particle can never be used with periphrastic verbs. Affirmative polarity focus thus is coded in most dialects of Basque simply by placing the ‘unified’ finite verb in focus position, with any and all rhematic complements being postverbal or topicalized. In other words assertion focus can be coded the same way as verb focus. In

some dialects, however, as we saw in Chapter 6, there is an alternative (optional) construction for coding verb focus, namely the *egin*-construction, which provides a way for distinguishing verb and polarity focus with analytic verbs.

If the Basque affirmative polarity focus construction was analogous to the negative polarity focus construction we would have expected to find sentences such as the one in (7.69).

- (7.69) Jon_{ERG} badu liburua irakurri.
 Jon:ERG EMPH:he.has.it book:DEF read:PFV
 “Jon has read the book.” (~ “Jon has too read the book.”)

It turns out that this construction is indeed possible in some north-eastern dialects in some specific contexts. Assertions such as this one are used when the assertion in question contrasts with its negative counterpart (a negation), that is, in counterpropositional contexts or contradictions, one of the three major contexts in which emphatic affirmations are found.²⁷

Arejita (1980) mentions that this construction is also possible in the Bizkaian (westernmost) dialect, but as a very marked way to refute a negative statement, as in (7.70B) (Arejita 1980:362).

- (7.70) A: Ez-da etorri.
 not-s/he.is come:PFV
 B: Ba-da etorri, neuk ikusi dot eta.
 EMPH:s/he.is-come:PFV, I:EMPH:ERG see:PFV I.have.it and
 “A: She didn’t come. / B: She did too come, cause I saw her.”

On the other hand, he admits that it would be much more common in this case to use a construction such as the one in (7.71), with *bai* “yes” instead of *ba* and without main verb inversion (*ibid.*).

- (7.71) *Bai etorri-da.*
 yes come:PFV-s/he.is
 “Indeed she came.”

This construction is identical in form to an equivalent Spanish construction, cf. *Sí ha venido* “she did come”. We saw too that the Spanish construction could also have the subordination marker *que* “that” following the affirmative particle, cf. *Sí que ha venido* “She did come”. (Here the *sí* acts as a fragment assertion, with the whole of the rheme being placed in the assertion-tail, as it were.) Notice that an identical construction is also used in Basque, as we can see in the following two sentences in (7.72b&c) from the Written Basque Corpus.

- (7.72) *Behi euskaldun baten memoriak*, p. 16
- a. *Gogora datorkit, zentzu honetan nire bizitzako aurreneko neguan gertutakoa.*
 mind:DEF:ALL it.is.coming.to.me sense that:LOC my life:GEN
 front:AUGMENTATIVE:DEF:GEN winter:DEF:LOC happen:PFV:GEN:DEF
 - b. *Orduan bai egin-zidala konpainia!*
 then yes do:PFV-it.did.it.to.me company
 - c. *Orduan bai izan-zitzaidala lagun!*
 then yes be:PFV-it.was.to.me:COMPL friend
- “In that respect, it comes to mind what happened during the first winter of my life. / Then it really kept me company. / Then it really was a friend.” (A cow referring to her soul.)

This construction may or may not have been borrowed, but it does fully conform to the Basque grammatical and information structure system and thus it would be a highly motivated borrowing. Borrowing of such a minor construction seems to be much more likely than borrowing of much more basic features, such as postverbal focus, especially if the construction does not go against basic structural principles of the language. Of course, the direction of borrowing may have gone the other way. And, indeed they may be the result of independent, parallel developments in these languages.

7.4.4 Polarity focus in all-new assertions

In this section we have seen some of the uses of polarity focus assertions, and that polarity focus is found when the proposition in question is quite accessible. The proposition doesn't have to be presupposed or even discourse given in order for the polarity to be the focus. It is sufficient that it be easy to accommodate from the context and not surprising. In Basque this results in VO-type sentences which have puzzled investigators. Thus, for example, the main clause in (7.73A) below has SVO order, even though both parts of the rheme, the verb and the object, are new.

(7.73) Q: Etxekoek erosi-duzue zerbait Amari?

house:GEN:DEF:PL:ERG buy:PFV-you.have.it something mother:DEF:DAT

A: Ez, bainan Aitak ikusi-du liburu bat, Amaren gustokoa izan-daitekeena.

not but Father:DEF:ERG see:PFV-he.has.it book one, Mother:GEN pleasing:DEF
be:PFV-it.may.be:REL:DEF

Q: "Has anybody in your family bought anything for your mother? A: "No, but Father did see a book, which Mother may like."

The reason for SVO order here is that the assertion is emphatic or contrastive and has 'polarity focus'. Notice that this assertion (*Father did see a book*) contrasts somewhat with another assertion with different but related content (*The family hasn't bought a present*). In other words, from the assertion that nobody had bought anything for mother one could have inferred that nobody had been looking for a present, and it is that assertion that the emphatic assertion contrasts with. Notice that the English translation could also have *do*-support, thus showing that the assertion is emphatic.

7.4.5 Polarity focus and word order change

The fact that emphatic assertions, such as the one we have just seen, result in assertions with (S)VO order, suggests that this construction might play a role in word order change from OV to VO order. The reason for this is that, although usually the fact that an assertion has polarity focus depends on intrinsic properties of the context of the assertion, it is also possible for an assertion to be extrinsically emphatic if the speaker chooses to make it so. If there are no obvious candidates for the focus role in the clause, that is, elements which are intrinsically focal, because they are all accessible, it is always possible for an assertion to display extrinsic assertion focus (and thus, in Basque, VO order).

Furthermore, if extrinsic assertion focus were to acquire specialized rhetorical functions in some contexts, or in some constructions, as I suggested had happened with topic inversion in Basque, the result could be an increase in the frequency of polarity focus assertions and thus of VO-type structures. This, of course, would be accompanied by a concomitant loss in markedness for this construction. After all, making statements emphatic doesn't result in ungrammatical sentences, but only at worst in inappropriately emphatic ones. However, there is probably enough variation in the use of extrinsic emphatic assertions among different speech styles in the language as to guarantee that minor increases in their use would not be immediately apparent. Notice that Basque polarity focus sentences have exactly the same structure as similar sentences in VO languages, namely (S)VO-type order, since in such languages a focus verb is always placed in rheme-initial position.

A further factor may become involved in this equation to make the construction more common and more likely to become a vehicle for word order change. I am referring to the influence of non-native Basque speakers and imperfect bilinguals. Such speakers might be more likely to misanalyze these assertions given their superficial similarity with unmarked constructions in Spanish, and to use them in a greater variety of contexts. We could compare such increased use of assertion focus constructions to the increased use of *do*-support sentences in English, which seems to be a feature of some dialects of English.

I do not believe that this phenomenon is as important or as crucial for word order change, if it plays a role at all, as the phenomenon of the increased use of extraposition of focus constituents discussed in Chapter 6. This is one area, however, that should be monitored as we attempt to ascertain any possible ongoing changes in word order.

7.4.6 BA in subordinate clauses

Finally, before leaving the topic of emphatic affirmations I would like to mention the occasional presence of the particle *ba* in non-asserted clauses, as has been reported by different investigators. This particle, as a marked polarity particle, seems to be restricted to a large extent to asserted sentences, such as main and completive statements and polarity questions (see below), as well as dependent assertions in sentence-tails. Since background sentences, such as relative (modifier) clauses and setting clauses, are not asserted, we would not expect to find this particle there (although we do commonly find the negative particle *ez* in them). Thus, for example, a relative clause containing a single synthetic verb does not (typically) display this particle, which would be obligatory in an asserted sentence. Below we can see some examples which reflect this difference

between asserted and non-asserted clauses: In (7.74) the synthetic verb is in a relative clause and cannot bear the *ba-* particle, and in (7.75) the synthetic verb is in a statement, where it must bear the particle (cf., e.g., de Rijk 1972a, 1972b; Oihartzabal 1984, Oyharçabal 1985a; Osa 1990:165).

(7.74) (?Ba)daukadan liburua.
(EMPH:)I.have.it:REL book:DEF
“The book (that) I have.”

(7.75) *(Ba)daukat liburua.
(EMPH:)I.have.it book:DEF
“I (do) have the book.”

Occasionally, however, *ba-* is indeed found in such non-asserted clauses, including relative clauses. In these cases the presence of *ba-* has to do with polarity *contrast* and not with polarity or assertion focus, which is a property of asserted clauses only. As we saw in Chapter 3, we occasionally find instances of contrastive elements inside non-asserted clauses which contrast with elements in similarly unasserted clauses. That is, contrastive elements are not found only in asserted clauses. The pragmatic relation focus, on the other hand, is not applicable to non-asserted clauses. It is thus not surprising that if the affirmative polarity of a non-asserted clause contrasts with the negative polarity of a similar, non-asserted clause, the particle *ba-* would be used for coding this fact, much like the negative particle *ez* would be used in the inverse situation.

This means that whereas the relative clause in (7.74) without *ba-* means “The book (that) I have”, the one with *ba-* means “The book (that) I do have” (as opposed to the one I do not have). Notice that in English, *do*-support, which is typically only found in asserted clauses, is used in this context as well.

7.5 INTENSITY ADVERBIALS

As we have seen, emphasis or contrast of an entire assertion is typically coded by treating the head of the rheme, the finite verb, as the focus constituent ('polarity focus'). Assertion emphasis, however, can also be coded by means of emphatic adverbials such as *really*, *indeed*, and *never* in English. As I mentioned in Section 7.1, Basque has several 'intensity adverbials' or 'operators' which are used to code assertion emphasis. The ordering characteristics of such elements are quite interesting.

The most common of these adverbials in Basque is perhaps *oso* "very (much)". which can also be used as an intensifier of adjectives, e.g. *oso (etxe) handi* "very big (house)". Others include *arras* "completely, totally"; *orok* "all; every, everything", and *zeharo*, *erabat*, *guztiz*, all meaning "completely, totally".

These emphatic adverbials can be used in two major ways. On one hand, they can be used as focus adverbials which are placed in pre-verbal focus position, causing other (non-topicalized) complements to be postposed, e.g. (7.76) (from Zubiri and Zubiri 1995:262).

(7.76) *Nik oso maite dut euskara.*
I:ERG very love I.have.it Basque:DEF
"I really love Basque."

This position is used when the rest of the clause consists of accessible ideas, as in this case, in other words, when there isn't a highly focal element in the assertion. In (7.77) we can see another example of an intensity adverbial (*benetan* "truly") in focus position, which causes the non-verbal predicate (*sakona* "deep") which would otherwise have normally been in focus position, as well as the object (*ezkutupe illun ura* "that dark hiding

place”), to be postposed. (All the following examples from classical authors are from Agirre Berezibar 1991:634).

(7.77) Bai, **benetan** dute sakona ezkutupe illun ura! (Lizardi 1934:72)
yes truly they.have.it deep:DEF hiding.place dark that
“Yes, that dark hiding place of theirs is really deep.”

Again, the ideas expressed by the postverbal complements must be accessible or easy to accommodate from the context.

A variant of this emphasis coding mechanism consists in placing the emphatic particle as a modifier of the head of the focus phrase. This alternative is found, for instance, when the focus element is an adjectival predicate, cf., English *The house is very big*. This enhances the focality of the focus and thus that of the whole rheme, something which may have formal consequences, such as topic inversion, as we can see in (7.78).

(7.78) Qso handia da etxe hori!
very big:DEF it.is house that
“That house is very big!”

When the ideas in the rheme are less accessible, on the other hand, these particles, which add emphasis to the whole rheme, are placed *after* the rheme proper, in a separate intonation unit, much like assertion extensions in sentence-tails. In (7.79) we see an example of one such sentence with the intensifier *oso* “very (much)”.

(7.79) Ama serio zegoen oso. (Lertxundi 1984:86)
mother:DEF serious she.was very
“Mother was very serious.”

This sentence has two accents: one on the focus constituent (*serio* “serious”) and one on the intensifier *oso*. In addition, there must be a final intonational contour after the verb

and before the emphatic particle. An alternative to the sentence in (7.79) could be the one in (7.80), in which the intensifier *oso* is a modifier of the rhematically salient predicate adjective.

- (7.80) *Ama oso serio zegoen.*
mother very serious she.was
“Mother was very serious.”

The sentence in (7.81) also has a post-rhematic intensifier, in this case the adverbial *erabat* “completely”. Here the verb itself is the focus, as marked by the new-focus *egin* verb-focus construction.

- (7.81) *Bere kanta ttikia ixildu egin zen erabat.* (Atxaga 1985:25)
its song small:DEF become.quiet:PFV do:PFV it.was totally
“Its small song ceased completely.”

Notice that the English version of this sentence would also have two accents, one on the verb and one on the adverb, indicating that this is more than a simple assertion.

Post-rhematic intensifiers of this sort may become ‘conventionalized’, and even cliticized to the rheme, eventually becoming operators, that is, particles which have scope over the whole rheme. This seems to be, for instance, the source of the French negative operator *pas* (< “step”), which used to be added to negative assertions for emphatic purposes, but which has become the main negation operator in spoken French and is not accented anymore.

Unlike preverbal intensity adverbials, ‘extraposed’ ones allow asserted clauses to have a different focus and to be verb-final. Thus a sentence with a post-rhematic intensifier in Basque may look rather different from the equivalent sentence in Romance, since it may have a preverbal focus complement. The alternative, in which the emphatic

adverbial is in focus position, on the other hand, produces sentences which are identical to their Romance counterparts. This is perhaps the reason why the post-rhematic intensifier construction has been said by Basque grammarians to be ‘typically Basque’ or ‘authentic’, even though the same strategy is available in neighboring languages. The post-rhematic intensifier alternative is rather rare in text counts, however. It is not clear how prevalent it is, as compared to the alternative in which the emphatic adverbial is in focus position. There are no occurrences of this construction in the written or spoken corpora, for instance.²⁸ It might be more common in conversation, however, or at least in some conversational styles.

It would seem that which of these two alternatives is used should depend solely on the accessibility of the ideas expressed inside the rheme. On the other hand, it could also be that the extraposition alternative is falling into disuse among Basque speakers. One possible reason for this could be that the focus intensifier strategy results in assertions with orders which are identical to those of similar clauses in, say, Spanish.

7.6 POLARITY QUESTIONS

Polarity (*yes-no*) questions constitute inquiries as to the polarity of a proposition, i.e. as to whether it is true or not. In other words, they are used to request the addressee to make an affirmation or a negation about the proposition. As in the case of negative assertions, the ‘polarity’ or, rather, the whole assertion, ‘represented’ by the finite verb, is the default focus in this type of assertions. The form of polarity question constructions in all languages reflects this default situation, although something other than the ‘polarity’

may be the focus of a polarity question under certain circumstances, much like in the case of negative statements.

For this reason, we find that in Basque the finite verb is typically placed in focus position in yes-no questions. It isn't even necessary for the other focus candidates in the asserted clause to be discourse given, as long as they can be accommodated. This is what we find in the question in (7.82b) in the following excerpt. Notice that the synthetic finite verb, bearing the *ba-* particle, is in rheme-initial position. (The dialectal question particle *al* is optional.)²⁹

(7.82) *Argia egutegia* (May 11, 1994)

a. Jatetxean bezeroak zerbitzariari:

restaurant:DEF:LOC customer:DEF:ERG servant:DEF:DAT

b. "Gazte! **Ba-al-duzu igel-hankarik?**"

youth EMPH-QN-you.have.it frog-leg:PART

c. "Ez jauna! **Erreumak** eragiten-dit ibilera hau."

no mister rheumatism:ERG it.causes.it way.of.walking this

"In the restaurant the customer (says) to the waiter: / 'Young man! **Do you have frog-legs?** / No sir, it's because of rheumatism that I walk like this."

Notice that the postposed, and unaccented, object (*frog-legs*) does not express a discourse-given idea, and thus in a statement it would have been the prime focus candidate. Here, however, the fact that the object's referent is easy to accommodate in this particular context (a restaurant), allows the finite verb to be in focus position. Actually, if the object had been in focus position it would have to be interpreted as highly contrastive.

When the idea expressed by a complement of the verb is less accessible, the favored solution in Basque seems to be to topicalize the object constituent. Thus in (7.83) we can see an equivalent sentence to the one in (7.82b) with the object fronted.

- (7.83) Igel-an~~k~~arik ba-al-duzu?
 frog-leg:PART EMPH-QN-you.have.it
 “Do you have frog legs?” (Lit. “Frog legs, do you have any?”; cf., Spanish,
 “Ancas de rana ¿tienen ustedes?”)

There are some contexts in which a complement (or the verb itself) is more focal than the polarity, and thus is the focus and is placed in focus position. This typically happens when the complement is contrastive and when the idea expressed by the predicate minus the focus idea is not in question, as we can see in example (7.84).

- (7.84) Trenez (al-)datoz?
 train:INSTR (QN-)they.are.coming
 “Are they coming by train?”

Here the speaker already knows that certain people *are* coming, a fact which is not in question, and the question ‘focuses on’ the means of transportation. The lack of the emphatic particle *ba* here with the synthetic verb *datoz* “they are coming” indicates that the preverbal constituent is indeed in focus position.

7.7 IMPERATIVE AND DESIDERATIVE SENTENCES

Imperative and desiderative assertions are another type of assertion which displays a strong tendency in Basque for the verb to be in rheme-initial (focus) position. In fact, the vast majority of imperatives in texts are verb-initial, e.g. (7.85) (from Zubiri and Zubiri 1995:596).

- (7.85) Goazen beste taberna batera.
 let.us.go other bar one:ALL
 “Let’s go to another bar.”

Verb-initial order in this imperative assertion is undoubtedly unmarked, even though the referent of the allative complement is new (though easy to accommodate) and would be the focus in a declarative sentence. In VO languages, in which a focus verb is already in a focus position, imperatives do not differ in word order from declaratives. Notice, however, that the English version of the Basque imperative in (7.85) has the same focus structure as the Basque imperative, since the accented element is the verb and not the complement. In other words, in VO languages, the ‘choice’ of the verb as the focus doesn’t have repercussions for word order, but in an OV language it always does.

There are only two possible reasons why these assertions would have the verb in focus position: either because the verbal idea itself is the focus or because these assertions are somehow emphatic and thus the finite verb acts as the focus of the assertion. The former possibility would be motivated by the fact that imperatives are action oriented assertions, i.e. they are commands the function of which is to impel the addressee to act. But we could also see imperatives as emphatic assertions, much like exclamations, and as such, the whole assertion could be seen as the focus.

Imperatives do not just have the verb in focus position, but also in clause initial position, since the ‘topic’ of imperative clauses, actually a vocative, in the vast majority of cases is not overtly expressed (this is true even in English). When the vocative ‘topic’ is overt, it is typically clause initial, as it would be expected *and* dislocated, that is, in a separate intonation unit (cf. *You, come here*), although it can be inverted (cf. below). The tendency is thus for imperative sentences to be verb initial. In Basque this applies equally to finite imperatives, as well as to non-finite imperatives, which are much more common in actual spoken Basque.

The tendency to place the verb first in these assertions is not frowned upon by prescriptive grammarians, the way it is with statements. Zubiri and Zubiri 1995 is typical in this respect. They tell us, for instance, that “in affirmative imperative sentences, normally, the verb goes first (behind the subject if there is one), and the rest of the complements follow” (Zubiri and Zubiri 1995:596).³⁰ This includes cases in which both the verb and the complements are discourse-new (e.g. *Open the window*). In other words, the default, or unmarked, word order in these clauses is the opposite of what we find in statements, but for reasons having to do with the differences in focus structure, and not because that is how imperative sentences are coded.

Zubiri and Zubiri admit that sometimes a complement does go before the verb, and the examples they give are very revealing. They are all cases in which the complement is intrinsically and significantly more focal than the verb, all of them involving a non-verbal predicate in preverbal position and a copula or copula-like verb. In three of the four examples they give the verb is the copula verb *egon* “be” and the fourth one has the semantically weak verb *ibili* “go, walk”, used as a copula. We can see two of these examples in (7.86a&b) below (they are both non-finite imperatives).³¹ (Notice that in the English versions of these sentences it is the non-verbal predicate that is accented, thus showing that they have an identical focus structure to the Basque sentences.)

- | | | | |
|--------|-------------------------|-------------------|-----------------------|
| (7.86) | a. <u>Kontuz</u> ibili! | care:INSTR go:PFV | “Be <u>careful</u> .” |
| | b. <u>Geldi</u> egon! | still be:PFV | “Be <u>still</u> .” |

This suggests that when the verb is in focus position in imperatives, the reason is that the verb is the focus, and not because they are emphatic, although both factors may be

actually involved. When the verb is finite, and especially if it is synthetic, however, it may have an even stronger tendency to be initial, cf. (7.87a&b).

- (7.87) a. Zaude ixilik!
 you.be quiet:ADV
 b. Isilik zaude!
 quiet:ADV you.be
 “Be quiet!”.

An examination of a long list of imperative and desiderative sentences in Agirre Berezibar (1991:II.V), extracted from a large number of written sources, also shows that the vast majority of these clauses are verb-initial (or single verbs). Complements are routinely, though not always, postverbal, in both finite and non-finite imperatives. The cases in which complements are most likely to be preverbal (rheme-initial) are those in which they are very focal, i.e. surprising or unexpected, emphatic, or contrastive, although many such complements are found in postverbal position as well.

The other major type of preverbal complements in these imperatives consists of those which are non-referential. As we already saw in the case of statements, non-referential complements, such as non-verbal predicates, tend to form a tighter bond with the verb and are much more likely to be rheme-initial and focal (as compared to the semantically weak verbs they are typically associated with). In addition to being very focal in the contexts they are typically found in (with verbs with low semantic content), non-referential ideas are always low in topicality, and topicality seems to be a major requirement for a nominal to be placed in assertion tail position and to be unaccented.

Below we can see some examples of imperatives with preverbal (focus) complements. In (7.88) the preverbal nominal (*eskerrak* “thanks”) is non-referential and

in (7.89) it seems to be contrastive (*ber xilotik* “through the same hole”), or at least it seems that the verb (*sar* “enter”) should be very accessible (Agirre Berezibar 1991:815, 811, respectively).

(7.88) *Eskerrak emazkik.* (Azkue 1989/1947:385)
thanks:DEF give:IMPER:PL
“Give thanks.”

(7.89) *Ber xilotik sar.* (Cerquand 1874:II-27)
same hole:DEF:ABL enter:ROOT
“Enter through the same hole.”

These complements do not necessarily have to be preverbal and they could easily postverbal (and unaccented) as well. Notice that the accent in the English translations show that the English versions of these sentences have the same focus structure as the Basque ones.

The sentence in (7.90) below, also from the same collection, is a desiderative formula, but we can discern the same phenomenon at work: the preverbal element is a (non-referential) predicate adjective and it is always placed in preverbal position in this case (*on* “good”, as in *do good*) (from Agirre Berezibar 1991:815).

(7.90) *Egun on mutilak eta on egin beizuela.* (Bilbao 1954:163)
day good boys:DEF and good do/make:PFV it.do.it.to.you:COMPL
“Good day boys, and good appetite.” (Lit. “May it do you good.”)

When the verbal idea is active, a new complement is always the focus of an imperative however. Thus, for instance, in a context in which A asks B the question: *What should I buy?*, and thus the verbal idea (*buy*) is active, the object of an imperative assertion, such as *Buy a book*, would be the focus, in Basque, just as in English. In this case we could find the focus object in (preverbal) focus position in Basque, as in (7.91a)

below. But it could also be extraposed, as in (7.91b). There doesn't seem to be any preferences for one or the other order.

- (7.91) a. Liburu bat eros-ezazu. (OV)
book one buy:ROOT-you.have.it:IMPER
b. Eros-ezazu liburu bat. (V,O)
buy:ROOT-you.have.it:IMPER book one
“Buy a book.”

Even if average imperatives are not emphatic assertions, there are also more marked imperatives which are indeed emphatic. In English emphatic imperatives are coded sometimes by means of *do* support, e.g. *Do close the door*. (When the verb is negative it may even display another characteristic of emphatic assertions, namely subject inversion, e.g. *Don't you close the door!*).

The Basque emphatic marker *ba* cannot be used to code imperative assertions as emphatic. Although the functions of *ba* and *do*-support overlap, as we have seen above, the particle *ba* cannot be used in this context. (The reason for this difference probably has to do with the different source of the emphasis markers, an action verb, *do*, and an affirmative particle, *bai* “yes”.)

On the other hand, Basque has another means of indicating that assertions are emphatic, namely by means of topic postposing. This, however, is not a viable alternative in imperatives, since the ‘topic’ is typically elided. However, as I mentioned in Chapter 4, topic inversion is the norm with desideratives. In the next section I will look at exclamations and other emphatic statements, which also display this property.

7.8 EMPHASIS, EXCLAMATIONS, AND SUBJECT INVERSION

7.8.1 The sources of language inversion crosslinguistically

As I argued in Chapter 3 and subsequent chapters, there is a strong correlation crosslinguistically between rheme-initial position and very salient (focal) foci (and thus very salient rhemes). This is true even in 'VO languages', which place 'regular' foci in postverbal position. In addition, and especially in OV languages, in which rheme-initial is the standard position for all foci, topic inversion is also often used with very salient foci (and thus very salient rhemes). (Topic inversion, however, is also found without rheme-initial foci in VO languages.) This results in the rheme and/or the focus coming first in the clause and/or in the sentence.

In Chapter 5 I mentioned that Sasse argued that one of the functions of 'thetic sentences', by which he meant sentences with *subject inversion* (of any type) in languages such as Italian, is to code *exclamations*. It has often been noticed that exclamations display subject inversion in many languages, even in English in some cases.³² The inversions found in many exclamations, however, seem to be of the antitopic variety and the are found with verbs which under normal circumstances must have topics and cannot have a thetic configuration.

It is true that some of the most clear-cut examples of thetic clauses are also very often emphatic assertions with rheme-focus. These clauses display an unaccented subject inside the assertion (after the verb in most languages) which is hard to distinguish from an inverted topic. Out-of-the-blue thetic announcements, for instance, are very often

exclamatory in nature, such as the classical example *The phone's ringing!* Presentations can also be emphatic if they are (presented as) surprising, as we saw in Chapter 5.

Exclamations actually seem to be nothing more than a special type of emphatic assertion which have a focus that is extrinsically very salient. And exclamations, just all emphatic assertions in general, may be 'at heart' either topicless or not. Just like not all thetic (topicless) sentences are exclamations, it is also true that not all exclamations are thetic sentences. But comment-topic exclamations can be easily confused with thetic sentences, especially with emphatic thetic sentences, since in many languages both sentence types may have an unaccented subject after the verb.

To recapitulate, the reason for the confusion seems to be that we are dealing with three possible reasons for subjects to invert, cross-linguistically. First, a subject may invert because it is the focus and in many, perhaps most, VO languages all focus complements are placed after the verb, including the subject. This kind of inversion does not typically apply to OV languages, since in these languages a focus complement goes in preverbal position (unless, of course, it's extraposed, which, as we saw is common enough in spoken Basque). Remember that these subjects are always foci and thus accented.

The second reason for inversion is found when the focus (and thus the rheme) is very salient and the topic inverts with the verb to allow the rheme to come first. This we find with predicates which under normal circumstances must have topics, but which in this unusual circumstance allow what would should be the topic (along with other setting elements) to come after the verb (i.e. topics become antitopics). Pragmatically (informationally) speaking, these subjects still retain (at least some of) the properties of

topics. This kind of topic inversion is perhaps particularly common in OV languages, since in these languages this is one of the main ways of coding that the rheme is very salient, since foci are typically in rheme-initial position, whether very salient or not.

The third source of inverted subjects we find in the case of emphatic assertions with predicates which typically do not have a topic ('unaccusative' predicates). Because in these cases the finite verb acts as the focus (there are typically no other focus candidates besides the subject and the verb), the subject must be placed in the only place in which non-topics and non-foci go (in these languages), namely after the verb. These are not antitopics, but they share some properties with antitopics, properties which derive from their status as additional complements which are not topics or foci, such as their position and their accessibility requirements.

The facts about these three possible causes of subject inversion and their applicability in typical VO languages (with flexible subjects), such as Spanish, and flexible OV languages, such as Basque can be seen summarized in Table 7-4.

	Non-emphatic Assertions		Emphatic Assertions	
VO language	Topic-Comment	S[VO]	Topic-Comment	S[OV], S[VO]
	Thetic	[VS]	Thetic	[VS]
	Comment-Topic	n/a	Comment-Topic	VSO, QVS
OV language	Topic-Comment	S[OV]	Topic-Comment	S[VO]
	Thetic	[SV]	Thetic	[VS]
	Comment-Topic	n/a	Comment-Topic	VSO, QVS

Table 7-4: Some typical orderings found in normal and emphatic assertions, whether thetic or topicful, in VO and OV languages.

In the next two sections I we will look at some examples of emphatic statements with topic inversion. I will present some examples from Spanish first which show the

connection that exists between heightened focality of the focus (and thus the rheme), preverbal position of the focus, and optional inversion of the subject. Next I will show similar phenomena with examples from the Basque corpus.

7.8.2 Spanish exclamations

In Spanish, ‘inverted’ order (focus-verb-topic) is common whenever there is a very focal focus constituent, which is typically in rheme-initial position. Remember that in Spanish, as in VO languages in general, only very focal focus complements go in rheme-initial position, such as contrastive ones (which are intrinsically very salient) or emphatic ones (which are extrinsically very salient). The resulting assertions are thus contrastive or emphatic too (cf. Silva-Corvalán 1983).

In (7.92) we see several variations of a transitive asserted clause with an overt subject and object, sentences which under normal circumstances must have a topic. The one in (7.92a) is an unmarked SVO sentence, where the overt subject is the topic and the object is the focus. With the right intonation and in the right context the object could be contrastive or emphatic with this same focus and ordering configurations.

- (7.92) a. Mi madre ha comprado un libro. (SVQ)
my mother she.has bought one book
b. Mi madre un libro ha comprado. (SQV)
my mother one book she.has bought
c. Un libro ha comprado mi madre. (QVS)
one book she.has bought my mother
d. Ha comprado un libro mi madre. (VQS)
she.has bought one book my mother
e. Ha comprado un libro mi madre. (VOS)
she.has bought one book my mother
“My mother bought a book.”

However, a very focal focus object would most likely be fronted, and then the subject would either be dislocated, as in (7.92b), or inverted as in (7.92c). Subject inversion is also possible without object fronting, as in (7.92d). This inversion in a way *codes* the heightened focality of the focus. Subject inversion without object fronting also obtains when the whole rheme is the focus and the finite verb acts as the focus constituent, as we can see in (7.92e).

Inverted order is also found in assertions without a finite verb, which are typically assertions in which a non-verbal predicate is the focus and the copula is ‘elided’. Below we can see an example from a common children’s expression, which comes in two varieties: with and without inversion, as shown in (7.93a&b).

- (7.93) a. El último que llegue tonto. (Topic-Focus)
 the last.one who s/he.arrives dumb
 b. Tonto el ultimo que llegue. (Focus-Topic)
 dumb the last.one who s/he.arrives
 “The last one to arrive is dumb!”

The first assertion in (7.93a) has Topic-Focus order and the second one in (7.93b) has ‘inverted’ Focus-Topic order, in both cases without a finite (copula) verb. The second version in (7.93b) is more emphatic, or exclamatory-like.³³ This Focus-Topic version seems unusually well fitted for issuing a challenge, for example, whereas only the Topic-Comment version would probably be used to provide ‘information’ (although it could also be used to issue a challenge with the right intonation).

7.8.3 Inverted order in Basque exclamations

In Basque too, (non-focus) subject inversion is the result of intrinsic or extrinsic heightened focality of the focus of the assertion. As we saw in Chapter 4, there are few examples of antitopic inversion in the Spoken Basque Corpus, most of them having identificational foci in copula clauses. This may be due to the fact that this ‘disruptive’ type of assertion is not common in narrative. Topic inversion is more common in lively conversation and in some (non-narrative) written genres, as we have seen. In novels, for example, they appear mostly in dialogue.

In (7.94d) we can see an exclamation from a Basque novel, a direct quote, with a copula verb, a rather long preverbal focus constituent, and an inverted topic (antitopic).

(7.94) *Behi euskaldun baten memoriak*, p. 102

- a. Azkenean,
last:DEF:LOC
- b. izpi guzti haiekin lokatz figura bat lortu-nuen,
fragment all those:COM mud/clay figure one I.obtained.it
- c. esaldi bat, egia bat:
sentence/phrase one, truth one
- d. —Oraindik errenditu ez den ejerzitoaren almazena da **Balantzategi!**
still surrendered:PFV not-it.is:REL army:DEF:GEN warehouse:DEF it.is
Balantzategi

“(a) Finally, / (b) with all those fragments I obtained a clay figure, / (c) a statement, a truth: / (d) ‘Balantzategi is the warehouse of the army that hasn’t surrendered yet!’”

Notice that the assertion was coded by the writer as an exclamation. The order of this exclamation is Focus-Copula-Topic, rather than the more common, non-emphatic order Topic-Focus-Copula. The inverted subject in this case feels like a topic. In other words it seems to be a statement about *Balantzategi*.

In the following example, from the same novel, the ongoing discourse topic in the section leading to these two clauses was a pack of wolves which had been following the protagonist of the novel, a Basque cow.

(7.95) *Behi euskaldun baten memoriak*, p. 26

- a. uhurika hasten zen bat,
howl:ADV begin:IMPFV it.was one
- b. uhurika hasten ziren denak.
howl:ADV begin:IMPFV they.were all:PL

“One started to howl, / (and) they all started to howl.”

In these clauses too the inverted subjects have very topical referents. Here, however, perhaps because of the low specificity of the referents (*one* and *all*), they seem to be less like topics, pragmatically speaking, and thus the assertions have some of the ‘feel’ ofthetic sentences. That is, the point of these assertions seems to be, not to say something about one or all of the wolves, respectively, but to present a sudden, unexpected or surprising event, or series of events, in its entirety. (The verb here is *hasi* “begin” is often used in sentences which express a sudden and unexpected event.) Notice, however, that copula sentences are notthetic under normal circumstances and normally must have a topic. However, in the English versions of these sentences, the subjects could receive the main accent of the assertion, cf. *One started to howl, and they all started to howl*. This is perhaps due to the fact that this seems to be somewhat of an idiomatic construction. In Basque, the postverbal subjects could not be foci, since the focus in both cases is the preverbal adverbial predicate and the subjects are unaccented.

These two sentences exemplify a phenomenon that I have already discussed, namely thatthetic (topicless) sentences are in some ways pragmatically similar to antitopic sentences, since they both seem to be more ‘event centered’ than ‘referent

centered'. In other words, in both types of sentences the idea is to present a whole event and not to predicate something of a referent. That is, in neitherthetic nor emphatic antitopic assertions is a referent (the topic) used to ground the assertion the way it is in most assertions in discourse, which are predications.

It is possible that not all inverted subjects found in sentences which have verbs which are typically found in assertions with topics, are antitopics. In some cases in which a subject is inverted (and unaccented) there seems to be another argument filling the topic role, as is the case in example (7.96) below.

- (7.96) *Kuba triste dago*, p. 32
'Harritu egin-nau honek, ...'
surprise/shock:PFV do:PFV-he.has.me this:ERG
"This guy really shocked me!" (Cf. "He really shocked me, this guy.")

Here the verb is obviously the focus, coded by the new-verb focus *egin* construction. The topic, however, would seem to be the speaker, which is covertly coded as the object (absolute argument) of the verb. Of course, it is also possible that here there is both a topic *and* an antitopic, just as we have seen occasionally sentences may have two topics instead of one.

As we saw in Chapter 3, with some so-called unaccusative predicates, the subject may be the topic if it's topical, but the focus if it is new. *Come*, *appear*, and *arrive*, for instance, are of this type. In OV languages like Basque, in both cases the subject may be preverbal (unless the focus is extraposed or the topic inverted). In VO languages like Spanish, when the subject is the focus it is always inverted ($V\bar{S}$) and when it is the topic it precedes the rheme ($S\bar{V}$). However, we find that in non-continuous contexts, assertions with these verbs, in which the subject seems to be the topic, are often emphatic assertions

since they announce some event. In these cases the topic subject is often inverted, both in VO and OV languages like the ones we've been discussing. These sentences, however, are typically identical to emphaticthetic sentences in which the finite verb acts as the focus of the assertion. Take, for example, the Basque sentence in (7.97), from the written corpus.

- (7.97) *Behi euskaldun baten memoriak*, p. 30
iritsi-da garaia!
arrive:PFV-it.is time/period:DEF
“The time has come!” (cf., Spanish, “¡Ha llegado la hora!”)

Is this an emphaticthetic (topicless) sentence, or a comment-topic one? The only way to ascertain this fact might seem to be to ask whether the postverbal subject is pragmatically the topic and whether it could be pre-rhematic in this context, in which case it would be a topic. In this case it seems that a topic-comment configuration would not be very likely (cf. *Garaia iritsi-da*), perhaps because of the low topicality of the subject's referent (*the time*). In English, however, the idiomatic phrase *the time has come*, seems to have just such a configuration.

When the subject of the sentence is the focus in sentences with these predicates, however, the subject is clearly in preverbal focus position in Basque, as in the case of the sentence in (7.98) below.

- (7.98) *Liburu bat iritsi-da.*
book:DEF arrive:PFV-it.is
“A book has arrived.” (cf., Spanish, “Ha llegado un libro.”)

That is, the subject is rheme-initial *unless* it is extraposed, in which case the subject is also ‘postverbal’, as we can see in (7.99), but in this case the accentuation and the intonation make it clear that the subject is not postverbal but rheme final.

- (7.99) *Iritsi-da, liburu bat*
arrive:PFV-it.is book:DEF
“A book has arrived.” (cf., Spanish, “Ha llegado un libro.”)

7.9 CONCLUSIONS

To summarize, in this chapter we have seen that in some cases the affirmation or negation of a proposition may be what is focal about those assertions. This is typically found when the proposition (and thus all the ideas expressed therein) are accessible from the context. In these cases the finite verb, as head of the assertion, acts as the focus constituent. This focus configuration is also possible in some cases when the ideas in the proposition are not strictly speaking accessible, but can be accommodated.

It has been said that this type of focus configuration is unmarked (or common) for negative sentences, but marked (emphatic) for affirmative ones. There seems to be some truth to this claim in the case of negative sentences, at least those which are uttered with minimal context, though perhaps it is not true of negations in narrative, where focus complements seem to be the norm. Emphatic affirmations may not be as marked as it is said either, especially with some predicates and in some contexts, such as in announcements and exclamations.

In this chapter we have also seen that the negation construction in Basque reflects a situation in which the finite verb acts as the focus constituent, and thus it is placed in

rheme-initial position. This creates a conflict when the finite verb is not the focus, which is often the case in narrative, since the focus constituent is not in the expected rheme-initial position (unless, of course, it is outside the scope of negation). In addition, in negative assertions the non-finite element of periphrastic verbs is always in a position where it is not expected to be in an OV language, namely after the verb (rheme finally). This phenomenon is equivalent to the verb-second phenomenon found in many erstwhile OV languages and it has been suggested that it may be an intermediate stage in the shift from OV to VO order.

Finally, we have seen that, not only in negative assertion constructions, but also in other asserted sentence types, such as polarity questions, imperatives, and certain types of declarative assertions, the most common focus structure is also different from the one we expect in 'typical' or 'canonical' affirmative statements, in which a complement of the verb is typically the focus. The reasons for these differences have to do with the meanings and uses of these constructions, which call for a different focus structure from the one found in declarative assertions.

NOTES TO CHAPTER 7

1. Lambrecht also notices that this is a well known fact when he says that: “It has been observed that in natural language negative sentences are ordinarily uttered only if the speaker assumes that the addressee believes or at least entertains the possibility, that the corresponding affirmative sentence is true (see Givón 1975b, Gazdar 1979:67; Horn 1989: Ch. 3)” (Lambrecht 1994:63-64). This proposition is sometimes said to be *presupposed* (to be true). I prefer to use the term *accessible* (or ‘pragmatically presupposed’) since the truth of the proposition is not at issue here.

² In other words, the *affirmation* or *negation* itself may be the focus in three situations. First of all, when the polarity **contrasts** with the polarity of the same proposition that is available from the context, such as in contradictions. Secondly, when the polarity is the only **new** element of the sentence, as in answers to yes-no polar questions. Given the fact that there are only two main possible values for the polarity, it would seem that all cases in which the polarity is the new element, it is also a contrastive element. Finally, the assertion (affirmation or negation) is the focus when the whole assertion is presented **emphatically** by the speaker, because it is surprising or unexpected, or otherwise forcefully presented.

3. The fact that *ez* is written as a separate word, whereas *ba* is written as part of the finite verb is just an oddity about Basque orthography. In both cases the particle is cliticized on to the finite verb that follows it. The negative particle *ez* used to be written together with the finite verb until fairly recently (cf. Villasante 1980:33). The emphatic affirmative particle is still written as part of the verb in Standard Basque, just like other verbal clitics, cf. conditional *ba* and causal *bait*.

4. Remember that most of the verbs in the texts are periphrastic, since many verbs do not even have a synthetic conjugation. It is true that the most common verbs do have a synthetic conjugation, but then again, only for present and past tense continuous imperfective aspect forms.

5. For some non-main constructions where there is (supposed to be) no inversion, see e.g. Euskaltzaindia 198b7:493, 497; 1993:437). Mitxelena and Villasante both mention that some earlier authors, such as Cardaberaz, Lardizabal, and Añibarro, didn't always invert the finite verb in main (asserted) negative clauses (cf. Michelena 1978:224, Villasante 1980:30).

6. This includes causatives and conditional sentences for instance. A popular learning grammar for instance, chastises the use of inversion in conditional clauses (Zubiri 1989). Under the heading “typical errors”, Zubiri mentions that “a very frequent error consists in placing the verb behind the auxiliary in negative sentences. For instance [in a conditional clause]: *If you don't do it: * Ez BADUZU EGITEN; EGITEN ez BADUZU*” (Zubiri 1989:263-4, my translation). (The conditional *ba* prefix is phonologically identical to the affirmative *ba* prefix.)

7. According to Altube “Sentences in which the negative *ez* is prefixed to the finite verb, have almost always as the inquired element [i.e. the focus] that same morpheme *ez*, that is, the negative quality of the verb” (Altube 1929:48; my translation, J.A.) [“Las oraciones en cuya flexión verbal va prefijado el negativo *ez*, presentan casi siempre como elemento inquirido, ese morfema *ez*, o sea la cualidad negativa del verbo”].

8. As Osa stresses, the cliticized focal pronoun *or* “there” in *ordáukat ori* (< *hor daukat hori* there I.have.it that “that’s where I have that”), can also be fully accented, whereas *ba* and *ez* must always be ‘proclitic’ and thus unaccented (unless it forms a two syllable word with the verb, in which case it automatically receives the ‘accent’ for the word (Osa 1990:293fn5). A similar thing happens with conditional *ba*, argues Osa, which since it has become proclitic it cannot be emphasized, which is why it is often reinforced by *balain* (Osa 1990:294; cf. Chapter 6).

9. The English negative particle *not* cannot stand for the assertion in this way and acts always as an operator on the salient element, either the finite verb or some other element, cf. *I didn’t; Not mé; Not yésterday*. In other languages, such as Basque and Spanish, in which the negative operator and the negation word (*no*) are the same word, this word can also be used to stand for the assertion/predicate.

10. In negative finite imperative sentences the non-finite verb inverts with the finite verb, just like in main clauses.

11. According to Osa 1990, Oyharçabal is another author who holds this position. It seems to me, however, that Oyharçabal (1985a:104) believes that only one element is the focus here. He does mention the two-focus possibility, but doesn’t seem to subscribe to it. (Osa’s reference is actually to Oyharçabal 1984, which doesn’t figure in his bibliography. His actual source seems to be Oyharçabal 1985. He is not referring to Oihartzabal 1984, which is by the same author, but with Basque spelling.)

12. Oyharçabal (1985a) disagrees with Lafitte’s choice: “I do not think, then, that Lafitte is right when he says that, when there is more than one element between the auxiliary and the main verb, the first (the one closest to the auxiliary) is the main focus. *In my opinion, negative sentences are like affirmative ones in this respect: the closest to the right the more rhematic something is, all the way to the focus*” (Oyharçabal 1985a:110; my italics, J.A.) (The original says: “Ez dut uste, bada, Lafitte-k arrazoi duen aditz laguntzailearen eta aditz nagusiaren artean elementu bat baino gehiago delarik, lehena (aditz laguntzaileari hurbilena) galdegai nagusia izaten dela dioenean. *Ene eritziz, ezezko perpausak baiezkoen ereduari darraizkio puntu horretan: eskuinago eta irazkinkoiago, galdegai izateraino*”.)

13. Oyharçabal also mentions that exceptions to this tendency are found, just like there are exceptions to the preverbal positioning of foci: “What I am putting forth here is nothing but a tendency. As in affirmatives, with the help of intonation, the usual order

may be broken. We know for instance that sometimes the non-verbal predicate is placed to the right of the copula (although some find that construction unauthentic. Well, the same thing may happen in negatives” (Oyharçabal 1985a:111; my translation, J.A.) [“Hemen azaltzen duguna joera baizik ez da. Baiezkoetan bezala intonazioa lagun, usaiako ordena hauts daiteke. Adibidez, badakigu batzutan, atributoa kopularen eskuinean ematen dela (nahiz zenbaiti ez zaion itzuli hori biziki jatorra. Bada, gauza bera gerta daiteke ezezkoetan”].

14. Odriozola and Zabala 1992:40 also mention that the element under the scope of negation doesn't have to come right after the finite verb, but that that is where it is most salient.

15. Speech Act = *Body_Statement**; Polarity = *Neg*; Other Codes = **Brace_Intact**; or **Brace_Irrelevant_With_Fronting**; or **Exbraciation**; or **Partial_Exbraciation**.

16. 108 brace-irrelevant negative clauses; 66 with no overt topic; 17 A/ez...; 14 S/ezV...; 3 O/ezV...; 1 I/ezV...; 4 with partitive topic: K/ezV...; 3 with other fronted elements. Rhemes: 12 ezV, 2 ezVezer, 2 ezVnahi, 14 ezVO, 2 ezVS, 2 ezVSX, 46 ezVV, 7 ezVX, 1 ezVxla, 18 ezVXn, 1 ezVX/S, 1 ezVXX.

17. 25 clauses with an intact brace: 5 with ergative topics, 1 with other topic; and 19 with no overt topic. Rheme orders: ezVAV, 5 ezVezerV, 5 ezVOV, 1 ezVVbehar, 3 ezVVnahi, 7 ezVXV, 3 ezVXVnahi.

18. There are 25 negative clauses displaying complement exbraciation. 20 of these have no overt topics, 5 have ergative overt topics. Rheme orders: ezVbeharV, ezVnahiO, 3 ezVnahiV, 3 ezVnahiXtzea, ezVnahiXV, ezVVezer, 3 ezVVO, 4 ezVVX, 4 ezVVXla, 1 ezVVXn, 3 ezVVXtzen.

19. The two sentences are the following (exbraciated in bold, in-brace in italics): 93C1B06:155 eta ez du egon bearrik kartzelan, ez? “and he doesn't have to **be in jail**”; 93C2A05:24 Ez zion ezer uzten egiten “he doesn't let him **do anything**”.

²⁰ Again, the fronted element does not have to be a ‘topic’, i.e. an argument of the verb outside the rheme and what the rheme-predicate is about, but rather it belongs to the more general class of ‘settings’, as described in Chapters 3 and 4, to which topics belong, which typically precede the rheme.

²¹ Perhaps it wouldn't be surprising if this phrase *behin ere* lit. “once also” were to become a focus constituent meaning, together with the negative particle that it is next to, a single constituent meaning *never*. I do not have any evidence that this has happened yet, however.

22. *Soil* means “mere, lone, alone”, just like *bakar*. The ending *-ik* used in both cases is an adverbial ending. Thus *soilik* may in fact be a cognate of Spanish *sólo* “alone; only”,

and *bakarrik*, used in this sense, may be but a calque from Spanish. Cf. French *seulement*.

23. There are 65 instances of *besterik ez* in the three novels in the extended written corpus (*Hil ala bizi*, *Otto Pette*, and *Obabakoak*), and 196 cases of *besterik total*.

24. The *do* construction in English is used for polarity emphasis, although there are other possible alternatives to make the polarity more emphatic, perhaps because overuse of the *do* construction (or the fact that it has other uses) has taken away some of this emphasis: “The use of the emphatic *do*-construction has become a conventionalized grammatical way of expressing emphasis. The word *do* in (2.24) acts as a mere intensifier, equivalent to an adverb like *really*, so that *I do hope* is equivalent to *I really hope*” (Lambrech 1994:72).

25. About this, Sasse says: “if the subject is placed in front of the verb, the utterances are okay in the respective contexts but lose much of their “verum focus” character even if uttered with the intonational peak on the predicate. This reveals that VS order has a bearing on their specific focus interpretation; non-fronted verbal focus is interpreted as completive rather than polar focus” (Sasse 1995b:152-53).

26. Other Codes = **Polarity_Ba**.

27. For a study of the parallelisms between the Basque negative construction and this dialectal affirmative emphatic construction, from a GB perspective, see the very interesting studies in Ortiz de Urbina 1989 and Laka 1994:76, for example.

28. In Atxaga’s (1991) book from which part of the written corpus is taken there are a few examples, although none is in Chapter 1, the one included in the written corpus. There is a particular excerpt in pg. 160 which contains two tokens of this construction, one with *oso* “very” and one with *zeharo* “completely”.

29. The *al* ‘particle’, derived from *ahal* “be able/possible”, is used in some dialects in polar questions Another such question ‘particle’ is the rhetorical question particle *ote* “perhaps, ...”. These particles are not rhematically salient and are compatible with polarity markers *ba* and *ez*, cf. Euskaltzaindia 1987b:501ff.

30. The original quote says: “Baiezko aginte-perpausetan, normalean, aditza aurrean jarri ohi da (subjektua dagoenean, beronen atzean), gainerako osagarriak atzean jartzen direlarik” (Zubiri and Zubiri 1995:596).

31. In western (non-north-eastern) Basque dialects the perfective form of the verb is used to make non-finite imperative assertions. In north-eastern dialects the root of the verb is typically used instead.

32. Cf., e.g., McCawley 1973, Green 1980, Huddleston 1993. Green (1980) for instance mentions a number of English constructions in which the subject is inverted, although she mixes examples of postverbal subject foci with postverbal subjects which are not foci, but

'antitopics'. Green (1980) mentions the following constructions: (a) Preposed AP inversion: e.g. *but most important are their fascinating detail*; (b) *Such ... that* inversion: e.g. *Súch was my respect for him, that ...*; (c) *So ... that* inversion: e.g. *but só little does that matter that ...*; (d) Negative adverb inversion: e.g. *Rarely did I hear such ...*; *Only when ... does ...*; *Not since ... had ...*; (e) *No sooner ... than* inversion: e.g. *No sooner was the door closed than . . .*; (f) Past participle inversion: *Expécted to draw considerable interest are ...* (Green 1980).

33. The Spanish newspaper *El Diario Vasco*, June 16, 1996, carried the following excerpt about the replacement of the draft in Spain by an all-volunteer army: "Francisco Castañón, director de la Oficina del Defensor del Soldado (ODS), dice que la situación es clara. «Hay una carrera con el lema '**tonto el último que haga la mili**'», explica. «Nadie quiere ser el último soldado...» Translation: "FC, director of the Office of the Defender of Soldiers says the situation is very clear. "There is a race on with the motto: 'the last one to be drafted is dumb', he explains. 'Nobody wants to be the last soldier...'" (my translation, J.A.).

Bibliography

- Abaitua, Joseba and Trask, Robert L. 1987. Accusativity in Basque: Reply to Bossong (short note). *Linguistics* 25, pp. 395-401.
- Abbott, Miriam. 1991. Macushi. In: Desmond Derbyshire and Geoffrey K. Pullum, eds., *Handbook of Amazonian languages, Vol. 3*, pp. 23-160. Berlin: Mouton de Gruyter.
- Abraham, Werner and de Meij, Sjaak, eds. 1986. *Topic, focus, and configurationality*. Amsterdam and Philadelphia: John Benjamins.
- Agirre Badiola, Domingo de. 1966/1912. *Garaa*. Arantzazu: Arantzazuko Frantziskotar Argitaldaria.
- Agirre Berezibar, Jesus Maria. 1991. *Euskal gramatika deskriptiboa*. Bilbo: Labayru Ikastegia.
- Aissen, Judith L. 1992. Topic and focus in Mayan. *Language* 68.1, pp. 43-80.
- Aitchison, Jean. 1987. The language lifegame: Prediction, explanation and linguistic change. In: Willem Koopman, Frederike van der Leek, Olga Fischer, and Roger Eaton, eds., *Explanation and linguistic change* (Amsterdam studies in the theory and history of linguistic science, Series IV - Current issues in linguistic theory, 45), pp. 11-32. Amsterdam and Philadelphia: John Benjamins. (From a workshop on 'Explanation and Linguistic Change' held at the University of Amsterdam on April 9 and 10, 1985.)
- _____. 1989. Spaghetti junctions and recurrent routes: Some preferred pathways in language evolution. *Lingua* 77, pp. 151-71.
- _____. 1992. Relative clauses in Tok Pisin: Is there a natural pathway? In: Marinel Gerritsen and Dieter Stein, eds., *Internal and external factors in syntactic change* (Trends in linguistics. Studies and monographs, 61), pp. 295-316. Berlin: Walter de Gruyter. (A selection of papers that were presented at the workshop ... held during the Ninth International Conference on Historical Linguistics at Rutgers University in August 1989.)
- Altube, Sebero (Altube'tar Seber). 1920. *De sintaxis euskérica*. San Sebastián: Imprenta de Martín Mena y Cía.

- _____. 1929. *Erderismos* (Euskera Euskaltzaindia' ren lan eta agiriak 10.1-4). Bermeo: Euskaltzaindia (Gaubeka' ren Irarkolan). (Facsimile edition: Bilbao: Indauchu Editorial, 1975.)
- _____. 1934. *Observaciones al Tratado de "Morfología Vasca" de Don R. M. de Azkue*. Bermeo, Bizkaia: Gaubeka Idaztegia Irarkola. (Bound with the original work by Azkue and a short appendix by Eguskitza in a facsimile edition: Edición popular, 10. Bilbo, Editorial La Gran Enciclopedia Vasca, 1969.)
- _____. 1956. Literatur-euskera. *Euskera* 1, pp. 3-13. (Reply to Mitxelena 1953; followed by Villasante 1956.)
- Altuna Bengoechea, Francisco María (Patxi). 1979. *Versificación de Dechepare: Métrica y pronunciación*. Bilbo: Editorial Mensajero.
- Ammann, H. 1928. *Die Menschliche Rede 2*. Der Satz, Lahr i. B. (Cited in Sgall 1987.)
- Amuriza, Xabier. 1983. *Hil ala bizi* (Euskal literatura). Donostia: Elkar.
- Andersen, Henning. 1988. Center and periphery: Adoption, diffusion and spread. In: Jacek Fisiak, ed., *Historical dialectology*, pp. 39-83. Berlin: Mouton de Gruyter. (Cited in Trudgill 1992.)
- _____. 1989. Understanding linguistic innovations. In: Leiv Egil Breivik and Ernst Håkon Jahr, eds., *Language change: Contributions to the studies of its causes* (Trends in linguistics. Studies and monographs, 43), pp. 5-27. Berlin: Mouton de Gruyter. (Mostly papers presented at the symposium "The causes of language change: Do we know them yet?" held at the School of Languages and Literature, University of Tromsø, October 15-17, 1987.)
- Andersen, Roger W. 1982. Determining the linguistic attributes of language attrition. In: Richard D. Lambert and Barbara F. Freed, eds., *The loss of language skills*, pp. 83-118. Rowley, MA: Newbury House Publishers. (Based on a conference held May 1980 at the University of Pennsylvania.)
- _____. 1983. Transfer to somewhere. In: S. Gass and L. Selinker, eds., *Language transfer in language learning* Cambridge, MA: Newbury House/Harper & Row.
- _____. 1984. The One to One Principle of interlanguage construction. *Language Learning* 34, pp. 77-95.
- _____. 1989. The "up" and "down" staircase in secondary language development. In: Nancy D. Dorian, ed., *Investigating obsolescence: Studies in language contraction and death* (Studies in the Social and Cultural Foundations of Language, 7), pp. 385-94. Cambridge: Cambridge University Press.

- Anderson, Stephen R. 1977. On mechanisms by which languages become ergative. In: Charles N. Li, ed., *Mechanisms of syntactic change*, pp. 317-63. Austin: University of Texas Press.
- _____. 1992. *A-Morphous morphology* (Cambridge Studies in Linguistics). Cambridge: Cambridge University Press.
- _____. 1993. Wackernagel's revenge: Clitics, morphology, and the syntax of second position. *Language* 69.1, pp. 68-98.
- Andersson, Anders-Borje and Dahl, Östen. 1974. Against the penthouse principle. *Linguistic Inquiry* 5, pp. 451-3.
- Antinucci, Francesco; Duranti, Alessandro, and Gebert, Lucyna. 1979. Relative clause structure, relative clause perception, and the change from SOV to SVO. *Cognition* 7, pp. 145-76.
- Anttila, Raimo. 1972. *Introduction to historical and comparative linguistics*. New York: Macmillan. (Cited in Bolinger 1979.)
- Araujo, Frank P. 1975. Counting sheep in Basque. *Anthropological Linguistics* 17.4, pp. 139-45.
- Arejita, Adolfo. 1980. Aditzaren galdegai funtzinoaz. *Euskera* 25.2, pp. 355-69.
- Aristar, Anthony Rodrigues. 1991. On diachronic sources and synchronic pattern: An investigation into the origin of linguistic universals. *Language* 67.1, pp. 1-33.
- Aske, Jon. 1987. The accusativity/ergativity balance in a non-split ergative language: The case of Euskara (also known as Basque). In: Jon Aske, Natasha Beery, Laura Michaelis, and Hana Filip, eds., *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society* (General Session and Parasession on Grammar and Cognition), pp. 1-14. Berkeley, CA: Berkeley Linguistics Society.
- _____. 1989. Path predicates in English and Spanish: A closer look. In: Kira Hall, Michael Meacham, and Richard Shapiro, eds., *Proceedings of the Fifteenth Annual Meeting of the Berkeley Linguistics Society* (General Session and Parasession on Theoretical Issues in Language Reconstruction), pp. 1-14. Berkeley, CA: Berkeley Linguistics Society.
- Atxaga, Bernardo. 1985. *Bi anai*. Donostia: Erein. (Cited in Agirre Berezibar 1991.)
- _____. 1988. *Obabakoak*. Donostia: Erein.
- _____. 1991. *Behi euskaldun baten memoriak*. Iruñea: Pamiela.
- Axular, Pedro de. 1977/1643. *Gero*. Oñati: Jakin. (Originally published in Bordeaux.)

- Azkue, Resurrección M. 1891. *Euskal-izkindea/Gramática éuskara*. Bilbo: Astui-ko Jose-n Moldegintzan / Tipografía de José de Astuy.
- _____. 1923-5/1969. *Morfología vasca* (Euskera IV-VI). Bilbo: Euskaltzaindia. (Reprinted in three volumes: Bilbo, La Gran Enciclopedia Vasca, 1969; the reprinted version has appendices by Alube, Eguskitza, Azkue.)
- _____. 1935. Observaciones acerca de *Euskaltzaindiari* de Altube' tar' S. *Euskera* 16, pp. 144-49.
- _____. 1989/1966/1947. *Euskaleriaren yakintza / Literatura popular del Pas Vasco II: Ipuin ta irakurgaiak*. Madrid: Espasa-Calpe.
- Ball, Catherine N. 1991. *The historical development of the it-cleft*. Philadelphia: University of Pennsylvania.
- _____. 1994. The origins of informative-presupposition *it-cleft*. *Journal of Pragmatics* 22, pp. 603-28.
- Barandiaran, Jose Migel. 1973. *Obras completas, II*. Bilbo.
- Bates, Elizabeth. 1978. Functionalism and the biology of language. *Papers and reports on child language development* 15, p. K1-K26.
- _____. 1994. Re: What is meaning? Resent by: COGLING: cogling@ucsd.edu. Author's address: bates@crl.ucsd.edu. Date sent: Mon, 25 Apr 94 09:19:09 PDT.
- Bates, Elizabeth and MacWhinney, Brian. 1979. A functionalist approach to the acquisition of grammar. In: Elinor Ochs and Bambi B. Schieffelin, eds., *Developmental pragmatics*, pp. 167-211. New York: Academic Press.
- _____. 1987. Competition, variation, and language learning. In: Brian MacWhinney, ed., *Mechanisms of language acquisition*, pp. 157-93. Hillsdale, NJ: Lawrence Erlbaum Associates.
- _____. 1989. Functionalism and the competition model. In: Brian MacWhinney and Elizabeth Bates, eds., *The crosslinguistic study of sentence processing*, pp. 3-76. Cambridge: Cambridge University Press.
- Bates, Elizabeth; McNew, Sandra; MacWhinney, Brian; Devescovi, Antonella, and Smith, Stan. 1982. Functional constraints on sentence processing. *Cognition* 11, pp. 245-99.
- Bean, Marian C. 1983. *The development of word order patterns in Old English* (Croom Helm linguistics series). London: Croom Helm; and Totowa, NJ: Barnes & Noble.

- Behaghel, Otto. 1932. *Deutsche Syntax: Eine geschichtliche Darstellung. 4: Wortstellung, Periodenbau*. Heidelberg: Winter.
- Bernstein, Basil. 1971. *Class, codes and control, Volume 1*. London: Routledge and Kegan Paul. (Cited in Trudgill 1992.)
- Bever, Thomas G. 1970. The cognitive basis for linguistic structures. In: John R. Hayes, ed., *Cognition and the development of language*, pp. 279-352. New York: John Wiley.
- _____. 1975. Functional explanations require independently motivated functional theories. In: Robin E. Grossman, L. James San, and Timothy J. Vance, eds., *Papers from the parasession on functionalism: April 17, 1975* (Eleventh regional meeting of the Chicago Linguistic Society), pp. 580-609. Chicago: Chicago Linguistic Society.
- Biber, Douglas. 1984. *A model of textual relations within the written and spoken modes*. Ph. D. dissertation, University of Southern California.
- _____. 1985. Investigating macroscopic textual variation through multi-feature/multi-dimensional analyses. *Linguistics* 23, pp. 337-60.
- _____. 1986a. On the investigation of spoken/written differences. *Studia Linguistica* 40, pp. 1-21.
- _____. 1986b. Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language* 62, pp. 384-414.
- _____. 1987. A textual comparison of British and American writing. *American Speech* 62, pp. 99-119.
- _____. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Bickerton, Derek. 1993. Subject focus and pronouns. In: Francis Byrne and Donald Winford, eds., *Focus and grammatical relations in Creole languages* (Creole language library, a companion series to the "Journal of Pidgin and Creole Languages", volume 12), pp. 189-212. Amsterdam and Philadelphia: John Benjamins.
- Bilbao, Felix. 1973/1954. *Ipuin barreka*. Arantzazu: Editorial Franciscana. (Cited in Agirre Berezibar 1991.)

- Birnbaum, Henrik. 1984. Notes on syntactic change: Co-occurrence vs. substitution, stability vs. permeability. In: Jacek Fisiak, ed., *Historical syntax* (Trends in linguistics. Studies and monographs, 23), pp. 25-46. Berlin: Mouton. (Papers presented at the Third International Conference on Historical Linguistics, devoted to historical syntax, held at Adam Mickiewicz University, at Błażejewko, Poland, 31 March - 3 April, 1981; Cited in Fischer 1992.)
- Birner, Betty J. 1994. Information status and word order: An analysis of English inversion. *Language* 70.2, pp. 233-59.
- Bolinger, Dwight. 1954. English prosodic stress and Spanish sentence order. *Hispania* 37, pp. 152-56.
- _____. 1961a. Syntactic blends and other matters. *Language* 37.3, pp. 366-81. (A reply to Lees 1960.)
- _____. 1961b. Contrastive accent and contrastive stress. *Language* 37.1, pp. 83-96.
- _____. 1967. Adjectives in English: Attribution and predication. *Lingua* 18, pp. 1-34.
- _____. 1972a. Adjective position again. *Hispania* 55.1, pp. 91-94.
- _____. 1972b. Accent is predictable (if you're a mind-reader). *Language* 48.3, pp. 633-44.
- _____. 1977a. Another glance at main clause phenomena. *Language* 53.3, pp. 511-19.
- _____. 1977b. *Meaning and form*. London: Longman.
- _____. 1979. Meaning and memory. In: George G. Haydu, ed., *Experience forms: Their cultural and individual place and function* (World anthropology), pp. 95-111. The Hague: Mouton. (Papers from the IXth International Congress of Anthropological and Ethnological Sciences. "This paper was originally delivered as a lecture at Brown University, April 29, 1974. It has been revised slightly from the previous published version in *Forum Linguisticum*, 1976, 1:1-14.)
- _____. 1986. *Intonation and its parts*. Stanford, CA: Stanford University Press.
- _____. 1989. *Intonation and its uses*. Stanford, CA: Stanford University Press.
- _____. 1992. The role of accent in extraposition and focus. *Studies in Language* 16.2, pp. 265-324.
- Bolkestein, A. Machtelt. 1995. Functions of VS order in Latin. In: Hans-Jürgen Sasse and Yaron Matras, eds., *Sprachtypologie und Universalienforschung = Verb-subject order and theticity in European languages* 48.1-2, pp. 32-43.

- Bollenbacher, John. 1977. The Basque passive. In: William A. Douglass, Richard W. Etulain, and William H. Jacobsen, Jr., eds., *Anglo-American contributions to Basque studies: Essays in honor of Jon Bilbao* (Desert Research Institute publications in the Social Sciences, no. 13), pp. 181-92. Reno, NV: Desert Research Institute.
- Boost, K. 1956. *Neue Untersuchungen zum Wesen und Struktur des Deutschen Satzes*. Berlin: Akademie-Verlag. (Cited in Sgall 1987.)
- Bossong, Georg. 1984. Ergativity in Basque. *Linguistics* 22, pp. 341-92.
- Bréal, Michel. 1897/1900/1964. *Essai de sémantique*. Paris. (English translation: *Semantics: Studies in the science of meaning*, London: Heinemann, 1900; London and New York: Dover Publications, 1964.)
- Breivik, Leiv Egil. 1991. On the typological status of Old English. In: Dieter Kastovsky, ed., *Historical English syntax* (Topics in English linguistics, 2), pp. 31-50. Berlin: Mouton de Gruyter. (Proceedings of the Kellner-Festival, held in 1988: at Schloss Liechtenstein.)
- Bresnan, Joan. 1994. Locative inversion and the architecture of Universal Grammar. *Language* 79.1, pp. 72-131.
- Brody, Jill. 1984. Some problems with the concept of basic word order. *Linguistics* 22, pp. 711-36.
- Brown, Gillian and Yule, George. 1983. *Discourse analysis*. Cambridge: Cambridge University Press.
- Bybee, Joan L. 1985a. Diagrammatic iconicity in stem-inflection relations. In: John Haiman, ed., *Iconicity in syntax* (Typological studies in language, 6), pp. 11-48. Amsterdam and Philadelphia: John Benjamins. (Proceedings of a symposium on iconicity in syntax, Stanford, June 24-6, 1983.)
- _____. 1985b. *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.
- _____. 1988a. The diachronic dimension in explanation. In: John A. Hawkins, ed., *Explaining language universals*, pp. 350-379. Oxford: Basil Blackwell.
- _____. 1988b. Morphology as lexical organization. In: Michael Hammond and Michael Noonan, eds., *Theoretical morphology: Approaches in modern linguistics*, pp. 119-41. San Diego: Academic Press. (Cited in Pinker and Prince 1994.)

- Bybee, Joan L.; Pagliuca, William; and Perkins, Revere D. 1990. On the asymmetries in the affixation of grammatical material. In: William Croft, Keith Denning, and Suzanne Kemmer, eds., *Studies in typology and diachrony: Papers presented to Joseph H. Greenberg on his 75th birthday* (Typological studies in language, 20), pp. 1-42. Amsterdam and Philadelphia: John Benjamins.
- _____. 1991. Back to the future. In: Elizabeth Closs Traugott and Bernd Heine, *Approaches to grammaticalization. Volume II: Focus on types of grammatical markers* (Typological studies in language, 19:2), pp. 17-58. Amsterdam and Philadelphia: John Benjamins.
- Bybee, Joan L., Perkins, Revere, and Pagliuca, William. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago and London: University of Chicago Press.
- Campbell, Lyle. 1980. Explaining universals and their exceptions. In: Elizabeth Closs Traugott, Rebecca LaBrum, and Susan Shepherd, eds., *Papers from the Fourth International Conference on Historical Linguistics*, pp. 17-26. Amsterdam: John Benjamins. (Cited in Campbell, Bubenik, and Saxon 1988.)
- Campbell, Lyle; Bubenik, Vit, and Saxon, Leslie. 1988. Word order universals: Refinements and clarifications. *Canadian Journal of Linguistics* 33.3, pp. 209-30.
- Campbell, Lyle and Muntzel, Martha C. 1989. The structural consequences of language death. In: Nancy D. Dorian, ed., *Investigating obsolescence: Studies in language contraction and death* (Studies in the Social and Cultural Foundations of Language, 7), pp. 181-96. Cambridge: Cambridge University Press.
- Cerquand, Jean François. 1986/1874. *Ipar Euskal Herriko legendea eta ipuinak* (2 volumes). Donostia: Txertoa. (Cited in Agirre Berezibar 1991.)
- Chafe, Wallace. 1974. Language and consciousness. *Language* 50.1, pp. 111-33.
- _____. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In: Charles N. Li, ed., *Subject and topic*, pp. 25-55. New York: Academic Press.
- _____. 1980. The deployment of consciousness in the production of a narrative. In: Wallace Chafe, ed., *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production* (Advances in Discourse Processing, 3), pp. 9-50. Norwood, NJ: Ablex.
- _____. 1982. Integration and involvement in speaking, writing, and oral literature. In: Deborah Tannen, ed., *Spoken and written language: Exploring orality and literacy*, pp. 35-54. Norwood, NJ: Ablex.

- _____. 1984a. Speaking, writing, and prescriptivism. In: Deborah Schiffrin, ed., *Meaning, form, and use in context: Linguistic applications* (GURT 84: Georgetown University round table conference in linguistics, 1984), pp. 95-103. Washington, DC: Georgetown University Press.
- _____. 1984b. How people use adverbial clauses. In: Claudia Brugman and Monica Macaulay, eds., *Proceedings of the Tenth Annual Meeting of the Berkeley Linguistics Society* (General Session and Parasession on Subordination), pp. 437-49. Berkeley: Berkeley Linguistics Society.
- _____. 1985. Linguistic differences produced by differences between speaking and writing. In: David R. Olson, Nancy Torrance, and Angela Hildyard, eds., *Literacy, language, and learning: The nature and consequences of reading and writing* Cambridge: Cambridge University Press.
- _____. 1986a. Academic speaking. In: Vassiliki Nikiforidou, Mary VanClay, Mary Niepokuj, and Deborah Feder, eds., *Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society* (General Session and Parasession on Semantic Typology), pp. 26-40. Berkeley, CA: Berkeley Linguistics Society.
- _____. 1986b. Writing in the perspective of speaking. In: Charles R. Cooper and Sidney Greenbaum, eds., *Studying writing: Linguistic approaches* (Written communication annual, an international survey of research and theory, vol. 1), pp. 12-39. Beverly Hills, CA: Sage Publications.
- _____. 1987. Cognitive constraints on information flow. In: Russell S. Tomlin, ed., *Coherence and grounding in discourse*, pp. 21-52. Amsterdam: John Benjamins.
- _____. 1988a. Linking intonation units in spoken English. In: John Haiman and Sandra A. Thompson, eds., *Clause combining in grammar and discourse* (Typological studies in language, 18), pp. 1-28. Amsterdam and Philadelphia: John Benjamins.
- _____. 1988b. Punctuation and the prosody of written language. *Written Communication* 5, pp. 395-426.
- _____. 1991. Grammatical subjects in speaking and writing. In: Rosalind Horowitz, ed., *Studies of orality and literacy: Critical issues for the practice of schooling* = *Text* 11.1, pp. 45-72.
- _____. 1992a. Information flow in speaking and writing. In: Pamela Downing, Susan D. Lima, and Michael Noonan, eds., *The linguistics of literacy* (Typological studies in language, 21), pp. 17-29. Amsterdam and Philadelphia: John Benjamins. (From the Seventeenth Annual University of Wisconsin-Milwaukee Linguistics Symposium, which was held in Milwaukee on April 8-10, 1988.)

- _____. 1992b. The flow of ideas in a sample of written language. In: William C. Mann and Sandra A. Thompson, eds., *Discourse description: Diverse linguistic analyses of a fund-raising text* (Pragmatics and Beyond, new series, 16), pp. 267-94. Amsterdam and Philadelphia: John Benjamins.
- _____. 1994a. *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: The University of Chicago Press.
- _____. 1994b. Review of Jan Firbas, Functional sentence perspective in written and spoken communication, 1992. *Language* 70.2, pp. 350-353.
- Chafe, Wallace, ed. 1980. *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production* (Advances in Discourse Processing, 3). Norwood, NJ: Ablex.
- Chafe, Wallace and Danielewicz, Jane. 1987. Properties of spoken and written language. In: Rosalind Horowitz and S. Jay Samuels, eds., *Comprehending oral and written language* New York: Academic Press.
- Chafe, Wallace and Tannen, Deborah. 1987. The relation between written and spoken language. *Annual Review of Anthropology* 16, pp. 383-407.
- Chao, Yuen Ren. 1968. *A grammar of spoken Chinese*. Berkeley and Los Angeles: University of California.
- Chomsky, Noam. 1964/1961. On the notion "Rule of Grammar". In: Jerry A. Fodor and Jerrold J. Katz, eds., *The structure of language: Readings in the philosophy of language*, pp. 119-36. Englewood Cliffs, NJ: Prentice-Hall. (Originally published in *Structure of language and its mathematical aspects, Proceedings of the Twelfth Symposium in Applied Mathematics*, R. Jakobson, ed., pp. 6-24. Providence, RI: The American Mathematical Society, 1961.)
- _____. 1965. *Aspects of the theory of syntax*. Cambridge, MA: The MIT Press.
- _____. 1971/1970. Deep structure, surface structure, and semantic interpretation. In: Danny D. Steinberg and Leon A. Jakobovits, eds., *Semantics: An interdisciplinary reader in philosophy, linguistics, and psychology*, pp. 183-216. Cambridge: Cambridge University Press. (Original version: In: Roman Jakobson and Shigeo Kawamoto, eds. 1970. *Studies in general and Oriental linguistics presented to Shiro Hattori on the occasion of his sixtieth birthday*. Tokyo: TEC Co. Ltd, pp. 52-91. Reprinted also in Noam Chomsky. 1972. *Studies on semantics in generative grammar*. The Hague: Mouton, pp. 62-119.)

- _____. 1992/1987. On the nature, use and acquisition of language. In: Martin Pütz, ed., *Thirty years of linguistic evolution: Studies in honour of René Dirven on the occasion of his sixtieth birthday*, pp. 3-29. Amsterdam and Philadelphia: John Benjamins. (Originally published in Kusakabe, Tokuji, ed., *Generative grammar: Its basis, development and prospects*, Kyoto University Press, 1987.)
- Clark, Eve V. and Clark, Herbert H. 1978. Universals, relativity, and language processing. In: Joseph H. Greenberg, Charles A. Ferguson, and Edith A. Moravcsik, eds., *Universals of human language, vol. 1: Method and theory* Stanford, CA: Stanford University Press.
- Clark, Herbert H. and Haviland, Susan E. 1977. Comprehension and the Given-New contract. In: Roy O. Freedle, ed., *Discourse production and comprehension*, pp. 1-40. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Comrie, Bernard. 1981/1989. *Language universals and linguistic typology* (1st edition). Chicago: University of Chicago Press.
- _____. 1984. Some formal properties of focus in Modern Eastern Armenian. *Annual of Armenian Linguistics* 5, pp. 1-21.
- _____. 1989/1981. *Language universals and linguistic typology* (Second edition). Oxford: Blackwell.
- _____. 1993. Language universals and linguistic typology: Data-bases and explanations. *Sprachtypologie und Universalienforschung (STUF)* 46.1, pp. 3-14.
- Comrie, Bernard and Matthews, Stephen. 1990. Prolegomena to a typology of Tough Movement. In: William Croft, Keith Denning, and Suzanne Kemmer, eds., *Studies in typology and diachrony: Papers presented to Joseph H. Greenberg on his 75th birthday* (Typological studies in language, 20), pp. 43-58. Amsterdam and Philadelphia: John Benjamins.
- Connolly, John H. 1991. *Constituent order in functional grammar: Synchronic and diachronic perspectives* (Functional Grammar series, 14). Berlin: Foris.
- Cooper, Charles R. and Greenbaum, Sidney, eds. 1986. *Studying writing: Linguistic approaches* (Written communication annual, an international survey of research and theory, vol. 1). Beverly Hills, CA: Sage Publications.
- Coopmans, Peter. 1989. Where stylistic and syntactic processes meet: Locative inversion in English. *Language* 65, pp. 728-51.
- Cooreman, Ann. 1992. The pragmatics of word order variation in Chamorro. In: Doris L. Payne, ed., *Pragmatics of word order flexibility* (Typological studies in language, 22), pp. 265-78. Amsterdam and Philadelphia: John Benjamins.

- Corrigan, Roberta L. and Lima, Susan D. 1994. Introduction. In: Susan D. Lima, Roberta L. Corrigan, and Gregory K. Iverson, eds., *The reality of linguistic rules* (Studies in language companion series, 26), p. xiii-xxiii. Amsterdam and Philadelphia: John Benjamins. (Selected papers from the 21st annual University of Wisconsin, Milwaukee, linguistics symposium, April 10-12, 1992.)
- Creider, Chet. 1977. Functional sentence perspective in a verb-initial language. In: Paul F. A. Kotey and Haig Der-Houssikian, eds., *Language and linguistic problems in Africa*, pp. 330-343. Columbia, SC: Hornbeam Press. (Proceedings of the VII Conference on African Linguistics, University of Florida, Gainesville, April 23-25, 1976; selected papers.)
- Croft, William. 1990. *Typology and universals* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.
- _____. 1991. The evolution of negation. *Journal of Linguistics* 27.1, pp. 1-27.
- _____. 1993. Functional-typological theory in its historical and intellectual context. *Sprachtypologie und Universalienforschung* 46.1, pp. 15-26.
- _____. 1995. Modern syntactic typology. In: Masayoshi Shibatani and Theodora Bynon, eds., *Approaches to language typology*, pp. 85-144. Oxford: Clarendon Press (Oxford University Press).
- Croft, William, Denning, Keith, and Kemmer, Suzanne. 1990. Typology and diachrony in the work of Joseph H. Greenberg. In: William Croft, Keith Denning, and Suzanne Kemmer, eds., *Studies in typology and diachrony: Papers presented to Joseph H. Greenberg on his 75th birthday* (Typological studies in language, 20), p. ix-xviii. Amsterdam and Philadelphia: John Benjamins.
- _____, eds. 1990. *Studies in typology and diachrony: Papers presented to Joseph H. Greenberg on his 75th birthday* (Typological studies in language, 20). Amsterdam and Philadelphia: John Benjamins.
- Cruttenden, Alan. 1986. *Intonation*. Cambridge: Cambridge University Press.
- Crystal, David. 1972. The intonation system of English. In: Dwight Bolinger, ed., *Intonation*. Harmondsworth, England: Penguin.
- _____. 1992. The changing English language: Fiction and fact. In: Martin Pütz, ed., *Thirty years of linguistic evolution: Studies in honour of René Dirven on the occasion of his sixtieth birthday*, pp. 119-30. Amsterdam and Philadelphia: John Benjamins.
- Crystal, David and Davy, D. 1975. *Advanced conversational English*. London: Longman. (Cited in Crystal 1992.)

- Cumming, Susanna. 1995. Agent position in the *Sejarah Melayu*. In: Pamela Downing and Michael Noonan, eds., *Word order in discourse* (Typological studies in language, 30), pp. 51-83. Amsterdam and Philadelphia: John Benjamins. (Selection of papers presented at a symposium entitled *Word Order in Discourse* held at the University of Wisconsin-Milwaukee, April 12-14, 1991, UWM Linguistics Symposium, 20, chaired by Edith Moravcsik and Michael Noonan.)
- Cutler, Anne; Hawkins, John A., and Gilligan, Gary. 1985. The suffixing preference: A processing explanation. *Linguistics* 23, pp. 723-58.
- Dahl, Östen. 1979. Typology of sentence negation. *Linguistics* 17, pp. 79-106.
- Danchev, Andrei. 1984. Translation and syntactic change. In: Jacek Fisiak, ed., *Historical syntax* (Trends in linguistics. Studies and monographs, 23), pp. 47-60. Berlin: Mouton. (Papers presented at the Third International Conference on Historical Linguistics, devoted to historical syntax, held at Adam Mickiewicz University, at Błażejewko, Poland, 31 March - 3 April, 1981; Cited in Fischer 1992.)
- _____. 1991. Language change typology and some aspects of the SVO development in English. In: Dieter Kastovsky, ed., *Historical English syntax* (Topics in English linguistics, 2), pp. 103-24. Berlin: Mouton de Gruyter. (Proceedings of the Kellner-Festival, held in 1988: at Schloss Liechtenstein.)
- Danielewicz, Jane and Chafe, Wallace. 1985. How 'normal' speaking leads to 'erroneous' punctuation. In: Sarah Warshauer Freedman, ed., *The acquisition of written language*, pp. 213-25. Norwood, NJ: Ablex.
- Daugherty, Kim G. and Seidenberg, Mark S. 1994. Beyond rules and exceptions: A connectionist approach to inflectional morphology. In: Susan D. Lima, Roberta L. Corrigan, and Gregory K. Iverson, eds., *The reality of linguistic rules* (Studies in language companion series, 26), pp. 353-88. Amsterdam and Philadelphia: John Benjamins. (Selected papers from the 21st annual University of Wisconsin, Milwaukee, linguistics symposium, April 10-12, 1992.)
- de Rijk, Rudolf P. G. 1969. Is Basque an S.O.V. language? *Fontes Linguae Vasconum* 1.3, pp. 319-51.
- _____. 1972a. *Studies in Basque syntax: Relative clauses*. Ph. D. dissertation, Massachusetts Institute of Technology (Cambridge, MA).
- _____. 1972b. Relative clauses in Basque: A guided tour. *The Chicago which hunt: Papers from the parasession of the eighth regional meeting of the Chicago Linguistic Society* (CLS 8) Chicago: Chicago Linguistic Society.
- _____. 1978. Topic fronting, focus positioning, and the nature of the verb phrase in Basque. In: Frank Jansen, ed., *Studies on fronting*, pp. 81-112. Lisse: Peter De Ridder.

- _____. 1988. Basque syntax and universal grammar. In: Eusko Jaurlaritza, *Conference on the Basque language of the 2nd Basque World Congress, Volume 1: Language Description*, pp. 69-88. Gasteiz: Eusko Jaurlaritza/Basque Government.
- _____. 1993. Basque hospitality and the suffix -ko. In: José Ignacio Hualde and Jon Ortiz de Urbina, eds., *Generative studies in Basque linguistics* (Amsterdam studies in the theory and history of linguistic science, Series IV: Current issues in linguistic theory, 105), pp. 145-62. Amsterdam and Philadelphia: John Benjamins.
- Deane, Paul D. 1988. Which NPs are there unusual possibilities for extraction from? *Papers from the general session of the twenty fourth regional meeting of the Chicago Linguistic Society* (CLS 24/1), pp. 89-100. Chicago: Chicago Linguistic Society.
- _____. 1991. Limits to attention: A cognitive theory of island phenomena. *Cognitive Linguistics* 2.1, pp. 1-63.
- _____. 1992. *Grammar in mind and brain: Explorations in cognitive syntax* (Cognitive linguistics research, 2). Berlin: Mouton de Gruyter.
- DeLancey, Scott. 1981. An interpretation of split ergativity and related patterns. *Language* 57.3, pp. 626-57.
- Derbyshire, Desmond C. 1977. Word order universals and the existence of OVS languages. *Linguistic Inquiry* 8.4, pp. 590-9.
- Derbyshire, Desmond C. and Pullum, Geoffrey K. 1981. Object-initial languages. *International Journal of American Linguistics (IJAL)* 47.3, pp. 192-214.
- _____, eds. 1986. *Handbook of Amazonian languages, volume 1*. Berlin: Mouton de Gruyter.
- Derwing, Bruce L. 1977. Is the child a little linguist? In: John Macnamara, ed., *Language learning and thought* (Perspectives in neurolinguistics and psycholinguistics), pp. 79-84. New York: Academic Press.
- Derwing, Bruce L. and Skousen, Royal. 1989. Real-time morphology: Symbolic rules or analogical networks? In: Kira Hall, Michael Meacham, and Richard Shapiro, eds., *Proceedings of the Fifteenth Annual Meeting of the Berkeley Linguistics Society* (General Session and Parasession on Theoretical Issues in Language Reconstruction), pp. 48-62. Berkeley, CA: Berkeley Linguistics Society.
- _____. 1994. Productivity and the English past tense: Testing Skousen's Analogy Model. In: Susan D. Lima, Roberta L. Corrigan, and Gregory K. Iverson, eds., *The reality of linguistic rules* (Studies in language companion series, 26), pp. 193-218. Amsterdam and Philadelphia: John Benjamins. (Selected papers from the 21st annual University of Wisconsin, Milwaukee, linguistics symposium, April 10-12, 1992.)

- Dik, Simon C. 1978. *Functional Grammar* (North-Holland linguistic series, 37). Amsterdam: North-Holland.
- _____. 1981/1978. *Functional Grammar* (Publications in language sciences, 7). Dordrecht, NL: Foris. (Third printing with minor changes; otherwise, identical type and pagination; first printing published by North-Holland, Amsterdam, 1978.)
- _____. 1989. *The theory of Functional Grammar; Part I: The structure of the clause*. Dordrecht: Foris.
- Dik, Simon C., et al. 1981. On the typology of focus phenomena. In: Teun Hoekstra, Harry van der Hulst, and Michael Moortgat, eds., *Perspectives on functional grammar*, pp. 41-74. Dordrecht: Foris. (Originally it appeared as *Leids Taalkundig Bulletin GLOT* 3, 41-74, 1980.)
- Dinnsen, Daniel A. 1980. Phonological rules and phonetic explanation. *Journal of Linguistics* 16, pp. 171-91.
- Dixon, Robert M. W. 1991. *A new approach to English grammar, on semantic principles*. Oxford: Clarendon Press; New York: Oxford University Press.
- Dorian, Nancy C. 1977. The problem of the semi-speaker in language death. *International Journal of the Sociology of Language* 12, pp. 23-32.
- _____. 1981. *Language death: The life cycle of a Scottish Gaelic dialect*. Philadelphia: University of Pennsylvania Press. (Cited in Giacalone Ramat 1992.)
- _____. 1982. Language loss and maintenance in language contact situations. In: Richard D. Lambert and Barbara F. Freed, eds., *The loss of language skills*, pp. 44-59. Rowley, MA: Newbury House. (Based on a conference held May 1980 at the University of Pennsylvania.)
- Downing, Pamela. 1995. Word order in discourse: By way of introduction. In: Pamela Downing and Michael Noonan, eds., *Word order in discourse* (Typological studies in language, 30), pp. 1-27. Amsterdam and Philadelphia: John Benjamins. (Selection of papers presented at a symposium entitled *Word Order in Discourse* held at the University of Wisconsin-Milwaukee, April 12-14, 1991 (UWM Linguistics Symposium, 20), chaired by Edith Moravcsik and Michael Noonan.)
- Downing, Pamela, Lima, Susan D., and Noonan, Michael, eds. 1992. *The linguistics of literacy* (Typological studies in language, a companion series to the journal "Studies in Language", volume 21). Amsterdam and Philadelphia: John Benjamins. (From the Seventeenth Annual University of Wisconsin-Milwaukee Linguistics Symposium, which was held in Milwaukee on April 8-10, 1988.)

- Downing, Pamela and Noonan, Michael, eds. 1995. *Word order in discourse* (Typological studies in language, 30). Amsterdam and Philadelphia: John Benjamins. (Selection of papers presented at a symposium entitled *Word Order in Discourse* held at the University of Wisconsin-Milwaukee, April 12-14, 1991 (UWM Linguistics Symposium, 20), chaired by Edith Moravcsik and Michael Noonan.)
- Dressler, Wolfgang U. 1988. Language death. In: Frederick J. Newmeyer, ed., *Language: The socio-cultural matrix* (Linguistics: The Cambridge survey, 4), pp. 184-92. Cambridge: Cambridge University Press.
- Dressler, Wolfgang U. and Wodak-Leodolter, R. 1977. Language preservation and language death in Brittany. *International Journal of the Sociology of Language* 12, pp. 33-44. (Cited in Giacalone Ramat 1992.)
- Dryer, Matthew S. 1980. The positional tendencies of sentential noun phrases in universal grammar. *Canadian Journal of Linguistics* 25.2, pp. 123-95.
- _____. 1988a. Universals of negative position. In: Michael Hammond, Edith Moravcsik, and Jessica Wirth, eds., *Studies in syntactic typology* (Typological studies in language, 17), pp. 93-124. Amsterdam and Philadelphia: John Benjamins.
- _____. 1988b. Object-verb order and adjective-noun order: Dispelling a myth. In: John A. Hawkins and Heather K. Holmback, eds., *Lingua* 74, pp. 185-217.
- _____. 1989a. Discourse-governed word order and word order typology. *Belgian Journal of Linguistics* 4, pp. 69-90.
- _____. 1989b. Large linguistic areas and language sampling. *Studies in Language* 13.2, pp. 257-92.
- _____. 1991. SVO languages and the OV:VO typology. *Journal of Linguistics* 27, pp. 443-82.
- _____. 1992a. The Greenbergian word order correlations. *Language* 68.1, pp. 81-138.
- _____. 1992b. Review of Doris L. Payne, *The pragmatics of word order: Typological dimensions of verb initial languages* (Empirical approaches to language typology). Berlin: Mouton de Gruyter, 1990. *Canadian Journal of Linguistics* 37.1, pp. 82-84.
- _____. 1992c. Review of William Croft, Keith Denning, and Suzanne Kemmer, eds., *Studies in typology and diachrony: Papers presented to Joseph H. Greenberg on his 75th birthday* (Typological studies in language). Amsterdam: John Benjamins, 1990. *Canadian Journal of Linguistics* 37.1, pp. 79-82.

- _____. 1994. Grammatical Relations. Resent by: FUNKNET@OREGON.UOREGON.EDU. Author's address: Matthew Dryer <LINDRYER@ubvms.cc.buffalo.edu>. Date sent: Tue, 27 Sep 1994 00:05:38 0400 (EDT).
- _____. 1995. Frequency and pragmatically unmarked word order. In: Pamela Downing and Michael Noonan, eds., *Word order in discourse* (Typological studies in language, 30), pp. 105-35. Amsterdam and Philadelphia: John Benjamins. (Selection of papers presented at a symposium entitled *Word Order in Discourse* held at the University of Wisconsin-Milwaukee, April 12-14, 1991 (UWM Linguistics Symposium, 20), chaired by Edith Moravcsik and Michael Noonan.)
- Du Bois, John W. 1981. Ergativity and preferred argument structure. Ms., University of California, Santa Barbara. (Cited in Lambrecht 1994.)
- _____. 1985. Competing motivations. In: John Haiman, ed., *Iconicity in syntax* (Typological studies in language, 6), pp. 343-65. Amsterdam and Philadelphia: John Benjamins. (Proceedings of a symposium on iconicity in syntax, Stanford, June 24-6, 1983.)
- _____. 1987. The discourse basis of ergativity. *Language* 63.4, pp. 805-55.
- Du Bois, John W.; Cumming, Susanna; and Schuetze-Coburn, Stephan. 1988. Discourse transcription. In: Sandra A. Thompson, ed., *Santa Barbara papers in linguistics, vol. 2: Discourse and grammar*, pp. 1-72. Santa Barbara, CA: Department of Linguistics.
- Du Bois, John W., Cumming, Susanna, Schuetze-Coburn, Stephan, and Paolino, Danae. 1992. *Discourse transcription* (Santa Barbara Papers in Linguistics, 4). Santa Barbara, CA: UCSB Department of Linguistics.
- Edwards, Jane A. and Lampert, Martin D., eds. 1993. *Talking data: Transcription and coding in discourse research*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Eguzkitza, Andolin. 1981. On the so-called passive in Basque. *Fontes Linguae Vasconum* 13 (No. 37), pp. 233-53.
- Elexpuru, Juan Martin. 1994. *Kuba triste dago*. Donostia: Elkar.
- Emonds, Joseph E. 1970. *Root and structure preserving transformations*. Ph. D. dissertation, Massachusetts Institute of Technology.
- _____. 1976. *A transformational approach to English syntax: Root, structure-preserving, and local transformations*. New York: Academic Press.
- Enkvist, Nils Erik. 1982. Report on Topic 2: Aspects of text analysis and generation. In: Sture Allén, ed., *Text processing: Text analysis and generation, text typology and attribution* (Data linguistica, 16), pp. 617-35. Stockholm, Sweden: Almqvist and Wiksell International. (Proceedings of Nobel Symposium 51.)

- Erguvanli, Eser Emine. 1984. *The function of word order in Turkish grammar*. Berkeley: University of California Press.
- Erteschik-Shir, Nomi. 1979. Discourse constraints on dative movement. In: T. Givón, ed., *Discourse and syntax* (Syntax and semantics, 12), pp. 441-68. New York: Academic Press.
- Erteschik-Shir, Nomi and Lappin, Shalom. 1979. Dominance and the functional explanation of island phenomena. *Theoretical Linguistics* 6, pp. 41-85.
- Euskaltzaindia (The Basque Language Academy). 1987a. *Euskal gramatika: Lehen urratsak - I: Eraskina* (Basque grammar: first steps, vol. 1 - Appendix). Bilbo: Euskaltzaindia. (Appendix to volume 1; by Euskaltzaindia's Grammar Committee, Gramatika Batzordea, headed by Patxi Altuna.)
- _____. 1987b. *Euskal gramatika: Lehen urratsak - II* (Basque grammar: first steps, vol. 2). Bilbo: Euskaltzaindia. (By Euskaltzaindia's Grammar Committee, Gramatika Batzordea, headed by Patxi Altuna.)
- _____. 1990. *Euskal gramatika: Lehen urratsak - III: Lokailuak* (Basque grammar: first steps, vol. 3: Connectors). Bilbo: Euskaltzaindia. (By Euskaltzaindia's Grammar Committee, Hizkuntz Batzordea, headed by Patxi Altuna.)
- _____. 1993. *Euskal gramatika laburra: Perpaus bakuna* (Short Basque grammar: Main clauses). Bilbo: Euskaltzaindia. (by Euskaltzaindia's Grammar Committee, headed by Patxi Altuna.)
- Faarlund, Jan Terje. 1981. Obligatory fronting in a verb initial language: An attempt at pragmatic syntax. *Papers from the seventeenth meeting of the Chicago Linguistic Society* (CLS 17) Chicago: Chicago Linguistic Society. (Cited in Gerritsen 1984.)
- Farkas, Donka F. 1984. The status of VP in Hungarian. *Papers from the twentieth regional meeting of the Chicago Linguistic Society* (CLS 20), pp. 65-76. Chicago: Chicago Linguistic Society.
- Ferguson, Charles A. 1990. From esses to aitches: Identifying pathways of diachronic change. In: William Croft, Keith Denning, and Suzanne Kemmer, eds., *Studies in typology and diachrony: Papers presented to Joseph H. Greenberg on his 75th birthday* (Typological studies in language, 20), pp. 59-78. Amsterdam and Philadelphia: John Benjamins.
- Fillmore, Charles J. 1968 [1967]. The case for case. In: Emmon Bach and Robert T. Harms, eds., *Universals in linguistic theory*, pp. 1-88. New York: Holt, Rinehart, and Winston.
- _____. 1969 [1966]. Toward a modern theory of case. In: David A. Reibel and Sanford A. Schane, eds., *Modern studies in English: Readings in Transformational Grammar*, pp. 361-75. Englewood Cliffs, NJ: Prentice-Hall.

- _____. 1976. Frame semantics and the nature of language. In: Steven R. Harnard et al., eds., *Origins and evolution of language and speech*, pp. 20-32. New York: New York Academy of Sciences.
- _____. 1977. The case for case reopened. In: Peter Cole and Jerry M. Sadock, eds., *Syntax and semantics 8: Grammatical relations* New York: Academic Press.
- _____. 1982. Frame semantics. In: Linguistic Society of Korea, ed., *Linguistics in the morning calm*, pp. 111-38. Seoul: Hanshin.
- _____. 1985a. Frames and the semantics of understanding. *Quaderni di Semantica* 6.2, pp. 222-54. («Quaderni di Semantica» Round Table on Frame Semantics I; from a workshop on “the use of script/frame in linguistic semantics” at the 13th International Congress of Linguists, Tokyo, Aug. 29-Sept. 4, 1982.)
- _____. 1985b. Syntactic intrusions and the notion of grammatical construction. In: Mary Niepokuj, Mary VanClay, Vassiliki Nikiforidou, and Deborah Feder, eds., *Proceedings of the Eleventh Annual Meeting of the Berkeley Linguistics Society* (General Session and Parasession on Poetics, Metrics, and Prosody), pp. 73-86. Berkeley, CA: Berkeley Linguistics Society.
- _____. 1988. The mechanisms of “Construction Grammar”. In: Shelley Axmaker, Annie Jaissner, and Helen Singmaster, eds., *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society* (General Session and Parasession on Grammaticalization), pp. 35-55. Berkeley, CA: Berkeley Linguistics Society.
- _____. 1989. Grammatical Construction theory and the familiar dichotomies. In: Rainer Dietrich and Carl F. Graumann, eds., *Language processing in social context* (North-Holland linguistic series, 54), pp. 17-38. Amsterdam: North-Holland.
- _____. 1994. A Construction Grammar tutorial. Abstract for the Cognitive Science Seminar of the Berkeley Cognitive Science Program, Friday, January 28, 1994.
- Fillmore, Charles J.; Kay, Paul, and O'Connor, Mary Catherine. 1988. Regularity and idiomaticity in grammatical constructions: The case of ‘let alone’. *Language* 64.3, pp. 501-38.
- Firbas, Jan. 1964. On defining the theme in functional sentence analysis. In: Josef Vachek and E. Pauliny, eds., *Travaux linguistiques de Prague 1: L'Ecole de Prague d'aujourd'hui*, pp. 267-80. Prague: Editions de l'Académie Tchèqueoslovaque des Sciences.
- _____. 1966. Non-thematic subjects in contemporary English. In: Josef Vachek and M. Komarek, eds., *Travaux linguistiques de Prague 2: Les problèmes du centre et de la périphérie du système de la langue*, pp. 239-56. University, AL: University of Alabama Press.

- _____. 1986. On the dynamics of written communication in the light of the theory of Functional Sentence Perspective. In: Charles R. Cooper and Sidney Greenbaum, eds., *Studying writing: Linguistic approaches* (Written communication annual, an international survey of research and theory, vol. 1), pp. 40-71. Beverly Hills, CA: Sage Publications.
- _____. 1987. On two starting points of communication. In: Ross Steele and Terry Threadgold, eds., *Language topics: Essays in honour of Michael Halliday* (Volume I), pp. 23-46. Amsterdam and Philadelphia: John Benjamins.
- _____. 1992. *Functional sentence perspective in written and spoken communication*. Cambridge: Cambridge University Press.
- Fischer, Olga. 1992. Syntactic change and borrowing: The case of the accusative-and-infinitive construction in English. In: Marinel Gerritsen and Dieter Stein, eds., *Internal and external factors in syntactic change* (Trends in linguistics. Studies and monographs, 61), pp. 17-88. Berlin: Walter de Gruyter. (A selection of papers that were presented at the workshop ... held during the Ninth International Conference on Historical Linguistics at Rutgers University in August 1989.)
- Fishman, Joshua A. 1972. *The sociology of language: An interdisciplinary social science approach to language in society*. Rowley, Massachusetts: Newbury House. (Cited in Seliger and Vago 1991.)
- Fodor, Janet Dean. 1984. Constraints on gaps: Is the parser a significant influence? In: Brian L. Butterworth, Bernard Comrie, and Östen Dahl, eds., *Explanations for language universals* (Published simultaneously as a special issue of *Linguistics*: 21.1), pp. 9-34. Berlin: Mouton.
- Fodor, Jerry A. 1983. *The modularity of mind*. Cambridge, MA: The MIT Press.
- Fodor, Jerry A., Bever, Thomas G., and Garrett, Merrill F. 1974. *The psychology of language: An introduction to psycholinguistics and generative grammar*. New York: McGraw-Hill.
- Foley, William A. and Van Valin, Robert, Jr. 1984. *Functional syntax and universal grammar* (Cambridge studies in linguistics, 38). Cambridge: Cambridge University Press.
- Fox, Barbara A. 1985. Word-order inversion and discourse continuity in Tagalog. *Special issue: Quantified studies in discourse = Text 5.1-2*, pp. 39-54.
- _____. 1995. The category 'S' in English conversation. In: Werner Abraham, T. Givón, and Sandra A. Thompson, eds., *Discourse grammar and typology: Papers in honor of John W. M. Verhaar* (Studies in language companion series, 27), pp. 153-78. Amsterdam and Philadelphia: John Benjamins.

- Fox, Barbara A. and Hopper, Paul J., eds. 1994. *Voice: Form and function* (Typological studies in language, 27). Amsterdam and Philadelphia: John Benjamins.
- Frazier, Lyn. 1979/1978. *On comprehending sentences: Syntactic parsing strategies*. Bloomington, IN: Indiana University Linguistics Club. (Ph. D. dissertation, University of Connecticut, 1978.)
- _____. 1979. Parsing and constraints on word order. *University of Massachusetts occasional papers in linguistics* 5, pp. 177-98.
- _____. 1985. Syntactic complexity. In: David R. Dowty, Lauri Karttunen, and Arnold M. Zwicky, eds., *Natural language parsing: Psychological, computational, and theoretical perspectives* (Studies in natural language processing, 1), pp. 129-89. Cambridge: Cambridge University Press.
- Frazier, Lyn and Fodor, Janet D. 1978. The sausage machine: A new two-stage parsing model. *Cognition* 6, pp. 291-326.
- _____. 1987. Theories of sentence processing. In: J. L. Garfield, ed., *Modularity in knowledge representation and natural-language understanding* Cambridge, MA: The MIT Press.
- Frazier, Lyn and Rayner, Keith. 1988. Parameterizing the language processing system: Left- vs. right-branching within and across languages. In: John A. Hawkins, ed., *Explaining language universals*, pp. 247-79. Oxford: Basil Blackwell.
- Fujii, N. 1989. *Right dislocation in Japanese discourse*. M.A. thesis, University of Oregon at Eugene. (Cited in Givón 1990, Herring 1994.)
- Gal, Susan. 1979. *Language shift*. New York: Academic Press. (Cited in Giacalone Ramat 1992.)
- _____. 1984. Phonological style in bilingualism: The interaction of structure and use. In: Deborah Schiffrin, ed., *Meaning, form and use in context* (GURT 84: Georgetown University round table conference in linguistics, 1984), pp. 290-302. Washington, DC: Georgetown University Press.
- Garrett, Andrew. 1990. The origin of NP split ergativity. *Language* 66.2, pp. 261-96.
- Gary, Norman. 1976. *A discourse analysis of certain root transformations in English*. Bloomington, IN: Indiana University Linguistics Club.
- Gazdar, Gerald. 1979. *Pragmatics: Implicature, presupposition, and logical form*. New York: Academic Press.
- Geeraerts, Dirk. 1993. Review of Savas L. Tsohatzidis, ed., *Meanings and prototypes. Studies in linguistic categorization*. Routledge: London and New York, 1990. *Lingua* 90.4, pp. 379-84.

- Gerritsen, Marinel. 1980. An analysis of the rise of SOV patterns in Dutch. In: Elizabeth Closs Traugott et al., eds., *Papers from the Fourth International Conference on Historical Linguistics*, pp. 123-36. Amsterdam and Philadelphia: John Benjamins.
- _____. 1984. Divergent word-order developments in Germanic languages: A description and a tentative explanation. In: Jacek Fisiak, ed., *Historical syntax* (Trends in linguistics. Studies and monographs, 23), pp. 107-36. Berlin: Mouton. (Papers presented at the Third International Conference on Historical Linguistics, devoted to historical syntax, held at Adam Mickiewicz University, at Błażejewko, Poland, 31 March - 3 April, 1981.)
- _____. 1992. Internal and external factors in the stabilization of verb-last order in Dutch infinitive clauses. In: Marinel Gerritsen and Dieter Stein, eds., *Internal and external factors in syntactic change* (Trends in linguistics. Studies and monographs, 61), pp. 355-94. Berlin: Walter de Gruyter. (A selection of papers that were presented at the workshop ... held during the Ninth International Conference on Historical Linguistics at Rutgers University in August 1989.)
- Giaccalone Ramat, Anna. 1992. The pairing of structure and function in syntactic development. In: Marinel Gerritsen and Dieter Stein, eds., *Internal and external factors in syntactic change* (Trends in linguistics. Studies and monographs, 61), pp. 317-39. Berlin: Walter de Gruyter. (A selection of papers that were presented at the workshop ... held during the Ninth International Conference on Historical Linguistics at Rutgers University in August 1989.)
- Gibson, Edward. 1992. On the adequacy of the Competition Model. *Language* 68.4, pp. 812-30. (Review article of Brian MacWhinney and Elizabeth Bates, *The crosslinguistic study of sentence processing*. Cambridge: Cambridge University Press, 1989.)
- Gilligan, Gary. 1987. *A cross-linguistic approach to the Pro-Drop parameter*. Ph. D. dissertation, University of Southern California. (Cited in Lambrecht 1994.)
- Givón, T. 1975a. Serial verbs and syntactic change: Niger-Congo. In: Charles N. Li, ed., *Word order and word order change*, pp. 47-112. Austin: University of Texas Press.
- _____. 1975b. Negation in language: Pragmatics, function, ontology. *Working papers in linguistic universals* 18. Stanford, CA: Stanford University Press.
- _____. 1976. Topic, pronoun, and grammatical agreement. In: Charles N. Li, ed., *Subject and topic*, pp. 149-88. New York: Academic Press.
- _____. 1977. The drift from VSO to SVO in Biblical Hebrew: The pragmatics of tense-aspect. In: Charles N. Li, ed., *Mechanisms of syntactic change*, pp. 181-254. Austin: University of Texas Press.

- _____. 1978. Negation in language: Pragmatics, function, ontology. In: Peter Cole, ed., *Syntax and Semantics 9: Pragmatics*, pp. 69-112. New York: Academic Press.
- _____. 1979. *On understanding grammar* (Perspectives in neurolinguistics and psycholinguistics). New York: Academic Press.
- _____. 1983a. Topic continuity in discourse: An introduction. In: T. Givón, ed., *Topic continuity in discourse: A quantitative cross-language study* (Typological studies in language, 3), pp. 1-41. Amsterdam and Philadelphia: John Benjamins.
- _____. 1983b. Topic continuity in spoken English. In: T. Givón, ed., *Topic continuity in discourse: A quantitative cross-language study* (Typological studies in language, 3), pp. 343-63. Amsterdam and Philadelphia: John Benjamins.
- _____. 1983c. Topic continuity and word-order pragmatics in Ute. In: T. Givón, ed., *Topic continuity in discourse: A quantitative cross-language study* (Typological studies in language, 3), pp. 141-212. Amsterdam and Philadelphia: John Benjamins.
- _____. 1984a. *Syntax: A functional-typological introduction, vol. 1*. Amsterdam and Philadelphia: John Benjamins.
- _____. 1984b. Direct object and dative shifting: Semantic and pragmatic case. In: Frans Plank, ed., *Objects: Towards a theory of grammatical relations*, pp. 151-82. London: Academic Press.
- _____. 1987. Beyond foreground and background. In: Russell S. Tomlin, ed., *Coherence and grounding in discourse* (Typological studies in language, 11), pp. 175-88. Amsterdam and Philadelphia: John Benjamins.
- _____. 1988. The pragmatics of word-order: Predictability, importance and attention. In: Michael Hammond, Edith Moravcsik, and Jessica Wirth, eds., *Studies in syntactic typology* (Typological studies in language, 17), pp. 243-84. Amsterdam and Philadelphia: John Benjamins.
- _____. 1989. *Mind, code and context: Essays in pragmatics*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- _____. 1990. *Syntax: A functional-typological introduction, vol. 2*. Amsterdam and Philadelphia: John Benjamins.
- _____. 1991a. Serial verbs and the mental reality of 'event': Grammatical vs. cognitive packaging. In: Elizabeth Closs Traugott and Bernd Heine, *Approaches to grammaticalization. Volume I: Focus on theoretical and methodological issues* (Typological studies in language, 19:1), pp. 81-127. Amsterdam and Philadelphia: John Benjamins.

- _____. 1991b. Markedness in grammar: Distributional, communicative and cognitive correlates of syntactic structures. *Studies in Language* 15.2, pp. 335-70.
- _____. 1991c. Isomorphism in the grammatical code: Cognitive and biological considerations. *Studies in Language* 15.1, pp. 85-114.
- _____. 1992a. On interpreting text-distributional correlations: Some methodological issues. In: Doris L. Payne, ed., *Pragmatics of word order flexibility* (Typological studies in language, 22), pp. 305-20. Amsterdam and Philadelphia: John Benjamins.
- _____. 1992b. The grammar of referential coherence as mental processing instructions. In: Norbert Dittmar, ed., *Topic--From grammar to discourse = Linguistics* 30.1, pp. 5-55.
- _____. 1994. The pragmatics of de-transitive voice: Functional and typological aspects of inversion. In: T. Givón, ed., *Voice and inversion* (Typological Studies in Language, 28; A companion series to the journal *Studies in Language*), pp. 3-44. Amsterdam and Philadelphia: John Benjamins.
- _____. 1995. *Functionalism and grammar*. Amsterdam and Philadelphia: John Benjamins.
- _____, ed. 1983. *Topic continuity in discourse: A quantitative cross-language study* (Typological studies in language, 3). Amsterdam and Philadelphia: John Benjamins.
- _____, ed.. 1994. *Voice and inversion* (Typological Studies in Language, 28). Amsterdam and Philadelphia: John Benjamins.
- Goenaga, Patxi. 1980/1978. *Gramatika bideetan*. Erein.
- Goldberg, Adele E. 1995. *Constructions: A Construction Grammar approach to argument structure* (Cognitive theory of language and culture). Chicago: The University of Chicago Press.
- Goldsmith, John. 1985. A principled exception to the Coordinate Structure Constraint. In: William H. Eilfort, Paul D. Kroeber, and Karen L. Peterson, eds., *Papers from the general session at the twenty-first regional meeting of the Chicago Linguistic Society* (CLS 21, 1985, part 1), pp. 133-43. Chicago: Chicago Linguistic Society.
- _____. 1994. Grammar within a neural network. In: Susan D. Lima, Roberta L. Corrigan, and Gregory K. Iverson, eds., *The reality of linguistic rules* (Studies in language companion series, 26), pp. 95-113. Amsterdam and Philadelphia: John Benjamins. (Selected papers from the 21st annual University of Wisconsin, Milwaukee, linguistics symposium, April 10-12, 1992.)

- Goodall, Grant. 1993. Spec of IP and Spec of CP in Spanish Wh-questions. In: William J. Ashby, Marianne Mithun, Giorgio Perissinotto, and Eduardo Raposo, eds., *Linguistic perspectives on the Romance languages: Selected papers from the 21st Linguistic Symposium on Romance Languages (LSRL XXI), Santa Barbara, California, 21-24 February 1991* (Amsterdam studies in the theory and history of linguistic science. Series IV, Current issues in linguistic theory, volume 103), pp. 199-209. Amsterdam and Philadelphia: John Benjamins.
- Green, Georgia M. 1976. Main clause phenomena in subordinate clauses. *Language* 52.2, pp. 382-97.
- _____. 1980. Some wherefores of English inversion. *Language* 56.3, pp. 582-601.
- Green, Georgia M. and Morgan, Jerry L. 1981. Writing ability as a function of the appreciation of differences between oral and written communication. In: Carl H. Frederiksen and Joseph F. Dominic, eds., *Writing: The nature, development, and teaching of written communication, vol. 2: Writing: Process, development and communication*, pp. 177-88. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Greenbaum, Sidney. 1969. *Studies in English adverbial usage*. London: Longmans.
- Greenberg, Joseph H. 1966/1963. Some universals of grammar with particular reference to the order of meaningful elements. In: Joseph H. Greenberg, ed., *Universals of language* (Second edition), pp. 73-111/58-90. Cambridge, MA: The MIT Press. (Report of a conference on Language Universals held at Dobbs Ferry, New York, April 13-15, 1961. Originally published in 1963, and reedited and retypeset in 1966.)
- _____. 1969. Some methods of dynamic comparison in linguistics. In: Jaan Puhvel, ed., *Substance and structure of language*, pp. 147-203. Berkeley and Los Angeles: University of California Press. (Lectures delivered before the Linguistic Institute of the Linguistic Society of America, University of California, Los Angeles, June 17-August 12, 1966.)
- _____. 1979. Rethinking linguistics diachronically. *Language* 55.1, pp. 275-90.
- Grimes, Joseph E. 1975. *The thread of discourse*. The Hague: Mouton.
- _____. 1982. Reference spaces in text. In: Sture Allén, ed., *Text processing: Text analysis and generation, text typology and attribution* (Data linguistica, 16), pp. 381-414. Stockholm, Sweden: Almqvist and Wiksell International. (Proceedings of Nobel Symposium 51.)
- Gruber, Jeffrey. 1965. *Studies in lexical relations*. Ph. D. dissertation, M.I.T.
- _____. 1976. *Lexical structures in syntax and semantics*. New York, Amsterdam: American Elsevier, North Holland. (Revised Ph. D. dissertation, *Studies in lexical relations*, M.I.T., 1965.)

- Gumperz, John J. and Wilson, Robert. 1971. Convergence and creolization: A case from the Indo-Aryan/Dravidian border in India. In: Dell Hymes, ed., *Pidginization and creolization of languages* (Proceedings of a conference held at the University of the West Indies, Mona, Jamaica, April 1968), pp. 151-67. Cambridge: Cambridge University Press.
- Gundel, Jeanette K. 1977/1974. *Role of topic and comment in linguistic theory*. Bloomington, IN: Indiana University Linguistics Club. (Ph. D. dissertation, University of Texas, 1974.)
- _____. 1985. 'Shared knowledge' and topicality. *Journal of Pragmatics* 9.1, pp. 83-107.
- _____. 1987. *Universals of topic-comment structure* (Another IULC Twentieth Anniversary Volume). Bloomington, IN: Indiana University Linguistics Club.
- _____. 1988. Universals of topic-comment structure. In: Michael Hammond, Edith Moravcsik, and Jessica Wirth, eds., *Studies in syntactic typology* (Typological studies in language, 17), pp. 209-39. Amsterdam and Philadelphia: John Benjamins.
- Gundel, Jeanette K.; Hedberg, Nancy; and Zacharski, Ron. 1990. Givenness, implicature, and the form of referring expressions in discourse. In: Kira Hall, Jean-Pierre Koenig, Michael Meacham, Sondra Reinman, and Laurel A. Sutton, eds., *Proceedings of the Sixteenth Annual Meeting of the Berkeley Linguistics Society* (General Session and Parasession on the Legacy of Grice), pp. 442-53. Berkeley, CA: Berkeley Linguistics Society.
- _____. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69.2, pp. 274-307.
- Gurrutxaga, Antton, *et al.*, compilers. 1994. *Euskal hiztegi modernoa*. Usurbil: Elhuyar Kultur Elkartea; Donostia: Elkar, S.L.
- Haider, Hubert. 1986. Who is afraid of typology. *Folia Linguistica* 20, pp. 109-45.
- Haiman, John. 1974a. Concessives, conditionals and verbs of volition. *Foundations of Language* 11, pp. 342-60. (Cited in Croft 1995.)
- _____. 1974b. *Targets and syntactic change*. The Hague: Mouton.
- _____. 1976. Agentless sentences. *Foundations of Language* 14, pp. 19-53. (Cited in Croft 1995.)
- _____. 1978a. Conditionals are topics. *Language* 54.3, pp. 564-89.
- _____. 1978b. A study in polysemy. *Studies in Language* 2, pp. 1-34. (Cited in Croft 1995.)

- _____. 1980. The iconicity of grammar: Isomorphism and motivation. *Language* 56.3, pp. 515-40.
- _____. 1983. Iconic and economic motivation. *Language* 59.4, pp. 781-819.
- _____. 1985a. Symmetry. In: John Haiman, ed., *Iconicity in syntax* (Typological studies in language, 6), pp. 73-95. Amsterdam and Philadelphia: John Benjamins. (Proceedings of a symposium on iconicity in syntax, Stanford, June 24-6, 1983.)
- _____. 1985b. *Natural syntax: Iconicity and erosion*. Cambridge: Cambridge University Press.
- _____. 1986. Constraints on the form and meaning of the protasis. In: Elizabeth Closs Traugott, Alice ter Meulen, Judy Snitzer Reilly, and Charles A. Ferguson, eds., *On conditionals*, pp. 215-27. Cambridge: Cambridge University Press. (From a symposium on Conditionals and Cognitive Processes, Stanford University in December 1983.)
- _____. 1991a. The bureaucratisation of language. In: H. Christof Wolfart, ed., *Linguistic studies presented to John L. Finlay* (Algonquian and Iroquoian linguistics. Memoir, 8), pp. 45-70. Winnipeg: Algonquian and Iroquoian Linguistics.
- _____. 1991b. From V/2 to subject clitics: Evidence from Northern Italian. In: Elizabeth Closs Traugott and Bernd Heine, eds., *Approaches to grammaticalization. Volume II: Focus on types of grammatical markers* (Typological studies in language, 19:2), pp. 135-58. Amsterdam and Philadelphia: John Benjamins.
- _____. 1993. Life, the universe, and human language: A brief synopsis. In: Philip W. Davis, ed., *Alternative modes in language: Reinterpreting language, Part 1 of 2 = Language Sciences* 15.4, pp. 293-322. (Papers from a Rice University Symposium on 'Alternative modes in language: Reinterpreting language' (April 1991).)
- _____. 1994. Ritualization and the development of language. In: William Pagliuca, ed., *Perspectives on grammaticalization* (Amsterdam studies in the theory and history of linguistic science, Series IV - Current issues in linguistic theory, 109), pp. 3-28. Amsterdam and Philadelphia: John Benjamins. (Selection from the papers presented at the Nineteenth Annual Linguistics Symposium of the University of Wisconsin-Milwaukee.)
- _____. 1995. Grammatical signs of the divided self: A study of language and culture. In: Werner Abraham, T. Givón, and Sandra A. Thompson, eds., *Discourse grammar and typology: Papers in honor of John W. M. Verhaar* (Studies in language companion series, 27), pp. 213-34. Amsterdam and Philadelphia: John Benjamins.

- Haiman, John, ed. 1985. *Iconicity in syntax* (Typological studies in language, 6). Amsterdam and Philadelphia: John Benjamins. (Proceedings of a symposium on iconicity in syntax, Stanford, June 24-6, 1983.)
- Hajičová, Eva. 1991. "Free" word order described without unnecessary complexity. *Theoretical Linguistics* 17.1/2/3, pp. 99-106.
- Hajičová, Eva and Sgall, Petr. 1987. The ordering principle. *Journal of Pragmatics* 11, pp. 435-54.
- Hale, Kenneth L. 1970. The passive and ergative in language change: The Australian case. In: Steven A. Wurm and Donald Laycock, eds., *Pacific Linguistic studies in honour of Arthur Capell*, pp. 757-81. Canberra: Pacific Linguistics.
- _____. 1971. Deep-surface canonical disparities and language change: An Australian case. *Trends in Linguistics* 11, (Cited in Haiman 1993.)
- _____. 1982. Preliminary remarks on configurationality. In: James Pustejovsky and Peter Sells, eds., *Proceedings of the 12th Annual Meeting of the North Eastern Linguistics Society* (NELS, 12), pp. 86-96.
- _____. 1983. Warlpiri and the grammar of non-configurational languages. *Natural Language and Linguistic Theory* 1, pp. 5-47.
- _____. 1989. On nonconfigurational structures. In: László Marácz and Pieter Muysken, eds., *Configurationality: The typology of asymmetries*, pp. 293-300. Dordrecht: Foris.
- _____. 1992. Basic word order in two "free word order" languages. In: Doris L. Payne, ed., *Pragmatics of word order flexibility* (Typological studies in language, 22), pp. 63-82. Amsterdam and Philadelphia: John Benjamins.
- Hale, Kenneth L.; LaVerne, Jeanne; and Platero, Paul. 1977. Three cases of overgeneralization. In: Peter W. Culicover, Thomas Wasow, and Adrian Akmajian, eds., *Formal syntax*, pp. 379-416. New York: Academic Press. (Cited in Campbell, Bubenik, and Saxon 1988.)
- Hall, Christopher J. 1992. *Morphology and mind: A unified approach to explanation in linguistics* (Theoretical linguistics). New York: Routledge.
- Halliday, Michael A. K. 1967a. Notes on transitivity and theme in English, Part 1. *Journal of Linguistics* 3.1, pp. 37-81.
- _____. 1967b. Notes on transitivity and theme in English, Part 2. *Journal of Linguistics* 3.2, pp. 199-244.
- _____. 1968. Notes on transitivity and theme in English, Part 3. *Journal of Linguistics* 4.2, pp. 179-215.

- _____. 1985. *Introduction to functional grammar*. London: Edward Arnold.
- _____. 1987. Spoken and written modes of meaning. In: Rosalind Horowitz and S. Jay Samuels, eds., *Comprehending oral and written language*, pp. 55-82. San Diego and London: Academic Press.
- Hannay, Mike. 1991. Pragmatic function assignment and word order variation in a functional grammar of English. In: A. Machtelt Bolkestein, ed., *Syntax and pragmatics = Journal of Pragmatics* 16, pp. 131-55.
- Harris, Alice C. 1985. *Diachronic syntax: The Kartvelian case* (Syntax and semantics, 18). New York: Academic Press. (Cited in Campbell, Bubenik, and Saxon 1988.)
- Harris, Martin B. 1984. On the causes of word order change. In: Jan G. Kooij, ed., *Special issue on word order = Lingua* 63.3-4, pp. 175-204. (Cited in Campbell, Bubenik, and Saxon 1988.)
- _____. 1985. Divergent patterns of word order in contemporary French. In: Jacek Fisiak, ed., *Papers from the 6th International Conference on Historical Linguistics* (Amsterdam studies in the theory and history of linguistic science. Series IV, Current issues in linguistic theory, 34), pp. 235-49. Amsterdam and Philadelphia: John Benjamins. (Poznan', Poland, 22-26 August 1983.)
- Harris, Randy Allen. 1993. *The linguistics wars*. New York & Oxford: Oxford University Press.
- Haspelmath, Martin. 1994. Functional categories, X-bar theory, and grammaticalization theory. *Sprachtypologie und Universalienforschung (STUF)* 47.1, pp. 3-15.
- Hawkins, John A. 1980. On implicational and distributional universals of word order. *Journal of Linguistics* 16, pp. 193-235.
- _____. 1982. Cross-category harmony, X-bar and the predictions of markedness. *Journal of Linguistics* 18.1, pp. 1-35.
- _____. 1983. *Word order universals*. New York: Academic Press.
- _____. 1986. *A comparative typology of English and German: Unifying the contrasts*. London: Croom Helm; Austin: University of Texas Press.
- _____. 1988. Explaining language universals. In: John A. Hawkins, ed., *Explaining language universals*, pp. 3-30. Oxford: Basil Blackwell.
- _____. 1990. A parsing theory of word order universals. *Linguistic Inquiry* 21.2, pp. 223-61.

- _____. 1993. Heads, parsing and word order universals. In: Greville G. Corbett, Norman M. Fraser, and Scott McGlashan, eds., *Heads in grammatical theory*, pp. 231-65. Cambridge: Cambridge University Press.
- _____. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.
- Hawkins, John A. and Cutler, Anne. 1988. Psycholinguistic factors in morphological asymmetry. In: John A. Hawkins, ed., *Explaining language universals*, pp. 280-317. Oxford: Basil Blackwell.
- Heath, Jeffrey. 1977. Remarks on Basque verbal morphology. In: William A. Douglass, Richard W. Etulain, and William H. Jacobsen, Jr., eds., *Anglo-American contributions to Basque studies: Essays in honor of Jon Bilbao* (Desert Research Institute publications in the Social Sciences, no. 13), pp. 193-201. Reno, NV: Desert Research Institute.
- Heine, Bernd; Claudi, Ulrike; and Hünemeyer, Friederike. 1991. From cognition to grammar: Evidence from African languages. In: Elizabeth Closs Traugott and Bernd Heine, *Approaches to grammaticalization. Volume I: Focus on theoretical and methodological issues* (Typological studies in language, 19:1), pp. 149-87. Amsterdam and Philadelphia: John Benjamins.
- Heine, Bernd and Reh, Mechthild. 1984. *Grammaticalization and reanalysis in African languages*. Hamburg: Helmut Buske Verlag.
- Hengeveld, Kees. 1989. Layers and operators in functional grammar. *Journal of Linguistics* 25, pp. 127-57.
- _____. 1990. The hierarchical structure of utterances. In: Jan Nuyts, A. Machtelt Bolkestein, and Co Vet, eds., *Layers and levels of representation in language theory: A functional view* (Pragmatics and Beyond New Series, 13), pp. 1-24. Amsterdam and Philadelphia: John Benjamins.
- Herring, Susan C. 1994. Afterthoughts, antitopics, and emphasis: The syntactization of postverbal position in Tamil. In: Miriam Butt, Tracy Holloway King, and Gillian Ramchand, eds., *Theoretical perspectives on word order in South Asian languages* (CSLI lecture notes, 50) Stanford, CA: Center for the Study of Language and Information.
- Hetzron, Robert. 1971. Presentative function and presentative movement. *Studies in African Linguistics, Supplement 2*, pp. 79-105.
- _____. 1975. The presentative movement, or why the ideal word order is VSOP. In: Charles N. Li, ed., *Word order and word order change*, pp. 345-88. Austin: University of Texas Press.
- Hewson, John. 1991. Determiners as heads. *Cognitive Linguistics* 2.4, pp. 317-38.

- Hock, Hans Henrich. 1991/1986. *Principles of historical linguistics* (Second revised and updated edition). Berlin: Mouton de Gruyter.
- Hockett, Charles A. 1958. *A course in modern linguistics*. New York: The Macmillan Company.
- Hoekstra, Teun and Mulder, René. 1990. Unergatives as copular verbs: Locational and existential predication. *Linguistic Review* 7, pp. 1-79. (Cited in Bresnan 1994.)
- Hooper, Joan B. and Thompson, Sandra A. 1973. On the applicability of root transformations. *Linguistic Inquiry* 4, pp. 465-97.
- Hopper, Paul J. 1979a. Aspect and foregrounding in discourse. In: T. Givón, ed., *Discourse and syntax* (Syntax and semantics, 12), pp. 213-41. New York: Academic Press.
- _____. 1979b. Some observations on the typology of focus and aspect in narrative language. *Studies in Language* 3, pp. 1-37.
- _____. 1985. Review of Dirk G. J. Panhuis, *The communicative perspective in the sentence: A study of Latin word order*. (Studies in language, companion series, 11.) Amsterdam and Philadelphia: John Benjamins, 1982. *Language* 61.2, pp. 466-70.
- _____. 1986a. Discourse function and word order shift: A typological study of the VS/SV alternation. In: Winfred P. Lehmann, ed., *Language Typology 1985*, pp. 123-40. Amsterdam and Philadelphia: John Benjamins.
- _____. 1986b. Some discourse functions of classifiers in Malay. In: Colette Craig, ed., *Noun classes and categorization* (Typological studies in language, 7), pp. 309-26. Amsterdam: John Benjamins. (Proceedings of a symposium on categorization and noun classification, Eugene, Oregon, October 1983; Cited in Chafe 1994a.)
- _____. 1987a. Emergent grammar. In: Jon Aske, Natasha Beery, Laura Michaelis, and Hana Filip, eds., *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society* (General Session and Parasession on Grammar and Cognition), pp. 139-57. Berkeley, CA: Berkeley Linguistics Society.
- _____. 1987b. Stability and change in VN/NV alternating languages: A study in pragmatic and linguistic typology. In: Jef Verschueren and Marcella Bertuccelli-Papi, eds., *The pragmatic perspective (Selected papers from the 1985 International Pragmatics Conference)* (Pragmatics and Beyond Companion Series, 5), pp. 455-75. Amsterdam and Philadelphia: John Benjamins.
- _____. 1988. Emergent Grammar and the A Priori Grammar Postulate. In: Deborah Tannen, ed., *Linguistics in context: Connecting observation and understanding: Lectures from the 1985 LSA/TESOL and NEH Institutes* (Advances in Discourse Processes, 29), pp. 117-34. Norwood, NJ: Ablex.

- _____. 1991. On some principles of grammaticization. In: Elizabeth Closs Traugott and Bernd Heine, *Approaches to grammaticalization. Volume II: Focus on types of grammatical markers* *Approaches to grammaticalization. Volume I: Focus on theoretical and methodological issues* (Typological studies in language, 19:1), pp. 17-35. Amsterdam and Philadelphia: John Benjamins.
- Hopper, Paul J. and Thompson, Sandra A. 1980. Transitivity in grammar and discourse. *Language* 56.2, pp. 251-99.
- _____. 1993. Grammatical universals, discourse pragmatics, and semantics. In: Philip W. Davis, ed., *Alternative modes in language: Reinterpreting language, Part 1 of 2 = Language Sciences* 15.4, pp. 357-76. (Papers from a Rice University Symposium on *Alternative modes in language: Reinterpreting language*, April 1991.)
- Hopper, Paul J. and Traugott, Elizabeth Closs. 1993. *Grammaticalization* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.
- Horn, Laurence R. 1989. *A natural history of negation*. Chicago: The University of Chicago Press.
- Horowitz, Rosalind and Samuels, S. Jay, eds. 1987. *Comprehending oral and written language*. San Diego and London: Academic Press.
- Horvath, Julia. 1981. *Aspects of Hungarian syntax and the theory of grammar*. Ph. D. dissertation, University of California, Los Angeles.
- Hualde, José Ignacio, Lakarra, Joseba A., and Trask, R. L., eds. 1995. *Towards a history of the Basque language* (Amsterdam studies in the theory and history of linguistic science. Series IV, Current issues in linguistic theory ; v. 131). Amsterdam and Philadelphia: John Benjamins.
- Huck, Geoffrey J. and Goldsmith, John A. 1995. *Ideology and linguistic theory: A reappraisal of Chomsky and semantics*. London and New York: Routledge.
- Huck, Geoffrey J. and Na, Younghee. 1990. Extraposition and focus. *Language* 66.1, pp. 51-77.
- Huddleston, Rodney. 1993. On exclamatory-inversion sentences in English. *Lingua* 90.3, pp. 259-69.
- Hudson, Richard. 1984. *Word grammar*. Oxford: Basil Blackwell.
- _____. 1990. *English word grammar*. Oxford: Basil Blackwell.
- Hyman, Larry M. 1971. Consecutivization in Fe? Fe? *Journal of African Studies* 10, pp. 29-43. (Cited in Givón 1991a.)

- _____. 1975. On the change from SOV to SVO: Evidence from Niger-Congo. In: Charles N. Li, ed., *Word order and word order change*, pp. 113-48. Austin: University of Texas Press.
- _____. 1984. Form and substance in language universals. In: Brian L. Butterworth, Bernard Comrie, and Östen Dahl, eds., *Explanations for language universals* (Published simultaneously as a special issue of *Linguistics*: 21.1), pp. 67-85. Berlin: Mouton.
- Irigoyen, Alfonso. 1980. Sebero Altube eta euskal hizkuntzaren pleguak. *Euskera* 25.2, pp. 325-53.
- Jacobsen, Wesley M. 1992. Are conditionals topics? The Japanese case. In: Diane Brentari, Gary N. Larson, and Lynn A. MacLeod, eds., *The joy of grammar: A festschrift in honor of James D. McCawley*, pp. 131-60. Amsterdam and Philadelphia: John Benjamins.
- Jacobsen, William H., Jr. 1974. Norminative-ergative syncretism in Basque. *Papers from the Basque Linguistics Seminar, University of Nevada, Summer, 1972* = *Anuario del seminario de filología vasca «Julio de Urquijo»* 6, pp. 67-109. (San Sebastián: Diputación de Guipúzcoa.)
- _____. 1977. The Basque locative suffix. In: William A. Douglass, Richard W. Etulain, and William H. Jacobsen, Jr., eds., *Anglo-American contributions to Basque studies: Essays in honor of Jon Bilbao* (Desert Research Institute publications in the Social Sciences, no. 13), pp. 163-68. Reno, NV: Desert Research Institute.
- _____. 1985. The analog of the passive transformation in ergative-type languages. In: Johanna Nichols and Anthony C. Woodbury, eds., *Grammar inside and outside the clause: Some approaches to theory from the field*, pp. 176-91. Cambridge: Cambridge University Press.
- Jacobson, Sven. 1975. *Factors influencing the placement of English adverbs in relation to auxiliaries. A study in variation* (Acta Universitatis Stockholmiensis, Stockholm Studies in English, 33). Stockholm: Almqvist and Wiksell International.
- Jahr, Ernst Håkon, ed. 1992. *Language contact: Theoretical and empirical studies* (Trends in linguistics, studies and monographs, volume 60). Berlin: Mouton de Gruyter. (Based on the Fifth International Tromsø Symposium on Language held at the University of Tromsø, Norway, in Sept. 1989, the theme of which was "Language Contact".)
- Jansen, Frank. 1978. *Studies on fronting*. Lisse: Peter De Ridder. (Cited in Gerritsen 1984.)

- _____. 1981. *Syntaktische konstrukties in gesproken taal*. Amsterdam: Huis aan de drie grachten. (Cited in Gerritsen 1984.)
- Jelinek, Eloise. 1984. Empty categories, case, and configurationality. *Natural Language and Linguistic theory* 2, pp. 39-76.
- Jespersen, Otto. 1922. *Language, its nature, development and origin*. London: Allen and Unwin.
- _____. 1924. *The philosophy of grammar*. London: Macmillan. (Reprinted in 1955 by Allen and Unwin, London.)
- _____. 1949. *A modern English grammar on historical principles, vol. 7*. London: Allen and Unwin. (Reprinted in 1974.)
- _____. 1964/1933. *Essentials of English grammar*. University of Alabama Press.
- Junghare, Indira Y. 1985. The functions of word order variants in Indo-Aryan. In: E. Bashir, M Deshpande, and P. Hook, eds., *Select papers from SALA-7: South Asian Languages Analysis Roundtable Conference*, pp. 236-53. Bloomington, IN: Indiana University Linguistics Club. (Cited in Herring 1994.)
- König, Ekkehard. 1991. *The meaning of focus particles: A comparative perspective* (Theoretical linguistics). London and New York: Routledge.
- _____. 1993. Review of Terttu Nevalainen, BUT, ONLY, JUST: Focusing adverbial change in Modern English 1500-1700. *Mémoires de la Société Néophilologique de Helsinki*, 51. Helsinki, 1991. *Lingua* 91.4, pp. 361-64.
- Kachru, Braj B. 1979. The Englishization of Hindi: Language rivalry and language change. In: Irmengard Rauch and Gerald F. Carr, eds., *Linguistic method: Essays in honor of Herbert Penzl* (Janua linguarum. Series maior, 79), pp. 199-212. The Hague: Mouton. (Cited in Cowan 1995.)
- Kay, Paul and Fillmore, Charles J. 1994. Grammatical constructions and linguistic generalizations: The What's X doing Y? construction. Ms., Department of Linguistics, University of California, Berkeley.
- Keenan, Edward L. 1976a. Towards a universal definition of "subject". In: Charles N. Li, ed., *Subject and topic*, pp. 303-34. New York: Academic Press.
- _____. 1976b. Remarkable subjects in Malagasy. In: Charles N. Li, ed., *Subject and topic*, pp. 247-302. New York: Academic Press.
- _____. 1978. The syntax of subject-final languages. In: Winfred P. Lehmann, ed., *Syntactic typology: Studies in the phenomenology of language*, pp. 267-328. Austin: University of Texas Press.

- _____. 1984. Semantic correlates of the ergative/absolutive distinction. *Linguistics* 22, pp. 197-223.
- Keenan, Edward L. and Comrie, Bernard. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8, pp. 63-99.
- Keenan, Elinor Ochs. 1977. Why look at planned and unplanned discourse? In: Elinor Ochs Keenan and Tina L. Bennett, eds., *Discourse across time and space* (Southern California Occasional Papers in Linguistics, 5), pp. 1-41. Los Angeles: Department of Linguistics, University of Southern California.
- _____. 1979. Planned and unplanned discourse. In: T. Givón, ed., *Discourse and syntax* (Syntax and semantics, 12), pp. 51-81. New York: Academic Press.
- Kim, Alan Hyun-Oak. 1988. Preverbal focusing and type XXIII languages. In: Michael Hammond, Edith Moravcsik, and Jessica Wirth, eds., *Studies in syntactic typology* (Typological studies in language, 17), pp. 147-69. Amsterdam and Philadelphia: John Benjamins.
- _____. 1995. Word order at the noun phrase level in Japanese: Quantifier constructions and discourse functions. In: Pamela Downing and Michael Noonan, eds., *Word order in discourse* (Typological studies in language, 30), pp. 199-246. Amsterdam and Philadelphia: John Benjamins. (Selection of papers presented at a symposium entitled *Word Order in Discourse* held at the University of Wisconsin-Milwaukee, April 12-14, 1991 (UWM Linguistics Symposium, 20), chaired by Edith Moravcsik and Michael Noonan.)
- Kim, Alan Hyun-Oak and Shin, Hyon-Sook. 1992. Postposing in Korean. Paper presented at the XVth International Congress of Linguists, Québec, August 9-14, 1992. (Cited in Herring 1994.)
- Kiss, Katalin É. 1987. *Configurationality in Hungarian*. Dordrecht: Reidel; Budapest: Akademiai Kiado.
- Kryk-Kastovsky, Barbara. 1993. Review of Ekkehard König, *The meaning of focus particles: A comparative perspective*. London and New York: Routledge, 1991. *Journal of Pragmatics* 20.6, pp. 608-14.
- Kuno, Susumu. 1971. The position of locatives in existential sentences. *Linguistic Inquiry* 2, pp. 333-78.
- _____. 1972. Functional sentence perspective: A case study from Japanese and English. *Linguistic Inquiry* 3, pp. 269-320.
- _____. 1973. Constraints on internal clauses and sentential subjects. *Linguistic Inquiry* 4.3, pp. 363-85.

- _____. 1974. The position of relative clauses and conjunctions. *Linguistic Inquiry* 5.1, pp. 117-36.
- _____. 1975. Three perspectives in the functional approach to syntax. In: Robin E. Grossman, L. James San, and Timothy J. Vance, eds., *Papers from the parasession on functionalism: April 17, 1975* (Eleventh regional meeting of the Chicago Linguistic Society), pp. 276-336. Chicago: Chicago Linguistic Society.
- _____. 1976. Subject, theme, and the speaker's empathy: A reexamination of relativization phenomena. In: Charles N. Li, ed., *Subject and topic*, pp. 417-44. New York: Academic Press.
- _____. 1978. Japanese: A characteristic OV language. In: Winfred P. Lehmann, ed., *Syntactic typology: Studies in the phenomenology of language*, pp. 57-138. Austin: University of Texas Press.
- _____. 1987. *Functional syntax: Anaphora, discourse, and empathy*. Chicago: University of Chicago Press.
- Kuroda, S.-Y. 1972. The categorial and thethetic judgement: Evidence from Japanese syntax. *Foundations of Language* 9, pp. 153-85. (Originally written for the French journal *Langages*, where it appeared at a later date.)
- Kuter, Lois. 1989. Breton vs. French: Language and the opposition of political, economic, social, and cultural values. In: Nancy D. Dorian, ed., *Investigating obsolescence: Studies in language contraction and death* (Studies in the Social and Cultural Foundations of Language, 7), pp. 75-89. Cambridge: Cambridge University Press.
- Labov, William. 1972. Some principles of linguistic methodology. *Language in Society* 1, pp. 97-120.
- _____. 1989. The child as linguistic historian. *Language Variation and Change* 1, pp. 85-97.
- Ladefoged, Peter. 1972. Phonetic prerequisites for a distinctive feature theory. In: Albert Valdman, ed., *Papers in linguistics and phonetics in memory of Pierre Delattre* The Hague: Mouton. (Cited in Bolinger 1979.)
- Lafitte, Pierre. 1962/1944. *Grammaire Basque (Navarro-Labourdin Littéraire)* (Second edition). Baiona (Bayonne): Amis du Musée Basque & Ikas. (First edition: Bourdeaux: Edit. Delmas, 1944; Facsimile reproduction of second edition: Donostia: Elkar, 1979.)
- Lafon, René. 1943. *Le système du verbe basque au XVIe. siècle (2 vol)* (Publications de l'Université de Bordeaux). Bordeaux: Edit. Delmas. (Facsimile reproduction: Donostia: Elkar, 1980.)

- Laka, Itziar. 1993. The structure of inflection: A case study in X⁰ syntax. In: José Ignacio Hualde and Jon Ortiz de Urbina, eds., *Generative studies in Basque linguistics* (Amsterdam studies in the theory and history of linguistic science, Series IV: Current issues in linguistic theory, 105), pp. 21-70. Amsterdam and Philadelphia: John Benjamins.
- _____. 1994. *On the syntax of negation* (Outstanding Dissertations in Linguistics). New York & London: Garland Publishing. (A slightly updated version of her Ph.D. dissertation, M.I.T., 1990.)
- Lakoff, George. 1984. Performative subordinate clauses. In: Claudia Brugman and Monica Macaulay, eds., *Proceedings of the Tenth Annual Meeting of the Berkeley Linguistics Society* (General Session and Parasession on Subordination), pp. 472-80. Berkeley: Berkeley Linguistics Society.
- _____. 1986. Frame semantic control of the coordinate structure constraint. In: Ann M. Farley, Peter T. Farley, and Karl-Erik McCullough, eds., *Papers from the parasession on pragmatics and grammatical theory at the Twenty Second Regional Meeting of the Chicago Linguistic Society* (CLS 22/2), pp. 152-67. Chicago: Chicago Linguistic Society.
- _____. 1987. *Women, fire and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
- _____. 1991. Cognitive versus generative linguistics: How commitments influence results. *Language and Communication* 11.1/2, pp. 53-62. (Commentary to the paper by Frederick J. Newmeyer.)
- Lakoff, Robin. 1972. Another look at drift. In: Robert P. Stockwell and Ronald K. S. Macaulay, eds., *Linguistic change and generative theory*, pp. 172-98. Bloomington: Indiana University Press.
- Lambrecht, Knud. 1981. *Topic, antitopic and verb-agreement in non-standard French* (Pragmatics and beyond, vol II:6). Amsterdam: John Benjamins.
- _____. 1986. *Topic, focus, and the grammar of spoken French*. Ph. D. dissertation, University of California, Berkeley.
- _____. 1987. Aboutness as a cognitive category: The thetic-categorical distinction revisited. In: Jon Aske, Natasha Beery, Laura Michaelis, and Hana Filip, eds., *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society* (General Session and Parasession on Grammar and Cognition), pp. 366-82. Berkeley, CA: Berkeley Linguistics Society.
- _____. 1988a. Presentational cleft constructions in spoken French. In: John Haiman and Sandra A. Thompson, eds., *Clause combining in grammar and discourse* (Typological studies in language, 18), pp. 135-80. Amsterdam and Philadelphia: John Benjamins.

- _____. 1988b. There was a farmer had a dog: Syntactic amalgams revisited. In: Shelley Axmaker, Annie Jaissner, and Helen Singmaster, eds., *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society* (General Session and Parasession on Grammaticalization), pp. 319-229. Berkeley, CA: Berkeley Linguistics Society.
- _____. 1990. "What, me worry?" - 'Mad Magazine sentences' revisited. In: Kira Hall, Jean-Pierre Koenig, Michael Meacham, Sondra Reinman, and Laurel A. Sutton, eds., *Proceedings of the Sixteenth Annual Meeting of the Berkeley Linguistics Society* (General Session and Parasession on the Legacy of Grice), pp. 215-28. Berkeley, CA: Berkeley Linguistics Society.
- _____. 1994. *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents* (Cambridge studies in linguistics, 71). Cambridge: Cambridge University Press.
- Langacker, Ronald W. 1977a. Syntactic reanalysis. In: Charles N. Li, ed., *Mechanisms of syntactic change*, pp. 57-139. Austin: University of Texas Press.
- _____. 1977b. An overview of Uto-Aztecan grammar. *Studies in Uto-Aztecan grammar, vol. 1* Arlington, TX: Summer Institute of Linguistics, University of Texas at Arlington Press. (Cited in Campbell, Bubenik, and Saxon 1988.)
- _____. 1987. *Foundations of cognitive grammar, Vol. 1*. Stanford: Stanford University Press.
- _____. 1988. An overview of cognitive grammar. In: Brygida Rudzka-Ostyn, ed., *Topics in cognitive linguistics* (Current issues in linguistic theory, 50), pp. 3-48. Amsterdam and Philadelphia: John Benjamins.
- _____. 1991. *Foundations of cognitive grammar, Vol. 2: Descriptive application*. Stanford: Stanford University Press.
- Langendoen, D. Terence. 1973. The problem of grammatical relations in surface structure. In: Kurt R. Jankowsky, ed., *Language and international studies* (Georgetown University Round Table on Languages and Linguistics 1973), pp. 27-37. Washington, DC: Georgetown University Press.
- Lardizabal, F. Ignacio de, transl. 1855. *Testamentu zarreco eta berrico condaira*. Tolosa (Basque Country).
- Lass, Roger. 1980. *On explaining language change*. Cambridge: Cambridge University Press.
- _____. 1984. *Phonology: An introduction to basic concepts*. Cambridge: Cambridge University Press.

- Lavandera, Beatriz. 1978. Where does the sociolinguistic variable stop? *Language in Society* 7, pp. 171-83.
- Lehiste, Ilse. 1988. *Lectures on language contact*. Cambridge, MA: The MIT Press.
- Lehmann, Christian. 1982. *Thoughts on grammaticalization: A programmatic sketch* (Arbeiten des Kölner Universalien-Projekts, AKUP, 48). Cologne: Institut für Sprachwissenschaft der Universität zu Köln.
- _____. 1985a. Grammaticalization: Synchronic variation and diachronic change. *Lingua e Stile* 20.3, pp. 303-18.
- _____. 1985b. The role of grammaticalization in linguistic typology. In: Hansjakob Seiler and Gunter Brettschneider, eds., *Language invariants and mental operations* Tübingen: Gunter Narr Verlag. (International interdisciplinary conference held at Gummersbach/Cologne, Germany, Sep. 18-23, 1983.)
- _____. 1990. Towards lexical typology. In: William Croft, Keith Denning, and Suzanne Kemmer, eds., *Studies in typology and diachrony: Papers presented to Joseph H. Greenberg on his 75th birthday* (Typological studies in language, 20), pp. 161-86. Amsterdam and Philadelphia: John Benjamins.
- _____. 1992. Word order change by grammaticalization. In: Marinel Gerritsen and Dieter Stein, eds., *Internal and external factors in syntactic change* (Trends in linguistics. Studies and monographs, 61), pp. 395-416. Berlin: Walter de Gruyter. (A selection of papers that were presented at the workshop ... held during the Ninth International Conference on Historical Linguistics at Rutgers University in August 1989.)
- Lehmann, Winfred P. 1973. A structural principle of language and its implications. *Language* 49, pp. 47-66.
- _____. 1976. Surface and underlying structure in typological study. In: Alphonse Juillard, ed., *Linguistic studies offered to Joseph Greenberg on the occasion of his sixtieth birthday, vol. 3: Syntax* (Studia linguistica et philologica, 4, 3 vols.), pp. 517-26. Saratoga, CA: Anna Libri.
- _____. 1978. The great underlying ground-plans. In: Winfred P. Lehmann, ed., *Syntactic typology: Studies in the phenomenology of language*, pp. 3-56. Austin: University of Texas Press.
- Lertxundi, Anjel. 1984. *Goiko kale*. Donostia: Erein. (Cited in Agirre Berezibar 1991.)
- _____. 1994. *Otto Pette: hilean bizian bezala* (Narrazioa, 3). Alberdania.
- Levin, Beth C. 1983. Unaccusative verbs in Basque. *ALNE/NELS* 13, pp. 129-44.

- _____. 1989. The Basque verbal inventory and configurationality. In: László Marácz and Pieter Muysken, eds., *Configurationality: The typology of asymmetries*, pp. 39-62. Dordrecht: Foris.
- _____. 1991. Locative inversion: A problem for the unaccusative hypothesis? Evanston: IL: Northwestern University, Ms. (cited in Bresnan 1994).
- Levin, Beth C. and Rappaport Hovav, Malka. 1995. *Unaccusativity: At the syntax-lexical semantics interface* (Linguistic Inquiry monograph, 26). Cambridge, MA: The MIT Press.
- Levinson, Stephen C. 1987. Pragmatics and the grammar of anaphora. *Journal of Linguistics* 23, pp. 379-434.
- _____. 1991. Pragmatic reduction of the Binding Conditions revisited. *Journal of Linguistics* 27, pp. 107-61.
- Lewis, David. 1979. Scorekeeping in a language game. In: Rainer Bäuerle, Urs Egli, and Arnim von Stechow, eds., *Semantics from different points of view*, pp. 172-87. Berlin: Springer-Verlag.
- Li, Charles N. and Thompson, Sandra A. 1973. Serial verb constructions in Mandarin Chinese. In: Claudia Corum, T. Cedric Smith-Stark, and Ann Weiser, eds., *You take the high node and I'll take the low node. Papers from the Comparative Syntax Festival: The differences between main and subordinate clauses, 12 April 1973* (Paravolume to Papers from the ninth regional meeting of the Chicago Linguistic Society), pp. 96-103. Chicago: Chicago Linguistic Society.
- _____. 1976a. Subject and topic: A new typology of language. In: Charles N. Li, ed., *Subject and topic*, pp. 457-89. New York: Academic Press.
- _____. 1976b. On the issue of word order in a synchronic grammar: A case against 'Movement Transformations'. *Lingua* 39, pp. 169-81.
- _____. 1981. *Mandarin Chinese: A functional reference grammar*. Berkeley and Los Angeles: University of California Press.
- Lima, Susan D., Corrigan, Roberta L., and Iverson, Gregory K., eds. 1994. *The reality of linguistic rules* (Studies in language companion series, 26). Amsterdam and Philadelphia: John Benjamins. (Selected papers from the 21st annual University of Wisconsin, Milwaukee, linguistics symposium, April 10-12, 1992.)
- Lindblom, Björn; MacNeilage, Peter; and Studdert-Kennedy, Michael. 1984. Self-organizing processes and the explanation of phonological universals. In: Brian L. Butterworth, Bernard Comrie, and Östen Dahl, eds., *Explanations for language universals* (Published simultaneously as a special issue of *Linguistics*: 21.1), pp. 181-203. Berlin: Mouton.

- Lizardi, Xabier. 1972/1934. *Itz-lauz*. Donostia: Edili. (Cited in in Agirre Berezibar 1991.)
- Loepfe, A. 1940. *Die Wortstellung im Griechischen Sprechsatz*. Dissertation: Freiburg in der Schweiz. (Cited in Sgall 1987.)
- Longacre, Robert E. 1982. Discourse typology in relation to language typology. In: Sture Allén, ed., *Text processing: Text analysis and generation, text typology and attribution* (Data linguistica, 16), pp. 457-86. Stockholm, Sweden: Almqvist and Wiksell International. (Proceedings of Nobel Symposium 51.)
- _____. 1983. *The grammar of discourse* (Topics in language and linguistics). New York: Plenum Press. (Based on an earlier work, *An anatomy of speech notions*, 1976.)
- _____. 1985. Sentences as combinations of clauses. In: Tim Shopen, ed., *Language typology and syntactic description, vol. 2* Cambridge: Cambridge University Press.
- Lord, Carol. 1973. Serial verbs in transition. *Studies in African Linguistics* 4, pp. 269-96. (Cited in Givón 1991a.)
- Lyons, John. 1968. *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press.
- MacWhinney, Brian. 1987. The competition model. In: Brian MacWhinney, ed., *Mechanisms of language acquisition*, pp. 249-308. Hillsdale, NJ: Lawrence Erlbaum Associates.
- _____. 1989a. Competition and connectionism. In: Brian MacWhinney and Elizabeth Bates, eds., *The crosslinguistic study of sentence processing*, pp. 422-57. Cambridge: Cambridge University Press.
- _____. 1989b. Competition and lexical categorization. In: Roberta Corrigan, Fred Eckman, and Michael Noonan, eds., *Linguistic categorization*, pp. 195-242. Amsterdam and Philadelphia: John Benjamins.
- Malkiel, Yakov. 1981. Drift, slope and slant: Background of, and variations upon, a Sapirian theme. *Language* 57, pp. 535-70.
- Mallinson, Graham and Blake, Barry J. 1981. *Language typology: Cross-linguistic studies in syntax*. Amsterdam: North Holland.
- Mann, William C. and Thompson, Sandra A. 1986a. Relational propositions in discourse. *Discourse Processes* 9.1, pp. 57-90.

- _____. 1986b. Rhetorical Structure Theory: Description and constructino of text structures. In: Gerard Kempen, ed., *Natural language generation: New results in artificial intelligence, psychology and linguistics* (Proceedings of the Third International Workshop on Text generation, Nijmegen, The Netherlands, August 1986.)
- _____. 1987. Rhetorical Structure Theory: A framework for the analysis of texts. *IPRA Papers in Pragmatics* 1, pp. 79-105.
- _____. 1988. Rhetorical Structure Theory: Toward a functional theory of text organization. *Text* 8.3, pp. 243-81.
- _____. 1989. Rhetorical Structure Theory: A theory of text organization. In: Livia Polanyi, ed., *The structure of discourse* Norwood, NJ: Ablex.
- Mann, William C. and Thompson, Sandra A., eds. 1992. *Discourse description: Diverse linguistic analyses of a fund-raising text* (Pragmatics and Beyond, new series, 16). Amsterdam and Philadelphia: John Benjamins.
- Marácz, László and Muysken, Pieter, eds. 1989. *Configurationality: The typology of asymmetries*. Dordrecht: Foris.
- Marty, Anton. 1897. Über die Scheidung von Grammatischem, Logischem und Psychologischem Subject resp. Prädicat. *Archiv für Systematische Philosophie* 3.2, pp. 294-333. (Cited in Sgall 1987.)
- _____. 1897. Über die Scheidung von grammatischem, logischem und psychologischem Subjekt resp. Prädikat. *Archiv für systematische Philosophie* 3, pp. 174-90, 294-333.
- Mathesius, Vilém. 1928/1964. On linguistic characterology with illustrations from modern English. *Actes du Premier Congrès International de Linguistes à La Haye* (Reprinted: J. Vachek, ed., 1964, *A Prague School reader in linguistics*. Bloomington: Indiana University Press, pp. 59-67.)
- _____. 1929. Zur Satzperspektive im modernen Englisch. *Archiv für das Studium der neueren Sprachen und Literaturen* 155, pp. 202-10.
- _____. 1983/1929. Functional linguistics. In: Josef Vachek, ed., *Praguiana*, pp. 121-42. Amsterdam and Philadelphia: John Benjamins.
- Matisoff, James A. 1991. Areal and universal dimensions of grammatization in Lahu. In: Elizabeth Closs Traugott and Bernd Heine, *Approaches to grammaticalization. Volume II: Focus on types of grammatical markers* (Typological studies in language, 19:2), pp. 383-454. Amsterdam and Philadelphia: John Benjamins.

- Matthews, Peter H. 1993. *Grammatical theory in the United States: From Bloomfield to Chomsky* (Cambridge Studies in Linguistics 67). Cambridge: Cambridge University Press.
- McCawley, Noriko. 1973. Boy! Is syntax easy! *Proceedings of the ninth annual meeting of the Chicago Linguistic Society* (CLS 9), pp. 369-77. Chicago: Chicago Linguistic Society.
- _____. 1977. What is the 'emphatic root transformation' phenomenon? *Proceedings of the thirteenth regional meeting of the Chicago Linguistic Society* (CLS 13), pp. 384-400. Chicago: Chicago Linguistic Society.
- McCloskey, James. 1983. A VP in a VSO language? In: Gerald Gazdar et al., eds., *Order, concord and constituency*, pp. 9-55. Dordrecht, Holland: Foris.
- McKnight, George. 1897. The primitive Teutonic order of words. *Journal of English and Germanic Philology* 1, pp. 136-219.
- Meillet, Antoine. 1912. L'évolution des formes grammaticales. *Scientia (Rivista di scienza)* 12 ,
- _____. 1918. Convergence des développements linguistiques. *Revue Philosophique de la France et de l'Étranger* 85, pp. 97-110. (Reprinted in Meillet 1921, *Linguistique historique et linguistique générale*, Paris: Klincksieck; Meillet 1948, Paris: Champion.)
- Michelena, Luis (=Mitxelena, Koldo). 1953. Arnaut Oihenart. *Boletín de la Real Sociedad Vasca de Amigos del País (BRSVAP)* 9, pp. 460-462.
- _____. 1957. Las antiguas consonantes vascas. In: D. Catalán, ed., *Miscelánea: Homenaje a André Martinet* (vol. I), pp. 113-57. La Laguna, Spain.
- _____. 1978. Miscelánea filológica vasca I. *Fontes Linguae Vasconum* 10.2 (Issue No. 29), pp. 205-28.
- _____. 1987. *Palabras y textos*. Euskal Herriko Unibertsitatea (University of the Basque Country). (Collection of articles on Basque language and typology; contains *Miscelánea Filologica Vasca I-IV*, as well as an all-new 26 page addendum titled *Rectificaciones y Enmiendas*.)
- Miller, George A. and Chomsky, Noam. 1963. Finitary models of language users. In: R. Duncan Luce, Robert R. Bush, and Eugene Galanter, eds., *Handbook of Mathematical Psychology, vol. II* (Chapter 13), pp. 419-191. New York: Wiley. (vols. 1-3: 1963-1965.)
- Milroy, James and Milroy, Leslie. 1985. Linguistic change, social network, and speaker motivation. *Journal of Linguistics* 21, pp. 339-84.

- Mithun, Marianne. 1985. When speakers write. In: Mary Niepokuj, Mary VanClay, Vassiliki Nikiforidou, and Deborah Feder, eds., *Proceedings of the Eleventh Annual Meeting of the Berkeley Linguistics Society (General Session and Parasession on Poetics, Metrics, and Prosody)*, pp. 259-72. Berkeley, CA: Berkeley Linguistics Society.
- _____. 1986. Disagreement: The case of pronominal affixes and nouns. In: Deborah Tannen and James E. Alatis, eds., *Languages and linguistics: The interdependence of theory, data, and application* Washington, DC: Georgetown University Press.
- _____. 1987. Is basic word order universal? In: Russell S. Tomlin, ed., *Coherence and grounding in discourse* (Typological studies in language, 11), pp. 281-328. Amsterdam and Philadelphia: John Benjamins. (Reprinted in Payne, Doris L., ed., *Pragmatics of word order flexibility*, Typological studies in language, 22, Amsterdam and Philadelphia: John Benjamins, 1992.)
- _____. 1989. The acquisition of polysynthesis. *Journal of Child Language* 16, pp. 285-312.
- _____. 1990. Language obsolescence and grammatical description. *International Journal of American Linguistics* 56.1, pp. 1-26.
- _____. 1995. Morphological and prosodic forces shaping word order. In: Pamela Downing and Michael Noonan, eds., *Word order in discourse* (Typological studies in language, 30), pp. 387-423. Amsterdam and Philadelphia: John Benjamins. (Selection of papers presented at a symposium entitled *Word Order in Discourse* held at the University of Wisconsin-Milwaukee, April 12-14, 1991 (UWM Linguistics Symposium, 20), chaired by Edith Moravcsik and Michael Noonan.)
- Mitxelena, Koldo. 1977 [1975]. Euskal hizkera eta euskal neurkera. *Euskera* 22.2, pp. 721-33.
- _____. 1981. Galdegaia eta mintzagaia euskaraz. *Euskal linguistika eta literatura: Bide berriak* (Euskal Herria Saila, 4), pp. 57-81. Bilbo: Deustuko Unibertsitateko Argitarazioak.
- Moore, Timothy E. 1972. Speeded recognition of ungrammaticality. *Journal of Verbal Learning and Verbal Behavior* 11, pp. 550-560. (Cited in Hall 1992.)
- Moravcsik, Edith A. 1978. Universals of language contact. In: Joseph H. Greenberg, Charles A. Ferguson, and Edith A. Moravcsik, eds., *Universals of human language, vol. 1*, pp. 93-122. Stanford, CA: Stanford University Press.
- _____. 1988. Review of 'Topic, focus and configurationality' by Werner Abraham and S. de Meih, eds. *Studies in Language* 12.2, pp. 471-83.

- Myhill, John. 1992. *Typological discourse analysis: Quantitative approaches to the study of linguistic function*. Oxford UK and Cambridge USA: Basil Blackwell.
- Myhill, John and Xing, Zhiqun. 1993. Towards an objective definition of contrast. Ms., University of Michigan.
- Nelde, Peter. 1992. Multilingualism and contact linguistics. In: Martin Pütz, ed., *Thirty years of linguistic evolution: Studies in honour of René Dirven on the occasion of his sixtieth birthday*, pp. 379-97. Amsterdam and Philadelphia: John Benjamins.
- Nevalainen, Terttu. 1987. Adverbial focusing and intonation. *Lingua* 73, pp. 141-65.
- _____. 1991. *But, only, just: Focusing adverbial change in Modern English, 1500-1900* (Mémoires de la Société Néophilologique de Helsinki). Helsinki: Société Néophilologique.
- Newmeyer, Frederick J. 1987. Presentation there-insertion and the notions 'root transformation' and 'stylistic rule'. In: Barbara Need, Eric Schiller, and Anna Bosch, eds., *Papers from the 23rd annual regional meeting of the Chicago Linguistic Society, Part one: The general session (CLS 23/1)* Chicago: Chicago Linguistic Society.
- _____. 1992. Iconicity and generative grammar. *Language* 68.4, pp. 756-96.
- _____. 1994. Competing motivations and synchronic analysis. *Sprachtypologie und Universalienforschung* 47.2, pp. 67-77.
- Nichols, Johanna. 1984. Functional theories of grammar. *Annual Review of Anthropology* 13, pp. 97-117.
- _____. 1986. Head marking and dependent-marking grammar. *Language* 62.1, pp. 56-119.
- Ochs, Elinor. 1979. Planned and unplanned discourse. In: T. Givón, ed., *Discourse and syntax* (Syntax and semantics, 12), pp. 51-80. New York: Academic Press.
- _____. 1988. *Culture and language development: Language acquisition and language socialization in a Samoan village* (Studies in the social and cultural foundations of language, 6). Cambridge: Cambridge University Press.
- Odriozola, Juan Carlos and Zabala, Igone. 1992. *Hitz-ordena, galdegaia eta komaren erabilera* (Idazkera teknikoa, 1). Bilbo: Euskal Herriko Unibertsitatea.
- Ogle, Richard. 1981. Redefining the scope of root transformations. *Linguistics* 19, pp. 119-46.

- Ohala, John J. 1987. Explanation, evidence, and experiment in phonology. In: Wolfgang U. Dressler, Hans C. Luschützky, Oskar E. Pfeiffer, and John R. Rennison, eds., *Phonologica 1984*, pp. 215-25. Cambridge: Cambridge University Press. (Cited in Ohala 1989.)
- _____. 1989. Sound change is drawn from a pool of synchronic variation. In: Leiv Egil Breivik and Ernst Håkon Jahr, eds., *Language change: Contributions to the studies of its causes* (Trends in linguistics. Studies and monographs, 43), pp. 173-98. Berlin: Mouton de Gruyter. (Mostly papers presented at the symposium "The causes of language change: Do we know them yet?" held at the School of Languages and Literature, University of Tromsø, October 15-17, 1987.)
- Oihartzabal, Beñat (= Oyharçabal, Beñat/Bernard). 1984. BA-baiezkoa aurrekia. *Euskera* 29.2, pp. 351-71.
- Olson, David R. 1994. *The world on paper: The conceptual and cognitive implications of writing and reading*. Cambridge: Cambridge University Press.
- Ortiz de Urbina, Jon. 1989. *Parameters in the grammar of Basque* (Studies in generative grammar, 33). Dordrecht: Foris.
- _____. 1991. Deskribapenatik azalpenetara euskal gramatikan. In: Jon Ortiz de Urbina et al., eds., *Euskal syntaxiaz*, pp. 11-27. Bilbo: Labayru Ikastegia.
- _____. 1993. Feature percolation and clausal pied-piping. In: José Ignacio Hualde and Jon Ortiz de Urbina, eds., *Generative studies in Basque linguistics* (Amsterdam studies in the theory and history of linguistic science, Series IV: Current issues in linguistic theory, 105), pp. 189-220. Amsterdam and Philadelphia: John Benjamins.
- Osa Unamuno, Eusebio. 1990. *Euskararen hitzordena: Komunikazio zereginaren arauera*. Gasteiz: Euskal Herriko Unibertsitatea.
- Osgood, Charles E. and Tanz, Christine. 1976. Will the real direct object in bitransitive sentences please stand up? In: Alphonse Juilland, ed., *Linguistic studies offered to Joseph Greenberg on the occasion of his sixtieth birthday, vol. 3: Syntax* (Studia linguistica et philologica, 4, 3 vols.), pp. 537-90. Saratoga, CA: Anna Libri.
- Oyharçabal, Beñat (= Oyharçabal/Oihartzabal, Bernard). 1985a. Behako bat ezezko esaldieri. *Euskera* 30.2, pp. 103-15.
- _____. 1986. Erlatibakuntza: Perpaus erlatiboaren egitura; izen erlatibatuaren betekizuna haren barnean. In: Georges Rebuschi, Pello Salaburu, Itziar Laka, Beñat Oyharçabal, and Patxi Goenaga, eds., *Euskal sintaxiaren zenbait arazo* (Donostiako Udako IV Ikastaroak, 1985), pp. 129-53. Bilbo: Euskal Herriko Unibertsitatea.

- Oyharçabal, Bernard. 1985b. *Les relatives en Basque*. Département des Recherches Linguistiques DRL, Université de Paris.
- _____. 1987. *Etude descriptive de constructions complexes en Basque: Propositions relatives, temporelles, conditionnelles et concessives*. Ph. D. dissertation, Université de Paris 7.
- _____. 1989. Pro-drop and the resumptive pronoun strategy in Basque. In: László Marác and Pieter Muysken, eds., *Configurationality: The typology of asymmetries*, pp. 63-84. Dordrecht: Foris.
- _____. 1993. Verb agreement with not arguments: On allocutive agreement. In: José Ignacio Hualde and Jon Ortiz de Urbina, eds., *Generative studies in Basque linguistics* (Amsterdam studies in the theory and history of linguistic science, Series IV: Current issues in linguistic theory, 105), pp. 89-114. Amsterdam and Philadelphia: John Benjamins.
- Pagliuca, William, ed. 1994. *Perspectives on grammaticalization* (Amsterdam studies in the theory and history of linguistic science, Series IV - Current issues in linguistic theory, 109). Amsterdam and Philadelphia: John Benjamins. (Selection from the papers presented at the Nineteenth Annual Linguistics Symposium of the University of Wisconsin-Milwaukee.)
- Panhuis, Dirk G. J. 1982. *The communicative perspective in the sentence: A study of Latin word order* (Studies in language, Companion series, 11). Amsterdam and Philadelphia: John Benjamins.
- Payne, Doris L. 1985. Review of *Word Order Universals*, by John A. Hawkins, 1983. *Language* 61, pp. 462-66. (Cited in Campbell, Bubenik, and Saxon 1988.)
- _____. 1990. *The pragmatics of word order: Typological dimensions of verb initial languages* (Empirical approaches to language typology, 7). Berlin: Mouton de Gruyter.
- _____. 1992a. Introduction. In: Doris L. Payne, ed., *Pragmatics of word order flexibility* (Typological studies in language, 22), pp. 1-13. Amsterdam and Philadelphia: John Benjamins.
- _____. 1992b. Nonidentifiable information and pragmatic order rules in 'O'odham. In: Doris L. Payne, ed., *Pragmatics of word order flexibility* (Typological studies in language, 22), pp. 137-66. Amsterdam and Philadelphia: John Benjamins.
- _____. 1995. Verb initial languages and information order. In: Pamela Downing and Michael Noonan, eds., *Word order in discourse* (Typological studies in language, 30), pp. 449-85. Amsterdam and Philadelphia: John Benjamins. (Selection of papers presented at a symposium entitled *Word Order in Discourse* held at the University of Wisconsin-Milwaukee, April 12-14, 1991 (UWM Linguistics Symposium, 20), chaired by Edith Moravcsik and Michael Noonan.)

- Perkins, Revere D. 1989. Statistical techniques for determining language sample size. *Studies in Language* 13.2, pp. 293-315.
- Perlmutter, David M. 1971. *Deep and surface structure constraints in syntax*. New York: Holt, Rinehart and Winston. (Revised version of Ph. D. dissertation, M.I.T., 1968.)
- _____. 1978. Impersonal passives and the Unaccusative Hypothesis. In: Jeri J. Jaeger, Anthony C. Woodbury et al., eds., *Proceedings of the fourth annual meeting of the Berkeley Linguistics Society* (BLS 4), pp. 157-89. Berkeley: Berkeley Linguistics Society.
- Pinker, Steven. 1991. Rules of language. *Science* 253, pp. 530-534.
- Pinker, Steven and Prince, Alan. 1988. On language and connectionism. *Cognition* 28, pp. 73-194.
- _____. 1994. Regular and irregular morphology and the psychological status of rules of grammar. In: Susan D. Lima, Roberta L. Corrigan, and Gregory K. Iverson, eds., *The reality of linguistic rules* (Studies in language companion series, 26), pp. 321-51. Amsterdam and Philadelphia: John Benjamins. (Selected papers from the 21st annual University of Wisconsin, Milwaukee, linguistics symposium, April 10-12, 1992.)
- Pollock, Jean-Yves. 1989. Verb Movement, Universal Grammar and the structure of IP. *Linguistic Inquiry* 20.3, pp. 365-424.
- Prideaux, Gary D. 1994. Review of Anna Siewierska, *Functional Grammar*, London, Routledge, 1991. *Language* 70.1, pp. 166-69.
- Prince, Ellen F. 1978. A comparison of Wh-clefts and It-clefts in discourse. *Language* 54.4, pp. 883-906.
- _____. 1981a. Topicalization, focus-movement, and Yiddish-movement. In: Danny K. Alford et al., eds., *Proceedings of the seventh annual meeting of the Berkeley Linguistics Society* (BLS 7), pp. 249-64. Berkeley: Berkeley Linguistics Society.
- _____. 1981b. Toward a taxonomy of given-new information. In: Peter Cole, ed., *Radical pragmatics*, pp. 223-55. New York: Academic Press.
- _____. 1989. On pragmatic change: The borrowing of discourse functions. In: A. Kasher, ed., *Cognitive aspects of language use*, pp. 1-14. Amsterdam: Elsevier.
- _____. 1992a. On syntax in discourse in language contact situations. In: Claire J. Kramsch and Sally McConnell-Ginet, eds., *Text and context : Cross-disciplinary perspectives on language study* (Foreign language acquisition research and instruction, 2), pp. 98-112. Lexington, MA and Toronto: D.C. Heath & Co.

- _____. 1992b. The ZPG letter: Subjects, definiteness, and information-status. In: William C. Mann and Sandra A. Thompson, eds., *Discourse description: Diverse linguistic analyses of a fund-raising text* (Pragmatics and Beyond, new series, 16), pp. 295-325. Amsterdam and Philadelphia: John Benjamins.
- Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey N., and Svartvik, Jan. 1972. *A grammar of contemporary English*. London: Longman / Harcourt, Brace, Jovanovich.
- Ramat, Paolo. 1995. Typological comparison: Towards a historical perspective. In: Masayoshi Shibatani and Theodora Bynon, eds., *Approaches to language typology*, pp. 27-48. Oxford: Clarendon Press (Oxford University Press).
- Rebuschi, Georges. 1983. A note on focalization in Basque. *Journal of Basque Studies* 4.2, pp. 29-42.
- _____. 1984/1982. *Structure de l'énoncé en Basque*. Paris: Société d'études linguistiques et anthropologiques de France. (Ph.D. dissertation, Université de Paris 7, 1982.)
- _____. 1985a. Positions, configurations et classes syntaxiques, aspects de la construction de la phrase simple en Basque. *Euskera* 30.2, pp. 117-28.
- _____. 1985b. Theorie du liage et langues non-configurationnelles: Quelques données du Basque Navarro-Labourdin. *Euskera* 30.2, pp. 389-433.
- _____. 1986. Sujeto, ergatividad y (no-)configuracionalidad; lugar del euskara en la tipología generativa. In: Georges Rebuschi, Pello Salaburu, Itziar Laka, Beñat Oyharçabal, and Patxi Goenaga, eds., *Euskal sintaxiaren zenbait arazo* (Donostiako Udako IV Ikastaroak, 1985), pp. 7-56. Bilbo: Euskal Herriko Unibertsitatea.
- _____. 1989. Is there a VP in Basque? In: László Marácz and Pieter Muysken, eds., *Configurationality: The typology of asymmetries*, pp. 85-116. Dordrecht: Foris.
- Reichman, Rachel. 1978. Conversational coherency. *Cognitive Science* 2.4, pp. 283-327.
- Reinhart, Tanya. 1981. Pragmatics and linguistics: An analysis of sentence topics. *Philosophica* 27, pp. 53-94.
- _____. 1982/1981. *Pragmatics and linguistics: An analysis of sentence topics*. Bloomington, IN: Indiana University Linguistics Club. (First published in *Philosophica* 27:1, 1981.)
- _____. 1983. *Anaphora and semantic interpretation*. London: Croom Helm; Chicago: The University of Chicago Press.

- Richards, Jack C. 1973. Error analysis and second language strategies. In: John W. Oller, Jr. and Jack C. Richards, eds., *Focus on the learner*, pp. 114-35. Rowley, MA: Newbury House. (Cited in Dorian 1982.)
- Rochemont, Michael S. and Culicover, Peter W. 1990. *English focus constructions and the theory of grammar* (Cambridge studies in linguistics, 52). Cambridge: Cambridge University Press.
- Romaine, Suzanne. 1981. On the problem of syntactic variation: A reply to Beatriz Lavandera and William Labov. *Working Papers in Sociolinguistics* 81 Austin, TX.
- _____. 1989. Pidgins, creoles, immigrant, and dying languages. In: Nancy D. Dorian, ed., *Investigating obsolescence: Studies in language contraction and death* (Studies in the Social and Cultural Foundations of Language, 7), pp. 369-83. Cambridge: Cambridge University Press.
- Romaine, Suzanne and Martin-Jones, Marilyn. 1986. Semilingualism: A half-baked theory of communicative competence. *Applied Linguistics* 6, pp. 105-17.
- Ross, John Robert. 1967/1968. *Constraints on variables in syntax*. Ph. D. dissertation, Massachusetts Institute of Technology. (Distributed by the Indiana University Linguistics Club, Bloomington, IN, 1968.)
- _____. 1973. The penthouse principle and the order of constituents. In: Claudia Corum, T. Cedric Smith-Stark, and Ann Weiser, eds., *You take the high node and I'll take the low node. Papers from the Comparative Syntax Festival: The differences between main and subordinate clauses, 12 April 1973* (Paravolume to Papers from the ninth regional meeting of the Chicago Linguistic Society), pp. 397-422. Chicago: Chicago Linguistic Society.
- _____. 1975. Parallels in phonological and semantic organization. In: James F. Kavanagh and James E. Cutting, eds., *The role of speech in language* (Syntax and semantics, 12), pp. 283-304. Cambridge, MA: The MIT Press. (Proceedings of a conference entitled "The role of speech in language", 5th in the series "Communicating by language", sponsored by the National Institute of Child Health and Human Development, National Institutes of Health; held at the Urban Life Center, Columbia, MD, Oct. 7-10, 1973.)
- Sadock, Jerrold M. 1971. Queclaratives. *Papers from the 7th regional meeting of the Chicago Linguistics Society* (CLS 7), pp. 223-31. Chicago: Chicago Linguistic Society.
- Sankoff, David and Brown, P. 1976. The origins of syntax in discourse. *Language* 52.
- Sapir, Edward. 1921. *Language: An introduction to the study of speech*. New York: Harcourt, Brace.

- Sasse, Hans-Jürgen. 1987. Thethetic/categorical distinction revisited. *Linguistics* 25, pp. 511-80.
- _____. 1995a. Theticity and VS order: A case study. In: Hans-Jürgen Sasse and Yaron Matras, eds., *Sprachtypologie und Universalienforschung = Verb-subject order and theticity in European languages* 48.1/2, pp. 3-31.
- _____. 1995b. A contrastive study of VS clauses in Modern Greek and Hungarian. In: Hans-Jürgen Sasse and Yaron Matras, eds., *Sprachtypologie und Universalienforschung = Verb-subject order and theticity in European languages* 48.1-2, pp. 142-88.
- Sasse, Hans-Jürgen and Matras, Yaron, eds. 1995. *Sprachtypologie und Universalienforschung = Verb-subject order and theticity in European languages* 48.1-2.
- Saussure, Ferdinand de. 1959/1916. *Course in general linguistics* (translated by Wade Baskin). New York: Philosophical Library; McGraw-Hill. (Original French edition: *Cours de linguistique générale*. Paris: Payot, 1916.)
- Saxton, Dean. 1982. Papago. *Studies in Uto-Aztecan grammar*, vol. 3 Arlington, TX: Summer Institute of Linguistics, University of Texas at Arlington Press. (Cited in Campbell, Bubenik, and Saxon 1988.)
- Saxton, Dean and Saxton, Lucille. 1969. *Dictionary, Papago and Pima to English, English to papago and Pima*. Tucson, AZ: University of Arizona Press. (Cited in Campbell, Bubenik, and Saxon 1988.)
- Schachter, Paul. 1976. The subject in Philippine languages: Topic, actor, actor-topic, or none of the above. In: Charles N. Li, ed., *Subject and topic*, pp. 491-518. New York: Academic Press.
- _____. 1977. Reference-related and role-related properties of subjects. In: Peter Cole and Jerry M. Sadock, *Syntax and Semantics 8: Grammatical relations*, pp. 279-306. New York: Academic Press.
- Schlobinski, Peter and Schütze-Coburn, Stephan. 1992. On the topic of topic and topic continuity. In: Norbert Dittmar, ed., *Topic--From grammar to discourse = Linguistics* 30.1, pp. 89-121.
- Schmerling, Susan F. 1976. *Aspects of English sentence stress*. Austin: University of Texas Press.
- Schuetze-Coburn, Stephan; Shapley, Marian, and Weber, Elizabeth G. 1991. Units of intonation in discourse: A comparison of acoustic and auditory analyses. *Language and Speech* 34, pp. 207-34.

- Schwartz, Arthur. 1975. Verb anchoring and verb-movement. In: Charles N. Li, ed., *Word order and word order change*, pp. 437-62. Austin: University of Texas Press. (Cited in Danchev 1991.)
- Seliger, Herbert W. and Vago, Robert M. 1991. The study of first language attrition: An overview. *First language attrition*, pp. 3-16. Cambridge: Cambridge University Press.
- Sgall, Petr. 1987. The position of Czech linguistics in theme-focus research. In: Ross Steele and Terry Threadgold, eds., *Language topics: Essays in honour of Michael Halliday* (Volume I), pp. 47-55. Amsterdam and Philadelphia: John Benjamins.
- Sgall, Petr, Hajičová, Eva, and Benešová, E. 1973. *Topic, focus and generative semantics*. Kronberg/Taunus: Scriptor.
- Sheintuch, Gloria. 1981. On the notion verbal unit and its relevance to certain syntactic constructions. *Linguistics 19*, pp. 325-55.
- Shibatani, Masayoshi. 1991. Grammaticization of topic into subject. In: Elizabeth Closs Traugott and Bernd Heine, *Approaches to grammaticalization. Volume II: Focus on types of grammatical markers* (Typological studies in language, 19:2), pp. 93-134. Amsterdam and Philadelphia: John Benjamins.
- Shibatani, Masayoshi and Bynon, Theodora. 1995. Approaches to language typology: A conspectus. In: Masayoshi Shibatani and Theodora Bynon, eds., *Approaches to language typology*, pp. 1-25. Oxford: Clarendon Press (Oxford University Press).
- Siewierska, Anna. 1988. *Word order rules*. London: Croom Helm.
- Silva-Corvalán, Carmen. 1983. On the interaction of word order and intonation: Some OV constructions in Spanish. In: Flora Klein-Andreu, ed., *Discourse perspectives on syntax*, pp. 117-40. New York: Academic Press.
- _____. 1984. Topicalización y pragmática en español. *Revista Española de Lingüística* 14.1, pp. 1-19. (Expanded and revised version of paper given at the VI Congreso Internacional de la Asociación de Lingüística y Filología de la América Latina, 1981. A similar English version of this paper has appeared in Klein-Andreu, F., ed., 1983, *Discourse perspectives on syntax*. New York: Academic press.)
- _____. 1986. Bilingualism and language change: The extension of *estar* in Los Angeles Spanish. *Language* 62.3, pp. 587-608.
- _____. 1991. Spanish language attrition in a contact situation with English. In: Herbert W. Seliger and Robert M. Vago, eds., *First language attrition: Structural and theoretical perspectives*, pp. 151-71. Cambridge: Cambridge University Press.

- _____. 1993. On the permeability of grammars: Evidence from Spanish and English contact. In: William J. Ashby, Marianne Mithun, Giorgio Perissinotto, and Eduardo Raposo, eds., *Linguistic perspectives on the Romance languages: Selected papers from the 21st Linguistic Symposium on Romance Languages (LSRL XXI), Santa Barbara, California, 21-24 February 1991* (Amsterdam studies in the theory and history of linguistic science. Series IV, Current issues in linguistic theory, volume 103), pp. 19-43. Amsterdam and Philadelphia: John Benjamins.
- _____. 1994. *Language contact and change: Spanish in Los Angeles* (Oxford studies in language contact). Oxford: Oxford University Press.
- Silverstein, Michael. 1976. Hierarchy of features and ergativity. In: Robert M. W. Dixon, ed., *Grammatical categories in Australian languages* (AIAS linguistic series, 22), pp. 112-71. Canberra: Australian Institute of Aboriginal Studies. (Reprinted in *Features and projections*, ed. by Pieter Muysken and Henk van Riemsdijk, 163-232. Dordrecht: Foris.)
- Simone, Raffaele, ed. 1995. *Iconicity in language* (Amsterdam studies in the theory and history of linguistic science. Series IV, Current issues in linguistic theory, v. 110). Amsterdam and Philadelphia: John Benjamins. (Papers presented at a conference held Oct. 1992, Rome, Italy.)
- Skousen, Royal. 1990. *Analogical modeling of language*. Dordrecht, NL: Kluwer Academic Publishers.
- _____. 1992. *Analogy and structure*. Dordrecht: Kluwer Academic Publisher.
- Slobin, Dan I. 1973. Cognitive prerequisites for the development of grammar. In: Charles A. Ferguson and Dan I. Slobin, eds., *Studies of child language development*, pp. 175-208. New York: Holt, Rinehart & Winston.
- _____. 1977. Language change in childhood and history. In: John Macnamara, ed., *Language learning and thought* (Perspectives in neurolinguistics and psycholinguistics), pp. 185-214. New York: Academic Press.
- _____. 1982 [1978]. Universal and particular in the acquisition of language. In: Eric Wanner and Lila R. Gleitman, eds., *Language acquisition: The state of the art*, pp. 128-70. Cambridge: Cambridge University Press.
- _____. 1985. The child as a linguistic icon-maker. In: John Haiman, ed., *Iconicity in syntax* (Typological studies in language, 6), pp. 221-48. Amsterdam and Philadelphia: John Benjamins. (Proceedings of a symposium on iconicity in syntax, Stanford, June 24-6, 1983.)
- Slobin, Dan I. and Bever, Thomas G. 1982. Children use canonical sentence schemas: A cross-linguistic study of word order and inflections. *Cognition* 12, pp. 229-65.

- Smith, Neil V. 1981. Consistency, markedness and language change: On the notion 'consistent language'. *Journal of Linguistics* 17, pp. 39-54. (Cited in Campbell, Bubenik, and Saxon 1988.)
- Stenson, Nancy. 1974. Induced errors. In: John H. Schumann and Nancy Stenson, eds., *New frontiers in second language learning*, pp. 54-70. Rowley, MA: Newbury House. (Cited in Dorian 1982.)
- Stockwell, Robert P. 1977. Motivations for exbraciation in Old English. In: Charles N. Li, ed., *Mechanisms of syntactic change*, pp. 291-316. Austin: University of Texas Press.
- _____. 1984. On the history of the verb-second rule in English. In: Jacek Fisiak, ed., *Historical syntax* (Trends in linguistics. Studies and monographs, 23), pp. 575-92. Berlin: Mouton. (Papers presented at the Third International Conference on Historical Linguistics, devoted to historical syntax, held at Adam Mickiewicz University, at Błażejewko, Poland, 31 March - 3 April, 1981.)
- Stockwell, Robert P. and Minkova, Donka. 1991. Subordination and word order change in the history of English. In: Dieter Kastovsky, ed., *Historical English syntax* (Topics in English linguistics, 2), pp. 367-408. Berlin: Mouton de Gruyter. (Proceedings of the Kellner-Festival, held in 1988: at Schloss Liechtenstein.)
- Sweetser, Eve. 1986. Polysemy vs. abstraction: Mutually exclusive or complementary. In: Vassiliki Nikiforidou, Mary VanClay, Mary Niepokuj, and Deborah Feder, eds., *Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society* (General Session and Parasession on Semantic Typology), pp. 528-38. Berkeley, CA: Berkeley Linguistics Society.
- Talmy, Leonard. 1985. Lexicalization patterns: Semantic structure in lexical forms. In: Tim Shopen, ed., *Language typology and syntactic description, vol. 3*, pp. 57-149. Cambridge: Cambridge University Press.
- _____. 1988. The relation of grammar to cognition. In: Brygida Rudzka-Ostyn, ed., *Topics in cognitive linguistics* (Current issues in linguistic theory, 50), pp. 165-206. Amsterdam and Philadelphia: John Benjamins.
- Tannen, Deborah. 1982a. Oral and literate strategies in spoken and written discourse. *Language* 58, pp. 1-20.
- _____. 1982b. The oral/literate continuum in discourse. In: Deborah Tannen, ed., *Spoken and written language: Exploring orality and literacy* (Advances in discourse processing, 9), pp. 1-16. Norwood, NJ: Ablex.
- _____. 1989. *Talking voices: Repetition, dialogue, and imagery in conversational discourse* (Studies in interactional sociolinguistics, 6). Cambridge: Cambridge University Press.

- _____. 1992. How is conversation like literary discourse? The role of imagery and details in creating involvement. In: Pamela Downing, Susan D. Lima, and Michael Noonan, eds., *The linguistics of literacy* (Typological studies in language, 21), pp. 31-46. Amsterdam and Philadelphia: John Benjamins. (From the Seventeenth Annual University of Wisconsin-Milwaukee Linguistics Symposium, which was held in Milwaukee on April 8-10, 1988.)
- Tannen, Deborah, ed. 1982. *Spoken and written language: Exploring orality and literacy* (Advances in discourse processing, 9). Norwood, NJ: Ablex.
- Thomason, Sarah Grey and Kaufman, Terrence. 1988. *Language contact, creolization and genetic linguistics*. Berkeley: University of California Press.
- Thompson, Sandra A. 1978. Modern English from a typological point of view: Some implications of the function of word order. *Linguistische Berichte* 54, pp. 19-35.
- _____. 1990. Information flow and dative shift in English discourse. In: Jerold A. Edmondson, Crawford Feagin, and Peter Mühlhäusler, eds., *Development and diversity: Language variation across time and space. A Festschrift for Charles-James N. Bailey*, pp. 239-53. Arlington, TX: The Summer Institute of Linguistics and The University of Texas at Arlington.
- Thompson, Sandra A. and Longacre, Robert E. 1985. Adverbial clauses. In: Timothy Shoppen, ed., *Language typology and syntactic description, vol. 2*, pp. 171-233. Cambridge: Cambridge University Press.
- Thompson, Sandra A. and Mulac, Anthony. 1988. A quantitative perspective on the grammaticalization of evidential parentheticals in English. In: Sandra A. Thompson, ed., *Santa Barbara papers in linguistics, vol. 2: Discourse and grammar*, pp. 165-81. Santa Barbara, CA: Department of Linguistics.
- _____. 1991a. A quantitative perspective on the grammaticization of epistemic parentheticals in English. In: Elizabeth Closs Traugott and Bernd Heine, *Approaches to grammaticalization. Volume II: Focus on types of grammatical markers* (Typological studies in language, 19:2), pp. 313-30. Amsterdam and Philadelphia: John Benjamins.
- _____. 1991b. The discourse conditions for the use of the complementizer *that* in conversational English. *Journal of Pragmatics* 15, pp. 237-51.
- Tomlin, Russell S. 1986. *Basic word order*. London: Croom Helm.
- Tomlin, Russell S. and Rhodes, Richard. 1979. An introduction to information distribution in Ojibwa. In: Paul R. Clyne, William F. Hanks, and Carol L. Hofbauer, eds., *Papers from the Fifteenth Regional Meeting of the Chicago Linguistic Society* (CLS 15), pp. 307-20. Chicago: Chicago Linguistic Society. (Reprinted in Doris L. Payne, ed., 1992.)

- Trask, L. R. (= Trask, Robert L.). 1995. On the history of the non-finite verb forms in Basque. In: José Ignacio Hualde, Joseba A. Lakarra, and R. L. Trask, eds., *Towards a history of the Basque language* (Amsterdam studies in the theory and history of linguistic science. Series IV, Current issues in linguistic theory ; v. 131), pp. 207-34. Amsterdam and Philadelphia: John Benjamins.
- Trask, Robert L. 1977. Historical syntax and Basque verbal morphology: Two hypotheses. In: William A. Douglass, Richard W. Etulain, and William H. Jacobsen, Jr., eds., *Anglo-American contributions to Basque studies: Essays in honor of Jon Bilbao* (Desert Research Institute publications in the Social Sciences, no. 13), pp. 203-17. Reno, NV: Desert Research Institute.
- _____. 1979. On the origins of ergativity. In: Frans Plank, ed., *Ergativity: Towards a theory of grammatical relations*, pp. 385-404. London: Academic Press.
- _____. 1984. *Synchronic and diachronic studies in the grammar of Basque*. Ph.D. dissertation, University of London.
- _____. 1985a. The Basque passive: A correct description. *Linguistics* 23, pp. 985-91.
- _____. 1985b. -ko atzizkia euskaraz. *Euskera* 30.1, pp. 165-73.
- Traugott, Elizabeth Closs and Heine, Bernd, eds. 1991a. *Approaches to grammaticalization. Volume I: Focus on theoretical and methodological issues* (Typological studies in language, 19:1). Amsterdam and Philadelphia: John Benjamins.
- _____. 1991b. *Approaches to grammaticalization. Volume II: Focus on types of grammatical markers* (Typological studies in language, 19:2). Amsterdam and Philadelphia: John Benjamins.
- Traugott, Elizabeth Closs and König, Ekkehard. 1991. The semantics-pragmatics of grammaticalization revisited. In: Elizabeth Closs Traugott and Bernd Heine, eds., *Approaches to grammaticalization. Volume I: Focus on theoretical and methodological issues* (Typological studies in language, 19:1), pp. 189-218. Amsterdam and Philadelphia: John Benjamins.
- Traugott, Elizabeth Closs and Romaine, Suzanne. 1985. Some questions for the definition of "style" in socio-historical linguistics. In: Suzanne Romaine and Elizabeth Closs Traugott, eds., *Papers from the workshop on socio-historical linguistics = Folia Linguistica Historica* 6.1, pp. 7-39.
- Trudgill, Peter. 1978. Creolization in reverse: Reduction and simplification in the Albanian dialects of Greece. *Transactions of the Philological Society* 1976/1977, pp. 32-50.
- _____. 1986. *Dialects in contact*. Oxford: Basil Blackwell. (Cited in Trudgill 1989.)

- _____. 1989. Contact and isolation in linguistic change. In: Leiv Egil Breivik and Ernst Håkon Jahr, eds., *Language change: Contributions to the studies of its causes* (Trends in linguistics. Studies and monographs, 43), pp. 227-37. Berlin: Mouton de Gruyter. (Mostly papers presented at the symposium "The causes of language change: Do we know them yet?" held at the School of Languages and Literature, University of Tromsø, October 15-17, 1987.)
- _____. 1992. Dialect typology and social structure. In: Ernst Håkon Jahr, ed., *Language contact: Theoretical and empirical studies* (Trends in linguistics, studies and monographs, volume 60), pp. 195-211. Berlin: Mouton de Gruyter. (Based on the Fifth International Tromsø Symposium on Language held at the University of Tromsø, Norway, in Sept. 1989, the theme of which was "Language Contact".)
- Tuggy, David. 1993. Ambiguity, polysemy, and vagueness. *Cognitive Linguistics* 4.3, pp. 273-90.
- Txillardegi (= Alvarez Enparantza, José Luis). 1984. *Euskal azentuaz*. Donostia: Elkar.
- Vachek, Josef, ed. 1964. *A Prague School reader in linguistics*. Indiana University Press.
- Vallduví, Enric. 1990. *The informational component*. Ph. D. dissertation, University of Pennsylvania.
- _____. 1993. Catalan as VOS: Evidence from information packaging. In: William J. Ashby, Marianne Mithun, Giorgio Perissinotto, and Eduardo Raposo, eds., *Linguistic perspectives on the Romance languages: Selected papers from the 21st Linguistic Symposium on Romance Languages (LSRL XXI), Santa Barbara, California, 21-24 February 1991* (Amsterdam studies in the theory and history of linguistic science. Series IV, Current issues in linguistic theory, volume 103), pp. 335-50. Amsterdam and Philadelphia: John Benjamins.
- van Langendonck, Willy. 1994. Determiners as heads? *Cognitive Linguistics* 5.3, pp. 243-59.
- Van Valin, Robert D., Jr. 1985. Case marking and the structure of the Lakota clause. In: Johanna Nichols and Anthony C. Woodbury, eds., *Grammar inside and outside the clause: Some approaches to theory from the field*, pp. 363-413. Cambridge: Cambridge University Press.
- _____. 1986. Pragmatics, island phenomena, and linguistic competence. In: Ann M. Farley, Peter T. Farley, and Karl-Erik McCullough, eds., *Papers from the parasession on pragmatics and grammatical theory at the Twenty Second Regional Meeting of the Chicago Linguistic Society (CLS 22/2)*, pp. 223-33. Chicago: Chicago Linguistic Society.
- _____. 1987. The unaccusative hypothesis vs. lexical semantics: Syntactic vs. semantic approaches to verb classification. *North Eastern Linguistic Society* 17, pp. 641-61.

- _____. 1990a. Functionalism, anaphora and syntax. Review article of Susumu Kuno. *Functional syntax: Anaphora, discourse and empathy*. University of Chicago Press, 1987. *Studies in Language* 14.1, pp. 169-219.
- _____. 1990b. Layered syntax in Role and Reference Grammar. In: Jan Nuyts, A. Machtelt Bolkestein, and Co Vet, eds., *Layers and levels of representation in language theory: A functional view* (Pragmatics and Beyond New Series, 13), pp. 193-232. Amsterdam and Philadelphia: John Benjamins.
- _____. 1992. Incorporation in Universal Grammar: A case study in theoretical reductionism. *Journal of Linguistics* 28, pp. 199-220. (Review article of Mark C. Baker, *Incorporation: A theory of grammatical function changing*. Chicago: University of Chicago Press, 1988.)
- _____. 1993a. Review of William Croft, *Syntactic categories and grammatical relations: The cognitive organization of information*. Chicago: University of Chicago Press, 1991. *Language* 69.1, pp. 155-62.
- _____. 1993b. A synopsis of Role and Reference Grammar. In: Robert D. Van Valin, Jr., ed., *Advances in Role and Reference Grammar* Amsterdam and Philadelphia: John Benjamins.
- _____. 1994. Extraction restrictions, competing theories and the argument from the Poverty of the Stimulus. In: Susan D. Lima, Roberta L. Corrigan, and Gregory K. Iverson, eds., *The reality of linguistic rules* (Studies in language companion series, 26), pp. 243-59. Amsterdam and Philadelphia: John Benjamins. (Selected papers from the 21st annual University of Wisconsin, Milwaukee, linguistics symposium, April 10-12, 1992.)
- Vennemann, Theo. 1973. Explanation in syntax. In: John P. Kimball, ed., *Syntax and Semantics* 2, pp. 1-50. New York & London: Seminar Press.
- _____. 1974a. Analogy in generative grammar: The origin of word order. In: Luigi Heilmann, ed., *Proceedings of the 11th International Congress of Linguists, 1972, Bologna-Florence, vol. II*, pp. 79-83. Bologna: Il Mulino.
- _____. 1974b. Topics, subjects and word order: From SXV to SVX via TVX. In: J. M. Anderson and C. Jones, eds., *Historical Linguistics I*, pp. 339-76. Amsterdam: North Holland.
- _____. 1975. An explanation of drift. In: Charles N. Li, ed., *Word order and word order change*, pp. 269-306. Austin: University of Texas Press.
- _____. 1976. Categorical Grammar and the order of meaningful elements. In: Alphonse Juillard, ed., *Linguistic studies offered to Joseph Greenberg on the occasion of his sixtieth birthday, vol. 3: Syntax* (Studia linguistica et philologica, 4, 3 vols.), pp. 615-34. Saratoga, CA: Anma Libri.

- _____. 1984. Verb-second, verb late, and the brace construction in Germanic: A discussion. In: Jacek Fisiak, ed., *Historical syntax* (Trends in linguistics. Studies and monographs, 23), pp. 627-36. Berlin: Mouton. (Papers presented at the Third International Conference on Historical Linguistics, devoted to historical syntax, held at Adam Mickiewicz University, at Błażewko, Poland, 31 March - 3 April, 1981; Cited in Danchev 1991.)
- Vennemann, Theo and Harlow, Ray. 1977. Categorical grammar and consistent basic VX serialization. *Theoretical Linguistics* 4, pp. 227-54. (Cited in Payne 1990.)
- Villasante, Luis. 1956. Comunicación del P. Villasante sobre la réplica del Sr. Altube al Sr. Michelena acerca de sintaxis vasca. *Euskera* 1, pp. 14-18. (The journal *Euskera* was restarted in 1956; 1931= Vol. XII. 1956 = Vol. I.)
- _____. 1980. *Sintaxis de la oración simple* (Serie «Luis de Eleizalde» sobre unificación del euskera escrito, 6). Oñate, Basque Country: Editorial Franciscana Aránzazu.
- von der Gabelentz, Georg. 1869. Ideen zu einer vergleichenden Syntax: Wort und Satzstellung. *Zeitschrift für Völkerpsychologie und Sprachwissenschaft* 6, pp. 376-84.
- _____. 1881/1960. *Chinesische Grammatik: Mit Ausschluß der niederen Stils und der heutigen Umgangssprache*. Halle (Saale): VEB Niemeyer.
- _____. 1972/1901/1891. *Die Sprachwissenschaft, ihre Aufgaben, Methoden und bisherigen Ergebnisse*. Tübingen: Narr. (Original place of publication: Leipzig; cited in Sgall 1987, and Schlobinski and Schutze-Coburn 1992.)
- Wackernagel, Jacob. 1892. Über ein Gesetz der indogermanischen Wortstellung. *Indogermanische Forschungen* 1, pp. 333-436.
- Ward, Gregory L. 1988. *The semantics and pragmatics of preposing*. New York: Garland Publishing, Inc. (Ph. D. dissertation, University of Pennsylvania, 1985.)
- _____. 1990. The discourse function of VP preposing. *Language* 66.4, pp. 742-63.
- Ward, Gregory L. and Prince, Ellen F. 1991. On the topicalization of indefinite NPs. In: A. Machtelt Bolkestein, ed., *Syntax and pragmatics = Journal of Pragmatics* 16, pp. 167-77.
- Watson, Seosamh. 1989. Scottish and Irish Gaelic: The giant's bed-fellows. In: Nancy D. Dorian, ed., *Investigating obsolescence: Studies in language contraction and death* (Studies in the Social and Cultural Foundations of Language, 7), pp. 41-59. Cambridge: Cambridge University Press.
- Wegener, P. 1885. *Untersuchungen über die Grundfragen des Sprachlebens*. Halle/S. (Cited in Sgall 1987.)

- Weil, Henri. 1844. *De l'ordre des mots dans les langues anciennes comparées aux langues modernes*. Paris. (English translation: 1887; Most recent English edition: 1978. The order of words in the ancient languages compared with that of the modern languages. (Amsterdam classics in linguistics, 14). Amsterdam: John Benjamins.)
- Weinreich, Uriel. 1974/1953. *Languages in contact: Findings and problems*. The Hague: Mouton. (8th printing; 1st printing: 1963; Originally published as Number 1 in the series *Publications of the Linguistic Circle of New York*, New York, 1953.)
- Weinreich, Uriel; Labov, William; and Herzog, Marvin I. 1968. Empirical foundations for a theory of language change. In: Winfred P. Lehmann and Yakov Malkiel, eds., *Directions for historical linguistics*, pp. 95-189. Austin: University of Texas Press. (Cited in Croft 1991.)
- Werth, Paul. 1993. Accommodation and the myth of presupposition: The view from discourse. *Lingua* 89.1, pp. 39-95.
- Wierzbicka, Anna. 1986. The semantics of 'internal dative' in English. *Quaderni di Semantica* 7.1, pp. 121-35. (Comments follow by Ray Cattell, Scott DeLancey, Georgia M. Green, Robert S. Kirsner, and Timothy Shopen.)
- Wilbur, Terence H. 1979. *Prolegomena to a grammar of Basque* (Amsterdam studies in the theory and history of linguistic science, Series IV - Current issues in linguistic theory, 8). Amsterdam: John Benjamins.
- Wright, Suzanne and Givón, T. 1987. The pragmatics of indefinite reference: Quantified text-based studies. *Science of Language* 11.1, pp. 1-33.
- Yngve, Victor H. 1960. A model and an hypothesis for language structure. *Proceedings of The American Philosophical Society* 104.5, pp. 444-66.
- _____. 1961. The depth hypothesis. In: R. Jakobson, ed., *Structure of language and its mathematical aspects* (Proceedings of Symposia in Applied Mathematics, 12), pp. 131-38. Providence, RI: Americal Mathematical Society.
- Zepeda, Ofelia. 1983. *A Papago grammar*. Tucson: University of Arizona Press. (Cited in Campbell, Bubenik, and Saxon 1988.)
- Zimmer, Karl E. 1975. Some thoughts on likely phonologies for non-ideal speakers. In: Robin E. Grossman, L. James San, and Timothy J. Vance, eds., *Papers from the parasession on functionalism: April 17, 1975* (Eleventh regional meeting of the Chicago Linguistic Society), pp. 556-67. Chicago: Chicago Linguistic Society.

Ziv, Yael. 1975. On the relevance of content to the form-function correlation: An examination of extraposed relative clauses. In: Robin E. Grossman, L. James San, and Timothy J. Vance, eds., *Papers from the parasession on functionalism: April 17, 1975* (Eleventh regional meeting of the Chicago Linguistic Society), pp. 568-79. Chicago: Chicago Linguistic Society.

_____. 1994. Left and right dislocations: Discourse functions and anaphora. *Journal of Pragmatics* 22, pp. 629-45.

Ziv, Yael and Cole, Peter. 1974. Relative extraposition and the scope of definite descriptions in Hebrew and English. In: Michael W. La Galy, Robert A. Fox, and Anthony Bruck, eds., *Papers from the tenth regional meeting of the Chicago Linguistic Society* (CLS 10), pp. 772-86. Chicago: Chicago Linguistics Society.

Zubiri Ibarrodo, Ilari. 1989. *Gramática didáctica del euskera*. Donostia: HABE.

Zubiri, Ilari and Zubiri, Entzi. 1995. *Euskal gramatika osoa*. Bilbo: Didaktiker.

Appendices

Appendix 2.1: Finite dependent clauses in the Written Basque Corpus
(The four right columns do not include negative dependent clauses)

Conjunction	Translation	No.	#V...	...V...	...V#	V
-n bezala	<i>as/like</i>	4	-	-	4	-
-n bitartean	<i>while</i>	1	-	-	1	-
-n arren	<i>although</i>	2	-	-	2	-
... bait-	<i>because/for ...</i>	16	-	5	6	-
... ba-	<i>if</i>	25	1	2	14	2
...ba- ... bezala	<i>as if</i>	3	-	-	3	-
#ea	opt. embedded qn. subordinator	13	-	-	6	-
-eta	<i>because (3neg. w/ inv.)</i>	7	-	-	4	-
-la eta	<i>because, since</i>	2	-	-	2	-
-la	<i>that ...</i>	61	-	13	40	2
-nik	complement of neg. verb	5	-	1	4	-
-lako	<i>because; since</i>	3	-	-	3	-
-la Adv	adverbial/predicate (<i>Kuba</i>)	4	-	-	4	-
-larik	adverbial/predicate (<i>Kuba</i>)	1	-	-	1	-
-n	Rel. Cl. (finite)	74	-	-	57	17
-na	Rel. Cl. (headless)	17	-	-	12	5
-nean	<i>when</i>	23	-	-	21	2
-netik	<i>since</i>	2	-	-	2	-
-nez	<i>how, according to</i>	8	-	-	2	6
<i>subjunctive</i>	<i>purpose, tails</i>	2	-	-	2	-
zeren ... bait-	<i>because, since</i>	7	-	2	2	1
TOTAL		291	1	23	192	35

Appendix 2.2: Finite dependent clauses in the Spoken Basque Corpus
(The four right columns do not include negative dependent clauses)

Conjunction	Meaning/function	No.	#V...	...V...	...V#	V
-n bitartean	<i>while</i>	6	-	-	6	-
ba- ... bezala	<i>as if</i>	4	-	-	4	-
-n bezala	<i>since (setting) (SXV)</i>	1	-	-	1	-
ba-	<i>if (XV, OV)</i>	2	-	-	2	-
#ea	<i>whether (only ex. is neg.)</i>	1	-	-	-	-
-eta	<i>because</i>	1	-	1	-	-
-la eta	<i>because, since (tail) (OV)</i>	1	-	-	1	-
-la	<i>complement</i>	238	33	40	124	14
-la	<i>predicate</i>	38	19	1	14	1
-lako	<i>because; since</i>	28	6	3	13	4
-lakoan	<i>thinking that ... (tail, XV)</i>	1	-	-	1	-
-larik	<i>Adv. (XV)</i>	1	-	-	1	-
-n	<i>(Relative Clause)</i>	65	-	-	65	-
-na	<i>(Headless relative clause)</i>	43	9	2	24	8
-na	<i>Predicate (= -la)</i>	14	3	2	8	1
-nean	<i>when ...</i>	70	13	5	43	9
-nez	<i>how, since</i>	4	-	-	4	-
-nik	<i>complement of neg. verb (OV)</i>	1	-	-	1	-
nola	<i>since (cf. Sp. como..., tail)</i>	21	4	6	9	1
#nola	<i>complement/statement</i>	19	4	5	6	4
#nola	<i>how (question), modal adv.</i>	4	-	-	3	1
<i>subjunctive</i>	<i>so that ... (OVXtzeko)</i>	1	-	1	-	-
#ze(ren)	<i>because, since</i>	17	5	4	4	1
#zergatik	<i>because</i>	10	2	3	4	-
TOTAL		591				

Appendix 2.3: Dependent non-finite clauses in the Spoken Basque Corpus by subordinating conjunctions

Conjunction	Translation	No.	#V...	...V...	...V#	V
-tu	complement / unmarked clause	80	5	4	54	18
-tuz	adverbial (gerundive)	28	10	1	14	3
-tu ahala	<i>as</i>	2	-	-	1	1
-tu eta gero	<i>after</i>	3	-	-	1	2
-tu beharrean	<i>instead of</i>	8	1	-	6	1
-tu baino lehen	<i>before</i>	3	-	-	3	-
-tu gabe/barik	<i>without</i>	8	-	-	4	4
-tu ordez	<i>instead of</i>	1	-	-	1	-
-tu ondoren	<i>after</i>	2	-	-	-	2
-tu <i>Noun</i>	noun complement (<i>nahian</i>)	1	-	-	1	-
-tu orduko	<i>before -ing</i>	1	-	-	-	1
-tuta	Adverbial Cl.: in-clause- and tail	30	1	-	27	2
-tutako	adjectival modifier	6	-	-	6	-
-tutakoan	<i>when going to</i>	1	-	-	-	1
-tzea	deverbal noun (abs. sg.)	17	2	1	11	2
-tzean	<i>upon -ing</i>	4	1	1	1	1
-tzen	complement	193	19	8	78	88
-tzen	predicate/adverbial	50	7	4	30	9
-tzera	<i>to</i> (purpose; with motion verbs)	43	2	-	15	26
-tzeagatik	for (reason) (OV)	1	-	-	1	-
-tzearren	<i>because of (for)</i>	1	-	-	1	-
-tzeko	purpose	73	14	-	38	20
-tzeko(a)	Noun/adj. complement	17	-	-	3	14
-tzeko	complement of verb of saying	24	5	-	12	7
-tzerakoan	when going to	5	-	-	2	2
-ra	compl./predicate with <i>begi</i>	13	1	-	11	1
-ka	adverbial (3 IV)	3	-	-	3	-
TOTAL		618				

Appendix 2.4: Dependent non-finite clauses in the Written Basque Corpus

Conjunction	Translation	No.	#V...	...V...	...V#	V
-tu	perfective (unmarked)	60	-	-	47	13
-tu ezker	<i>if</i>	3	-	-	2	1
-tu eta gero	<i>after -ing</i>	2	-	-	2	-
-tu orduko	<i>by when</i>	2	-	-	1	1
-tu gabe/barik	<i>without -ing</i>	11	-	-	9	2
-tu ezean	<i>other than by -ing</i>	1	-	-	1	-
-tu ondoren	<i>after -ing</i>	5	-	-	5	-
-tu aurretik	<i>before -ing</i>	1	-	-	-	1
-tu arte	<i>until -ing</i>	2	-	-	-	2
-tu baino hobeto	<i>better than -ing</i>	1	-	-	1	-
-tu berria	<i>newly -ed</i>	1	-	-	-	1
-tu arren	<i>despite -ing</i>	1	-	-	1	-
-tu bezala	<i>as</i>	1	-	-	1	-
-tu <i>Noun</i>	noun complement	11	-	-	9	2
-tu eta	causal	1	-	-	1	-
-tuz	3dverbial (gerundive)	33	-	-	32	1
-tua	Adj. Cl. (non-fin)	24	-	-	17	7
-tuta	Adv. Cl. (controlled undergoer)	52	-	-	19	1
-tutako	Rel. Cl. (non-fin)	21	-	-	19	2
-tutakoan	<i>after -ing</i>	1	-	-	1	-
-tuta bezala	<i>as</i>	1	-	-	1	-
-tzeke	Noun/Adj. complement	20	-	-	15	5
-tzeke	purpose (adverbial)	17	-	-	14	3
-tzeke	Speech verb complement	4	-	-	3	1
-tze-a/ak/ari	deverbal noun (abs/erg/dat/...)	42	-	-	33	9
-tzen	gerundive (not complement)	23	-	-	16	7
-tzera	conditional (neg.) / purpose	10	-	-	7	3
-tzeke	<i>without -ing (= -tu gabe)</i>	1	-	-	1	-
-ra	(begira)	7	-	-	7	-
-ka	adverbial	6	-	-	1	5
TOTAL		365				

Appendix 2.5: Background of speakers in the Spoken Basque Corpus
(native = learned at home; new Basque = non-native Basque speaker)

Speaker	Tape Sections	Story	
HEND-01	93C-1A01	Pear-1	Male; Age 10; Native French; Basque learned mostly at school; parents are new Basques.
HEND-02	93C-1A02	Chaplin-1	Male; Age 10; Basque native; fully bilingual; uses Basque primarily; father is old Basque; mother new Basque.
HEND-03	93C-1A03	Pear-2	Male; Age 9; native: Hebrew; at home Basque and French is spoken; good speaker of Basque; father non-native speaker; mother non-speaker.
HEND-04	93C-1A05, 93C-1B03	Chaplin-2, Pear-6	Male; Age 10; French and Spanish native; Basque learned at school; speaks all the same; average speaker; uses French more.
HEND-05	93C-1A09	Pear-3	Female; Age 10; native Basque, Spanish and French; speaks Basque with mother; other with father; uses all equally well.
HEND-06	93C-1A10	Chaplin-3	Female; Age 10; native French and Spanish; parents are non-fluent new Basques; uses all three languages equally well.
HEND-07	93C-1A15	Pear-4	Male; Age 10; Basque native; also speaks French and some Spanish; speaks Basque best; good speaker.
HEND-08	93C-1A16	Chaplin-4	Male; Age 10; Basque native; speaks some French and Spanish as well; excellent Basque.
HEND-09	93C-1A17	Pear-5	Male; Age 10; French native; also some Spanish; parents are non-Basques; uses French most; poor speaker.
HEND-10	93C-1B01	Chaplin-5	Male; Age 10; Spanish native; some French; Basque learned at school; good speaker; uses all languages equally well.
HEND-11	93C-1B04	Chaplin-6	Male; Age 10; French native; Basque learned some at home and at school; average speaker in both languages; uses Basque at school and some with siblings; parents speak some Basque.
IKAS-01	93C-1B05	Pear-7	Female; Age 13; Basque native; uses Basque at home; also speaks Spanish; speaks Basque best; uses Basque most.
IKAS-02	93C-1B06	Chaplin-7	Female; Age 13; Spanish native; parents do not speak Basque; speaks Basque well; uses mostly Spanish, also at school.

IKAS-03	93C-1B07	Pear-8	Male; Age 13; native Basque and Spanish; parents are new Basques but speak mostly Basque at home; uses Spanish most, even at school; good speaker in both languages, perhaps better in Spanish.
IKAS-04	93C-1B08	Chaplin-8	Male; Age 13; Spanish native; Basque at school (age 3); parents do not speak Basque; speaks Spanish better.
IKAS-05	93C-1B09	Pear-9	Male; Age 13; Spanish and Basque native; mother is new Basque; father doesn't speak Basque; speaks Basque well, but Spanish better.
IKAS-06	93C-1B10	Chaplin-9	Female; Age 13; Basque native; also speaks Spanish; good Basque speaker; speaks Basque better. Uses Basque most.
IKAS-07	93C-1B11	Pear-10	Male; Age 13; Basque native; also speaks Spanish; speaks Basque much better; speaks mostly Basque; very good Basque speaker.
IKAS-08	93C-2A01	Chaplin-10	Female; Age 13; Basque native; also speaks Spanish; speaks Basque better; very good Basque speaker; uses mostly Basque.
IKAS-09	93C-2A02	Pear-11	Male; Age 13; Basque and Spanish native; mother speaks some Basque; speaks both languages equally well; average speaker.
IKAS-10	93C-2A03	Chaplin-11	Male; Age 13; Basque native; also Spanish native: with grandmother at home; speaks Basque much better though; very rich Basque; poor Spanish; uses mostly Basque.
IKAS-11	93C-2A04	Pear-12	Male; Age 13; Basque native; speaks some Spanish, not very well; very good Basque speaker; speaks mostly Basque.
IKAS-12	93C-2A05	Chaplin-12	Male; Age 13; Basque and Spanish native; mother Basque native; father Spanish native; speaks both well; average speaker.
DEUS-01	93C-2A06	Pear-13	Female; Age 22; Basque and Spanish native; from Oihartzun; uses mostly Basque.
DEUS-02	93C-2A07	Chaplin-13	Female; Age 22; Basque native; from Errenteria; learned Basque in the street; uses both languages equally.
DEUS-03	93C-2A08	Pear-14	Female; Age 22; Basque and Spanish native; from Eibar; uses both languages equally; Basque with father; Spanish with mother.
DEUS-04	93C-2A09	Chaplin-14	Female; Age 23; Basque and Spanish native; from Oihartzun; uses Basque at home.

DEUS-05	93C-2A10	Pear-15	Female; Age 22; Basque native; from Ataun; Spanish learned later; uses mostly Basque.
DEUS-06	93C-2B01, 93C-2B02	Chaplin-15	Female; Age 23; Basque native; from Azpeitia; Spanish learned at school; uses mostly Basque.
DEUS-07	93C-2B03	Pear-16	Female; Age 23; Basque native; from Eibar; Spanish learned at school, in the street; uses Basque more.
DEUS-08	93C-2B04	Chaplin-16	Female; Age 23; Basque native; from Galdakao; Spanish learned outside the home; uses Basque more than Spanish.
DEUS-09	93C-2B05	Pear-17	Male; Age 24; Basque native; from Bergara; Spanish learned in the street; uses mostly Basque.
DEUS-10	93C-2B06	Chaplin-17	Male; Age 23; Basque native; from Hernani; Spanish learned mostly outside the home; uses Basque more; uses mostly Spanish with his friends though.
DEUS-11	93C-2B07	Pear-18	Male; Age 22; Spanish native; from Donostia; learned some Basque at home and then at school; uses mostly Spanish.
DEUS-12	93C-2B08	Chaplin-18	Male; Age 22; Basque native; from Azpeitia; learned Spanish at school; uses Basque 70%; with friends half and half.
LUR-01	93C-3A07	Chaplin-19	Female; Age 79; Basque native; from Azkoitia; Spanish learned late (age 10); uses mostly Basque.
LUR-02	93C-3A08	Pear-19	Female; Age 54; Basque native; from Azkoitia; learned Basque as a child; now she uses both languages equally: Basque at home and with friends; Spanish at work and teaching.
LUR-03	93C-3A09	Pear-20	Female; Age 24; Basque native; from Zumarraga; learned Spanish as a child in the street and at Basque school; uses Basque 50%; uses Spanish with most friends.
LUR-04	93C-3A10	Chaplin-20	Male; Age 47; Basque native; from Hendaia; French and Spanish learned later. Uses Basque 70-75%.
LUR-05	93C-3A11	Pear-21	Female; Age 23; Basque native; from Urnieta; parents from Azpeitia; Spanish learned at (Basque) school; uses Basque 80%.
LUR-06	93C-3A12	Chaplin-21	Male; Age 30; Basque native; from Eibar; Spanish at school and in the street; uses 90% Basque.

Appendix 4.1: Clauses with inverted topic in the Spoken Basque Corpus

Code	#	Text	Info structure
1. 1A02	22	Orduan, esaten dio indartsuak <joateko= e= goiat>~, <i>Then the strong guy tells him to go up</i>	...A <i>esan</i>
2. 1A02	27	Gero esaten dute poliziek <<jateko>~ ordua dela>~. <i>Then the police says it's time for lunch</i>	...A <i>esan</i>
3. 2A10	11	oandik # <biltzen>~ seitzen dau gizon orrek, udarak, <i>he's still gathering that guy, the pears</i>	...A
4. 2B01	276	ati iriki die, zeldako atie, bi gizon da polizi oiei Txarlot orrek, .. <i>he opened the door, the cell door, for those guys that Chaplin guy</i>	...A <i>that X guy</i>
5. 2B01	40	ta - bere banderakin segitzen du tipuek, <i>and he keeps on going with his flag, that guy</i>	...A <i>tipua</i>
6. 2B04	109	arrapau itten dittula Txaplin onek- <i>that he caught them, that Chaplin guy</i>	...A <i>that X guy</i>
7. 3A08	25	eta - deittu itte xobe(?) mutikoeak, <i>and they call him, those boys</i>	...A
8. 3A10	188	eta- esaten du= ba Txarlotek ezetz, <i>and he says "no way", Chaplin</i>	...A <i>esan; clarif.?</i>
9. 1B03	38	eben pasatzen dira aurre-aurretik iruak, <i>uh they go right by the three of them</i>	...S <i>Focus extrap.</i>
10. 1B06	60	eta aserretzen da pilo bat [X] gizona. ez? <i>and he gets very angry that man, right?</i>	...S <i>Focus extrap.</i>
11. 2A02	80	Ta orrela juten dira mutilak aurrerago, - <i>and thus continue the boys further</i>	...S
12. 2A07	124	Neskana juten da polizia, <i>the police goes to where the girl is</i>	...S
13. 2A09	333	Ta geo pixkat geixio kar- jarraitzen du pelikulak, <i>and then the movie continues a bit longer</i>	...S
14. 2B01	110	an indartu de Txarlot, <i>Chaplin got real strong</i>	...S
15. 2B01	309	jun de tipu une, <i>he went that guy</i>	...S <i>tipua</i>
16. 2B01	51	<"Punto-Kruz" eitten> o ai de - tipua. <i>He's sewing (?) or something, that guy</i>	...S <i>tipua</i>
17. 3A10	106	baino seko mareauta gelditzen da tipoa. ez? <i>but he gets all dizzy that guy, right?</i>	...S
18. 1B06	246	eta an sartzen dute neska. <i>and that's where they put the girl</i>	...O <i>object!</i>
19. 1B11	17	Bere atzean- gizonaren atzean dagolako umea. - <i>because the kid is behind him behind the man</i>	...Cop_Ident_Loc
20. 2A03	10+	~<or dagoela bandera> <i>that there is the flag</i>	...Cop_Ident_Loc <i>complementizer</i>
21. 2B01	304	=ta an zeon tipu erriez, <i>and there was the guy laughing!</i>	...Cop_Ident_Loc

22. 3A09	5	eta= an dau bea. <i>and there he is</i>	...Cop_Ident_Loc
23. 1A09	84	eiek direla= - ba- lapurtu diyotenek. <i>they are the ones who stole them from him</i>	...Cop_Ident_Ref complementizer
24. 1A10	158+	~<gizona dela <ogia= - e lapurtu dutena>~>> <i>the man is the one who stole the bread</i>	...Cop_Ident_Ref complementizer
25. 1B04	170	ebe neska zela= <artu- artu- lapurtu zuena= - ogia>~, <i>uh that the girl was the one who took- stole it - the bread</i>	...Cop_Ident_Ref complementizer
26. 1B06	18+	~<bera dela buruzagia> <i>that he is the leader</i>	...Cop_Ident_Ref complementizer
27. 1B06	193	*zu izan zara lapurra ez dakit zer?* <i>you were the thief and so on</i>	...Cop_Ident_Ref
28. 1B06	200b	~ *ni, ni izan naiz la- lapurra, ??? <i>I, I was the thief</i>	...Cop_Ident_Ref
29. 1B06	28+	~<Txarlot dela buruzagia> <i>Chaplin is the leader</i>	...Cop_Ident_Ref complementizer
30. 1B10	13	bueno # ori da protagonista, <i>well that is the protagonist</i>	...Cop_Ident_Ref
31. 2A01	15+	~<bera dela liderra> <i>that he is the leader</i>	...Cop_Ident_Ref complementizer
32. 2A03	60+	~<*ori da <daukana droga>~*> <i>that's the one who has drugs</i>	...Cop_Ident_Ref
33. 2A07	22+	~<bera zala .. manifestazioko buruzagia> <i>that he was the leader of the demonstration</i>	...Cop_Ident_Ref complementizer
34. 2A09	130+	~<bera dela best- e= - gaixtoa, > <i>that he is the oth- uh the bad guy</i>	...Cop_Ident_Ref complementizer
35. 2A09	287	Bera izan dela errudun guztiya, <i>that he is the only culprit</i>	...Cop_Ident_Ref complementizer
36. 2A09	68+	~<bera zela denen buru> <i>that he is the head of all of them</i>	...Cop_Ident_Ref complementizer
37. 2B01	382	ta neski dala= <e(a)man berrekue>. <i>and that the girl is the one they have to take away</i>	...Cop_Ident_Ref complementizer
38. 2B05	90	=Ori da - gutxi-gora-beera .. pelikulia. - <i>that's it pretty much the movie</i>	...Cop_Ident_Ref
39. 2B06	214	lapurra da oi!* <i>he's a thief, that one</i>	...Cop_Pred_Ref
40. 2B06	230	Ne- e- ni izan naiz lapurra*. - <i>... I was the thief</i>	...Cop_Ident_Ref
41. 3A07	141	neska ura izen dala ogixe lapurtu duna. <i>that girl was the one who stole the bread</i>	...Cop_Ident_Ref complementizer
42. 3A10	190	ta bera izan dela lapurra, <i>and that he was the thief</i>	...Cop_Ident_Ref complementizer
43. 3A10	32+	~<bera dela buru=> <i>that he was the head</i>	...Cop_Ident_Ref complementizer
44. 3A12	118	bera izan dela lapurra. - <i>that he was the thief</i>	...Cop_Ident_Ref complementizer
45. 3A12	24+	~<ba bera dela ori ba= manifestazio buruan dijoana> <i>well that he was uh the one at the head of the demonstration</i>	...Cop_Ident_Ref complementizer

46. 3A12	30+	~<bera dela .. ba manifestazio orren ba li- - liderra edo ola> <i>that he was well the leader of that demonstration or something</i>	...Cop_Ident_Ref <i>complementizer</i>
47. 2A09	298+	~<no- nola tx- neska izandu zen ba= - gaizkillea> <i>since the girl was the culprit</i>	...Cop_Ident_Ref <i>nola "since"</i>
48. 2B01	483	Ta ya nola= .. dana beteta zeuen kamiona, <i>and already since the truck was completely full</i>	...Cop_Pred <i>nola "since"</i>

Appendix 4.2: Clauses with a postverbal overt subject in the Written Basque Corpus
 It doesn't include 48 post-quote subject inversions or non-copula antitopic constructions
 (cf. Appendix 4.3)

No.	Book	Text	Type
1	<i>Behi</i> , 21	gauza bat da ikara	Copula
2	<i>Behi</i> , 27	nire nireak ziren bere ahoko ile haiek.	Copula
3	<i>Behi</i> , 13	* Oso mihi dotore eta leunekoa da nire barruko hori,	Copula
4	<i>Behi</i> , 26	"Hori da <pentsatzea>,"	Copula
5	<i>Behi</i> , 25	<koxkorrez beteta> zegoen harkaitz beltza;	Copula
6	<i>Behi</i> , 24	posible zen <beherago elurrik ez egotea>.	Copula
7	<i>Kuba</i> , 20	* Bel-beltza da Toni:	Copula
8	<i>Behi</i> , 23	"Hori da marrua <egin duzuna>,"	Copula
9	<i>Kuba</i> , 19	Debekatua dago <oinez edo bizikletaz zeharkatzea>	Copula
10	<i>Behi</i> , 21	izan dira <beraiek baino iheslari nagusiagoak>.	Copula
11	<i>Behi</i> , 19	Huraxe zen <jakin nahi nuena>,"	Copula
12	<i>Behi</i> , 18	Ni nintzen behi hori.	Copula
13	<i>Behi</i> , 18	Eta hantxe nengoen ni,	Copula
14	<i>Behi</i> , 16	holakoxea zara zu!" *	Copula
15	<i>Behi</i> , 14	Zima horretakoa naiz ni:	Copula
16	<i>Behi</i> , 24	<otsoz beteta> daudela mendiak. *	Copula
17	<i>Kuba</i> , 29	Baina garbi dago <ez naizela beren modukoa>.	Copula
18	<i>Kuba</i> , 35	Deigarria da <zenbat emakume ilehori dagoen>,"	Copula
19	<i>Kuba</i> , 35	Oso gazteak eta politikak dira gehienak.	Copula
20	<i>Kuba</i> , 32	* "Hor daude Txernobil-go umeak."	Copula
21	<i>Kuba</i> , 31	Ezagun da <Tonik lagun duela taxista>.	Copula
22	<i>Kuba</i> , 30	alproja samarra da Toni hau, eh"	Copula
23	<i>Kuba</i> , 30	agerian dago <Tonik ohetik atera duela>.	Copula
24	<i>Kuba</i> , 21	nire bizikletaz dira beren galderak:	Copula
25	<i>Kuba</i> , 26	Familia normalak dira <jaten ematen dutenak>,"	Copula
26	<i>Kuba</i> , 26	Oso gozoa dago dena.	Copula
27	<i>Kuba</i> , 26	erdian <<zintzilik dagoen> bonbila motela eta ategaineko leihatila estua> dira argi-iturri bakarrak. *	Copula
28	<i>Kuba</i> , 24	ezagun da <garagardoak bere bidea egin duela>. *	Copula
29	<i>Kuba</i> , 23	baina hor dago bizibidea"	Copula
30	<i>Kuba</i> , 23	baina bidaietan atseginagoa izaten da <entzutea> <<eztabaidan aritzea> baino>. *	Copula
31	<i>Kuba</i> , 22	"Duela gutxi izan dira hauteskunde libreak,	Copula
32	<i>Behi</i> , 28	Hori baldin bada kontua,	Copula
33	<i>Behi</i> , 21	<Ihes eginda> daudela denak,	Copula
34	<i>Behi</i> , 15	<zela <bizitzea>>	Copula
35	<i>Kuba</i> , 36	<putzuz beteta> baitago dena;	Copula
36	<i>Behi</i> , 21	zerbait bazela <gaikalde hartan batere ikararik gabe egoten jakitea>.	Copula

37	<i>Behi</i> , 16	<Aingeru Goardakoa omen zen> hari Setatsua deituko niola nik.	Overt_Topic_I
38	<i>Behi</i> , 27	eta <isatseko azken ileetatik> tiraka hasi zitzaidan otsoetako bat.	Covert_Topic_I
39	<i>Behi</i> , 14	"Poza eman beharko lizuke <<bera zure barruan dagoela> jakiteak>.	Covert_Topic_I
40	<i>Behi</i> , 16	Gogora datorkit, zentzu honetan <nire bizitzako aurreneko neguan gertutakoa>.	Covert_Topic_I
41	<i>Behi</i> , 21	eta luze, oso luze, oso oso luze iruditzen zitzaidan denbora.	Covert_Topic_I
42	<i>Behi</i> , 25	astindu batez mugitu zitzaidan isatsa,	Covert_Topic_I
43	<i>Kuba</i> , 19	Asko gustatzen omen zaio <turistekin jardutea>;	Covert_Topic_I
44	<i>Kuba</i> , 24	etxera datorkizu polizia bila.	Covert_Topic_I
45	<i>Kuba</i> , 22	kosta egiten zait <zer esaten duten aditza>,	Covert_Topic_I
46	<i>Kuba</i> , 27	<besoan duen> <sendatutako> zauri handi bat erakutsi dit Mirthak.	Covert_Topic_I
47	<i>Kuba</i> , 26	* Erreparu puxkat eman dit <ate-burua zeharkatzeak>,	Covert_Topic_I
48	<i>Kuba</i> , 37	<Kostata> etsi dio andreak. *	Covert_Topic_I
49	<i>Kuba</i> , 33	ezustean harrapatu du galderak;	Covert_Topic_O
50	<i>Kuba</i> , 32	Harritu egin nau honek,	Covert_Topic_O
51	<i>Kuba</i> , 27	aulki txiki-txiki batean eseri nau Mirthak paretaren kontra.	Covert_Topic_O
52	<i>Behi</i> , 30	iritsi da garaia! *	No_Topic??
53	<i>Behi</i> , 20	udazkenaren epeltasun hartan ibili ziren xatorrak han eta hemen	No_Topic??
54	<i>Behi</i> , 30	Heldu da ordua,	No_Topic??

Appendix 4.3: Non-copula antitopic construction clauses
in the Written Basque Corpus with semi-literal English translation
Information Structure = *Invert_Topic_A* or *Invert_Topic_S*.

No.	Story, Pg.	Text with semi-literal translation	Order
1	<i>Behi</i> , 14	holáxe hitzegiten du nire barruko ahotsak , <i>Just like that spoke my inner voice</i>	<u>XVA</u>
2	<i>Behi</i> , 14	oso aspégarri egiten dira bere kontu gehienak , <i>Very boring become (are made) most stories</i>	<u>XVS</u>
3	<i>Behi</i> , 26	arbóladi bateraino iritsi ginen denok . <i>Up to a tree grove arrived us all</i>	<u>XVS</u>
4	<i>Behi</i> , 26	uhúrika hasten zen bat , <i>Howling began one</i>	<u>XVS</u>
5	<i>Behi</i> , 26	uhúrika hasten ziren denak . <i>Howling began all</i>	<u>XVS</u>
6	<i>Behi</i> , 26	halaxe esan zion otso batek besteari: <i>just like that said one wolf to another</i>	<u>XVAI</u>
7	<i>Behi</i> , 27	Hántxe joan zen gizarajoa <intzirika>. <i>Right there left the poor guy screaming</i>	<u>XV SX</u>
8	<i>Behi</i> , 29	Géra dadila zure testigantza! (desiderative) <i>May your testimony remain</i>	VS
9	<i>Behi</i> , 30	Ikás dezala munduak behiaren haunditasuna! (desiderative) <i>May the world learn the greatness of cows</i>	VAO
10	<i>Kuba</i> , 22	Sántu aurpegia du Giovanik; <i>Saint face has Giovanni</i>	<u>OVA</u>
11	<i>Kuba</i> , 27	atéa jo du Tonik. <i>The door knocked Toni</i>	<u>OVA</u>
12	<i>Kuba</i> , 27	¶ Ahal den lekúan eseri dira Toni eta Giovanni, <i>Wherever they could sat Toni and Giovanni</i>	<u>XVS</u>
13	<i>Kuba</i> , 28	Bárre egin dugu denok. <i>Laughter made all of us (= We all laughed)</i>	<u>OVA</u>
14	<i>Kuba</i> , 28	hiru péso ordaindu du gizonak <i>Three pesos paid the man</i>	<u>OVA</u>
15	<i>Kuba</i> , 29	Larruzko beso-poltsa edér bat atera du Mirthak, <i>A beautiful leather bag brought out Mirtha</i>	<u>OVA</u>
16	<i>Kuba</i> , 31	solásean doaz biak; <i>Conversing go the two of them</i>	<u>XVS</u>
17	<i>Kuba</i> , 32	beréhala antzematen da <Kubak bizi duen> trajedietako bat: <i>Right away is noticeable a tragedy that Cuba lives</i>	<u>XVS</u>
18	<i>Kuba</i> , 33	zéra diote biek aho batez: <i>Thus say both of them in one voice</i>	<u>OVAX</u>
19	<i>Kuba</i> , 35	¶ Bínaka edo hirúnaka doaz neskok <i>In pairs or groups of three go these girls</i>	<u>XVS</u>

Appendix 4.4: Article published in *Diario Vasco*, 9/27/96, Internet edition
It contains a large number of postverbal subjects
Antitopics and possible antitopics have been underlined

El Diario Vasco, Internet Edition, Sociedad; Viernes, 27 de Setiembre de 1996

J. Ignazio Iztuetaren ‘Vocabulario de vascuence’ atera da lehen aldiz

FELIX IBARGUTXI/DV.DONOSTIA

Juan Ignazio Iztuetaren *Vocabulario de vascuence* argitaratu du Iztueta Fundazioak. 1832.eko eskuizkribu hau argitara eman gabe zegoen oraindik. Edizioa ez da faksimila.

Atzo egin zen liburuaren aurkezpena, Donostiako KMn. Hain zuzen ere, hantxe gordetzen da eskuizkribua, Fondo Julio de Urquijo-ko parte baita.

Iztueta Zaldibian jaio zen 1767an, eta 1845ean hil. *Guipuzcoaco dantza gogoangarrien condaira* (1824) eta *Guipuzcoaco provinciaren condaira* (1847) eman zituen argitara.

Oraingo hiztegi honen prestatzailea Patri Urkizu, UNEDeko irakaslea da. Bere hitzurrean haxe dio: «Nahi eta nahiezkoa dugu gure aitzindarien arreta eta pazientzia handiz bildutako altxorrek kontutan edukitzea. Eta Iztuetarena da horietakoa, ez baitira ahanzteko Mitxelenak idatzitako hitzok: *Así Iztueta, a pesar de su guipuzcoanismo, fue asiduo recolector de voces y frases vizcaínas—de ello da fe en su diccionario manuscrito aún inédito—y no desconocía el labortano*».

Aurkezpeneko beste pasarte batean gogotik goraiatu zuen Urkizuk KMn dagoen Julio de Urquijoren liburu-bilduma: «Badaramazkit 30 urte baino gehiago bertako altxorraz baliatzen eta aitortu beharra daukat liburu harrobi hori ez zaidala inoiz agortzen».

Irakasle lezotar honen iritziz, «Iztuetaren *Condaira*-k ere merezi du argitalpen berri bat, ohar eta zuzenketa beharrezkoen hornitua. Gaur egun ere gustoz irakurtzen den liburua baita».

Luis Mari Mujikak hiztegiaren azalpena egin zuen atzoko aurkezpenean. «Aski kritikatu izan da Iztueta Larramendiren zale izateagatik, eta Bonaparte printzeak ere ez zuen gogoko haren euskara, baina hala eta guztiz ere balio handia du hiztegiak. Gainera, beti da ona argitara eman gabe zegoen zerbait argitaratzea».

Mujikak komentatu zuenez, inguru-mingurutan galtzen da batzutan Iztueta, euskal hitz zehatza aurkitu eta eskaini beharrean.

Bestalde, Pello Esnaletan gutun bat irakurri zuen atzo Emeterio Sorazu, Iztueta Fundazioaren idazkariak. Bilduma «aurrerakoitzat» jo behar omen da hiztegi hau. Esnaletan iritziz, hizkuntzari etorkizuna ikusten ziolako egin zuen Iztuetak lan hau. Txinako esaera batekin bukatzen zuen Esnalet eskutitza: «Nola jakingo duzu non zauden ez badakizu nondik datozen? Eta nola arraio jakingo duzu nora zoazen ez badakizu non zauden?»

Appendix 5.1: Headlines from *La Vanguardia*, May 3, 1996

No.	Text
1	La retirada militar de Hebrón polariza la campaña electoral israelí
2	El asesinato de Rabin apela al Tribunal Supremo
3	El descenso de la violencia marca la segunda fase de las elecciones legislativas en India
4	Incidentes y detenciones en la primera jornada de huelga general prevista en Paraguay
5	La regionalización del país enfrenta al Gobierno y a la oposición portugueses
6	Yandarbiyev podría dialogar con Moscú "bajo condiciones"
7	Tudjman impone un comisario gubernamental como alcalde de Zagreb
8	Los "torios" pierden más de la mitad de los cargos municipales
9	La negociación sobre la nueva Constitución de Sudáfrica se halla a un paso del fracaso
10	Fraga retira el recurso contra la cesión del IRPF como "regalo" a la investidura de Aznar
11	El Consell Executiu eleva al quinto hijo de Pujol a la categoría de director general
12	Los pactos con los nacionalistas y la política social centrarán el debate
13	Un Dos de Mayo en tono patriótico
14	Molins presentará a CiU como garante de la continuidad del Estado del bienestar
15	González dedicó el último Consejo a preparar el debate
16	Anguita preguntará a Aznar si practicará la misma política "de derechas" que González
17	El presidente rectifica y anuncia que irá a la reelección
18	El nuevo papel de Aznar
19	Aznar tranquiliza a Europa con Matutes y Serra en el gobierno
20	"Estáte localizable todo el puente"
21	Relacionado con el entorno familiar
22	Una carrera meteórica
23	HB amenaza a los funcionarios de prisiones
24	Ortega Lara, "trasladado" a la cárcel de Soria
25	La Generalitat concede 42 Creus de Sant Jordi
26	Francia rebaja de 1.000 a 300 m2 la exigencia de autorización previa para abrir un comercio
27	La historia en cifras
28	Un estudiante español sufre una brutal agresión de una pandilla juvenil en el sur de Inglaterra
29	La condena del caso Bulger, cuestionada
30	Unas avionetas rocían por error con pesticida un colegio en Almería
31	Policías y gitanos buscan en Figueres al presunto autor de la muerte de la joven Andrea
32	"Hospíderman" cae por fin
33	La banca suiza acepta discutir la devolución del dinero de los judíos víctimas del nazismo
34	Marimon rechaza ceder competencias en gestión de la naturaleza a Medi Ambient
35	La juez pide 2.922 millones de fianza a Erkimia por contaminar el Ebro
36	Las rebajas podrán empezar el 1 de julio
37	La justicia francesa condena al Estado y autoriza a llevar velo en clase a las musulmanas
38	Entre arrullos
39	El PSC propone ahora elevar el listón ambiental para la tercera pista de El Prat
40	Desorganización y caos en la venta de entradas para Bruce Springsteen

- 41 Más vale prever...
- 42 La Diputación de Barcelona creará una nueva red de teatros municipales
- 43 El embrión de una futura compañía nacional de danza
- 44 La actriz María Luisa Ponte muere a los 77 años en Aranjuez
- 45 En el alambre
- 46 Una catalana en Oregón
- 47 Lombardero sigue sin poder rodar en Teruel
- 48 Chapertons y su original mundo de neumáticos, en el Villarroel
- 49 "No se han recuperado por completo los progresos que trajo la Segunda República"
- 50 Escritor de periódico
- 51 La Virreina exhibe una retrospectiva del artista pop americano Tom Wesselmann
- 52 Roma acoge el realismo de Antonio López y sus compañeros de generación
- 53 "Los intelectuales, como todo el mundo, se equivocan a veces", dice Haro Tecglen
- 54 El agujero de la Seguridad Social incrementa en 479.000 millones el déficit de este año
- 55 El gasto en pensiones se dispara
- 56 Opel España acepta una subida salarial del 4,5%
- 57 Juppé prepara una reducción del gasto público de 1,5 billones de pesetas para 1997
- 58 EE.UU. acelera su crecimiento al 2,8% en el primer trimestre
- 59 Renault estudia la instalación de un centro de diseño en Barcelona
- 60 Las cavas medianas plantan cara

Appendix 5.2: Headlines from *El Periódico*, May 3, 1996

No.	Text
1	La 'fiebre Springsteen' crea un gran caos en el Eixample
2	Bramidos contra el abuso
3	Empieza la matanza de las 'vacas locas'
4	Estado del bienestar y autonomías
5	Descalabro 'tory' en las municipales
6	El polvorín de Kosovo
7	La ONU roza un acuerdo sobre las minas antipersona
8	La violencia marca la segunda fase de los comicios indios
9	Aznar inicia su gran prueba de fuego
10	Fraga allana la investidura a Aznar
11	Un brindis y una foto
12	Josep Antoni DURAN LLEIDA : "Los políticos debemos apartarnos definitivamente del caso GAL"
13	HB amenaza a los funcionarios de las cárceles
14	Trias nombra director general a un hijo de Pujol
15	El crimen infantil angustia Gran Bretaña
16	El caos retrasa el inicio de la matanza de reses británicas
17	Phillip TOBIAS : "Nuestros antepasados pervivieron porque aprendieron a hacer fuego"
18	Aparece un tercer sospechoso del crimen de Alcàsser
19	Mazazo policial a los tejanos piratas
20	Dos sentencias consideran ilegal la junta del Coegi d'Infermeria
21	L'Hospitalet gana la batalla de la Gran Via
22	El fiscal pide 14 años por un homicidio en Gràcia
23	El PSC busca un gobierno estable para Badalona
24	Charlie Sheen: "Resulta más fácil relacionarse con una prostituta"
25	"Qué disgusto ver lo que soy ahora, aburrido, tonto, viejo, mala persona"
26	Arranca el boicot contra Francia
27	Un grupo francés compra la cadena de 'supers' Supeco
28	CCOO y UGT prevén reunirse con Aznar la próxima semana
29	Jornada irregular y puente en la Bolsa de Madrid
30	Francia acometerá un plan de reducción del gasto público en 1997
31	Capello quiere crear un 'dream team'
32	Prosinecki desea también cumplir su contrato
33	Camacho se tomará un año sabático si el Sevilla baja
34	Acuerdo de mínimos de nueve clubs
35	El Barça pone en juego su presente y parte del futuro
36	Sabonis cautiva a la NBA con su juego ante Utah
37	Stuart BAIRD : "Me gusta que me corra la adrenalina por el cuerpo, es como una adicción"
38	María Luisa Ponte fallece a los 77 años en una residencia de ancianos
39	El obispo de Teruel reitera su negativa a rodar en una ermita
40	Chapertons estrena un montaje de humor protagonizado por neumáticos
41	Jamie Walters: "La TV es un arma de doble filo"
42	TVE-1, Antena 3 TV y Canal 33 se vuelcan en la investidura de Aznar
43	Canal 33 recupera informativo propio tres años después Valmont Leolo

Appendix 5.3: Headlines from *La Jornada*, May 3, 1996

No.	Text
1	Pugna de agrupaciones por la nueva cultura laboral
2	No miren al pasado: Zedillo a sindicatos
3	Elogia el Presidente a organismos que han evitado cerrar empresas
4	Fidel Velázquez: “ya tenemos todo convenido con los patrones”
5	Renovación sindical es luchar y presentar alternativas: foristas
6	Hay para 48 años
7	En un año bajaron en 3 mil millones de barriles las reservas de crudo
8	Investigará la CIDH denuncias de abusos contra mexicanos en la frontera
9	Aprueba el Senado de EU proyecto antimigratorio
10	Advierte Clinton que no lo sancionará si no revisan restricciones a beneficios sociales
11	Delegados de la Comisión observarán in situ los flujos migratorios, a petición mexicana
12	Entzin, pena de 6 años
13	A Elorriaga, 13 años de cárcel por terrorismo y conspiración
14	Paredes: candidatos cosméticos o renovar alianzas populares, disyuntiva priísta
15	Jorge Mendoza
16	La respuesta de la tele
17	¿Y dónde queda el pasado? ¿En la explanada del CT o en el Zócalo?
18	Decomiso 5,600 kilos de coca en la operación Zorro II
19	Anuncia EU la captura de 130 presuntos narcos
20	Pesquisa de 7 meses contra grupos de Colombia y México
21	Detienen en la vía a Puebla a 3 personas con 604.3 kilos de cocaína
22	Casi 13 mil acciones antinarcóticas en México en 1995: PGR
23	Niños de la UCIC de Iztapalapa lanzan piedras al Departamento del Distrito Federal, en protesta porque no se les dieron los juguetes prometidos por el
24	El STC lo niega
25	Asambleístas: en el Metro, fallas de seguridad y mantenimiento
26	Las críticas en EU enfrían las intenciones de ampliarlo a otros países, señalan
27	La crisis mexicana descarriló el TLC: expertos
28	El tema del comercio, congelado en la época preelectoral, dicen en un seminario ante periodistas
29	Hallan sin vida a los 5 alpinistas perdidos en el Popocatepetl
30	En estudio, reducir a 2 años el periodo del presidente de la Suprema Corte
31	También, destituir al presidente del Consejo de la Judicatura en caso de “perdida de la confianza”
32	Abren 3 nuevas averiguaciones por daño patrimonial a Ruta 100
33	Acuerda asamblea indígena proponer iniciativas sobre regiones autónomas
34	Urge nueva base teórica para la izquierda: Bovero

Appendix 5.4: Headlines from *Euskaldunon Egunkaria*, May 3, 1996

No.	Text
1	Ertzaintzaren aurkako eraso
2	Agenteek zenbaki fiktizioak erabili dituztelako eten egin da epaiketa
3	Fiskala eta akusazio partikularra ere bat etorri ziren defentsaren eskariarekin
4	Intsumiso presoan hotz greba
5	Mendi magalean
6	Eroskik Sebastian de la Fuente katea erosteko asmoa du
7	Kartel bat dela-eta Ertzaintzak atxilotutako Roberto Sampedro, HBko komunikazio arduraduna, aske utzi du epaileak
8	Josu Urrutikoetxeak gaur beteko du Frantziak ezarritako espetxe zigorra
9	Herri Urrats: Ipar Euskal Herriko ikastolen festa hilaren 12an izango da Senpereko lakuaren inguruan
10	Gaur hasi eta bihar amaituko da Aznarren inbestidura saioa Espainiako Legebiltzarrean
11	Futbola: Espainiako lehen mailako hainbat taldeko presidenteez telebistei egin beharreko eskakizunak biribildu dituzte Bilbon
12	Jorge Oteizaren arkitektura-lanak biltzen dituen katalogo-liburua aurkeztu du Iruñan Pamiela argitaletxeak

Appendix 5.5: Sentences used for eliciting subject inversion

	Foreign version	Basque version (First choice by IHS)
1	Esta soplando el viento.	Haizea txistuka ari da
2	Alguien sabe que va a venir el jefe.	Norbaitek badaki nagusia/jefea etorriko dela
3	A: ¿Cuándo viene tu madre? B: Va a venir dentro de un par de semanas.	A: Noiz dator zure ama B: Bi aste barru etorriko da
4	Va a haber problemas.	Arazoak izango dira/Arazoak egongo dira
5	A: ¿Qué comiste? B: Comí la carne, pero no comí el pescado.	A: zer jan zenuen? B: Haragia jan nuen, arraia, ordea, ez nuen jan
6	Se abrió la puerta.	Atea ireki zen/atea ireki egin zen
7	¿Hay alguna manzana madura?	Ba al da sagar heldurik?/Ba al dago sagar heldurik?/Ba al dago heldutako sagarrik?
8	Había amargura en sus palabras.	Bere hitzetan * nabari zitekeen
9	¿Cuántos años tiene Juan?	Zenbat urte ditu Jonek?
10	Ha aparecido mi libro.	Azaldu da nere liburua
11	A: ¿Qué hay de nuevo? B: Que viene mi hermano hoy.	A: Berririk bai? B: Gaur nere anaia datorrela.
12	Fue una hamburguesa lo que yo comí.	Hamburgesa bat izan zen nik jan nuena
13	Vi un libro en la mesa pero no vi un periódico.	Liburu bat ikusi nuen, egunkaririk ez nuen ikusi, ordea.
14	Me aterra el jefe.	Nagusiak beldurra ematen dit/ nagusiaren beldur naiz
15	Me sangra la nariz.	Odola daukat sudurretik/ Sudurra odola dariola daukat (guk etxean: sudurre odola daixola dauket)/ Sudurretik odola dariola nago / Sudurrean odola daukat
16	Se desató la revolución.	Iraultza hasi zen
17	Me duele el cuello.	Lepoan mina dut/Lepoko mina dut/ Lepoko mina daukat
18	¡La cuenta la pago YO!	Kontua NIK ordainduko dut
19	Le entró pánico.	Izutu egin zen/Beldurtu egin zen
20	¡Está equivocado el jefe!	Nagusia oker dago!
21	Mira, estan sonando las campanas.	Begira kanpaiak joka ari dira
22	Dios existe.	Jaungoikoa bada.
23	Ha explotado una bomba.	Bonba bat lehertu da.
24	Hacen falta nuevos métodos.	Metodo berriak behar dira/metodo berriak beharrezkoak dira
25	Jon comió una hamburguesa	Jonek hanburgesa bat jan zuen
26	Caían gotas del techo.	Zapaitik tantak erortzen ari ziren
27	Se ha fundido el fusible.	Fusiblea erre da
28	No habrá ningún problema.	Ez da arazorik izango
29	No hay ningún perro.	Ez da txakurrik/ ez dago txakurrik
30	El jefe dijo: “Jamás cederé”	Nagusiak esan zuen: “ez dut inoiz amore emango”

31	A: ¿Cuándo viene tu madre? B: Dentro de un par de semanas va a venir.	A: Noiz dator zure ama? B: Bi aste barru etorriko da
32	También Jon hizo las tareas.	Jonek ere egin zituen zereginak
33	Está ladrando un perro.	Zakur bat zaunkaka ari da.
34	El libro ha aparecido.	Azaldu da liburua
35	Ha llamado un hombre.	Gizon batek deitu du
36	Está lloviendo.	Euria ari du
37	¡Ya lo sabe el jefe!	Badaki nagusiak!
38	Existen las hadas.	Badira laminak
39	Tuvo lugar una gran fiesta.	Jai handia izan zen
40	Era medianoche.	Gauerdia zen
41	Hace mucho frío.	Hotz handia egiten du /Hotz handia dago
42	Los ovnis existen. Estoy segura.	Badaude OVNIak. Zihur nago.
43	¡Que pague el jefe!	Ordain dezala nagusiak!
44	Me picó un mosquito.	Moskito bat heldu zidan
45	Se ha parado el reloj.	Erlojua gelditu egin da
46	Esta puerta chirría.	Ate honek zarata ateratzen du
47	Ha llamado tu madre.	Zure amak deitu du
48	Esta puerta hace mucho ruido.	Ate honek zarata handia ateratzen du
49	Me pinché con una espina.	Arantza batekin
50	La puerta se abrió y entró el jefe.	Atea zabaldu eta nagusia sartu zen
51	Cuidado que viene el jefe.	Kontuz, nagusia dator eta.
52	Se ha quemado la tostada.	Tostada erre egin da
53	¿Que quién pagó la cuenta? Mi hermano pags la cuenta.	Kontua nork ordainduko duen? Nere anaiak ordainduko du kontua.
54	Se rompis el jarrón.	Jarroia hautsi zen
55	¡Que si habra problemas!	Izango direla arazoak!
56	Les siguió molestando el perro.	Zakurrak molestatzen jarraitu zuen.
57	Brillaba el sol cuando llegamos.	Eguzkia distiraka ari zen iritsi ginenean
58	Está sonando el teléfono.	Telefonoa joka ari da.
59	Reina un silencio total.	Isiltasun osoa da
60	Yo comí una hamburguesa.	Nik hanburgesa bat jan nuen
61	¡Pago YO!	Nik ordaintzen dut!
62	Dejó de sonar el telefono.	Telefonoak jotzeari utzi zion
63	A: ¿Quién paga? B: ¡El jefe paga!	A: Nork ordaintzen du? B: Nagusiak ordaintzen du.
64	Va a venir el jefe en persona.	Nagusia BERA etorriko da
65	Alguien no sabe que va a venir el jefe.	Bada nagusia datorrela ez dakienik
66	A: ¿Por qué no fuiste a la cena el otro día? B: Sí que fui a la cena.	Zergatik ez zinen joan lehengo eguneko afarira? Ni afarira joan nintzen eta. / Ni BANINTZEN joan afarira (como llevando la contraria)
67	A: ¿Cuándo viene tu madre? B: ¡Hoy viene!	A: Noiz dator zure ama? B: Gaur dator!
68	No sabe nadie que va a venir el jefe.	Inorrek ez daki gaur nagusia datorrela
69	Jon hizo también las tareas, además de lo otro.	Jonek besteaz gainera etxeko lanak ere egin zituen

70	A: ¿Qué ha pasado? B: Se ha ahogado un perro.	A: Zer jazo da? B: Zakur bat ito da.
71	Se abrió la puerta y apareció el teniente.	Atea ireki eta Tenientea azaldu zen
72	A: ¿Cuándo viene tu madre? B: Viene hoy.	Noiz dator zure ama? Gaur dator.
73	A: ¿Pero qué te pasa? B: Me duele el cuello.	Zer gertatzen da baina? Lepoko mina dut
74	Venían unos niños por la calle.	Ume batzuk zetozen kaletik barrena.
75	A: ¿Cómo acabó la partida de cartas? B: Ganó Jon.	Nola bukatu zen karta jokoa? Jonek irabazi zuen.
76	A: ¿Has visto algún árbol con hojas? B: No, las hojas salen en la primavera.	Ikusi al duzu zuhaitzik ostoeekin? Ez, hostoak udaberrian ateratzen dira
77	Se ha perdido un niño de la guardería.	Haur bat galdu da haurtzaindegian
78	En Japón hay tres tipos de hoteles.	Japonian hiru eratako hotelak daude
79	“Violan la constitución los partidos de la coalición, según sus rivales”.	Arerioen arabera koalizioko alderdiek konstituzioa erasotu dute
80	“Jamás cederemos” dijo el Presidente del Gobierno	“Ez dugu amore emango” esan zuen Gobernuko Presidenteak
81	Física le enseñó una profesora del colegio	Fisika ikastetxeko irakasle batek irakatsi zion (eta matematika bere aitak)
82	A las cinco sonará el despertador.	Iratzargailuak bostetan joko du
83	[Estuve en casa todo el día y] no paró de sonar el teléfono.	[Egun osoa etxean egin nuen eta] telefonoak ez zion jotzeari utzi
84	A: ¿Pero qué te pasa? B: Tuve un accidente.	Baina zer gertatzen zaizu? Istripu bat izan nuen
85	Han aparecido nuevos métodos de diagnosticar esa enfermedad.	Eritasun hori atzemateko metodo berriak azaldu dira
86	El jefe también va a venir.	Nagusiak ere etorri behar du
87	También a mi hermano le dieron un premio.	Nere anaiari ere sari bat eman zioten
88	A: ¿Quién paga? B: Paga el jefe.	Nork ordaintzen/ko du? Nagusiak ordaintzen/ko du
89	Se estancó el asunto en el juzgado.	Arazoa epaitegian ez aurrera ez atzera gelditu zen
90	No he encontrado ningún libro.	Ez dut aurkitu libururik
91	Alguien ya sabe que va a venir el jefe.	Norbaitek badaki nagusia etorriko dela
92	A: Me tenían que dar el premio a mí porque yo llegué primero. B: Sí, pero mi hermano pagó la cuenta.	A: Saria niri eman behar zidaten ni iritsi baintzen lehenengo; B: Bai, baina nire anaiak ordaindu zuen kontua
93	Va a venir el jefe en persona.	Nagusia bera etorriko da
94	La pareja tuvo una pelea esa noche. Al día siguiente se levantó el hombre y se fue a trabajar como si nada.	Bikoteak greska izan zuen gau hartan. Hurrengo egunean gizonezkoa altxa eta lanera joan zen ezer jazo ez balitz bezala.
95	Había tres protagonistas principales en esa película.	Filme horretan hiru protagonista nagusi zeuden
96	A menudo aparecen artículos como ese en el periódico.	Hori bezalako artikulua sarritan ateratzen dira egunkarian

97	Mañana permanecerán abiertas las siguientes estaciones de servicio de la Compañía de Petroleo del Norte:	Bihar CPN-ren zerbitzu estazio hauek izango dira irekita...
98	A: ¿Qué ha pasado? B: Se ha muerto el vecino de arriba.	Zer gertatu da? Goiko bizilaguna hil da
99	En el salón se paseaban impacientemente los que esperaban.	Saloian betetik bestera pasioan zenbiltzan zain zeudenak.
100	A Miren le asustaba la inesperada visita.	Mireni beldur ematen zion espero ez zuen bisita harek.
101	Jon comió hamburguesas y vio la televisión.	Jonek hamburgesa jan zuen eta telebista ikusi zuen
102	El libro no ha aparecido.	Liburua ez da azaldu
103	A los huelguistas los arrestó la policía.	Greban ari zirenak poliziak atxilotu zituen
104	Se sentaron los invitados a la mesa.	Gonbidatuak mahaian eseri ziren
105	El perro del vecino esta ladrando.	Auzokoaren zakurra zaunkaka ari da
106	“Aceptó Israel la invitación a las negociaciones”.	Israelek negoziaketetarako gonbidapena onartu zuen
107	A: A mi madre no le queda bien la tortilla. ¿Y a la tuya? B: A mi madre le queda muy bien.	Nere amari tortilla ez zaio ondo ateratzen. Eta zureari? Nere amari oso ongi ateratzen zaio.
108	Queda un pequeño problema, pero lo resolveremos.	Arazo txiki bat gelditzen da baina konponduko dugu
109	Casi al mismo tiempo hablaron los dos trabajadores.	Ia aldi berean bi langileak mintzatu ziren
110	Se avalanzó un hombre sobre mí.	Gizon bat bota zitzaidan gainera
111	Erase una vez un rey muy perezoso.	Behin batean bazen oso alferra zen errege bat
112	Le sangraba la nariz al niño	Umeak sudurra odola zeriola zuen
113	En matemáticas le preparó su padre.	Matematiketan aitak lagundu zion
114	Estuve en casa todo el día y el teléfono no paró de sonar.	Etxean egon nintzen egun osoan eta telefonoak ez zion jotzeari utzi
115	No es el momento adecuado para discutir esa cuestión	Ez da arazo hau eztabaidatzeko une aproposa
116	El martes llamó el jefe para ver qué pasaba.	Asteartean nagusiak deitu zuen zer gertatzen zen ikusteko
117	No he encontrado el periódico.	Ez dut aurkitu egunkaria
118	A: ¿Qué ha pasado? B: Un niño, que le ha atropellado un coche.	Zer gertatu da? Kotxeak ume bat harrapatu duela
119	Por fin ha aparecido el periódico.	Azkenean azaldu da egunkaria
120	Cruzó la sala del hotel pero lo detuvo una voz.	[Hotelaren sala gurutzatu zuen baina] ahots batek gelditu zuen
121	No les molestó más el perro.	Zakurrak ez zion gehiago molestatu
122	“Tienen intención de volver al poder los social-demócratas”.	Sozialdemokratek boterera itzultzeko asmoa dute
123	Ha desaparecido el jefe de policía.	Polizien nagusia desagertu egin da ((Desagertu da polizien nagusia edo egin da, ez da gauza berdina))
124	La boda es el domingo que viene.	Ezkontza hurrengo igandean da

125	A mi madre no le queda bien la tortilla. ¡A mi tía le queda bien!	Nire amari tortilla ez zaio ondo ateratzen. Nere izebari ondo ateratzen zaio!
126	En aquel tiempo existía un rey muy temible.	Garai haietan bazen errege oso ikaragarri bat.
127	Traje una hamburguesa a casa, salí a la calle un rato y Jon se comió la hamburguesa.	Hanburgesa bat ekarri nuen etxera, kalera irten nintzen une batez eta Jonek hanburgesa jan zuen
128	En la punta del muelle se reunieron todos los invitados.	Nasaren puntan gonbidatu guztiak elkartu ziren
129	¿Tú por qué no te sentaste junto al jefe?	Zu zergatik ez zinen nagusiarekin eseri?
130	Por esa razón volvió a su casa Miren.	Arrazoi horregatik itzuli zen bere etxera Miren.
131	Se salvaron dos náufragos a la altura de Getaria.	Getaria parean bi itsas salbatu ziren
132	A mi hermano le dieron también un premio.	Nere anaiari ere sari bat eman zioten
133	El jefe leyó el periódico todo el día.	Nagusiak egun osoa eman zuen egunkaria irakurtzen
134	El perro ha estado ladrando todo el día.	Zakurrak egun guztia zaunkaka eman du
135	El jefe quiere que vengan todos los trabajadores el sábado.	Nagusiak [langile guztiak larunbatean etortzea] nahi du
136	“Vuelven a la escena los tradicionales metodos de espionaje.”	Berriz ere itzuli dira ohizko espioitza metodoak
137	A: ¿Te dio la paga tu padre? B: ¡Cómo no me va a dar la paga mi padre!	Aitak eman al zizun paga? Nola ez dit ba nire aitak pagarik emango!
138	Vino Miren y le contó un chiste al oído a mi madre.	Miren etorri zen eta nire amari txiste bat kontatu zion belarrira
139	Lo que yo comí fue una hamburguesa.	Nik jan nuena hanburgesa bat izan zen
140	Alguien seguro que sabe que va a venir el jefe.	Norbaitek ziur badakiela nagusia etorriko dela
141	También el jefe va a venir.	Nagusiak ere etorri behar du
142	A mi hermano también le dieron un premio.	Nere anaiari ere eman zioten sari bat
143	A Juan no voy a ver. Pero a María la veré el viernes.	Juan ez dut ikusi behar. Maria ordea ostiralean ikusiko dut
144	A: Siempre estás leyendo novelas. B: Me gusta leer novelas.	Beti elaberriak irakurtzen ari zara. Gogoko dut elaberriak irakurtzea
145	Entré en la habitación y vi a Miren. Estaba leyendo un libro.	Gelan sartu eta Miren ikusi nuen. Liburu bat irakurtzen ari zen
146	¿El jefe por qué no vino el otro día?	Nagusia zergatik ez zen lehengo egunean etorri?
147	A: ¿Has visto algún árbol con hojas? B: No, las hojas no salen hasta la primavera.	Ikusi al duzu zuhaitzen bat horriekin? Ez, horriak ez dira ateratzen udaberria arte
148	A: ¡Señora! B: ¿Sí? A: El bolso, que se le ha caído.	Emakume! Bai? Poltsa, erori egin zaizula
149	Se convirtió el agua en vino.	Ura ardo bihurtu zen

150	A: ¿Cómo tienes el cuello, y cuándo vas a volver al trabajo? B: El cuello me duele mucho, así que no sé cuando volveré.	Nola daukazu lepoa, eta noiz itzuliko zera lanera? Lepoko min handia dut, hortaz ez dakit noiz itzuliko naizen
151	Una hamburguesa es lo que yo comí.	Hanburgesa bat zen nik jan nuena
152	¡La Real es la que no ha ganado un partido!	Erreala izan da partidua irabazi duena!
153	A: Dicen que no comiste la carne. B: La carne sí comí. El pescado no comí.	A: Haragia ez zenuela jan diote B: Haragia jan nuen. Arraia ez nuen jan
154	La puerta se abrió sola.	Atea bakarrik ireki zen
155	A: ¿Fuiste a la cena? B: Claro que fui a la cena.	A: Joan al zinen afarira? B: Noski joan nintzela afarira
156	Dios no existe.	Jaungoikorik ez dago / ez da/dago jaungoikorik
157	Me asusta la situación económica.	Egoera ekonomikoak beldur ematen dit
158	Una cosa sí sé, y es que Miren no ha venido.	Gauza bat badakit, Miren ez dela etorri
159	Todavía tenemos nieve en el suelo y va a caer mas nieve todavía.	Oraindik lurtean elurra daukagu eta [elur gehiago egingo du oraindik]
160	En nuestro país existen formas muy diferentes de ver el mundo.	Gure herrian mundua ikusteko era asko daude
161	¡Me gusta escribir cartas!	Gogoko dut eskutitzak idaztea!
162	A: ¿Qué te pasa? B: Se me han perdido las llaves.	Zer gertatzen zaizu? Giltzak galdu zaizkit
163	A: ¿Qué ha pasado? B: Acaba de despedir el jefe a un trabajador.	Zer gertatu da? Nagusiak langile bat bota berri du
164	A: Creía que Juan no tenía dinero para pagar el viaje. B: No, dinero sí tenía, no tenía pasaporte.	A: Juanek bidaia ordaintzeko dirurik ez zuela uste nuen B: Ez, dirua bazuen, ez zuen pasaporterik
165	A: Pero ¿cómo es que están todas las vacas por la carretera? B: El casero, que se ha dejado la verja abierta.	A: Baina nolatan daude behi guzti horiek errepidean? B: Baserritarra, atea zabalik utzi duela
166	A: ¿Quién no fue a la reunión de ayer? B: Juan no fue.	A: Zein ez zen joan atzoko bilerara? B: Juan ez zen joan
167	Juan no le trajo un regalo a su madre este año.	Juanek ez zion bere amari oparirik ekarri aurtan
168	A: Tú ayer no viniste, ¿verdad? B: No, yo no vine antesdeayer.	Zu atzo ez zinen etorri, ez? Ez, ni herenegun ez nintzen etorri
169	Yo no traje el libro a clase ayer, sino antesdeayer.	Nik liburua ez nuen atzio ekarri klasera, baizik
170	Miren no le mandó una carta al director, sino un reporte.	Mirenek zuzendariari ez zion eskutitz bat idatzi, txosten bat baizik
171	A: ¿Por qué no le diste el libro a Mirene? B: ¡Sí que le di el libro a Mirene!	A: Zergatik ez zenion liburua eman Mireni? B: Eman nion eta liburua Mireni!// Bai, eman nion eta liburua Mireni!//
172	La Real sí que ha ganado el partido.	Errealak bai irabazi duela partidua
173	¡YO pago!	Nik ordaitzen dut!

174	¿Sabes qué? Un cura ha abierto una casa de juego en un avión.	Badakizu zer? Apaiz batek joko etxe bat jarri duela hegazkin batean
175	Jon no mató a la oveja a propósito.	Jonek ez zuen apropos hil ardia
176	Mira, un perro esta persiguiendo a un gato.	Begira, zakur bat katu baten atzetik ari da
177	Esta hirviendo el puchero.	Ontzia irekiten ari da
178	A: ¿Qué pasa ahí fuera? B: Miren, que esta cantando.	A: Zer gertatzen da hor kanpoan B: Miren, kantuan ari dela
179	A: ¿Ha cambiado algo por aquí? B: Miren ha dejado de fumar.	Zerbait aldatu al da hemendik? Mirenek erretzerari utzi diola
180	Llamé al fontanero ayer, porque había una tubería que goteaba.	(Iturkinari atzo hots egin nion,) hodi bat ura dariola zegoen eta
181	A: ¿Por qué hay una cola tan larga? B: Dicen que hay un autobús que se ha averiado.	A: Zergatik dago horrelako ilada luzea? B: Autobus bat hondatuta dagoela diote.

Appendix 5.6: Affirmative statements with *etorri* “come”

Speaker	No.	Text	WO
93C1A01	6	ta beste norbait etortzen da - bere= .. >>	SVX
93C1A01	25	Eta= iru- iru aur etortzen dira,	SV
93C1A02	9	eta manifestazioneko gizon batzuk etortzen dira atzetik.	SVX
93C1A02	12	bo= gero= poliziak etortzen dira,	SV
93C1A02	67	ta - neska bat etortzen da,	SV
93C1A02	68	ta e= boulanjera - bo okina - etortzen da,	SV
93C1A05	4	eta gero= kamioi bat etortzen da, -	SV
93C1A05	14	ebe pilo bat jende etortzen da bere atzean,	SVX
93C1A05	15	gero etortzen dira poliziak, -	VS
93C1A05	140	etortzen da= m kamioia ogiarekin,	VSX
93C1A09	34	=Ta orduan iru mutil datozte,	SV
93C1A10	32	Ta gero # badator polizia, ..	VS
93C1A10	178	eta —neska etortzen da. -	SV
93C1A15	6	Gero etortzen da= - e pertsona bat, bere auntzarekin, -	VSX
93C1A16	105	polizia etortzen da,	SV
93C1B01	14	Eta orduan e== # polizia dator,	SV
93C1B01	29	eta poliziko jefea etortzen da.	SV
93C1B01	98	etortzen dira aur batzuk,	VS
93C1B04	20	gero ba jende asko etorri ziren, ..	SV
93C1B04	23	gero poliziak etorri ziren, -	SV
93C1B04	92	Bo mutikoa etorri zen, -	SV
93C1B04	95+	~<mutikoa etorri zen>	SV
93C1B04	114	E? Orduan, e berriz e - # mutiko bat etorri da, -	SV
93C1B04	166	eta polize bat etorri da, -	SV
93C1B04	167	bo poliziak etorri diRA, -	SV
93C1B05	33	eta= - dator kontra neska bat. ez? -	VXS
93C1B06	25	etortzen da polizia,	VS
93C1B06	73	Eta orduan etortzen dira bi espia bezala,	VS
93C1B06	81	etortzen dira bi gizonak,	VS
93C1B06	190	etortzen da polizia,	VS
93C1B07	18	eta etortzen da mutil bat - e= bizikleta batekin.	VXS
93C1B08	156	Ta orduan # dator polizia berriro. -	VXS
93C1B10	19	eta orduan # polizia etortzen da,	SV
93C1B10	51	e= etortzen da inspektore bat,	VS
93C1B10	87	eta orduan # polizi batzuk etortzen dira,	SV
93C1B11	10+	~<or ume bat bizikletan datorrela>	SXV
93C1B11	46	urrutian datorrela - beste neska bat,	XVS
93C1B11	113	~<or datozela iru ume>	XVS
	+		
93C2A01	12	Ta etortzen da polizia,	VS
93C2A02	26	eta gero etortzen da tipo ba=t - auntz batekin. ez?	VXS
93C2A03	14	=ta= atzetikan # manifa bat etortzen da. ezta?	SV
93C2A03	20	eta orduan # etortzen da polizia, -	VS
93C2A03	62	polizia beregana datorrela,	SXV
93C2A03	135	Eta gero etortzen dia [nola= - <droga zeukan>~ gizona?	VS

93C2A04	8	ba etortzen da mutil bat,	VS
93C2A05	66	Orduan gero # etorri zien - beste manifestat- beste e manifestante batzuk,	VS
93C2A07	16	ta justo .. ba manifestazioko [- @@] jendia .. ba etortzen da. ez?	SV
93C2A08	58	beste neska bat dator - bisikletan, andik onuntz, -	SVXX
93C2A09	41	eta bapatian manifestaf- e= pen bat etortzen da atzetikan,	XSVX
93C2A09	51	Ba orduan polizia etortzen da,	SV
93C2A09	273	eta - etortzen da= - neska=- ..	VS
93C2A09	283	Ta polizia etortzen da momentu ortan jabiakin,	SVXX
93C2A09	320	Ordun etortzen da polizia,	VS
93C2B01	20	zer ta or etortze zako,	XVS
	21	bere atzetik manifestazyuba. -	
93C2B01	30	Ta geo poliziak etortzeie, ..	SV
93C2B01	74a	Geo poliziak etorri ~	SV
93C2B01	79	bi tipo etorri die,	SV
93C2B01	162	etorri dia bi gizon ola\	VS
93C2B01	179	eta= .. a etorri di bi= gizon oiek,	VS
93C2B01	243	etorri die. -	V
	244	Preso txiki ori leno bi gizonak eramán den oi ta beste bi preso=, -	
	245	bi gizon oiekin ta polizia batekin.	
93C2B01	341	andre bat azaldu e <eskiñetikan etorri a(?)>~, -	SVXta
93C2B03	21	zelan mutiko bat datorren bisikletas, ..	SVX
93C2B04	29	atzetik manifestapen bat dator. -	SV
93C2B04	34	zea\ polizia etorten da,	SV
93C2B04	46	Da beste= bueno polizia beste zea batzuk etortzen die/ <orren billé>. -	SVX
93C2B04	167	ta etorten da polizia, -	VS
93C2B06	24	manifa bat komunistena, *ruum!* dator be- bere atzetikan,	SVX
93C2B06	26	Etortzen da=- etortzen dia= ^zea - madero bat(?),	VS
93C2B06	74	Eta= - etortzen dia- sartzen dia zeak,	VS
	75	kartzelako= maderobak,	
93C2B06	200	Kamioi bat etortzen da panaderikua,	SVX
93C2B06	216	Ordun e- <THROAT> polizia bat etortzen da	SV
93C2B06	310	Gero- gero etor- etortzen dira bi mutil koxkor,	VS
93C2B08	28	Orduan polizia etortzen da, ..	SV
93C3A07	28	jateako- jateko ordu etorri da. ..	SV
93C3A07	45	Ta gero Txarlot etorri da, ..	SV
93C3A07	114	eta manifestazixo bat zetorren.	SV
93C3A08	19	e= etortze zaizko iru mutiko laguntzea~. -	VSXtzera
93C3A09	26	Ta gero badator mutiko bat, bixikletan, -	VSX
93C3A09	39	beste neska bat dator paren, -	SVX
93C3A10	15	Eta orduan # etortzen zaio burura zea Txarloti,	VXSI
93C3A10	21	atzetik etortzen da <manifestazio bat ikaragarria>.	XVS
93C3A10	24	eta etortzen da polizia,	VS
93C3A10	31	=e= etortzen zaio polizia,	VS
93C3A10	84	etortzen direla polizia batzuek - atera,	VSX
93C3A10	124	etortzen dira kanpotikan e bueno\ -	VXS
	125	<bere lagun astakerten ori libratea datozen> batzuk, ..	
93C3A10	213	Eta etortzen da berriz ere polizia, -	VXS

93C3A12	18	ba manifestazio bat dator. -	SV
93C3A12	26	Ordun polizia etortzen da=, ..	SV
93C3A12	40	Alako batean # ba= etortzen dia polizi batzuk, edo, ..	VS
93C3A12	47+	~<polizia datorrela>	SV
93C3A12	164a	etortzen da= ba espetxeko presondegiko kamioia, ~	VS

Appendix 6.1: AVO clauses in the Written Basque Corpus

Story, Pg.	AVO	Type
<i>Behi</i> , 14	berak adieraziko dizu aukerarik onena. <i>Hé will show you the best option</i>	A=Focus
<i>Behi</i> , 14	berak zuzenduko ditu zure pausoak.” * <i>hé will straighten your steps</i>	A=Focus
<i>Behi</i> , 18	baina ixiltasunak jan egin zuen nire dei hura, <i>but the darkness ate that call of mine</i>	V=Focus O=Access.
<i>Behi</i> , 18	hark ekarri zuen aldaketarik, <i>that one brought a change</i>	A=Focus
<i>Behi</i> , 22	eta elurrak ezabatu egin ditu bide-bidatzurrak, <i>and the snow erased all the paths</i>	V=Focus
<i>Behi</i> , 23	Hotzikara batek hartu zizkidan bizkar-isatsak, <i>A chill took over my back</i>	I=Topic
<i>Behi</i> , 29	Edadeko behiak ahaztu egiten du <bezperan gertatutakoa> , <i>An older cow forgets what happened the the day before</i>	V=Focus
<i>Behi</i> , 29	neronek esango dizut <otsoengandik aldegin egunean agindutakoa>. <i>I myself will tell you what happened the day you got away from the wolves</i>	A=Focus D=Topic
<i>Kuba</i> , 23	Horiek kontrolatzen dute turismoaren eta azukrearen dirua. <i>Théy (are the ones who) control the money from tourism and sugar</i>	A=Focus
<i>Kuba</i> , 27	Emakume mulata batek ireki digu atea. <i>A mulata woman opened us the door</i>	A=Focus D=Topic
<i>Kuba</i> , 29	“Nik ere badut bizikleta” <i>Me too, I do have a bicycle</i>	Pol.=Focus O=Non-spec.
<i>Kuba</i> , 30	Mirthak egin ditu aurkezpenak. <i>Mírtha made the introductions</i>	A=Focus O=Non-spec.

Appendix 6.2: AOV clauses in the Written Basque Corpus

Story, Pg.	AOV	Type
Behi, 22	baina <buru gainean sentitzen nuen> harlausa batek erantzun guztiak zapuzten zizkidan. <i>but a stone block I felt on my head blocked all the answers</i>	O=Accessible
Behi, 23	baina nik alde egingo nuke. <i>but I would get out of here</i>	O=Non-ref.
Kuba, 22	Bakoitzak geure kuadrakoak aukeratu genituen, <i>Each one chooses those [representatives] from our block</i>	O=Accom.
Kuba, 23	Bigarren garagardoak ere hondoa jo du <i>The second beer too has already hit [the] bottom</i>	O=Non-ref.
Kuba, 23	“... BERAK hogeita hemezortzi bizitoki desberdin ditu: <i>He [Fidel] has 38 different dwelling places</i>	O=Non-spec.
Kuba, 26	Tonik ate bat jo du <i>Toni knocked on a door</i>	O=New
Kuba, 28	eta olio botilak 100 balio ditu; <i>and a bottle of oil costs 100</i>	O=Non-ref.
Kuba, 28	Mirthak gelako argi guztia bereganatzen du, <i>Mirtha attracts all the the lights in the room</i>	O=Non-spec.
Kuba, 31	Mirthak eta biok helbideak trukatu ditugu. <i>Mirtha and I traded addresses</i>	O=Low spec.
Kuba, 34	Turistek kaxak eta kaxak eramaten dituzte.” <i>Tourists take away boxes and boxes</i>	O=Non-ref.
Kuba, 36	Hodiren batek jarioa izan behar du, <i>A pipe must have a leak</i>	O=Non-ref.
Kuba, 37	gaitzek egundoko beldurra ematen didate. <i>illnesses scare the hell out of me</i>	O=Non-ref.

Appendix 6.3: VO clauses in the Written Basque Corpus

Story, pg.	Text	Comments
Behi, 15	Deituko nion Aingeru Goardako, <i>Were I to call him Guardian Angel,</i>	Special setting construction.
Behi, 15	deituko nion Izpiritu, Mintzo, Ahots edo nahi dena, <i>were I to call him Spirit, Speech, Voice, or whatever</i>	Special setting construction.
Behi, 17	Aurreratzen nuen ahoa, <i>I put my mouth forward</i>	Emphatic
Behi, 17	Egiten nuen uxar <i>I dug (lit. I did digging-action)</i>	Emphatic
Behi, 18	“Esango dizut hitz bat,” <i>I’ ll tell you something</i>	Emphatic
Behi, 22	Egiten nituen ahaleginak, <i>I made a great effort (I did all I could)</i>	Emphatic
Behi, 22	<astindu egin beharra> daukazu zeure buru hori, <i>You’re gonna have to shake that head of yours</i>	egin verb focus
Kuba, 23	Hemen ere badituzte neskalgunak, <i>Here too they do have girlfriends</i>	emphatic <i>ba</i> -focus
Kuba, 28	utzi egin nuen artisau lana, <i>I left the work of artisan</i>	egin verb focus
Kuba, 28	Ezagutu behar duzu nire alaba, <i>You háve to meet my daughter</i>	Emphatic verb focus <i>behar</i>
Kuba, 29	Ederretsi dut lana, <i>I approve of the work</i>	Emphatic
Kuba, 33	Jaitsi ditut praka motzak <i>I put down my shorts</i>	Emphatic

Appendix 6.4: AOV clauses in the Spoken Basque Corpus

1A02:31	Orduan bo= indartsuak e= - e= ogi bat du, <i>then well the strong guy uh he has a (piece of) bread</i>	New (Accom?)
1A05:36	be= besteak bildurra dauka, <i>the other one is afraid (lit. have fear)</i>	Non-ref.
1A16:80	ta poliziak beldurra zuen, =ta_ - <i>and the policeman was afraid (lit. have fear)</i>	Non-ref.
2A01:137	baina neskak alde egin du. <i>but the girl takes off (lit.: make side/room)</i>	Non-ref.
2A06:35	aurrak udare bat lapurtuko dio. <i>the child will steal a pear (from him)</i>	New/Accom.
2A09:55	eta denak ies egiten dute, <i>and all of them take off (lit. make escape)</i>	Non-ref.
2B01:205	berrize poliziak txistu jo du, <i>again the policeman blows (the) whistle</i>	Non-spec.
2B01:291	Ta- <THROAT> ordun # radiyu zeoze esan du. - <i>and then the radio said something or other</i>	Non-spec.
2B05:57	Eta= ortaiko mutil batek txistua jote jotzak, <i>and one of those boys whistles at him (lit. play whistle)</i>	Non-spec./Accom
2B07:68	Mutiko=- lapurrak, bere bidea jarraitzen du, <i>the boy the thief continues his way</i>	Non-spec.
2B08:129	neska orrek ogi bat lapurtzen du, ... <i>that girl steals one (loaf of) bread</i>	NEW
2B08:164	oi kamioi orrek beste= geldit- e= geldiketa bat egitenen(?) du, <i>that truck makes another stop</i>	Non-spec.
3A11:20	Ordun mutillak buelta ematen du, <i>then the boy turns around (lit. make turn)</i>	Non-spec.
3A12:28	ta= jendeak- manifestaziokoak ies egiten dute, - <i>and the people the demonstrators take off (lit. make escape)</i>	Non-spec.

Appendix 6.5: AVO clauses in the Spoken Basque Corpus

1A01:48	eta= e=ta berak badute= a= e= badute sagar bat bakotxa, <i>and and he they do have an apple [sic] each</i>	Accom.
1A02:32	eta= gizonak e- e nai du ogi puska bat, <i>and the man he wants a piece of bread</i>	New/Accom. V
1A04:8	=neskak - artzen du= udare= <i>the girl [sic] takes pear (unfinished)</i>	Accom.
1A04:14	ta neskak e= kentzen dio= txapela, - <i>and the girl takes off his hat (lit. "the hat")</i>	Accom.?
1A05:142	eta be neska pobrea artzen du ogi bat, <i>and the poor girl takes one loaf of bread</i>	Accom.
1A10:179	Gizonak uzten dio bere lekua. <i>the man leaves her his space</i>	Accom.
1A10:136	Ta gizon orrek ematen dio= gauza bat, <i>and that man gives him one thing</i>	NEW (non-spec.)
1A15:20	ba e= batek badu - e <pala bat ta <pilota bat e gomarekin>>, <i>well uh one does have a paddle and a ball uh with a rubber</i>	NEW
1A17:27	iru mutikoak ematen dute txapela. .. <i>the three boys give him his had (lit. the hat)</i>	Given
1B01:82	berak e= artu zuen e= - ogia, - <i>hé took the bread</i>	ERG focus
1B04:38	eta= gizona nai zue=n= ogiba, ogia, <i>and the man wants the bread</i>	Accom.
1B04:49	mutiko bat artzen du= e=... [...] gatza, <i>a guy takes uh the salt</i>	Accom.
1B04:111	mutikoak, artu zuen e= - atea, <i>the guy he took uh the door</i>	Given-
1B04:158	eta= mutikoa artu zuen ogia, <i>and the guy took the bread</i>	Given
1B04:16	gizonak artzen du ikurriña, <i>the man picks up the flag</i>	Given V
1B05:72	ta orduan mutilak ematen dizkie= iru udare, <i>and then the boy gives them three pears</i>	Accom.
1B08:141	ba bera - jotzen dio- jotzen dio kartzelakoa, <i>and he hits the jail guy</i>	Given
1B08:75	Baino bapatea=n polizia ikutzen du pitua, <i>but suddenly the police blows the whistle</i>	Non-spec.
1B10:63	eta Txarlotek artzen du gatzontzia, <i>and Chaplin takes the salt shaker</i>	Accom
2A01:60	Eta gizon orrek ikusten (?) ori, <i>and that man sees that</i>	Given
2A01:124	Eta orduan gizonak artzen du ogia, <i>and then the man takes the bread</i>	Given
2A01:49	eta= gizo=n gizon orrek artzen du= ogia. ez? <i>and that man takes the bread, right?</i>	Given V
2A01:127	Eta= gizonak esaten du ezetz. <i>and theman says no way</i>	Non-ref

2A03: 288-:289	“Nik nai det, ta e- zigarro bat”. <i>I want, a cigar</i>	Accom.; delayed?
2A03:218	Andre orrek e dauka ogia! <i>that wóman uh has the bread</i>	ERG focus
2A05:98	e andre batek lapurtu zuen ogi bat, <i>uh a woman stole a loaf of bread</i>	NEW
2A07:44	eta orduan Txarlotek artzen du gatza, <i>and then Chaplin takes the salt</i>	Given
2A07:114	poliziak artzen du Txarlot, <i>the policeman takes Chaplin</i>	Given
2A07:101	Ta neskak artzen du ogia, <i>and the girl takes the bread</i>	Given V
2A09:288	eta berak dakala ogiya eta ori. <i>and that hé has the bread and stuff</i>	ERG focus
2A09:210	Gais- e= gai[]kill(e)ak sartzen dittuzte poliziak eta= nausiya, <i>the bad guys lock up the policemen and the boss</i>	Given (Accom.)
2A09:285	eta Txarlotek esan- esaten dio ez. <i>and Chaplin says no</i>	Non-ref.
2B06:305	Eta besteak ematen dio puro bat- <i>and the other guy gives him a cigar</i>	Accom. V
2B06:73	Eta arek zeukan e= papelina bat. o zerbaite. ez? .. <i>and he had a little envelope. or something. right?</i>	NEW
2B07:18	Ta ordun e= mutikoak ikusten ditu .. udariak, <i>and then uh the boy sees .. the pears</i>	Given
2B08:10	Eta= g- gizon onek artzen du bandera ori, - <i>and that man picks up that flag</i>	Given V
2B08:65	eta beak artzen du= gatz ori, - <i>and he takes that salt</i>	Given V
2B08:142	dendako= jefiak - esaten du ezetz. <i>and the boss at the store says no way</i>	Non-ref.
3A07:42	Eundoko= - batzuk arrapau dittube be(ra)ixe(n) jefiek/ \eta. - <i>the hell - some guys got their boss and stuff</i>	Accom.
3A07:121	Eta Txarlotek kendu dixo ogixe, <i>and Chaplin takes the bread from him</i>	Given
3A09:23	auntzak usaitu ditu madarik, <i>the goat smelled the pears</i>	Given
3A10:96	=ba berak probatzen du bere ez-dakit-zer zian\ <i>well he tates his whatever it was</i>	Given
3A10:69	Eta berak artzen du ogi- ogi puska ori, - <i>and he takes that piece of bread</i>	Given V
3A11:37	ta mutillak ematen dizkio ba\ - madari bana. ez? <i>and the boy gives him [sic] a pear each. Right?</i>	Accom.; delayed?
3A11:19	ta neskak kentzen dio sonbreroa. .. <i>and the girl takes off his hat</i>	Given
3A12:53	Ordun Txarlotek artzen du gatzontzia, <i>then Chaplin takes the saltshaker</i>	Given

Appendix 6.6: VO clauses in the *Lur* subcorpus of the Spoken Basque Corpus

20. 3A07:149	Ta ba in ditu bere tonterixek, <i>And well he did all his silly things</i>	O=New
21. 3A07:3	Badakizu Txarlot\ <i>You know Chaplin</i>	<i>ba</i> focus; O=Given
22. 3A08:30	artu detxau txapela, <i>he took the hat</i>	O=Given
23. 3A08:34	Eta ikuste ittu <arbola= arbolan txermenak - jasotzen daona>, <i>And he sees the guy who was picking pears on top of the tree</i>	O=Given Focus Extrap.
24. 3A09:74	Eta= ematen dio txapela, <i>And he gives him his hat</i>	O=Given
25. 3A10:130	eta libratzen dute gizon ori. ... <i>and they free that man</i>	O=Given
26. 3A10:136	eta artzen du ate bat, <i>and he takes a door</i>	O=New; Focus Extrap. unimportant
27. 3A10:149	ta artzen ditu giltzak, <i>and he takes the keys</i>	O=New; Accommodated
28. 3A10:150	eta - libratzen dittu/ ba\ polizia oriek eta abar. - <i>and he frees, well those policemen and stuff</i>	O=Given; Focus Extrap.
29. 3A10: 152-153	ematen diote, pixka bat e= faboreko e- tratamentu bat bezela/ <i>they give him, well like like special treatment or something</i>	O=New; Focus Extrap.; Hesitation
30. 3A10:162	Eta - ematen dio olako= txeke= (ed)o zea bat, <i>And he gives him a check or something</i>	O=New; Focus Extrap. Hesitation
31. 3A10:216	eta artzen du bere kafea, <i>and he takes his coffee</i>	O=New; Accommodated
32. 3A10:234	eta aurkitzen du ba asiento libre bat, <i>and he finds well a free seat</i>	O=New; Accommodated Focus Extrap. Hesitation
33. 3A10:36	Eta kartzelan dauka/ ba\ lagun bat bastante asto- astakertena. <i>And in jail he has well a friend, who's rather stupid</i>	O=New; Focus Extrap.; Hesitation?
34. 3A10:81	eta - zeukan/ bere eskubiko aldean bazeukan beste tipo bat, <i>and he had, on his other side he had another guy</i>	O=New; <i>ba;</i> (starts as Focus extrap.)
35. 3A10:86	eta seinalatzen dute tipo ori. - <i>and they point to that guy</i>	O=Given
36. 3A10:94	eta eramaten du eskubian zeukan tipo ori/. ... <i>and they take that guy he had to his right</i>	O=Given
37. 3A11:11	Ta ikusten ditu madariak. <i>and he sees the pears</i>	O=Given

38. 3A11:12	Ta ordun artzen du saski bat osoik, <i>and then he takes one whole basket</i>	O=New; Focus Extrap. Accommodated
39. 3A11:36	Ordun ematen diote= txapela, <i>then they give him his hat</i>	O=Given; Focus Extrap.; Hesitation
40. 3A11:8	ya bete ditu= bi saski. - <i>he has already filled two baskets</i>	O=New (rate); Focus Extrap. Non-specific; Hesitation
41. 3A12:14	artzen du bandera, - <i>he takes the flag</i>	O=Given
42. 3A12:29	eta atxilotzen dute= - ba Txarlot, <i>and they jail well Chaplin</i>	O=Given; Hesitation
43. 3A12:48	ba= artzen du s= droga=ren bat, <i>well he takes some kind of drug</i>	O=New; Focus Extrap. Hesitation
44. 3A12:51	Eta eramaten dute ori. <i>and they that that guy</i>	O=Given

Appendix 6.7: OV clauses in the Lur subcorpus of the Spoken Basque Corpus

1. 3A07:104	zeoze sari bat eman dixobe, - <i>they gave him some kind of prize</i>	O=New; Non-specific
2. 3A07:142	Ta neski eman debe. <i>and they take the girl</i>	O=Given; Contrastive
3. 3A07:70	jira buelta iman da. <i>he gave a turn (= he turned around)</i>	O=New; Non-specific
4. 3A07:95	Gero azkenien beintzat # denak bota dittu/ \te. - <i>Then in the end anyway he knocks everybody down</i>	O=Given; Contrastive
5. 3A07:96	Nausixek/ - he! - librau in ditu. .. <i>The bosses ha! he freed them</i>	O=Topic; egin
6. 3A08:31	eta iru txarmena\ - emate izkoe - bat. <i>and he gives them three pears - one (?)</i>	O=New; Accommodated?
7. 3A09:12	eta an iru saski dauzke, <i>and there he has three baskets</i>	O=New; Accommodated?
8. 3A09:27	eta orduan madarik ikuste ittu, <i>and then he sees the pears</i>	O=Given
9. 3A09:65	=txistu itten dioe, - <i>they whistle at him (lit. make whistle)</i>	O=New; Non-referential
10. 3A09:7	Ta mantal bat dauke, <i>and he has an apron</i>	O=New
11. 3A09:77	ba madari (?) ematen dizkio, <i>well he gives him some pears</i>	O=New Accommodated?
12. 3A10:103	ta ola muesa pila bat egiten du. <i>and he makes a bunch of faces like that</i>	O=New; Non-specific
13. 3A10:164	edo <ez dakit nik ze izango zan> ematen dio, <i>or he gave him I don't know what</i>	O=New; Non-specific
14. 3A10:53	eta beren artean borroka pixkat izaten dute, <i>and among them they have a little fight</i>	O=New; Non-specific
15. 3A10:59	Eta an # borroka pixka bat izaten dute, - <i>and there they have a little bit of a fight</i>	O=New; Non-specific
16. 3A12:110	ba= la- ogi bat lapurtze u. .. <i>well st- she steals a loaf of bread</i>	O=New
17. 3A12:120	Txarlot eramaten dute. [-] preso. - <i>They take Chaplin away. prisoner.</i>	O=Given; Contrastive
18. 3A12:126	Ordun Txarlot - e= askatzen du, poliziak, <i>Then he frees Chaplin, the policeman (does)</i>	O=Given
19. 3A12:127	=eta neska eramaten du. - <i>and he takes the girl</i>	O=Given; Contrastive
20. 3A12:130	janari pila artzen du, - <i>he takes a bunch of food</i>	O=New; Accommodated
21. 3A12:141	ta kontua eskatzen dio, .. <i>and he asks for the bill</i>	O=New; Accommodated
22. 3A12:159	eta or ere # ba gauzak eskatzen dittu, <i>and there too well he asks for things</i>	O=New; Non-specific

23. 3A12:75	ta polizi batzuk .. polizia batzuk arrapatzen dituzte. preso oiek. <i>and they catch some police officers. Those prisoners</i>	O=New; Accomdated; Non-specific
24. 3A12:86	Orduan pres- e= poliziak askatzen ditu, .. <i>then he frees the policemen</i>	O=Given
25. 3A12:98	biño ezetz esan zun. - <i>but he said no</i>	O=New; Non-referential

Appendix 6.8: Most obvious examples of the delayed focus construction in Spoken Basque Corpus

1. 1B06:1	Ba= zen, Txarloti buruzko filme bat. ez? <i>Well it was, a movie about Chaplin, right?</i>	VX
2. 2A03:288-:289	Nik nai det, ta e- zigarro bat <i>I want, and uh a cigar</i>	AV,O
3. 2A06:59	eta bapatean ikusten ditu, - iru mutil. <i>and suddenly he sees, three boys</i>	VO
4. 2A08:1	Ba pelikulie san, - txorrada bat. <i>Well the movie was, a dumb thing</i>	SVX
5. 2A08:7	eta seuan, madarixak batzen. <i>and he was, picking pears</i>	VXtzen
6. 2B04:207	=beak nai duna da, bea [berrize kartzelan sartzia], <i>what he wants is, for him to go back to jail</i>	SV:
7. 3A10:152	ematen diote, pixka bat e= faboreko e- tratametu bat bezela/ <i>they give him, a bit like favorable treatment or something</i>	VO
8. 3A10:175	Eta ateratzen da, - ogitegi batetikan, neska bat. ... <i>and there comes out, from a bakery, a girl</i>	VXS

Appendix 6.9: Written Basque Corpus, SXV sentences

Behi, 14	Bera ni baino nagusiagoa zen, <i>He was older than me</i>
Behi, 15	hau ailobrea [sic] duk <i>This is trigonella foenum-graecum</i>
Behi, 15	Bera izpirituzkoa da. <i>She is of the spirit (spiritual)</i>
Behi, 15	Nirekiko guztiak zure esku daude, <i>Everything about me is at your disposal</i>
Behi, 16	Den dena elur egun batean gertatu zen. <i>Everything happened on a snowy day</i>
Behi, 17	<<Ni <jaten> aritutako> belazea> txuri zegoen; <i>The grass I'd been eating was white</i>
Behi, 19	zu libre zara, <i>you are free</i>
Behi, 19	Baina Setatsua beste kontu batzuekin zebilen. <i>But Setatsua had other ideas (was with some other stories)</i>
Behi, 21	behi izatea gauza haundia da! <i>being a cow is a great thing!</i>
Behi, 22	Baina egoera oso fastikagarria zen, <i>but the situation was very tiresome</i>
Behi, 24	baina hau Euskalerra da, <i>but this is the Basque Country</i>
Behi, 24	“Denak belarridunak dira, gainera,” <i>All of them have ears too/besides (lit. they are all ear-having)</i>
Behi, 26	eta nire arnasaren lurruna haize hotzean galtzen zen, <i>and the vapor of my breath gets lost in the cold air</i>
Behi, 27	baina arboladiaren bukaera ere gero eta hurbilago zegoen. <i>but the end of the tree grove was getting closer and closer</i>
Kuba, 17	Gosari-gela ia betea dago; <i>The breakfast room is almost full</i>
Kuba, 18	<Lekuak eragiten duen> lehen inpresioa gogorra da. <i>The first impression from that place is hard/harsh</i>
Kuba, 18	<Lurrikararen bat <izan berria> dela> ematen du. <i>It seems like there's just been an earthquake (lit. It gives that...)</i>
Kuba, 18	kaleak garbiak daude <i>the streets are clean</i>
Kuba, 19	Aurpegia argi-argi jartzen zaio; <i>His face gets all bright (lit. the face gets all bright to him)</i>
Kuba, 19	Espainiako neskek oso politak eta jatorrak omen dira; <i>Spanish girls are said to be very pretty and nice</i>
Kuba, 19	ni Toni naiz, <i>I am Toni</i>
Kuba, 20	Toni zoriontsu dago. <i>Toni is happy</i>

Kuba, 21	nirea bezalakoak <kotxeak baino estimazio handiagoa omen du. <i>those like mine are better thought of than cars</i>
Kuba, 22	Gehienak <ezkonduak> dira <i>Most are married</i>
Kuba, 22	Angolarra oso jende txarra omen da. <i>Angolans are said to be very bad people</i>
Kuba, 23	Aurretiko presidente guztiak Malekoiko Jauregian bizi izan ziren <i>All the previous presidents lived in the Malecon's Palace</i>
Kuba, 23	solasa neska-kontuetara lerratu da. <i>the conversation turn to girl stories</i>
Kuba, 23	13tik 25 urtera bitarteko neska gehienak prostituzioan ari dira. <i>Most of the girls from 13 to 25 are into prostitution</i>
Kuba, 25	Toni <bata eta bestea agurtuz> doa; <i>Toni goes around greeting everything</i>
Kuba, 26	Gela txiki-txikia da; <i>The room is very small</i>
Kuba, 27	Ni artisaua naiz, <i>I am an artisan</i>
Kuba, 27	Nire senarra Kubako artisau onenatarikoa da. <i>My husband is one of the best artisans in Cuba</i>
Kuba, 28	soldata bat 200 peso dira <i>The salary is 200 pesos</i>
Kuba, 28	hau miseria da. <i>this is a miserable situation</i>
Kuba, 29	ni ere zuek bezalakoa naiz.” <i>Me too I'm like you</i>
Kuba, 32	Ni <paso egitekotan> nago, <i>I am about to tell them to forget it</i>
Kuba, 33	“Horiek trusas dira, <i>those are trusas (slip underwear)</i>
Kuba, 33	“Ba, hori <<gobernuak asmatutako> hitza> da” <i>Well, that is a word that the government made up</i>
Kuba, 33	Baina eurak <askoz ere txarragoak> dira; <i>But they are much worse</i>
Kuba, 35	nire harrotasuna eta artasuna oso erasanak geratzen dira. <i>my pride and manhood are left very affected</i>
Kuba, 36	pixa-usaina ere nabarmena da. <i>pee smell too can be noticed (cf. ere topicalizer)</i>
Kuba, 36	Toni <adinean aurrera doan> gizon batekin hizketatu da. <i>Toni is talking with a older man</i>
Kuba, 36	Emakumea García Lorca antzestaldeko aktorea da. <i>The woman is an actor in the GL theater company</i>
Kuba, 37	Gauza hauek berez sekretuak dira, <i>These things therefore are secrets</i>

Appendix 6.10: Written Basque Corpus, SVX sentences

Behi, 21	eta zu bakarrik zaude hemen, <i>and only you are here (you're the only one here)</i>	S-focus, thetic
Behi, 21	behia dago hemen. <i>the cow is here.</i>	S-focus, thetic
Behi, 27	sua bezalako zerbait etorri zitzaidan aitortu ez dudan leku horretara, <i>something like fire came to me to that place I haven't mentioned</i>	S-focus, D-topic
Kuba, 17	jan beharra dago zer gerta ere <i>one must eat no matter what (there is need to eat no matter what)</i>	S-focus, thetic
Kuba, 18	Jende dezente da kalean; <i>A lot of people is in the street</i>	S-focus, thetic
Kuba, 18	Hamahiru-hamalau urteko mutiko beltz bizikletadun bat jarri zait parean. <i>A 13-14 year-old boy on a bicycle got next to me</i>	S-focus, D-topic
Kuba, 22	hemen koka eta maria dabil bakarrik.” <i>here coke and dope are found only</i>	S-focus, thetic
Kuba, 23	Hitza joan zaio zerbitzaria datorrela ikustean. <i>He stopped talking upon seeing the waiter coming (lit. the word went from him ...)</i>	S-focus, D-topic
Kuba, 23	“Hemen jeneralak eta partiduko buruak bizi dira majo. <i>Here the generals and the party bosses live well</i>	S-focus, thetic
Kuba, 24	“Txarrena izaten da larunbatean kuadrako guardia tokatzen bazaizu; <i>The worst is if you get block guard duty on Saturdays</i>	Extrap.
Kuba, 28	Gizonezko bat da bi botila hutsekin. <i>It's (there is) a man with two empty bottles</i>	S-focus, thetic
Kuba, 29	zorabioa ere egin izan zait bizikletan, <i>dizziness too has happened to me on the bicycle</i>	V-focus, egin
Kuba, 30	sei kilometro luze duen itsasertzeko malekoa da habanarren bilgunea, <i>The 6 Km. long malecon along the ocean is the gathering place of the people of Havana</i>	Extrap.
Kuba, 30	Halako konplizitate bat somatzen da bien artean, <i>Some such complicity between them can be detected</i>	S-focus, thetic
Kuba, 31	Almita ere itzarri da bere lozorrotik <i>Almita too she woke up from her stupor</i>	V-focus?
Kuba, 31	Plaza Zaharrean kotxe ilara dago jausitako etxe baten aurrean. <i>In the old square there's a line of cars in front of a fallen down house</i>	Present. thetic
Kuba, 34	Excusatio non petita ... hasten den latinezko esaera etorri zait gogora. <i>The Latin saying that begins with Excusatio non petita came to my head</i>	S-focus, D- topic
Kuba, 34	berbaro ikaragarria heltzen da gugana: <i>an incredible sound of voices reaches us</i>	Present. thetic
Kuba, 35	Neskatan egitea bada nire bidaiaren helburuetariko bat - <i>Going out with girls is one of the goals of my trip</i>	Polarity focus.
Kuba, 37	edo gisako hauspokeriaren bat etorri zait ahora, <i>or some other dumb old thing came to my mouth</i>	S-focus, D-topic
Kuba, 37	“Atzerritar asko ezkontzen da neska kubatarrekin” <i>Many foreigners marry Cuban girls</i>	S-focus
Kuba, 37	Txeluren kontsigna etorri zait gogora. <i>Txelu's slogan came immediately to mind</i>	S-focus, D-topic

Appendix 6.11: Spoken Basque Corpus, SXV sentences

1A02:25	eta ogea bere gainera etor- erortzen da, <i>and the bread falls on (top of) him</i>
1A02:59	gero gizonak - indar- gizona indartsuago bilakatzen da, <i>then the man becomes stronger</i>
1A02:71	neska korrika doa, <i>the girl goes running</i>
1A02:73	ta neska korrika doa, <i>and the girl goes running</i>
1A02:83	e neska kartzelara doa, .. <i>uh the girl goes to jail</i>
1A05:129	[eta] ebe iruak [iruak] iruak lurtean daude, <i>and uh the three of them are on the ground</i>
1A10:127	eta gero bere gizona ondoan dena arrituta dago. .. <i>and then his- the man the one next to him is all amazed</i>
1A10:17	<@@ Eta bandera, - lurrera erortzen da. <i>And the flag falls to the ground</i>
1A10:177	gizona eserita da, <i>the man is sitting down</i>
1A10:89	Eta gizona <arritua> dago, <i>and the man is amazed</i>
1B01:115	=orduan - neska - zutik gelditu bear da. <i>then the girl has to remain standing</i>
1B01:20	Eta= .. polizi- e bera zulo batean erortzen da, <i>and the polic- uh he falls into a hole</i>
1B01:23	eta polizia or dago. <i>and the police is right there</i>
1B08:18	Bapatean poliziak aurretikan datoz. <i>Suddenly, the police are coming from behind</i>
1B10:147	Txarlot ya libre dago. ez? <i>Chaplin is already free. ok?</i>
2A03:78	au txarra dago, <i>this is bad</i>
2A05:23	Gizon ori oso txarra zen. ez? <i>That man was very bad. OK?</i>
2A05:65	=ya= zelda itxia zegoen. <i>already the cell was closed</i>
2A07:65	ya= .. zeak atiak itxiyak zeuden. <i>already those the doors they were closed</i>
2A07:77	Baina Txarlot drogatuta dago. e? <i>But Chaplin is all drugged up, uh?</i>
2A08:50	eta gizona ya arbola gainian seuan. <i>and the man already he 's on top of the tree</i>
2A09:280	ta biak lurrera - erortzen dia. <i>and the two of them fall to the ground</i>
2A09:3	pelikula au <zuk ikusi duzuna baino pixkat konplikatuxegoa> da. <i>this movie is a bit more complicated than the one you saw</i>

2A09:332	ta biyek elkarrekin ikusten dia. <i>and the two of them see each other</i>
2A09:84	Txarlo=t - ba= beira ai ziyon eta ola, <i>Chaplin well he was looking at him and stuff</i>
2A09:91	Biyek sesiyuan asten dia, - <i>The two of them begin quarreling</i>
2A09:96	eta denak lerro batian jartzen dia, <i>and all get into a line</i>
2A10:28	guzik ba lurra eroitzen die. <i>all of them well they fall to the ground</i>
2A10:46	Ta mutiko orik, biden aurea seitzen de, <i>And that boy continues down the road</i>
2B01:13	<makil oi e lotue> ze- zeuen, <i>that stick uh it was tied</i>
2B01:132	ta danak filan jarri die, .. <i>and all of them get in line</i>
2B01:274	ta au re konorte (ga)be g(er)au de(?). - <i>and this one too was left knocked out</i>
2B01:349b	ta biyek beira eroi die. - <i>and the two of them fall down to the ground</i>
2B01:350	Ta ordun ogiye Txarlotena= Txarlotena pasa da. - <i>And then the bread passes on to Chaplin</i>
2B01:368	ta neska oi ala= - arri-arri einda geau de, <i>and that girl thus= she gets all amazed</i>
2B01:39	baña ortik poliziak an daude, <i>but around there the policemen are there</i>
2B01:392	eske neskie an ge(r)au de. <i>cause the girl stayed there</i>
2B01:9	Ta ordun egur batzu luzeuak zien, - <i>and then some of the logs were long</i>
2B04:151	=bueno neska ori beartsu bat da. <i>well that girl is a beggar (poor)</i>
2B04:51	ta- bera= zea gatz= bertan gelditze da, <i>and he I mean the salt stays right there</i>
2B05:34	E= madari gustiak lurreti=k - lurretik ziar gelditzen die, <i>Uh= all the pears turn out on the ground all over the ground</i>
2B05:88	=Beste seose igual aastu in jatan, <i>Other stuff perhaps I forgot all about</i>
2B06:181	*oain zu lib- libre zaude*. .. <i>now you are fr- free</i>
2B06:219	e= Txarlot andik zebilen. <i>uh Chaplin was (going) somewhere around there</i>
2B06:238	lapurra=, neska zan*. .. <i>the thief was the girl</i>
2B07:74a	Ta tipoa ola gelditzen da <i>And the guy gets all like that</i>
2B08:131	eta biak lurrera erortzen dira. - <i>and the two fall to the ground</i>

2B08:21	Baino meni- manifestaldi ori= bere atzetik dator. ... <i>but that demonstration is coming behind him</i>
2B08:35	ba gizon ori= oso= oso fuertea eta oso= - gaiztoa da, .. <i>well that man is very strong and very bad</i>
2B08:98	ya zoramen ori= itxuraz kendu zaio, <i>already that craziness of his seemingly is gone away</i>
3A07:83	beax- gizonak kanpun geldiu die. - <i>right- the man stayed outside</i>
3A09:28	beste gizona <arbolan iyota> dau, - <i>the other man is up on the tree</i>
3A09:49	=madari dana=k - lurrera itten(?) zaizkio, - <i>all the pears go all over the ground</i>
3A10:177b	biak lurrera erortzen dia. - <i>the two of them fall to the ground</i>
3A10:58	eta litera eta guzti bera erortzen da. ez? .. <i>and the bunk and everything fall down. OK?</i>
3A12:171	Ta neska ordun zutik dijoa. <i>Ans so that girl is going standing up</i>
3A12:69	e= bakoitza bere zeldara sartzen da, <i>uh each one goes into their own cell</i>

Appendix 6.12: Spoken Basque Corpus, SVX sentences

2A07:4	Txarlot - e badijoa kaletik, ... <i>Chaplin uh he is going down the street</i>	ba-focus
2B01:348	Eta neska ori bazijun korrike=, <i>And that girl was going running</i>	ba-focus
3A10:2	Eta= Txarlot badijoa kalean, <i>And Chaplin is going down the street</i>	ba-focus
1A10:53	Eta- - gizona dago <or eserita>, <i>and the man is sitting there</i>	Foc-Extrap.
1A16:100	ta Txarlot Txaplin - joaten zen nunbaitetik, <i>and Charlot Chaplin was going along somewhere</i>	Foc-Extrap.
1A16:79	lapurrak zeuden poliziaren kontra, <i>the thieves were against the police</i>	Foc-Extrap.
1B05:95	ta gizona gelditzen deda(?) ola, <i>and the man is left like that</i>	Foc-Extrap.
1B06:148	Eta Txarlot dago super kontento. <i>And Chaplin is very happy</i>	Foc-Extrap.
1B06:178	Eta= kotxe gidaria sartzen da dendan, <i>And the car driver enters in the store</i>	Foc-Extrap.
1B06:197	Txarlot dago orrela eskuarekin, <i>Chaplin is like that with his hand</i>	Foc-Extrap.
1B06:56	Ta orduan ba gizona dago bere literan eserita, <i>And then well the man is sitting on his bunk</i>	Foc-Extrap.
1B06:7	ba= bueno banderita erortzen da lurrera, <i>well well the little flag falls down to the ground</i>	Foc-Extrap.
1B08:162	Ta neska gelditzen da lurtean. <i>And the girl is left on the floor</i>	Foc-Extrap.
1B08:163	Ta bera dijoa kafetegi batetara. <i>And he is going to a coffee house</i>	Foc-Extrap.
1B09:45	ta= beste iru umeak juten dia ortikan, <i>and the other three boys go along somewhere</i>	Foc-Extrap.
1B10:156	ba orduan polizia juten da telefonora, <i>well then the policeman goes to the telephone</i>	Foc-Extrap.
2A01:138	Eta= - <gizonak nai zuna=> zan <kartzelara jun>. Berriro. <i>And what the man wanted was to go to jail. Again</i>	Foc-Extrap.
2A01:169	eta bera juten da or zutik. <i>and he goes standing there</i>	Foc-Extrap.
2A05:1	Ba istoria onen protagonista da Txarlot. ez? <i>Well this story's protagonist is Chaplin, OK?</i>	Foc-Extrap.
2A05:6	Eta polizia zegon manifestazioaren aurka. <i>And the police was against the demonstration</i>	Foc-Extrap.
2A07:120	Orduan okiña juten da poliziana, <i>Then the baker goes to the policeman</i>	Foc-Extrap.
2A09:284	eta= neska gelditzen da etzakiyela zer esan eta ori, <i>and the girl's left not knowing what to say and stuff</i>	Foc-Extrap.

2A09:43	eta= Txarlot zijoan pues bandera aurretikan, <i>and Chaplin was going well in front of the flag</i>	Foc-Extrap.
2B03:2	pelikulie asten da mendixen- ... <i>the movie begins in the mountains</i>	Foc-Extrap.
2B06:32	Txarlot eroitzen da zulo batean, ... <i>Chaplin falls in a hole</i>	Foc-Extrap.
2B06:61	=e- aundi ori eta Txarlot daude maai berean <i>uh the big guy and Chaplin are at the same table</i>	Foc-Extrap.
1B05:62	bueno mutilak joaten dira beste alderuntz, ez? <i>well the boys go in the other direction, OK?</i>	Foc-Extrap.
2B08:2	pelikula asten da= gizon batekin, - <i>the movie begins with a man</i>	Foc-Extrap., Hesit.
2A09: 44-45	eta bea gelditzen da, aurrian balijoa bezela, banderakin, <i>and he is left,</i> <i>as if he was going in front, with the flag</i>	Foc-Extrap.; 2 IUs
2A09:74-76	Ta tipo oi zen, =ba= naiko= aundiya ze=n, <i>and that guy was,</i> <i>well</i> <i>he was quite big</i>	Foc-Extrap.; 2 IUs
2B01: 144-:146	eta Txarlot zeuen, <beakin batea zeldan zeuen> andi orren, ta beste txiki-txiki baten erdiyen. - <i>and Chaplin was,</i> <i>in between that big guy who was in the cell with him,</i> <i>and another tiny little guy</i>	Foc-Extrap.; 2 IUs
1A01:51	eta baserritarra - e= egoten da= e e= <ikaratua>. <i>and the farmer uh is shocked</i>	Foc-Extrap.; Hesit.
1A10:180	Eta gizona, joaten da= ba zutik. <i>And the man, he's going well standing up</i>	Foc-Extrap.; Hesit.
1A10:35	eta gizona ast- sartzen da e= <bersu - bouce d'egout ... ba - barruban>, <i>and the man beg- he enters uh inside a bouce d'egout</i>	Foc-Extrap.; Hesit.
1A16:27	ta Txarlot Txaplin eseritzen da= [-] e= txik- gizon txiki baten ondoan, <i>and Charlot Chaplin sits down uh sm- next to a small guy</i>	Foc-Extrap.; Hesit.
1B11:1	Ba= bideo au - au da - e= gizon batena, <i>Well this video is uh about a man</i>	Foc-Extrap.; Hesit.
2A07:126	Orduan Txarlot sartzen da= restaurante batian, <i>Then Chaplin enters in a restaurant</i>	Foc-Extrap.; Hesit.
2A08:1	Ba pelikulie san, - txorrada bat. <i>Well the movie was, a dumb thing</i>	Foc-Extrap.; Hesit.
2B01:83	preso oi zeuen= Txarloten onduen, <i>that prisoner was .. next to Chaplin</i>	Foc-Extrap.; Hesit.
2B03:70	ta bestioak ... juten dia= oñez, <i>and the others ... go on foot</i>	Foc-Extrap.; Hesit.

2B06:157	irugarren preso z zeo = n e .. e = osea zeldako .. burni oien ba kontra. <i>the third prisoner was uh I mean against the irons of the cell.</i>	Foc-Extrap.; Hesit.
2B08:125	gero, kanpoan, <beak nai dudena> da - - berri = z kartzela bueltatzea da. <i>then, outside, what he wants is - it is to go back to jail</i>	Foc-Extrap.; Hesit.
3A11:3	=Pelikula da= nekazari bati buruzkoa. ez? <i>the movie is about a farmer. OK?</i>	Foc-Extrap.; Hesit.
2A01:164	eta= @ gizona asten da *ooo*, <i>and the man goes "oh oh oh"</i>	Foc-Extrap.; Quote
1A05:85	eta bera joaten da m= - bueltak ematen, - <i>and he is going m= turning around</i>	Foc-Extrap.??, 2 nd pred?; Hesit.
1B01:118	ni egonen naiz zutik*. <i>I will be standing up</i>	Focus-S; Contrast.
1B06:200b	*ni, ni izan naiz la- lapurra, <i>I, I was the thief</i>	Focus-S; Contrast.
2A03:322	ni jungo naiz zutik.* <i>I will go standing up</i>	Focus-S; Contrast.
2A05:102	Txarlot gelditu zen ogiarekin. <i>Chaplin ends up with the bread</i>	Focus-S; Contrast.
2A09:175	Eta bera zizuan azkena. <i>and he was going last</i>	Focus-S; Contrast.
2B06:25	=eta bera dago aurrian. ez? banderarekin. - <i>and he is in front. OK? with the flag</i>	Focus-S; Emphasis
2B08:169	Ori sartzen da barruan, <i>that one goes inside</i>	Focus-S; Emphasis
1A05:14	ebe pilo bat jende etortzen da bere atzean, <i>uh lots of people are coming behind him</i>	Focus-S; Exist.
1B06:181	ogi geiago gelditzen da kotxean. <i>there is more bread in the car</i>	Focus-S; Exist.
2B01:328	eta= an ogi mordu(e) ta zeuden eskaparatin, <i>and there was a bunch of bread and stuff in the window</i>	Focus-S; Exist.
3A07:34	ogi zati bat zeon maixen, .. <i>there was a piece of bread on the table</i>	Focus-S; Exist.
3A09:2	Ba= gizon bat dau <zuaitz baten iyota>, <i>Well there is a man up on a tree</i>	Focus-S; Exist.
3A09:4	Ola zuaiza dau bide baten ertzen, <i>Thus there is a tree on the side of a road</i>	Focus-S; Exist.
1A02:9	eta manifestazioneko gizon batzuk etortzen dira atzetik. <i>and some men from the demonstration come from behind</i>	Focus-S; Present.
1A09:81	iru mutikoak pasatzen dira ortikan, <i>three boys go by there</i>	Focus-S; Present.
1B03:9	e mutiko bat pasatzen da bizikletarekin, <i>uh a boy goes by on a bicycle</i>	Focus-S; Present.
2A01:48	Eta= ogi zati bat zegon maian, <i>and there was a piece of bread on the table</i>	Focus-S; Present.

2A01:57	polizi bat dago atean, <i>there is a policeman at the door</i>	Focus-S; Present.
2A03:282	kabina bat zeon ola paretan, <i>there was a booth like that on the wall</i>	Focus-S; Present.
2A04:22	e iru mutil zeuden or inguruan, <i>uh there were three boys around there</i>	Focus-S; Present.
2A06:50	eta neska bat pasatzen da bizikletan. <i>and a girl goes by on her bicycle</i>	Focus-S; Present.
2A07:106	ta okiña ateatzen da neskan atzetik. <i>and the baker comes out behind the girl</i>	Focus-S; Present.
2A08:64	arrixa seuan parian, <i>there was a rock right there</i>	Focus-S; Present.
2B05:28	e neska bat pasatzeuk e= kontran. <i>uh a girl goes by in the opposite direction</i>	Focus-S; Present.
2B06:200	Kamioi bat etortzen da panaderikua, <i>A truck comes with the bread</i>	Focus-S; Present.
2B06:24	manifa bat komúnistena *rúum!* dator be- bere atzetikan, <i>A communist demonstration "boom" comes from behind him</i>	Focus-S; Present.
3A09:39	beste neska bat dator paren, - <i>another girl is coming next to him</i>	Focus-S; Present.
2B01:248	<Txarlotekin batea zeon> aundi ori atera da zeldatiken, <i>That big guy who was together with Chaplin has come out of his cell.</i>	Focus-S; Present.(2)
1A01:6	ta beste norbait etortzen da - bere= .. [...] (ardi batekin) <i>and someone else comes - with (... his sheep)</i>	Focus-S; Present.; Post-pres. predicate
1A09:7	Ta gero mutiko bat e= .. pasatzen da bizikletarekin, - <i>and then a boy goes by on his bicycle</i>	Focus-S; Present.; Post-pres. predicate
1A16:113	eta andre bat zegon - or <lapurtu duena>, <i>and there was a woman there who had stolen (something)</i>	Focus-S; Present.; 2 nd predicate
2B05:30	eta= - arri bat jók páre parian, <i>and there's a rock right there</i>	Focus-S; Present.; 2 nd predicate
3A10:18	Idea etortzen zaio burura. <i>An idea comes to his mind</i>	Focus-S; Present.; D-topic
2B06:118	Eta- preso guztiak sartzen dira kartzelara, <i>And all the prisoners enter into the jail</i>	Focus-S; Thetic
3A10:180	eta= i- bueno\ istilu pixka bat sortzen da an, ..	Focus-S; Thetic
3A10:80	beintzat ori konpontzen da orrela, - <i>at least that gets solved like that</i>	Focus-S; Thetic
2B06:115	denak dijoazten fila-indiaz. filan. - <i>all are going in single file. in line.</i>	Focus-S; Thetic (denak)
1B11:65	udare gu[θ]tiak erortzen zaizkio= - zakutik. <i>all his pears fall out from the basket</i>	Focus-S; Thetic; D-topic; afterthought?
2B04:46	Da beste= bueno polizia beste zea batzuk etortzen die, <orren billé>. <i>And another well the police some other guys come, looking for him</i>	Focus-S; X = afterthought?

2B08:14	Eta= ... kaletik b- kale erditik bera doa banderakin, <i>And along the stree he 's going with his flag</i>	Focus-S? Emphasis?
3A09:50	eta an ingurun ba beste iru mutiko dabiltz <jolasen>. <i>and around there well three other boys are (going) playing</i>	Focus-S??; Exist.; 2 nd predicate??
2B03:75	A jeisten da artian, - <i>he comes down meanwhile</i>	Setting-X; Topic-S
2A06:58	Mutilla altxatzen da lurretik, <i>the boy comes down from the ground</i>	V-focus?
2B01:446	ta Txarlot- Txarlot jun da ara, .. <i>and Chaplin- Chaplin goes there</i>	V-focus? Foc-Extrap.?