

UC Berkeley
Dissertations, Department of Linguistics

Title

The Interaction of Valence and Information Structure

Permalink

<https://escholarship.org/uc/item/4vj5j5tw>

Author

Ruppenhofer, Josef

Publication Date

2004

The interaction of valence and information structure

Ruppenhofer, Josef Karl

ProQuest Dissertations and Theses; 2004; ProQuest

pg. n/a

The Interaction of Valence and Information Structure

by

Josef Karl Ruppenhofer

M.A. (University of Colorado at Boulder) 1997

M.A. (University of California at Berkeley) 1999

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 Paul Kay

Professor Johanna Nichols

Professor Laura A. Michaelis

Fall 2004

The Interaction of Valence and Information Structure

Copyright © 2004

by

Josef Karl Ruppenhofer

Abstract

The Interaction of Valence and Information Structure

by

Josef Karl Ruppenhofer

Doctor of Philosophy in Linguistics

University of California, Berkeley

Professor Charles J. Fillmore, Chair

This dissertation investigates discourse-pragmatic differences between variably linked arguments appearing in alternating argument structure constructions in the sense of Goldberg (1995) and Kay (manuscript). The properties that are studied include givenness, pragmatic relation (topic/focus), salience of referents, animacy, and others. They derive from the literature on sentence-type constructions such as topicalization and from research on the referential properties of NP form types.

The research carried out here has multiple uses. At the most basic level, it serves as an empirical check on existing characterizations of the pragmatic properties of the relevant arguments that are the result of syntactic and semantic analysis based on introspection alone. For instance, for the epistemic raising alternation involving verbs like *seem*, the predicted topicality difference between the subjects of the raised and unraised constructions (Langacker 1995) could not be confirmed.

This dissertation also addresses the question what kinds of pragmatic factors, if any, are relevant to argument structure constructions. Based on the evidence of the dative alternation, it does not seem to be the case that the kind of pragmatic influences on argument structure constructions are different or limited compared to the ones found to be relevant to sentence-type constructions.

The kind of research undertaken here can also inform the syntactic and semantic analysis of constructions. In the case of the dative alternation, the discourse-pragmatic characteristics of the variably linked arguments provide evidence that

Basilico's (1998) analysis of the difference between the alternates in terms of VP-shells and a difference betweenthetic and categorical 'inner' predication, on the one hand does not account for all the data and on the other can be re-stated in pragmatic terms other than thethetic-categorical distinction.

In addition to studies of valence alternations, this dissertation also discusses various null instantiation phenomena, which provide further evidence for the need to specify discourse-pragmatic properties as part of argument structure constructions and lexical entries.

Finally, it is suggested that the use of randomly sampled corpus data and statistical modelling throughout this dissertation improves both empirical and analytical coverage.

Contents

List of Figures	iv
List of Tables	v
1 Introduction	1
1.1 Why study the discourse pragmatics of argument structure constructions?	3
1.2 Information structure and its operationalization	17
2 Dative alternation	27
2.1 Introduction	27
2.1.1 Discourse status	28
2.1.2 Grammatical weight	42
2.2 Give in the BNC	45
2.2.1 Possession transfer uses of <i>give</i>	46
2.2.2 Cases that contradict the predictions of weight and definiteness	61
2.2.3 Cases where no prediction is made	81
2.2.4 Sentences with two personal pronouns	90
2.2.5 Sentences with a personal pronoun Recipient and a demonstrative Theme	111
2.2.6 An unattested pattern: Theme <i>anything</i> plus Recipient <i>anybody</i>	121
2.2.7 Régression analysis	125
2.3 <i>Give</i> in the Switchboard corpus	135
2.3.1 Régression analysis	151
2.4 Non-concrete Themes with <i>give</i>	159
2.5 Other lexical units: <i>deny</i> and <i>donate</i> in the BNC	178
2.5.1 Deny	179
2.5.2 Donate	192
2.6 Summary	203
2.7 Application of pragmatics to syntax	205

3	Raising constructions	215
3.1	Epistemic raising	215
3.1.1	Langacker's active zone analysis	216
3.1.2	French raising and topicality	218
3.1.3	English data	222
3.1.4	Formal properties of raised and unraised constructions	227
3.1.5	Pragmatic properties of referents in raised and unraised constructions	235
3.1.6	Rhetorical structure	270
3.1.7	Semantics	273
3.1.8	Summary	282
3.2	Possessor raising	284
3.2.1	Possessor/Attribute factoring alternations	284
3.2.2	Body part possessor raising	288
3.2.3	English body part possessor raising	292
3.2.4	Morphology	294
3.2.5	Discourse properties of referents	295
3.3	Change property factoring alternation	332
3.3.1	Factoring as a non-basic sentence type	335
3.3.2	English corpus data	336
3.3.3	Formal properties	337
3.3.4	Discourse properties of referents	342
3.4	Summary	362
4	Null instantiation	364
4.1	Typology of non-overt arguments	367
4.1.1	Ellipsis versus omission	367
4.1.2	Obligatory versus optional omission	371
4.1.3	Constructionally or lexically licensed argument omission	372
4.1.4	Change in linking	375
4.1.5	Null instantiation	376
4.2	The interpretation omitted arguments	376
4.2.1	Constructionally licensed omissions with no change in linking	376
4.2.2	Valence changing constructions and null instantiation	383
4.2.3	Lexically licensed omissions	399
4.2.4	Representation of omitted arguments	421
4.3	Possible explanations	440
4.3.1	Lemma frequency	440
4.3.2	Degree of polysemy	441
4.3.3	Definiteness of the argument	443
4.3.4	Thematic role	445
4.3.5	Verb neutrality and basic level status of the object	446

4.3.6	Differential object marking	447
4.3.7	Taxonomic verb classes	449
4.3.8	Selectional strength	450
4.3.9	Aspect	453
4.4	A possible synthesis: semantic frames	457
4.4.1	Why aspectual classes are not enough	457
4.4.2	Frame semantics	464
4.4.3	Lexical pragmatics	471
4.4.4	Possible complications	476
4.4.5	Summary	477
5	Conclusion	479
A	Stativity and definiteness	498
B	Frames and Null Instantiation	500
B.1	Abundance	500
B.2	Health_response	501
B.3	Word_relations	501
B.4	Kidnapping	501
B.5	Filling	502
B.6	Manufacturing	502
B.7	Warrant	503
B.8	Measure_duration	503
B.9	Quitting	504
B.10	Damaging	504
C	Coding	505

List of Figures

1.1	Typology of argument-realization related constructions	7
1.2	Discourse statuses of referents	20
2.1	Results for Wasow's corpus study on the dative alternation	43
2.2	Results for Wasow's production experiment on the dative alternation	46
2.3	Thetic inner predication	207
2.4	Categorical inner predication	207
3.1	Subsequent mentions of Possessors in the conflated construction with <i>kiss</i>	307
3.2	Previous mentions of Attributes in the non-factoring construction with <i>increase</i>	347
3.3	Subsequent mentions of Possessors in the factoring construction . . .	348
3.4	Distance of closest subsequent mentions of Possessors in the non-factoring construction	355
4.1	Typology of unexpressed arguments	368
4.2	Minimal lexical entry for <i>marry</i>	431
4.3	Minimal lexical entry for <i>drink</i> ₂ 'drink alcohol'	433
4.4	Minimal lexical entry for <i>zubeissen</i> 'bite'	435
4.5	Basic lexical entry for <i>know</i>	436
4.6	Lexical entry for DNI-verb <i>know</i>	437
4.7	Bender 1999's HPSG analysis of instructional imperatives	438

List of Tables

2.1	Asymmetries between the two objects in the double object construction	35
2.2	Voice by construction	47
2.3	Realizations of Themes	48
2.4	Realization of Recipients	49
2.5	Prior mentions in the double object construction	49
2.6	Prior mentions in the caused-motion construction	50
2.7	Subsequent mentions in the double object construction	50
2.8	Subsequent mentions in the caused-motion construction	50
2.9	Prior evocation of Themes	51
2.10	Prior evocation of Recipients	52
2.11	Relative prior evocation of Recipients	53
2.12	Subsequent evocation of Recipients	53
2.13	Subsequent evocation of Themes	54
2.14	Relative subsequent evocation of Recipients	55
2.15	Relative length of frame elements by construction	56
2.16	Relative definiteness of frame elements by construction	57
2.17	Relative definiteness II of frame elements by construction	57
2.18	Relative definiteness III for Recipients by construction	58
2.19	Relative weight and relative definiteness for Recipients	59
2.20	Relative weight and relative definiteness for Recipients in active voice sentences	60
2.21	Relative weight and relative definiteness for Recipients in passive voice sentences	61
2.22	Orientation of purpose clauses with <i>give money</i>	70
2.23	Prior mentions for double object tokens	92
2.24	Subsequent mentions for double object tokens	92
2.25	Prior mentions for caused-motion tokens	93
2.26	Subsequent mentions for caused-motion tokens	93
2.27	Prior mentions for reverse double object tokens	94
2.28	Subsequent mentions for reverse double object tokens	94
2.29	Distance to prior mention in double object construction	96

2.30	Distance to prior mention in reverse double object construction	96
2.31	Distance to prior mention in caused-motion construction	97
2.32	Test statistics for distance to closest prior mention	97
2.33	Ranks for distance to closest prior mention	98
2.34	Number of prior mentions in caused-motion construction	98
2.35	Number of prior mentions in reverse double object construction	99
2.36	Number of prior mentions in double object construction	99
2.37	Ranks for number of prior mentions	100
2.38	Test statistics for number of prior mentions of Theme	100
2.39	Distance to closest subsequent mention in the double object construction	101
2.40	Distance to closest subsequent mention in the reverse double object construction	102
2.41	Distance to closest subsequent mention in the caused-motion construction	102
2.42	Test statistics for distance to next subsequent mention	103
2.43	Ranks for distance to closest subsequent mention	103
2.44	Number of subsequent mentions in the double object construction . . .	104
2.45	Number of subsequent mentions in the reverse double object construction	104
2.46	Number of subsequent mentions in the reverse double object construction	105
2.47	Test statistics for number of subsequent mentions	105
2.48	Ranks for number of subsequent mentions	106
2.49	Distribution of Recipient pronouns across the two constructions	108
2.50	Recipients occurring in double object construction with singular demon- strative Theme	111
2.51	Recipients occurring in caused-motion construction with singular demon- strative Theme	111
2.52	Prior mentions for the double object construction	113
2.53	Subsequent mentions for the double object construction	114
2.54	Prior mentions for the caused-motion construction	114
2.55	Subsequent mentions for the caused-motion construction	115
2.56	Properties of link referents in the double object construction (co-referential mentions or inference triggers)	123
2.57	Properties of link referents in the caused-motion construction (co-referential mentions or inference triggers)	124
2.58	Parameter coding	127
2.59	Classification table for model with relative length as only factor	128
2.60	Variables in the equation after the single regression step for the para- meter relative length	128
2.61	Classification table for model with relative definiteness as only factor	129
2.62	Classification table for model with two factors relative length and rel- ative definiteness	131
2.63	Classification table for model with four factors (absolute lengths of frame elements, relative definiteness, subject choice)	133

2.64	Uses of <i>give</i> in the Switchboard corpus	135
2.65	Constructions for all and only concrete uses of <i>give</i>	135
2.66	Distribution of constructions for different corpora	136
2.67	Voice by construction for all tokens	136
2.68	Voice by construction for concrete tokens	137
2.69	Realization of Recipients and Themes for all tokens	138
2.70	Realization of Recipients and Themes for concrete tokens	139
2.71	Realization of Theme frame elements across constructions	140
2.72	Realization of Recipient frame elements across constructions	141
2.73	Relative length by construction for all tokens	141
2.74	Relative length by construction for concrete tokens	142
2.75	Relative definiteness by construction for all tokens	142
2.76	Relative definiteness by construction for concrete tokens	143
2.77	Relative definiteness by relative length by construction for concrete tokens	144
2.78	Relative definiteness by relative length by construction for all tokens .	144
2.79	Prior evocation status of Recipients across constructions	145
2.80	Prior evocation status of Themes across constructions	146
2.81	Relative prior evocation status for all tokens	147
2.82	Prior evocation status of Themes by prior evocation status of Recipient	149
2.83	Number of subsequent mentions	150
2.84	Relative subsequent mentions for concrete tokens	150
2.85	Relative subsequent mention for all tokens	151
2.86	Classification table for model with six factors	151
2.87	Length of arguments in support verb uses of <i>give</i>	161
2.88	Relative length of arguments in support verb uses	161
2.89	Definiteness of frame elements for support verb uses	162
2.90	Percentages of definite and indefinite arguments with concrete uses of <i>give</i> in the BNC	163
2.91	Construction by frame element morphology for support verb uses of <i>give</i>	163
2.92	Construction by relative length for support verb uses of <i>give</i>	164
2.93	Relative length by relative definiteness	165
2.94	Relative length by relative definiteness measured at three levels	166
2.95	Voice by construction for support verb uses of <i>give</i>	167
2.96	Argument morphology for <i>give emphasis</i>	169
2.97	Length of arguments	172
2.98	Relative length of arguments for <i>give boost</i>	173
2.99	Definiteness of Recipient relative to Theme for <i>give boost</i>	173
2.100	Relative length versus relative definiteness	174
2.101	Relative length for non-concrete uses in the Switchboard corpus	176
2.102	Relative definiteness for non-concrete uses in the Switchboard corpus	176

2.103	Relative length by relative definiteness for non-concrete uses in the Switchboard	177
2.104	Uses of <i>deny</i>	179
2.105	Voice across constructions for <i>deny</i>	180
2.106	Subject choice for <i>deny</i>	181
2.107	Recipient morphology with caused-motion uses of <i>deny</i>	182
2.108	Recipient morphology with double object uses of <i>deny</i>	182
2.109	Theme morphology with caused-motion uses of <i>deny</i>	183
2.110	Theme morphology with double object uses of <i>deny</i>	183
2.111	Length of Themes and Recipients with <i>deny</i>	184
2.112	Longer frame element by construction for <i>deny</i>	184
2.113	Relative definiteness for <i>deny</i>	184
2.114	Relative length by relative definiteness for <i>deny</i>	185
2.115	Classification table for <i>deny</i> for model with relative length, definiteness status, and subject choice as factors	186
2.116	Morphology of frame elements with <i>donate</i>	194
2.117	Morphological realization of non-zero Recipients with <i>donate</i>	195
2.118	Length for Causes and Theme arguments with <i>donate</i>	195
2.119	Recipient length with <i>donate</i>	196
2.120	Relative lengths for frame elements of <i>donate</i>	196
2.121	Relative length for non-zero Recipients for <i>donate</i>	196
2.122	Prior evocation status with <i>donate</i>	197
2.123	Subsequent evocation status with <i>donate</i>	198
2.124	Number of prior mentions for <i>donate</i>	198
2.125	Distance to closest prior mention for <i>donate</i>	199
2.126	Properties of subsequent mentions for <i>donate</i>	200
2.127	Form of Recipient/Cause frame elements for <i>bequeath</i>	202
3.1	Achard's results for the referents of <i>sembler</i>	219
3.2	Non-raising uses of four verbal predicates	224
3.3	Raising uses of four verbal predicates	224
3.4	Uses of <i>likely</i>	225
3.5	Raising predicates in the Wall Street Journal Corpus	226
3.6	Copular subordinate predication	228
3.7	Presence of copula by construction	228
3.8	Construction by genre	229
3.9	Tense of raising predicates	230
3.10	Tense by construction	231
3.11	Results of two-tailed Fisher's exact and χ^2 -tests for tense by construction for raised and unraised constructions only	232
3.12	Temporal relationship between embedded predicate and raising predicate for <i>seem</i> in the written part of the BNC	232

3.13	Temporal relationship between embedded predicate and raising predicate for <i>likely</i> in the written part of the BNC	233
3.14	Temporal relationship between embedded predicate and raising predicate for <i>likely</i> recoded with fewer contrasts	234
3.15	Temporal relationship between embedded predicate and raising predicate for <i>appear</i> in the spoken part of the BNC	235
3.16	Animacy of subject referents	235
3.17	Animacy by construction	236
3.18	Subject length	238
3.19	Wilcoxon-Mann-Whitney tests for subject length between constructions	239
3.20	Definite and indefinite subjects in the raised and unraised constructions	240
3.21	Tests for definiteness by construction	240
3.22	Tests for the contrast between highly topical and less topical definites by construction	242
3.23	Subject form for <i>appear</i> in spoken part of the BNC	243
3.24	Subject form for <i>appear</i> in written part of the BNC	244
3.25	Subject form for <i>likely</i> in the written part of the BNC	245
3.26	Subject form for <i>seem</i> in the written part of the BNC	246
3.27	Subject form for <i>seem</i> in the spoken part of the BNC	247
3.28	Connection of subject to prior discourse	248
3.29	Evocation by construction	249
3.30	Evocation/inferability by construction	250
3.31	Subsequent evocation	252
3.32	Subsequent evocation (including inferable) by construction	253
3.33	Number of prior mentions	254
3.34	Prior mentions	255
3.35	Number of subsequent mentions	257
3.36	Subsequent mentions	257
3.37	Distance of prior mentions	258
3.38	Distance of prior mentions	259
3.39	Distance of subsequent mentions	260
3.40	Distance of subsequent mentions	260
3.41	Distance of last mentions	261
3.42	Distance of last mentions	262
3.43	Kinds of inferables	263
3.44	χ^2 -table for reduced set of inferables	264
3.45	Subject morphology for extended written <i>seem</i> sample	266
3.46	Wilcoxon-Mann-Whitney tests for contextual mention measures of extended sample for written <i>seem</i>	267
3.47	Variables not in the equation	269
3.48	Variables in the equation after step 1	270
3.49	Contrast scores for extended sample of written <i>seem</i>	272

3.50	Presence of Cognizer	275
3.51	<i>At first (glance/sight)</i> with <i>seem</i> and <i>appear</i>	277
3.52	Distribution of constructions for the sequences <i>would seem</i> and <i>would appear</i>	278
3.53	Distribution of constructions for the sequences <i>may/might seem</i> and <i>may/might appear</i>	279
3.54	Predicate affirmation and negation of <i>seem</i>	280
3.55	Affirmation and negation of the predicate in the clausal or verb phrase complement of <i>seem</i>	281
3.56	Distribution of tokens across constructions	292
3.57	Types of Body Part referents occurring in the constructions	294
3.58	Morphology of Possessors with <i>kiss</i>	296
3.59	Fisher's exact test for the morphology of Possessors	296
3.60	Morphology of Body Parts	297
3.61	Fisher's exact test for the morphology of Body Parts	297
3.62	Prior mention of the Possessor	299
3.63	Distribution of prior evocation status across the conflated and the raised constructions	299
3.64	Subsequent mention of the Possessor	300
3.65	Distribution of subsequent evocation status across the conflated and the raised constructions	301
3.66	Construction by prior evocation status of Body Part referents	303
3.67	Construction by subsequent evocation status of Body Part referents	303
3.68	Contextual Mentions of Body Part referents	304
3.69	Normal distribution of number of prior and subsequent mentions of Body Part referents across conflated and raised constructions for <i>kiss</i>	305
3.70	Normal distribution of number of prior and subsequent mentions of Body Part referents across conflated and raised constructions for <i>slap</i>	305
3.71	Normal distribution of number of prior and subsequent mentions of Body Part referents across conflated and raised constructions for <i>smack</i>	305
3.72	Contextual mentions of Possessors	306
3.73	Normal distribution of number of prior and subsequent mentions of Possessor referents across conflated and raised constructions for <i>kiss</i>	308
3.74	Normal distribution of number of prior and subsequent mentions of Possessors across conflated and raised constructions for <i>slap</i>	308
3.75	Normal distribution of number of prior and subsequent mentions of Possessors across conflated and raised constructions for <i>smack</i>	308
3.76	Within-construction comparison of participants for <i>kiss</i>	310
3.77	Within-construction comparison of participants for <i>slap</i>	311
3.78	Within-construction comparison of participants for <i>smack</i>	311
3.79	Comparison between <i>slap</i> and <i>kiss</i> in the raised construction	312
3.80	Comparison between <i>slap</i> and <i>kiss</i> in the conflated construction	312

3.81	Distance values for contextual mentions of participants in the conflated construction for <i>slap</i>	314
3.82	Distance values for contextual mentions of participants in the conflated construction for <i>kiss</i>	314
3.83	Distance values for contextual mentions of participants in the conflated construction for <i>smack</i>	315
3.84	Distance values for contextual mentions of participants in the raised construction for <i>slap</i>	315
3.85	Distance values for contextual mentions of participants in the raised construction for <i>kiss</i>	316
3.86	Distance values for contextual mentions of participants in the raised construction for <i>smack</i>	316
3.87	Comparison between participants in the conflated construction for <i>kiss</i>	317
3.88	Comparison between participants in the conflated construction for <i>slap</i>	317
3.89	Comparison between participants in the conflated construction for <i>smack</i>	318
3.90	Comparison between participants in the raised construction for <i>kiss</i> .	318
3.91	Comparison between participants in the raised construction for <i>slap</i> .	318
3.92	Comparison between participants in the raised construction for <i>smack</i>	319
3.93	Comparison between constructions for Body Part referents with <i>slap</i> .	320
3.94	Comparison between constructions for Possessor referents with <i>slap</i> .	320
3.95	Comparison between constructions for Body Part referents with <i>kiss</i> .	321
3.96	Comparison between constructions for Possessor referents with <i>kiss</i> .	321
3.97	Comparison between constructions for Body Part referents with <i>smack</i>	321
3.98	Comparison between constructions for Possessor referents with <i>smack</i>	322
3.99	Comparison between lexical units for Body Part referents in the conflated construction	323
3.100	Comparison between lexical units for Possessor referents in the conflated construction	323
3.101	Comparison between lexical units for Possessor referents in the raised construction	323
3.102	Comparison between lexical units for Body Part referents in the raised construction	324
3.103	Presence of other Body Part referents in the 10 sentence window of context	325
3.104	Mention of same kind of Body Part referent within 10 preceding sentences	326
3.105	Mention of same kind of Body Part referent within 10 following sentences	326
3.106	Mention of a Body Part referent within 2 preceding sentences	328
3.107	Mention of a Body Part referent within 2 following sentences for <i>kiss</i>	328
3.108	Grammatical function of Body Part mentions in preceding 2 sentences	329
3.109	Grouped grammatical functions for Fisher's exact test of Body Part mentions in preceding 2 sentences	329
3.110	Grammatical function of Body Part mentions in following 2 sentences	330

3.111	Grouped grammatical functions for Fisher's exact test of Body Part mentions in following 2 sentences	330
3.112	Tense of <i>increase</i>	338
3.113	Morphology of Attribute subject of <i>increase</i> in the non-factoring construction	339
3.114	Morphology of Possessor of <i>increase</i> in the factoring and non-factoring constructions	340
3.115	Evocation status of Attribute of <i>increase</i> in the non-factoring construction	343
3.116	Evocation status of Attribute in property factoring construction with <i>increase</i>	343
3.117	Fisher's exact test for evocations status of Attributes	343
3.118	Evocation status of Possessor of <i>increase</i> in the non-factoring construction	343
3.119	Evocation status of Possessor in property factoring construction with <i>increase</i>	344
3.120	Fisher's exact test for evocation status of Possessors	344
3.121	Prior and subsequent mentions for referents in the non-factoring construction	346
3.122	Prior and subsequent mentions for referents in the factoring construction	346
3.123	Kolmogorov-Smirnov test for normal distribution of number prior and subsequent mentions in the factoring construction	346
3.124	Kolmogorov-Smirnov test for normal distribution of number prior and subsequent mentions in the non-factoring construction	346
3.125	Presence of other Attributes and Possessors with <i>increase</i>	350
3.126	Type of closest clause-level argument in preceding two sentences with <i>increase</i>	351
3.127	Type of closest clause-level argument in following two sentences with <i>increase</i>	354
3.128	Grammatical function of closest clause-level argument in subsequent and preceding two sentences with <i>increase</i> in the non-factoring construction	354
3.129	Grammatical function of closest clause-level argument in two preceding and subsequent sentences with <i>increase</i> in the property factoring construction	356
3.130	Distance values for contextual mentions of participants in the non-factoring construction	357
3.131	Distance values for contextual mentions of participants in the factoring construction	357
3.132	Kolmogorov-Smirnov tests for distributions of closest contextual mentions of participants in the non-factoring construction	358
3.133	Difference between distance values for Possessors across constructions	358

3.134	Statistics for difference between distance values for Possessors across constructions	359
3.135	Difference between distance values for Attributes across constructions	359
3.136	Statistics for difference between distance values for Attributes across constructions	359
3.137	Difference between distance values for Attributes and Possessors in the factoring construction	360
3.138	Statistics for difference between distance values of Attributes and Possessors in the factoring construction	360
3.139	Difference between distance values for Attribute and Possessors in the non-factoring construction	361
3.140	Statistics for difference between distance values of Attributes and Possessors in the non-factoring construction	361
4.1	Percentage of null instantiating uses in given sense	423
4.2	Syntactic properties of various argument omission types	428
4.3	Lemma frequency in the British National Corpus, polysemy, and interpretation type for select null instantiating verbs	442
4.4	Null instantiation type and overall definiteness for select lexical units	444
4.5	Definiteness bias for some verb senses without the possibility of null instantiation	445
4.6	Realization of highly accessible referents for ordering sense of <i>follow</i> .	448
4.7	Realization of highly accessible referents for co-theme sense of <i>follow</i>	449
A.1	Distribution of predicate types in a random sample from the BNC . .	499
A.2	Relationship between stative sentence aspect and definiteness of NP objects	499

Acknowledgments

This dissertation has been very long in the making and it might well have taken longer to finish had my advisor, Charles Fillmore, not finally pushed me towards the exit, for which I am grateful. Much greater gratitude is owed, of course, for his intellectual contributions to this dissertation. This work could not have been done without the frame semantic grounding that I acquired in classes taken from him and, more importantly, in working with and for him on the FrameNet project. The work on null instantiation, in particular, builds on his ideas. Verbs and valences are one half of my major linguistic interests and discourse-pragmatics is the other. The interest in it was sparked by Laura Michaelis whom I met at the University of Colorado at Boulder while I was there on exchange from Germany. Her breadth of expertise, her enthusiasm for her research, and her efforts to engage her students were a major reason why I decided to stick around and study linguistics. Laura Michaelis was my advisor at Boulder and I am grateful that she continued working with me on discourse-pragmatic matters after I came to Berkeley.

I also say thanks to my other committee members, Paul Kay and Johanna Nichols. I was lucky that Paul Kay brought his NSF-project on color names to the International Computer Science Institute, thereby exposing himself to unscheduled consultations. I thank him for always being accessible and generous with his time as well as for being willing to join my committee relatively late in the game after I lost Andreas Kathol as a committee member, to whom I would like to express my deep gratitude for accompanying me on part of the journey. Johanna Nichols deserves my thanks for close readings and making careful comments on my chapters on a tight schedule. I also thank her for her encouragement of this line of research.

I also thank my past and present colleagues at the FrameNet project as well as the various visitors FrameNet has hosted through the years. They have provided a very stimulating, supportive, and playful atmosphere to work in and to indulge my love for the English lexicon. I particularly thank Collin Baker, Hans Boas, Michael Ellsworth, Susanne Gahl, Christopher Johnson, Miriam Petruck, Petra Steiner, Margaret Urban, Tess Wood, and Abby Wright. FrameNet is housed at the International

Computer Science Institute and part of a vibrant community of linguistically minded computational researchers. Of its members I want to specially acknowledge Johnno Bryant, Nancy Chang, Jane Edwards, Shweta Narayan, Srinu Narayanan, Madelaine Plauché, and Liz Shriberg. I wish I had been able to absorb more of what is to be learned from them.

In the context of my work at FrameNet I should also thank the National Science Foundation for providing funding to FrameNet under grant ITR/HCI #0086132. California taxpayers unwittingly spent money on my education in the form of some UC scholarships. California is truly a state of great opportunity for speakers of Bavarian dialects these days.

Life is not all work; some of it is also course-work. The Berkeley linguistics department has been a challenging but rewarding place to learn in. I am glad I had a chance to take classes from teachers such as Andrew Garrett and Eve Sweetser. The department also has an excellent administrative staff and I especially thank Belén Flores. Without her help I might well have missed various deadlines on the long way to graduation. My time in Berkeley was made much more enjoyable by fellow-students who became good friends. Special thanks are due Nurit Melnik who spent a lot of time with me on work and play and let me share in the life of her family. I also have fond memories of time spent with Ken Van Bik, Rosemary Beam de Azcona, and Michael Ellsworth, my close collaborator at FrameNet. I thank Tess Wood and Margaret Urban for enduring many games of Scrabble with this indecisive non-native player. For hospitality and friendship I also owe a great debt of gratitude to the whole Virsik family.

I also want to thank several wonderful people I know at Boulder. I have enjoyed the friendship and support of Michelle Gregory who has been an example of energy and dedication. I am also glad for having got to know Bill Raymond, a man of great erudition, with whom I have shared many silent cups of coffee. I thank Cynthia Girand for friendship, encouragement, and interesting book recommendations. In a pattern of reverse migration, my fellow Bavarian, former FrameNet colleague, and Gerhard Polt-aficionada Susanne Stadlbauer has gone from Berkeley to Boulder. Similarly, Norberto Lozoya-Escobar has gone from Boulder to Germany to replace me there. I

know that he is enjoying himself.

Last but not least I want to thank friends and family in my Bavarian fatherland—in particular Stefan and Laura Wagner, Ägidius Kreitmeier, and Christian Fleischmann—for staying in touch, providing friendship and support as well as distraction from my labors. This thesis is dedicated to my mother, Anna Ruppenhofer, and my favorite and only sister, Marianne Ruppenhofer, in love and gratitude.

Chapter 1

Introduction

In this dissertation I investigate, on the basis of randomly sampled corpus data, alternate argument linkings that are associated with different argument structure constructions in the sense of Goldberg (1995) and Kay (manuscript). Argument structure constructions are pairings of form and function that are used to express basic clauses encoding “semantic structures which reflect scenes basic to human experience” (Goldberg 1995:5). Although the focus of this dissertation is on argument structure constructions that participate in alternations such as the double object construction and the caused-motion construction, under the constructional view of grammar assumed here (Fillmore & Kay 1999, Goldberg 1995 Michaelis & Lambrecht 1996), the alternate constructions that, under certain circumstances, participate in paraphrase or near-paraphrase relationships are not assumed to actually be syntactically derived from each other (Goldberg 2002). In fact, the English *Way*-construction illustrated below in (1.1) (Goldberg 1996, Israel 1996) as well as the resultative constructions exemplified in (1.2) (Boas 2000, Goldberg & Jackendoff 2004) do not come paired with other argument structure constructions. Neither does the very basic transitive construction, exemplified in (1.3) (Goldberg 1995:117-119).

(1.1) City Councilman Uvaldo Martinez **ate** his way out of office on bogus meal-expense charges to city credit cards.

(1.2) Then along came a kangaroo and **boxed** her to death and then an alligator

came and ate the kangaroo and lived happily ever after.

- (1.3) Admittedly I am writing this from a very biased point of view, having just **devoured** a deluxe five course meal in Hall.

The focus of the investigation will be the discourse-pragmatic properties of the variably linked arguments. The methodology used in this enterprise is borrowed from the rich discourse-pragmatic literature on sentence-type constructions such as left-dislocation, topicalization, locative inversion, and preposing constructions, and on the referential properties of NP form types. The relevant properties that are studied include givenness, pragmatic relation (topic/focus), salience of referents, animacy, and others. They will be discussed in greater detail in section 1.2 below.¹

The first of the valence alternations that will be examined is the so-called dative alternation exemplified in sentences (1.4a) and (1.4b). In the double object construction in (1.4a) the Recipient (*Peter*) appears as the first object and the Theme as the second object, in the caused-motion construction in (1.4b) the Recipient appears as the complement of a preposition and the Theme as the direct object.

- (1.4) a. John **sent** *Peter* the package. (Double object)
 b. John **sent** the package *to Peter*. (Caused-motion)

The second type of constructional alternation that I examine is exemplified by the epistemic raising alternation in (1.5). In the unraised construction in (1.5a) a propositional argument of the raising predicate is iconically encoded as a single clausal argument. The unraised construction often combines with the extraposition to yield structures like the one in (1.5b). The two kinds of unraised syntactic constructs contrast with the constructs licensed by the raised construction, exemplified by (1.5c), where the 'subject' *Luke* of the clause expressing the propositional argument has, in a manner of speaking, been raised out of that clause to function as the subject of the raising predicate *likely*, leaving behind the *to*-marked infinitive phrase *to win*.

¹It should be noted that the methodology used is valid and an interpretation of the findings is needed regardless of whether one posits a lexical or a syntactic relationships or no relationship at all between the alternating constructions.

Thus, in raised constructs, the single propositional semantic argument of the raising predicate is realized as two syntactic clause-level arguments of the predicate.

- (1.5) a. That *Luke* will win is **likely**. (Unraised)
 b. It is **likely** that *Luke* will win. (Unraised+extraposed)
 c. *Luke* is **likely** to win. (Raised)

In addition to the epistemic raising alternation, I analyze two other members of the family of raising alternations, namely the possessor raising alternation, exemplified in (1.6)) and the property factoring alternation, exemplified in (1.7).

- (1.6) a. I feel like **punching** someone's nose. (Unraised)
 b. I feel like **punching** someone on the nose. (Raised)
- (1.7) a. The price of technology has been plummeting while the price of oil has been **skyrocketing**. (Unraised)
 b. Within 12 months of that conversation in early 2000, oil **skyrocketed** in price (faster than I expected) to the \$35 to \$40 area. (Raised)

1.1 Why study the discourse pragmatics of argument structure constructions?

What reason is there to look at discourse-pragmatic factors at all and what benefit can be expected from the kind of research proposed here? I will address these two questions in order. The constructional alternations in (1.4) and (1.5) have mostly been analyzed in terms of their semantic and syntactic properties. They pose problems for theories of *linking*, that is theories describing how phrases expressing the necessary conceptual participants in the eventuality denoted by predicates are mapped to the grammatical functions of the clauses which the predicates head. For the most part, such theories focus on the semantic generalizations that can be made across (classes

of) verbs. These generalizations are usually stated in terms of thematic roles (Fillmore 1968), in terms of macro- or proto-roles (Dowty 1991), in terms of a lexical decomposition into primitive predicates (Dowty 1979, Hale and Keyser 1993, Wunderlich 1997, *inter alia*), or in terms of a combination of thematic roles and lexical decomposition (Koenig and Davis 2001).

The problem that the constructional alternations described above pose for linking alternations is that two or more different argument frames need to be licensed with what appears to be the same set of thematic roles (or, alternatively, the same lexical decomposition). If both alternates are given the same semantic analysis, then one needs to have two different linking rules and has to explain under which circumstances they apply. One possibility is to have a marked and an unmarked linking rule, as proposed by Foley and Van Valin (1984). Foley and Van Valin treat the double object form as a marked linking violating the generalization that English allows at most two core arguments, subject and object. The double object construction is said to exceptionally involve three core arguments, the Recipient being the Undergoer and the Theme being 'a simple Theme, even though it is a core argument' (Foley & Van Valin 1984:89). The fact that the first object in the double object construction is an NP unmarked by a preposition is taken to be crucial evidence for the syntactic core status of the Recipient. By contrast, since the Recipient is prepositionally marked in the Caused-motion construction it can be treated as non-core. The unmarked status of the Caused-motion construction is seen to be supported, too, by the fact that only the Recipient may be null instantiated.² Foley & Van Valin do, however, not give an explanation for when speakers use the marked linking.

Alternatively, if one assigns different semantic representations to the alternates, one can apply a single, possibly universal linking rule to both of them. However, in certain cases the assignment of different semantic representations may be hard to justify. Langacker, for instance, suggests that the alternates of the dative alternation differ in what semantic notion they highlight: (1.8a) is said to highlight the path

²On the view laid out in the null instantiation chapter, this argument is actually mistaken as the omissibility of the Recipient is restricted to a particular charitable-donation sense of *give*. It is not a general property of ditransitive verbs.

that the Theme traverses whereas (1.8b) is said to emphasize the resulting state of possession (1987:51).

- (1.8) a. I **sent** a book to the library
 b. I **sent** the library a book

The semantic explanation is intuitively appealing in that it predicts that in the case of (1.8b) the library has to be understood as an institution capable of possessing something rather than as a mere location. However, it does not account for cases where no physical transfer occurs but still both constructions can be used, as in (1.9).

- (1.9) a. In 1963 Emperor Haile Selassie I **gave** 500 hectares of land to any African that wished to return to Ethiopia.
 b. The local administration **gave** them 100 hectares of land left vacant since the collapse of the collective farm system.

And even when there is concrete transfer involved, non-semantic conditions must be assumed to account for certain non-occurring combinations:

- (1.10) a. ***Give** that book me
 b. ***Give** that me.

Given these problems it seems desirable to explore additional factors that may differentiate the alternate linkings. The discourse-pragmatics factors considered here have the advantage of being conceptually tractable and their influence is relatively easily verifiable compared to some purely semantic or conceptual notions.

Another reason to look towards discourse-pragmatic factors in characterizing argument structure alternations is the fact that other construction types that relate to argument realization also often have primary or secondary pragmatic functions. This is essentially an argument of plausibility. For the purpose of exploring argument realization-related constructions, let us assume the typology presented in Figure 1.1. This typology largely follows the proposal of Dixon and Aikhenvald (1997) but also

includes argument structure constructions and contrasts argument realization constructions with what we may call discourse function constructions. Of course, the adequacy of the typology in Figure 1.1 is subject to debate. This author has in fact collaborated on a monograph treating the German *be*-prefix as an applicative argument structure construction, which in some cases simply provides an alternate linking, in others creates or augments valences (Michaelis & Ruppenhofer 2001). From the construction grammar point of view it is not by itself significant that one argument structure construction has one more core argument position than another one in which a given verb can appear. For the sake of the current discussion, however, it is not necessary to work out a typology that is adequate from a construction grammar point of view. It should also be noted that Dixon & Aikhenvald's is more complex than can be represented here and that they consciously based their classification of constructions on cross-linguistically proto-typical properties, fully aware of non-prototypical constructions.

Of Dixon and Aikhenvald's (1997) set of four argument realization related construction types, the argument transferring type is the most familiar to a speaker of an Indo-European language. Argument transferring constructions involve adding an argument to, or removing one from, the core of a transitive clause. In languages with nominative-accusative case systems, the valency-reducing constructions which remove an A-argument from the clausal core are called passives.^{3,4} In syntactically ergative languages, the valency-reducing constructions that remove an O-argument from the clausal core are called antipassives. The two major valency-adding constructions do not seem to correlate with a language's case system but only with the kind of argument they add to an intransitive clause: constructions that add an A-argument are called causatives and constructions that add an O-argument are called applicatives.

³In the typological tradition, especially work associated with R.M.W. Dixon, it is customary to refer to the subject of transitive clauses as A, to their objects as O and to the subjects of intransitive clauses. The abbreviation A is semantically inspired by the most prototypical cases of transitive clauses but is in fact not meant to be restricted to volitional human Agents. All subjects of transitive clauses are intended to be covered (Dixon 1994:6-8)

⁴Valency-reducing constructions are the proto-typical cases of voice constructions. Dixon and Aikhenvald (1997) in fact suggest that the term voice be restricted only to valency-reducing constructions.

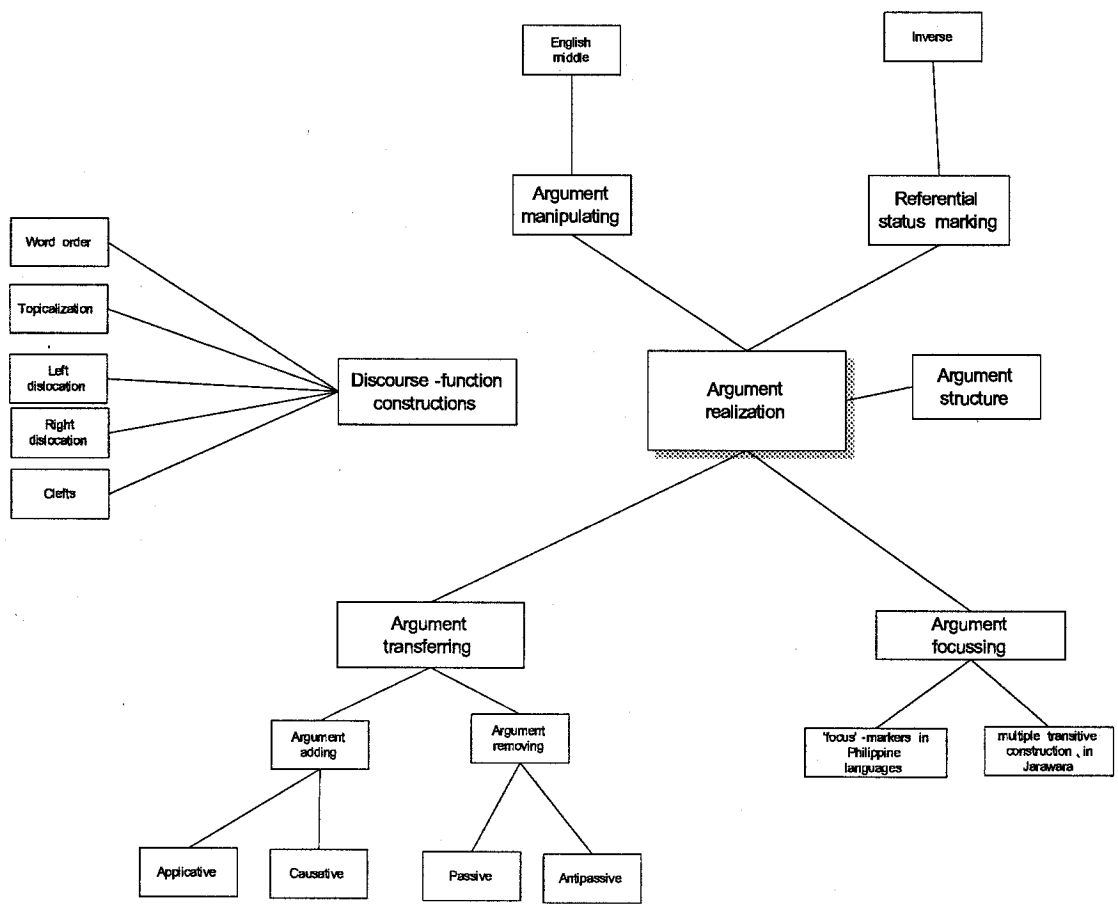


Figure 1.1: Typology of argument-realization related constructions

- b. Miotos_S [∅ ki-joma-ke-ka], ∅_O Watati_A [∅ ∅ hi-wa-hi-ka]

Miotos_S ∅
name(MASC) 3Sg+S
ki-joma-ke-ka, ∅_O
be in motion-through gap-COMING-DEC+FEM 3SG+O
Watati_A ∅ ∅ hi-wa hi-ka
name(FEM) 3SG+O 3SG+A Oc-see Oc-DEC+FEM

‘Mito came in and Watati saw him’ (= *ibid.*, (33))

Argument manipulating constructions are defined as those that bring a non-subject argument into surface subject position so that it has some but not all of the properties of a prototypical subject. English is said to exhibit this kind of construction in allowing instruments, patients, locatives and potentially any non-subject NP to become subjects, as is shown in (1.12b-c):

- (1.12) a. My steel-tipped boots **kick** (footballs) well.
b. It has its original leather seats and **drives** superbly.
c. Studio B **recorded** Oscar Peterson well.

The basic constraint seems to be that the element be part of the causal chain of bringing the event about. Typically, the sentences are modalized in some way, for instance, by an adverb, a modal, or by negation. What is usually called the English middle construction, exemplified in (1.12b) is said to be just a special case of a broader phenomenon.

The last construction type that Dixon and Aikhenvald (1997) identify are ones that serve to mark the referential status of arguments. In inverse systems, the alternating constructions do not reflect a difference in transitivity, the grammatical functions of the arguments are not changed, and neither of the constructions can be viewed as basic. The ‘point’ of the constructions is to mark the controller of the action: so-called direct marking is used when the relation between argument reference and

grammatical function are aligned in the expected way and inverse marking is used when this is not the case.⁵ An example from Cree is given in (1.13).

- (1.13) a. ni-wapam-aw
 1SG-see-DIRECT+1SG
 'I see him/her'
- b. ni-wapam-ik
 1SG-see-INVERSE
 he/she sees me (= Dixon and Aikhenvald 1997:99, (45a-b))

In Figure 1.1, I added discourse-function constructions to the typology suggested by Dixon and Aikhenvald (1997). The discourse-function constructions I have in mind—topicalization, left and right dislocation, clefts, etc.—serve to manage the discourse status of referents as well as to mark pragmatic relations. They typically do not involve marking of predicates and no change in valence or linking occurs, at least in familiar European languages such as English, French, Italian, and German. English examples are given in (1.14).

- (1.14) a. That kind of thing, I don't like. (Topicalization)
- b. You don't have to worry ... the children, they are **OK** (Left-dislocation)
- c. They're really **beautiful**, the clothes they sell. (Right-dislocation)
- d. It was Smithers who **gave** her a clue about the election. (*it*-cleft)

While discourse-function constructions by definition serve to manipulate the discourse status and the pragmatic relations of referents, some of the argument-realization constructions have discourse-related functions. Clearly, this is the case for argument focusing constructions and for most inverse systems. Argument transferring constructions, too, are not only used for semantic reasons but also to manage the flow of discourse. Haspelmath (1990), for instance, identifies agent backgrounding and

⁵The terms *inverse* and *direct* are used to refer to the markings typically appearing on verb stems in languages with inverse systems while the terms *proximate* and *obviative* refer to the respectively foregrounded and backgrounded discourse status of third-person referents.

patient foregrounding as two secondary functions of passive constructions. Dixon and Aikhenvald (1997) suggest the opposite for antipassive constructions: they background the patient and foreground the agent. Mithun (2000) argues that in Yup'ik causatives may be used to preserve topic continuity. Dixon & Aikhenvald argue that the use of applicatives, too, may be motivated by topic continuity, citing a 1981 study by Dixon of the Australian language Wargamay (2000:17).

The fact that a large number of construction types related to argument realization primarily or secondarily serve to fulfill pragmatic functions should make it at least plausible that argument structure constructions, too, do not solely have semantic content but may also impose pragmatic constraints.

I now turn to the question what benefits one can derive from engaging in discourse-pragmatic research on alternations. At a minimum, this research serves as an empirical check on existing characterizations of the pragmatic properties of the relevant arguments that are the result of syntactic and semantic analysis based on introspection alone. For instance, in the case of the epistemic raising alternation, it was predicted by Langacker (1995) that in sentences licensed by the raised construction, the referents of raised NP would tend to be more topical than their counterparts in sentences licensed by the unraised construction. My analysis of English corpus data for several raising predicates suggests that there is no evidence that this conclusion actually holds, which suggests that linguists' intuitive judgments about pragmatic properties are not always reliable and underscores the need for empirical grounding of theoretic analyses whenever possible. My results for the English data also contrast with Achard's (2000) results who concluded from a study of a small corpus of French news texts that Langacker's prediction held for the French epistemic raising verb *sembler* 'seem'. The different findings may be due either to incompatible methodologies, which cannot be ruled out, or they may point to the fact that generalizations about pragmatic properties of constructions have to be language-specific even if the constructions at issue share many syntactic and semantic characteristics.

The research carried out on the family of raising constructions also allows us to ask to what extent there are pragmatic family resemblances. Thompson has, for instance, asserted that with alternations "where a core-oblique OPTION exists in

a language that option is influenced by information flow parameters in discourse” (1997:76). The family of raising alternations that I investigated do not pattern in the same way. One of them, the change property factoring alternation shows consistent evidence that pragmatic factors play a role in choosing between the alternants, but the other two kinds of raising construction, epistemic raising and body part possessor raising, do not.

The research in this dissertation can also address the question what kinds of pragmatic factors, if any, are relevant to argument structure constructions. It might, for instance, be the case that either only pragmatic relations such as topic and focus or only givenness plays a role. If that was the case, then the range of pragmatic factors that influences argument structure constructions would be smaller than that for sentence-type constructions such as Left-dislocation, Topicalization, etc. However, the analysis of the dative alternation in chapter 2 shows that both accessibility and a preference for non-focality of the Recipient have an influence on the choice of argument structure construction (in addition to other factors, namely relative grammatical weight of Recipient and Theme, choice of sentence topic, and idiomaticity of the Theme).

Research into the pragmatic properties of variably linked arguments also helps us evaluate what role, if any, discourse pragmatic factors may play in the choices speakers make between the constructional alternates. If we find no discourse-pragmatic differences between the referents of NPs realizing any of the variable semantic role in the alternating constructions, we may conclude that speakers do not choose between the alternates on a pragmatic basis. If we do find differences, we need to further ask—at least in cases where this is relevant such as the dative alternation—whether the differences we do find are simply reflective of pragmatic properties associated with grammatical functions or whether they properly attach to specific mappings between semantic roles and grammatical functions.

The type of research undertaken here can also inform the syntactic and semantic analysis of constructions. In the case of the dative alternation I will argue that the discourse-pragmatic characteristics of the variably linked arguments provide evidence that Basilico’s (1998) syntactico-pragmatic analysis of the difference between the

alternates in terms of a difference between embedded focus structures that are aligned with syntactic layering of VPs does not explain all the data and should be re-stated in terms of discourse status rather than focus structure. As the argument is rather complicated, I will only give a brief summary here and refer the reader to section 2.7 for more detail.

Basilico's idea is that, paralleling the VP-internal subject hypothesis in the generative tradition, one can also assume VP-internal and -external positions for objects. These objects are further assumed to be subjects of so-called 'inner' predications consisting of the verb and any indirect complements. The VP-internal versus VP-external position of the object is taken to represent a difference in the type of focus structure of the inner predication. With categorical (topic-comment) predications objects are generated VP-externally; withthetic (presentational) predications, objects are generated VP-internally and then move outside of the VP. Thethetic-categorical distinction is further assumed to correspond to the stage-level v. individual-level distinction observed for subjects.

Basilico argues that the syntactic-semantic analysis in terms of object position allows an account of four kinds of asymmetries that can be observed between the alternates of the dative alternation (and also between the alternates of the locative spray/load and the creation/transformation alternations). These asymmetries have to do with quantifier scope relative to the second non-subject complement, scope of the object relative to frequency adverbs, ability of the object to appear in inversion constructions, and the interpretation of the object when part of *there*-insertion constructs as either individuated or part of a scene presented as an unanalyzed whole. Thethetic constructional alternates are taken to be more flexible in these asymmetries than the categorical constructional alternates. Thethetic constructions, for instance, allow the object to outscope the other non-subject complement or to be outscoped by it while the categorical constructions require that the object outscope the other non-subject complement; objects ofthetic predications can appear in inversion constructions but not objects of categorical predications; etc.

However, as Basilico himself points out, his account has a problem in that the objects of categorical constructions can be non-specific, which is unexpected if the

analogies between subjects and objects, and thetic versus categorical predication and stage versus individual-level predicates really hold: just as individual-level predications require specific subjects so objects occurring with categorical inner predications should be specific. Basilico observes that the members of the intransitive locative alternation—e.g., *Bees swarm in the garden* (thetic) versus *The garden swarms with bees* (categorical)—exhibit the same asymmetries as the members of the transitive locative alternation but that they do show the expected specificity restriction of the subject in the categorical construction. Based on this observation, Basilico reasons that what distinguishes the transitive from the intransitive locative verbs is that the former include a BECOME-operator whereas the latter as stative verbs lack such an operator. He further generalizes that with the categorical alternants of his three transitive alternations it is the presence of additional primitive predicates in their syntactic representations that allows for an overlay of theticity (via existential binding by an operator) over the inner categorical predication.

While superficially successful, Basilico's analysis is not complete. It is never spelled out when the higher BECOME-operator is called upon to provide existential binding. The account in terms of the thetic-categorical focus-structure opposition also neglects the third type of focus structure, argument focus as in (1.15).

(1.15) A: Who did you give the money to?—B: I **gave** it to Bill.

Another problem is that Basilico's account has no explanation why it should be the case that the caused-motion construction has to be used when the Theme is a pronoun. It does not follow from any pragmatic theory that the categorical double object construction should not be able to have a topical non-subject argument as part of its inner predication. A further problem of Basilico's account is that there are other categorical/individual-level intransitive predicates such as *turn* that have a BECOME-operator as part of their lexical representation but do not allow non-specific interpretation of their subjects.

(1.16) A frog **turned** into a prince.

The lack of a non-specific reading for the indefinite subject NP of (1.16) is unexplained. These and other problems argue that Basilico's account is not right. In section 2.7, the analysis is laid out in more detail and an explanation of the data in terms of discourse status rather than focus structure is suggested.

The study of the dative alternation also illustrates methodological benefits of corpus-based work. One is that the analyst very often finds patterns in the corpus that they are not aware of, or that they at least would not readily retrieve by introspection. The data for the verb *give*, for instance, showed that in British English it is also possible for Theme and Recipient to appear in that order as two bare NPs post-verbally as in (*He gave it me and I rushed out*). Moreover, the corpus data also immediately make it clear that this pattern only occurs when the Theme is a pronoun, a fact that a speaker who is generally aware of the existence of the 'reverse' double object pattern might not remember. Another benefit of corpus-based methodology combined with statistical modelling is that the analyst can keep track of how good the explanations provided really are. In the dative alternation, for instance, it was possible to eventually correctly predict the constructional choice for 96.3% of the sentences in a random sample from the Switchboard corpus of telephone conversations. Throughout the process of refining the model, inspection of sentences for which the prediction was wrong in several cases allowed the analyst to be alerted to factors not yet taken into account. The study of incorrectly predicted cases also showed that speakers sometimes choose what in the judgment of the analyst and colleagues was only the second or third-best option available. This last fact again underscores that while corpus-study is a very helpful methodology, the data analysis still requires a lot of careful judgment. Some tokens in a sample need not be explained by a linguistic model because they just represent poor linguistic choices. In other cases, patterns that are grammatical just happen not to be represented in a sample. In such a situation the analyst can retrieve instances of the undocumented pattern on the Internet, which I did for the sequences *give anything to anybody* and *give anybody anything*, or, as a last resort, construct their own examples, which I never found necessary.

In a somewhat separate strand of analysis, I also examined several types of argument omissions occurring in English, focusing in particular on the two types of

lexically licensed Null instantiation (Fillmore 1986). They are lexically governed, but they also require particular pragmatic contexts to be felicitous. In cases of Definite null instantiation (DNI) such as (1.17), the hearer has to be able to recover from prior context a specific referent for the omitted argument, whereas in cases of Indefinite null instantiation (INI) such as (1.18) the hearer needs to understand only that the missing argument is of some canonical type.

(1.17) Ike **contributed** \$20 $\emptyset_{\text{Recipient}}$. (DNI)

(1.18) Ike **contributed** \emptyset_{Theme} to the Red Cross. (INI)

One can think of null instantiation and other omissions as one of the alternates of overtness alternations. The focus of the investigation with the lexically licensed alternations is not so much on when either an overt or a null form is actually produced in discourse but on what kinds of regularities there are, if any, with respect to the predicates that allow null instantiation. I will argue that although the possibility of null instantiation for a particular argument of a particular predicate cannot be predicted, regularities about the interpretation of arguments omitted under lexical licensing are best captured in terms of verb classes, or more specifically semantic frames, rather than in terms of such notions as aspectual class, selectional strength, frequency and others. For instance, Recipients of transfer verbs are omissible only in the caused-motion construction and only if the verb has a charitable donation-meaning; *I gave a hammer* is either ungrammatical or it refers to a very unusual case of donation.

In addition to the lexically licensed omissions, I also examine omissions licensed by particular constructions such as the imperative (1.19) and the passive (1.20) as well as omissions that depend on particular genres such as instructional imperatives (1.21).

(1.19) \emptyset_{Agent} **Drop** it into your pocket or toss it to a friend.

(1.20) Among the more dramatic stories was that of a local official from the independent railway union who was **defenestrated** \emptyset_{Agent} from his ninth-floor flat.

(1.21) Add coffee and **stir** \emptyset_{Theme} until dissolved

While some of the constructional omission types share properties, overall they differ from each other in several respects, including, for instance, the interpretation of the omitted argument or the ability of the zero argument to antecede a reflexive. The idiosyncrasies found among the omission constructions prevent a general analysis covering all types and argue instead for constructional analyses that specify semantic, syntactic, and pragmatic properties for each of them separately. That the null instantiation constructions are different from each other is both theoretically predicted by, and easily accommodated within a constructional view of grammar that combines syntactic, semantic, and pragmatic properties in linguistic signs.

1.2 Information structure and its operationalization

In the remainder of this introduction, I will give a brief overview of the understanding of information structure assumed in this work. It is owed to a very large degree to the work of Knud Lambrecht (1994, 1995, 1997, 2001) and Ellen Prince (1981, 1992, 1997).

Following Lambrecht (1994), I take information to be a relational concept. Neither referential expressions nor predicates by themselves inform. It is only their combination that does. In particular, what is combined typically consists of a so-called pragmatic presupposition and a pragmatic assertion, which are defined as follows by Lambrecht:

PRAGMATIC PRESUPPOSITION: The set of propositions lexicogrammatically evoked in a sentence which the speaker assumes the hearer already knows or is ready to take for granted at the time the sentence is uttered.

PRAGMATIC ASSERTION: The proposition expressed by a sentence which the hearer is expected to know or take for granted as a result of hearing the sentence uttered. (1994:52)

Informationally, sentences basically come in three subtypes. The most basic kind of sentence informs the hearer of something about a topic. This is the basic topic-comment sentence in which the subject is the sentence topic and the predicate is the comment. An example is given in (1.22).

- (1.22) When we went to Disney Land in California several years ago, we were standing in line for the ridiculously overpriced admission tickets.

The second type consists of a presupposed open proposition, which may have been explicitly evoked or be simply inferable, and a variable. The clearest examples of this type are answers to *wh*-questions.⁶

- (1.23) A: Who did you vote for?
B: I voted for Smithers.

In (1.23), when B answers, the proposition that “B voted for X” is given in the context and speaker B simply supplies the filler of the variable slot.

The third type is illustrated by B’s answer in (1.24) and by example (1.25).

- (1.24) A: What happened, why are you late? B: My car broke down.
(1.25) There’s a zebra!

In these sentences, there is nothing presupposed, neither awareness of a participant nor of a proposition.

The three types are, in order, referred to as predicate focus, argument focus, and sentence focus. The naming indicates for each sentence type which part of the sentence encodes the semantic component of a pragmatically structured proposition whereby the assertion differs from the presupposition. The three sentence types also answer different questions: *What happened to this entity?*; *Some entity was involved in this eventuality: which entity?*; *What happened?*.

⁶The questions themselves arguably also involve presupposed open propositions. But they are presupposed only in the sense that the addressee is assumed to know the answer rather than being currently consciously aware of the open proposition.

The issue of how a sentence informs is independent of the discourse status of the referents involved. This is not to deny that there are correlations between the discourse status of a referent and whether that referent tends to be referred to in the presupposition or assertion of a proposition. But in principle the discourse status of referents and the focus type of a sentence are independent. This can be illustrated with a set of examples. A speaker can use a predicate focus sentence to say something about a referent that is textually or situationally given.

- (1.26) She did not confine her nursing to the clinical situation. **She** went out to the community and tried to solve the social problems that lead to disease.

But a given referent can also appear in argument focus sentences in the context of a presupposed open proposition.

- (1.27) A: Which of you did it?
B: **She** did.

And a textually or situationally given referent can also appear in sentence focus sentences:

- (1.28) **My** car broke down.

Maintaining the distinction between focus type and discourse status of referents will be particularly important in the analysis of the dative alternation.

The discourse statuses that referents, or, more precisely, mental representation of referents, can be characterized as shown in Figure 1.2, which follows Lambrecht's classification (1994:109).

Discourse statuses can be subdivided by two major criteria, identifiability and activation. Identifiability relates to a speaker's assessment of whether their interlocutor has a mental representation of the referent already stored in their mind or not. If it is assumed that the hearer does possess a representation for the referent, then the speaker also has to make an assessment of the activation status that the referent has in the mind of the hearer at speech time. The referent may be *inactive*, which Lambrecht intends to be equivalent to Prince's (1981) status *unused*. The referent may

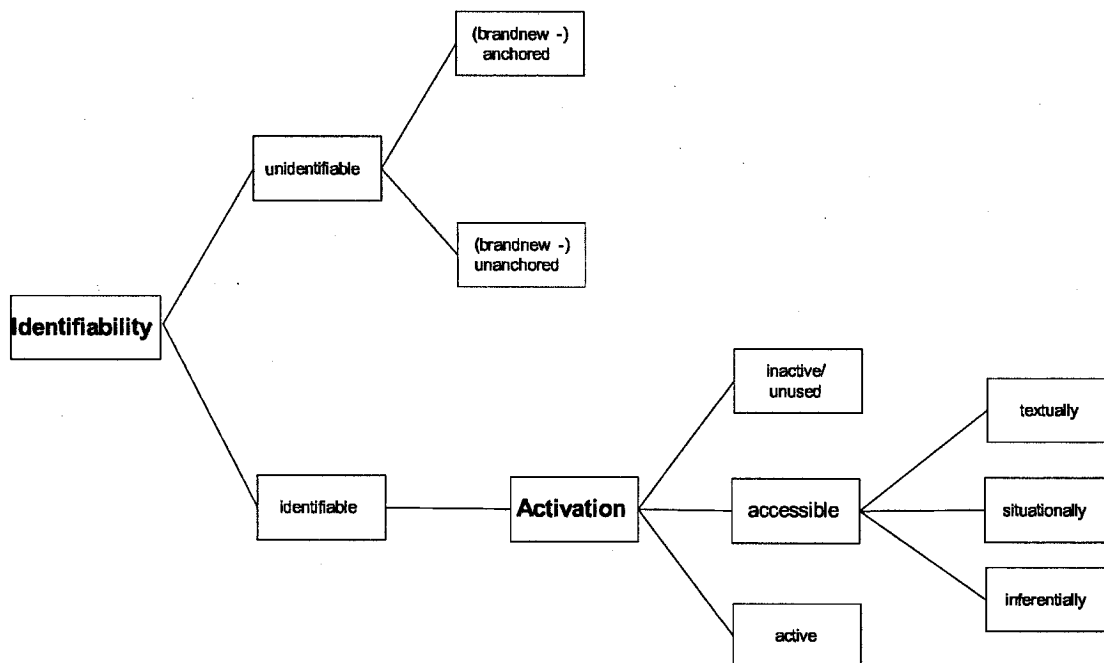


Figure 1.2: Discourse statuses of referents

be *active*, that is, it may be assumed to be currently lit up in the consciousness of the hearer. Or it may be *accessible*, a status that has three subtypes. Textually accessible referents are those that have been mentioned before in the text and have since become deactivated. Situationally accessible referents are those that are saliently present in the text-external world of the utterance context. And inferentially accessible referents are those that are inferrable via elements present in either the co-text or in the physical context. The discourse statuses that Lambrecht assumes are exemplified in (1.29).

- (1.29) a. I found *a billfold* on the street today. (brand-new unanchored)
- b. I was just with *a guy I know*, and we were doing 180 km/h (110 mp/h) down a back road, and then all of sudden there was a click and we couldn't get any faster. (brand-new anchored)
- c. Do you know why the traffic is so bad today?—*The president* is in town. (inactive/unused)
- d. Hey, I noticed there hasn't been any fox related posts here in a while, but now I have a reason to post one because I saw the neighborhood fox *last night* (situationally accessible)
- e. We had barely entered the restaurant when *the waiter* came to seat us and take our orders. (inferentially accessible)
- f. I noticed that *that waiter* was staring at us again. (textually accessible)
- g. It was her last big hit on Motown; after more than 20 years, *she* de-camped for RCA. (active)

As Lambrecht points out, identifiability and activation are categories of memory and consciousness whereas topic and focus are relational notions having to do with the pragmatic relations between denotata and propositions in given contexts (1994:76). Although discourse statuses are cognitive categories, they do have certain correlations with particular formal properties of the linguistic expressions that encode the referents possessing the statuses in questions.

- A referent of status unidentifiable is necessarily prosodically prominent; lack of prominence is reserved for constituents with active referents.
- In languages with definiteness marking, an unanchored brand-new (= unanchored unidentifiable) referent typically appears in the form of a simple indefinite noun phrase (*a car*), whereas an anchored brand-new (= anchored unidentifiable) referent is typically encoded by an indefinite NP containing a definite one (*a man I know*).

- All pronominal expressions have active referents (but not vice versa).
- Morphologically definite expressions tend to have identifiable referents. One exception to this consists of the demonstrative *this*, which can be used to determine NPs which encode referents that are speaker- but not hearer-identifiable, as is shown in (1.30).

(1.30) So, *this guy* calls me last week. Didn't know him but he knew my name through a common friend.

- Conversely, morphologically indefinite expressions tend to have unidentifiable referents. The major exception to this regularity are NPs in generic sentences: generic referents always seem to be on 'standby' and anaphoric uses of morphologically indefinite NPs are acceptable:

(1.31) We know that a lion is a carnivore and he is because the bulk of his diet is made up of meat. We also know that *a lion* will often eat grass and leaf material because he needs or wants to.

As is shown by the work of Gundel, Hedberg, and Zacharski (1993), the use of a given NP type usually requires a particular level of activation but it is in principle compatible with higher statuses. For instance, the use of definite NP requires that the referent be at least uniquely identifiable. It is, however, possible for a speaker to use a simple definite NP to encode a referent that is actually active, as is shown in (1.32)

(1.32) "Do you have the number?"—"Darn, no...I don't have *the number for the garage*."

The possibility for under-coding of cognitive status complicates attempts to record the discourse status of NP referents.

The discussion so far has treated discourse status for the most part in terms of whether and how a referent's identity is resolved—whether it is identifiable and if so

how active it is. But the issue of whether a referent is given or not is not all there is to discourse status. One additional notion that is needed is how salient a referent is in discourse. Saliency, which we may take to be mainly reflected by frequency of mention, is in principle a separate issue from givenness. A second notion that has to be kept distinct from givenness is the source of activation. In particular, constructions may be sensitive to the distinction between anaphora and deixis. Gregory and Michaelis show in an examination of the difference between the English topicalization and left-dislocation constructions that left-dislocation only permits non-anaphoric, that is deictic, active referents whereas about a quarter of the fronted pronouns in the topicalization construction are anaphoric (2001:1697-1698). Thus, in contexts like (1.33), one finds continuations with topicalization like (1.33a) but not tokens with left-dislocation like (1.33b).

(1.33) CONTEXT. A has just outlined some possible policies for local school board.

B: Uh huh. Thats some pretty good ideas. Why dont you do something with those? You should run for a local school board position.

- a. (TOP) A: That Im not so sure about . Ive got a lot of things to keep me busy.
- b. (LD) A: *That Im not so sure about it. Ive got a lot of things to keep me busy. (= Gregory & Michaelis (2001:1669), (5))

Similarly, Birner & Ward argue that the various preposing constructions of English form a natural class on pragmatic grounds, in that they are subject to a constraint requiring that “the referent of the preposed constituent be anaphorically linked to the preceding discourse” (1998:32). Note that there are not only constructions that are sensitive to how a referent was activated but also constructions that impose particular requirements on the kind of referents that they introduce into the discourse. Birner & Ward cite evidence that existential *there*-sentences require that the speaker believes that the information of the post-verbal NP’s denotatum is not

already familiar to the hearer; that is, the denotatum has to be hearer-new rather than merely discourse-new and possibly familiar to the hearer (1998:102-103).

I will now turn to a discussion of how I decided to get at the cognitive statuses, and, even less directly, at the pragmatic relations of discourse referents. Clearly, discourse status and pragmatic relation, especially topic-status, can only be established with certainty if one has access to speakers' mental states. Short of this, heuristics have to be used to get at these notions. The heuristics used here come from two traditions of research. One is research on the distribution of NP form types (Ariel 1988, 1990, 1996; Brown 1983; Chafe 1976, 1987; Clancy 1980; Fox 1987; Fraurud 1990, 1996; Gundel, Hedberg and Zacharski 1993; Löbner 1985; Toole 1996 *inter alia*). The other is research on discourse-function constructions like topicalization or left-dislocation; voice constructions; and grammatical functions (Cooreman 1983, 1988; Michaelis and Francis (in press); 1999; Givón 1983; Gregory & Michaelis 2001 *inter alia*).

The most obvious clue to the discourse status of a referent is the morpho-syntax of the noun phrase encoding that referent. I used a fairly fine-grained classification to preserve as much information as possible, distinguishing for instance, between NPs determined by the demonstrative *this* and ones determined by the demonstrative *that*.

Another major heuristic that I used to get at the discourse status of referents is inspection of contextual mentions. The most well-known metrics used in this area are two measures introduced and popularized by the work of Givón (1983a, 1983b, 1984a, 1984b, 1995), look-back and persistence. Look-back "assesses the gap between the previous occurrence in the discourse of a referent/topic and its current occurrence in a clause ... The gap is thus expressed in terms of *number of clauses to the left*" (Givón 1983a:13). Persistence is defined as the number of clauses in uninterrupted sequence following the current mention in which the referent is present as a semantic argument of the clause, in any role, and coded by any grammatical means (Givón 1983a:15). Lookback and persistence thus are operationalizations of topic status in terms of salience. These notions, enriched by additional information about the grammatical properties of the contextual mentions, are also part of the centering algorithm employed by Walker and Prince (1996) in characterizing pronoun use.

In my own work I do not use these measures in exactly the way Givón defined

them. The differences are as follows:

- While Givón measures the window of context that he inspects in clauses, I use a cruder measure by counting in sentences. Since many sentences consist of more than one clause in my data, I also limited my window of context to 10 sentences. Givón used a window size of 20 clauses.
- The look-back value I use is the same as Givón's but in addition I also record the number of prior mentions within the 10 sentence window of prior context.
- Givón's persistence value maps to two measures of mine. First, I record the distance to the next mention. Second, I record the distance from the current mention to the last subsequent mention within the window. For neither measure do I require an uninterrupted presence of the referent in any intervening clauses. As in the case of prior context, I also record the number of mentions of the referent within the window of context.

It should be noted that my deviation from Givón's measures is purely a matter of convenience that is in part motivated by the technical apparatus I used to explore corpus data. It is not based on any belief that my measures would yield substantially different results. While some of the individual measures may be cruder than Givón's due to the counting in sentences rather than clauses, I expect that the combination of all five measures related to contextual mentions yields about as much information as do Givón's measures.

As suggested earlier, the contextual mention measures are not able to distinguish between all the discourse statuses. For that reason, I also separately recorded how a referent was linked to prior discourse, distinguishing between (a) textually and situationally evoked referents, (b) inferrable referents, (c) unused referents, and (d) brand new. For inferable referents I made further distinctions according to what the inference was based on, distinguishing, for instance, between frame inferables and set/super-set/sub-set inferables. For a fuller list, consult Appendix C.

I also coded all the NPs for animacy as animacy is one of the factors contributing to topic-worthiness and has strong associations with particular semantic roles and

grammatical functions (Dahl & Fraurud 1996).

Another purely formal heuristic I used was the length of the constituents encoding a referent. This was counted simply as the number of space-separated words. Constituent length is most directly a measure of grammatical weight (Wasow 1997, 2002) and, indirectly, via the correlation between discourse status and weight also a measure of discourse status (Arnold et al. 2000).

Since grammatical function also has an at least probabilistic association with discourse status (Michaelis and Francis (in press), Givón 1984a), I also kept track of this feature indirectly for all the NPs participating in dative alternation sentences by recording voice for each sentence. In addition to voice, I also recorded other clausal features that were not related to NP reference in an obvious way such as tense and the presence of modals as well as aspectual features such as the use of the progressive or perfect construction. For a fuller list, see Appendix C.

For this study, I did not attempt to determine, as a matter of course, whether a given NP's referent bore one of the two pragmatic relations topic and focus to the proposition denoted by the clause. Most of the quantitative analyses do not make use of pragmatic relations. In the investigation of the dative alternation I did annotate the NPs of some data sets for the feature \pm focus. However, I conservatively marked only the most obvious cases as +focus, for instance, ones where an NP was used in an argument focus construction to correct a misunderstanding, or ones where focusing adverbs such as *only* were present.

The remainder of this dissertation is as follows. The first two chapters discuss the constructional alternations. Chapter 2 focuses on the dative alternation and chapter 3 provides an analysis of the three raising alternations, epistemic raising, body part possessor raising, and property change factoring. The focus shifts in chapter 4 to null instantiation. Chapter 5 provides a conclusion and outlook on other research that the work here suggests.

Chapter 2

Dative alternation

2.1 Introduction

In this chapter I will examine whether the choice between the alternates of the English dative alternation, exemplified (2.1), in can be modelled statistically in terms of discourse factors.¹

- (2.1) a. I **faxed** Sony all of my original receipts. (double object)
 b. I **faxed** all of my original receipts to Sony. (caused-motion)

The relevant factors can be divided up into two classes. One set is concerned with discourse status in a narrower sense, namely with notions such as topic-focus, given-new, etc. The other factor that is of interest to us here is the notion of grammatical weight, which is taken to be relevant to the ease or difficulty of processing and understanding sentences rather than discourse coherence. Before I begin to examine data sampled from the British National Corpus and the Switchboard corpus, I will present some discussion of existing literature on these subjects. The review will help us make

¹The terminology used in the literature to talk about these constructions varies quite a bit. The following conventional equivalencies are assumed: dative construction = double object construction; oblique-goal construction = caused-motion construction; indirect object = primary object of double object construction = first object of double object construction; direct object = second object in double object construction; but also: direct object = object of oblique-goal construction. To avoid confusion, I will usually speak in terms of frame semantic roles and constructions.

more precise the terms in which we discuss the data. We begin with a look at the pragmatically-oriented literature and then turn to the work of Wasow and colleagues on grammatical weight.

2.1.1 Discourse status

Dominance Erteschik-Shir (1979) represents one of the earliest studies of discourse-pragmatic conditioning of an argument structure alternation. Erteschik-Shir applies her concept of DOMINANCE to the dative alternation. Dominance is defined as follows:

- (2.2) DOMINANCE: A constituent C of a sentence S is dominant in S if and only if the speaker intends to direct the attention of his hearers to the intension of C by uttering S. (p. 443)

According to Erteschik-Shir, a dominant constituent ‘becomes the natural candidate for the topic of further conversation’. She proposes to test dominance by evaluating how appropriate a hearer response would be that assigns to the candidate for dominance in the speaker’s sentence a ‘truth, probability, or even interest value’ (p. 443). An example exchange is given in (2.3):

- (2.3) Speaker A: John said that Mary kissed Bill.
Speaker B: That’s a lie, she didn’t. (= (2a), p. 444)

According to ES, the appropriateness of B’s response in (2.3) shows that the complement clause used by speaker A was dominant in the sentence.² The application of a dominance test to the dative construction is illustrated in (2.4) and (2.5):

- (2.4) a. Speaker A: John gave a book to someone yesterday.
Speaker B: Oh yes, I know who it was. (= (20), p.451)

²Speaker A’s utterance in example 2.3 also allows a continuation with *he lied* which combines less well with the continuation *she didn’t*. The dominance of the speech clause over the reporting clause is thus not automatic.

- (2.5) a. Speaker A: John gave someone a book yesterday.

Speaker B:

- a. *Oh yes, I know who it was.
b. Oh yes, I know which one it was. (= (21), p. 451)

Erteschik-Shir notes that generally dominant and non-dominant NPs differ as to what kind of determiners occur in them. Indefinites are generally dominant and definites generally non-dominant. (Unstressed) Pronouns are said to have no dominant uses.

Erteschik-Shir applies the concept dominance to the dative alternation as follows:

- (2.6) In the structure ... V NP₁ NP₂ (derived from ... V NP₂ {to, for} NP₁), NP₁ is non-dominant and NP₂ is dominant.

This constraint is meant to account for the following (and other) contrasts:

- (2.7) a. John gave it to Mary
b. *John gave Mary it.
- (2.8) a. Who did John give the book to?
b. *Who did John give the book?
- (2.9) a. Mary was given the book.
b. *Mary was given the book to.

The contrast in (2.7) is said to follow from the prohibition against unstressed pronouns being dominant. The one in (2.8) is described as the result of conflicting dominance assignments: according to *wh*-question formation, the questioned element ought to be dominant but according to dative movement the recipient ought to be non-dominant, which results in a conflict in (2.8b). The problem in (2.9b) is also accounted for in terms of a conflict. Passive wants the subject to be non-dominant but

the caused-motion construction assigns dominance to the referent of the preposition's complement.

Though Erteschik-Shir's account can handle a large section of the relevant data, it relies on an intuitive understanding of what it means for a speaker to draw attention to a referent.

The suggested tests for dominance seem problematic. First, it is not clear in which sense the referent of a nominal, non-propositional constituent might be assigned a truth or probability value.

Second, the test seems to have more to do with whether a referent is assumed to be identifiable (in Lambrecht's terms: whether there is a knowledge presupposition concerning the identity of the referent) and discourse-active by the speaker than with any independent concept of dominance. While the test gives the right dominance assignments here:

- (2.10) Speaker A: John killed a cop.
 Speaker B: Oh yes, I know which one it is.(= (5), p. 447)
- (2.11) Speaker A: John killed the cop.
 Speaker B: ?Oh yes, I know which one it is.(= (7), *ibid.*)
- (2.12) Speaker A: John killed him.
 Speaker B: ** Oh yes, I know who he is. (= (10), *ibid.*)

it is less clear that it works correctly when talking about transfer scenarios where both arguments are equally identifiable and active:

- (2.13) a. Speaker A: I gave it to him.
 b. Speaker B:
 a. *Oh yes, I know who he is.
 b. *Oh yes, I know what it is.
 c. Speaker A: In the fifth act, I'm selling a guy a gun when the police find me.
 a. Oh yes, I know who/which character it is.
 b. Oh yes, I know which gun it is.

Further, the concept of dominance does not make clear reference to more primitive notions such as activation status, source of familiarity, or the pragmatic relation of the referent to the proposition. However, a more detailed analysis seems necessary given that Erteschik-Shir notices, e.g., that an expression's ability to be dominant may depend on whether the referent has been mentioned before. Examples such as (2.14) illustrate that referents such as *disease* in metaphorical (causative) transfer scenarios need not always be dominant when they are second mentions, even though they seem to require a dominant interpretation on first mention:

- (2.14) John gave Mary pneumonia and he gave it to Ted too.
 (2.15) *John gave Mary pneumonia and he gave Ted it too.
 (2.16) John gave Mary pneumonia.
 (2.17) *John gave pneumonia to Mary.

It seems that dominance is not a conceptual primitive that can be operationalized easily for studying the pragmatics of alternations.

Focus Goldberg (1995) relates some syntactic properties of the dative alternation to discourse pragmatics. Following Erteschik-Shir (1979), she assumes that the last

argument of each construction in (2.19-2.20) is specified as the focus (at least, in the active form).³ The following contrasts are said to follow from this distinction:

- (2.19) a. #She gave an old man it.
 b. (She gave it to an old man.) (= (33a-b), Goldberg 1995: 92)
- (2.20) a. #Mary gave a brand-new house to him.
 b. (Mary gave him a brand-new house.) (= (35a-b), Goldberg 1995: 92)

Importantly, these pragmatic constraints are taken to explain why metaphorically based support verb-like uses of *give* and other verbs occur only with the ditransitive patterns:

- (2.21) a. *She threw a parting glance to him.
 b. (She threw him a parting glance.) (= (42a-b), Goldberg 1995: 94)

However, Goldberg's characterization of the situation is somewhat problematic. One cannot interpret the placement of focal stress on the last constituent as an indication that that constituent is the extent of the focus domain. That would predict that the constructions are used only in cases of argument focus on the last constituent, which is clearly incorrect: sentence (2.19b), for instance, can be used in circumstances where it cannot be presupposed that the Donor will give the Theme to anybody.

In addition, there are counterexamples to the claim that in cases of metaphorical transfer the oblique-goal form is not acceptable:

- (2.22) Dirt encrusted on the bus window gave a grainy texture to the scenery.
 (BNC)
- (2.23) Precise and accurately placed enchaînements and éaulement give a clear definition to each step and pose as it takes its place in a perfectly balanced design. (BNC)

³Goldberg suggests that in cases like the following, where a physical rather than a metaphorical motion reading can be used, the constraint does not hold:

- (2.18) Sally threw a football to him. (= (36), Goldberg 1995: 92)

Finally, for cases of concrete transfer without clear physical motion which were claimed to require the ditransitive form, there seem to be counterexamples:

- (2.24) ... no nation capable of giving a Peace Prize to Henry Kissinger could possibly be accused of lacking a sense of humour. (BNC)
- (2.25) It has incorporated the 19.2Kbps modulation scheme on its high-end V.32bis Volante modem at no extra cost , and is giving a free upgrade to existing users. (BNC)
- (2.26) Four years prior to his death he gave a house to his daughter but lived in it. (BNC)

Like, Erteshik-Shir's account, Goldberg's characterization of the pragmatic difference between the alternates only in terms of focus basically misses the focus type versus discourse status distinction.

'Information flow' Thompson (1990) argues that the dative alternation is to a significant degree pragmatically conditioned. She analyzes tokens of the alternate constructions in terms of Givón's (1979, 1984a, 1984b) notion of TOPICWORTHINESS, which refers to a cluster of properties that make an argument a good topic. The features that Thompson attended to are related to the referents' discourse status rather than a sentence's focus type. The properties, listed in (2.27), are derived from the study of what typical subjects are like. Note that the order is not meant to reflect any weighting; none is given by Thompson.

- (2.27) Properties that increase topicworthiness
- Animacy
 - Pronominal Coding
 - Specific Reference: a noun is considered specific unless it 'refers to a class of entities or has no referent' (p. 244)

- Identifiability: If a noun is specific, it gets coded for identifiability. Pronouns, proper nouns, and definite NPs count as identifiable.
- Proper Name
- Length

Thompson also uses Chafe's (1987) activation states to differentiate the recipient and theme referents in the two alternate constructions:

- Active: If the referent is presumed to be in the addressee's consciousness
- Semiactive: if the referent is inferable from previous discourse or its previous mention was rather distant
- Inactive: if the referent is presumed not to be in the addressee's consciousness at all

Thompson finds the two following hypotheses strongly supported by her data:

- (2.28) recipients are typically much more topicworthy than patients
- (2.29) recipients in the double object construction are much more topicworthy than recipients in the oblique-goal construction

The results of my own corpus study agree with these findings. Nonetheless, there are some small concerns about the design of Thompson's study. It did not control for verb identity. Her sample of 196 clauses contained various ditransitive verbs. There is a danger that a small number of verbs such as give and tell may account for a great many tokens and these verbs may be more biased pragmatically than other alternating verbs. Another factor is that Thompson's data was taken from three English language narratives, which may not be representative of the other major variety, British English, and of spoken data. Thompson's study also does not control for semantic sub-class: recipients, beneficiaries, and beneficiaries all are collapsed in the class recipient.

Double object asymmetries and topicality Polinsky's 1998 paper provides a link between pragmatic and more narrowly syntactic studies of the dative alternation. The paper is concerned with the set of object asymmetries observed in the English double object construction. These asymmetries, summarized in Table 2.1, have traditionally been treated syntactically in the literature, e.g. by Larson (1988), but Polinsky argues that they can be given a pragmatic account.

EFFECT	SUPERIOR OBJECT	Example
Passivization	Recipient	<i>Mary was given a book</i> v. <i>A book was given Mary</i>
Scope Relations	Recipient	<i>I showed every critic two photos: every >> two; I showed to photos to every critic: two >> every</i> OR <i>every >> two</i>
Pronoun binding	Recipient	<i>What is giving everyone_i their_i biggest headache?</i> v. <i>*I denied its owner each paycheck_i</i>
Optional deletion	Recipient	<i>I gave/sent a book</i> v. <i>*I gave/sent John</i>
Pronominal expression	Recipient	<i>I gave them advice</i> v. <i>=*I gave John it</i>
Licensing of wh-elements	Patient	<i>What did John give Mary?</i> v. <i>?*Who did John give a book?</i>
Relativization	Patient	<i>the candy that John gave the little girl</i> v. <i>?/*the little girl that John gave (the) candy</i>
Licensing of negative polarity items	Patient	<i>I'll be amazed if Donald gives Ivana a red cent</i> v. <i>*/? I'll be amazed if Donald gives a red cent to Ivana</i>
Verb+Obj set phrase	Patient	<i>Fergie gave the royal family a headache</i> v. <i>*Fergie gave a headache to the royal family</i>

Table 2.1: Asymmetries between the two objects in the double object construction

The main claim that Polinsky makes is that the Recipient in the double object construction is more topical than the Theme. Polinsky relates the difference in topi-

cality to the semantics and presuppositional characteristics of the thematic roles that are part of the transfer scenarios. In particular, she argues that the Recipient is presupposed to exist whereas the Theme is not, which is taken to explain the set of contrasts below. The ditransitive sentence (2.30b) is unacceptable as compared to the benefactive sentence in (2.30a) because the ditransitive construction presupposes that the recipient independently exist before the transfer. Similarly, indefinite noun phrases encoding recipients are understood to be specific, i.e. to have a referent identifiable to the speaker (though not to the hearer).

- (2.30) a. I bought a ring for my wife in case I should decide to marry (= Polinsky (28a), p. 411)
- b. #I bought my wife a ring in case I should decide to marry (= Polinsky (28b), p. 411)
- (2.31) I sent a doctor a letter (= Polinsky (31), p. 413)

Note that Polinsky refers to topicality rather than topic-status in talking about why Recipients tend to behave the way they do. To re-iterate a point made before: it is not the case that in the double object construction Recipients are always the sentence topics—in the majority of cases, the Donor will have that pragmatic role—and Themes are always the foci—in the majority of cases, the predicate as a whole is the focus. The affinities that exist between the two linking constructions and other constructions such as voice or questioning are simply due to optimization: speakers tend to align constructions in such a way that the constraints of all constructions regarding referent accessibility are met at the same time.

The asymmetries that Polinsky discusses and her basic solution to their existence in terms of referent topicality are not only of interest with respect to the double object but they may well also have relevance to the paradigmatic choice between the double object construction and the caused-motion construction. Fortunately, Polinsky's claims can easily be tested with corpus data and the basic proposal can also be evaluated relative to other constructions that ought to interact with topicality.

But before we begin this evaluation, I would like to draw the reader's attention to some methodological considerations. The reader will have occasion to notice throughout the remainder of this work that sample sizes for various phenomena examined in the corpus differ. These differences are not the result of theoretical statistical decisions. In principle, bigger samples are better but above a certain sample size the gains in representativeness by enlarging the sample further become progressively smaller. The major concern in all cases has to be a decent minimum sample size. Size differences between samples, when comparisons are made, are not a problem as statistical procedures are set up to take sample sizes into account. The size differences mainly resulted from the fact that there is no simple way to extract a given number of uses of a word in a given sense and in a given construction from the BNC corpus. The analyst has to manually inspect a number of candidate sentences to identify the matches. A tree-parsed corpus like the Wall Street Journal corpus would not really be much better for my purposes since the parsing is not always right and there are no sense distinctions marked. Moreover, data marked up by an automatic system may be wrong in a way that systematically distorts the results. The different sample sizes then are due to my not being able to keep easy track while vetting candidate sentences of how many sense and constructional matches I had already identified. It should be noted, too, that I in some cases I made use of the Internet search engine Google when the corpus did not contain a significant number of examples of some phenomenon. To preview a result reported on later in this section, a search in the BNC for sentences with *who* or *whom* followed within 5 words by a form of *give* found very few tokens where the *wh*-word was understood as questioning the Recipient frame element. Most uses involved relative clauses with the Donor relativized on. The interrogative uses of *who* were mainly questions about the Donor. I therefore acquired data from the search engine to be able to address the question how Recipients with *give* are typically questioned.

Returning now to Polinsky's account of the object asymmetries, we will first consider whether her treatment correctly predicts on which argument explicit focus marking with *only* tends to appear. A random sample of 150 sentences containing a form of *give* and following it within 7 words the focusing adverb *only* contained 39 sentences

belonging to either the double object or the caused-motion construction in which the scope of *only* consisted of either the Recipient, as in (2.32) or the Theme (2.33).

- (2.32) Some went further and stated that the press must be free to say whatever needed saying; not surprisingly, opponents of this school of thought pointed out that this **gave** freedom only to those who had a newspaper at their disposal.
- (2.33) In 1987 there were 236,000 enquiries in England and Wales from people claiming to be homeless; 62,000 were found not to be homeless within the definition used in the legislation and 57,000 were found not to be in priority need and were **given** advice and assistance only .

In the 22 instances of the double object construction, *only* always had scope over the Theme while in the 17 instances of the caused-motion construction, *only* had scope over the Recipient 13 times and over the Theme 4 times. Of the 4 unexpected cases, 2 involve heavy NP shift of the object towards the end of the clause. Basically, the data for *only* are what Polinsky would predict: focal Recipients occur only in the caused motion construction.

Similarly, I extracted a random sample of 100 sentences in which *what* was followed by a form of *give* within 7 words. In 86 of the 100 sentences, *what* was used either as a *wh*-word in embedded or main questions or as the pronoun of a headless relative clause containing *give*. In 69 (80.2%) cases, the double object construction was used while the caused-motion construction was used in only 17 (19.8%) cases. Of these 17, 3 were passives with Theme subjects, which heavily favors the use of the caused-motion construction. Another 3 involved contrastive focus on the Recipient. 6 others involved (lexical) Recipients that had not been explicitly evoked, 3 of which were also formally indefinite. 1 case was the idiom *give rise*, which heavily favors the caused-motion construction.⁴ Thus, questioning of the Theme clearly favors the double object

⁴Though this may initially be doubted by some native speakers, double object constructs involving *give rise* are attested, if rarely. Consider: *Fandom, therefore, is divorced from the phenomenon which gave it rise and is seen as an island, without any connection to anything else.* It is not clear if there are solid tests that will reliably identify those combinations that speakers may feel are lexicalized.

construction. The reverse is true for questioning of the Recipient: here the caused-motion construction is heavily favored. Among a sample of 100 active form sentences containing the string *who did he give*, which was retrieved from the internet using the Google search engine, 88 (97.8%) sentences exhibited the caused-motion construction, 2 (2.2%) the double object construction. The remaining 10 were instances of particle verbs such as *give away* and *give up*. The data for *only* and for *wh*-questions that we have considered so far match Polinsky's predictions.

When it comes to passivization, we find another case of near-complementary distribution. Choosing the Theme as subject very heavily favors the caused-motion construction, whereas choosing the Recipient as subject favors the double object construction.

Let us now consider some corpus data on relativization. A random sample from the BNC of 200 sentences containing *which* as a relative pronoun followed by a form of *give* within 7 words yields 33 sentences where the Theme is relativized on. In 11 cases there is not Recipient expressed. Among the remaining 22 sentences, the linking construction chosen is the caused-motion construction in 12 cases and the double object construction in 9 cases. All 3 cases in which the clause was passive exhibited the caused-motion pattern. 8 of the 9 double object sentences had a pronominal Recipient, only 1 a lexical NP. For the caused-motion sentences the situation was reversed, only 1 Recipient was pronominal, all other 10 were lexical NPs.

The Recipient is relativized on in 12 cases.⁵ In 9 sentences, we find the caused-motion construction and in 3 the double object construction. Importantly, the former occur in the active voice, as in (2.34), while the latter occur in the passive voice, as in (2.35).

(2.34) It is an area to which the County Council has **given** priority in the past.

One test suggested to me by Charles Fillmore was relativization in the form of a reduced relative: *the priority given to ...* seems much more acceptable than *the rise given to ...*. Indeed, no instance of *the rise given to ...* could be found on the Google search engine. However, other strings such as *the effect given to ...* which do not seem to be much less idiomatic are well attested.

⁵The remaining tokens are accounted for as follows. In 115 cases the pronoun fills the *Donor* role, typically in metaphorical sentences. In 40 cases, the relativized semantic role was one of the more oblique ones such as Means or Reason, or the pronoun is an interrogative one, or *which* and *give* are in different clauses.

- (2.35) The dolphin received further protection in 1975, when the District Magistrate of Sukkur made trapping or killing of dolphins illegal under a Criminal Procedure Code, which was **given** extensive publicity by drums, through radio and newspapers, and by placing posters in key locations.

A parallel random sample of 400 sentences containing either *who* or *whom* followed by *give* within 7 words, contains 277 cases in which the Donor is relativized on. In another 80 cases, *who* is an interrogative pronoun or *who* and *give* do not belong to the same clause. This leaves 43 cases in which the Recipient is relativized on. In 30 the construction used in the double object construction. All 30 are in the passive voice. The remaining 13 sentences exhibit the caused-motion construction; 11 of these sentences are in the active voice.

I also extracted 200 sentences where *given* followed a head noun or *that* or *those*. 108 were indeed reduced relative clauses. In 10 cases, the Recipient was relativized on; as expected, all of them exhibit the double object construction. The remaining 98 cases have the Theme relativized on. In 60 cases, no Recipient is expressed at all. 37 clauses are instances of the caused-motion construction; 1 case is an instance of the double object construction (cf. (2.36)), notably with a pronominal Recipient.

- (2.36) Whatever the original guidelines **given** them, goes the argument, the agents get drawn into doing what is in the best interest of those they are regulating—they are, wittingly or unwittingly, captured by the people they are supposed to be policing.

That-relatives give a very similar picture. Of 300 inspected candidate sentences, 184 were actual that-relative clauses in which *that* was followed by *give* within 7 words. In 85 sentences, the Donor was relativized. The Recipient was relativized in only 3 sentences. 2 of these combined the double object construction with passive voice and 1 combined the caused-motion construction with active voice. The Theme was relativized on in 96 sentences. Of these, 37 did not express a Recipient. 30 sentences each occurred with caused-motion and double object linking. 25 of the double object sentences were active voice, 5 passive voice. Among the caused-motion

sentences, there was no such clear split, with 17 occurring in the active voice form and 13 in the passive.

The corpus data for relative clauses suggest that Polinsky's characterization of the situation was not detailed enough. She claims that generally, it is better to relativize on the Theme than on the Recipient in the double object construction. One factor that is not taken into account was voice. If the clause is passivized, then relativization on a Recipient in the double object construction is perfectly acceptable. Another factor influencing the choice of linking construction that Polinsky did not discuss is the morphological type of the Recipient in cases where the Theme is relativized on. Here, the corpus data suggests that the double object construction is robustly preferred only when the Recipient is pronominal, whereas her example in Table 2.1 above suggests that this would be true even for cases where the Recipient is lexical.

Importantly, the corpus data suggest that functional considerations do not carry as far as implied where relativization is concerned. Polinsky's account implies that the same criterion is used for choosing which linking construction to use when questioning an argument. The theoretical prediction therefore has to be that speakers should use that type of linking construction (or combination of linking construction and voice) in a relative clause which is conventionally associated with the least degree of topicality for the relativized-on argument. According to the criterion, Themes should be relativized on in the double object construction, and Recipients should be relativized on preferentially in the caused-motion construction. Yet, there clearly are types of relativization strategy that violate both of these theoretically derived predictions. Reduced relative clauses typically relativize on Themes in the caused-motion construction; these sentences are moreover always in passive voice, yielding a configuration in which Themes are more topic-worthy rather than less so in non-relative clauses. Likewise, among the *who/whom* relative clause relativizing on Recipients, the majority of cases (30/43) involves clauses in which passive voice co-occurs with the double object construction, a combination in which the Recipient is usually very topic-worthy in non-relative clauses.

The few small specialized corpus studies that we have carried out here suggest that while topicality is a good explanation for some of the observed asymmetries, it

does not cover all of them.

2.1.2 Grammatical weight

Wasow (2002) presents a summary of research carried out by the author and his colleagues on questions such as how weight is best measured and what role it plays in alternating constructions such as Heavy NP shift, Verb-particle ordering, and the dative alternation (Wasow 1997, Arnold et al. 2000).

One important finding of Wasow's that I used in doing my own research is that it is very hard to distinguish among various structural weight measures as predictors of weight effects: counting words, syntactic nodes, or phrasal nodes all work well. This allowed me to simply measure weight in terms of word count, which is by far the easiest method of the three.

Wasow (2002) also finds that graduated measures of weight do a better job of accounting for constituent ordering than do categorical definitions of heaviness. And he argues that ordering is sensitive to the relative weight of pairs of constituents, not just to the weight of one.

Wasow (2002) also presents data more specifically on the dative alternation. One of his analyses is concerned with 269 tokens of *give* taken from the Hansard corpus, which consists of Canadian parliamentary debates. These tokens were annotated for both weight and information status. The weight-related measure was the relative length of the Recipient compared to the Theme, classified three ways: 2 or more words shorter; shorter by one word, equally long, or longer by one word; and 2 or more words longer than the Theme. The information status categories were Prince's categories of Given, Inferable, and New.

A logistic regression performed on the data showed that both weight and information status were needed as predictors of constructional choice. Wasow (2002) notes, however, that unlike might be suggested by the graph summarizing his findings, which is reproduced here in Figure 2.1, there is no significant interaction between weight and information status because at the ends of the weight spectrum, when there is a difference of more than 2 words, weight more or less categorically determines the

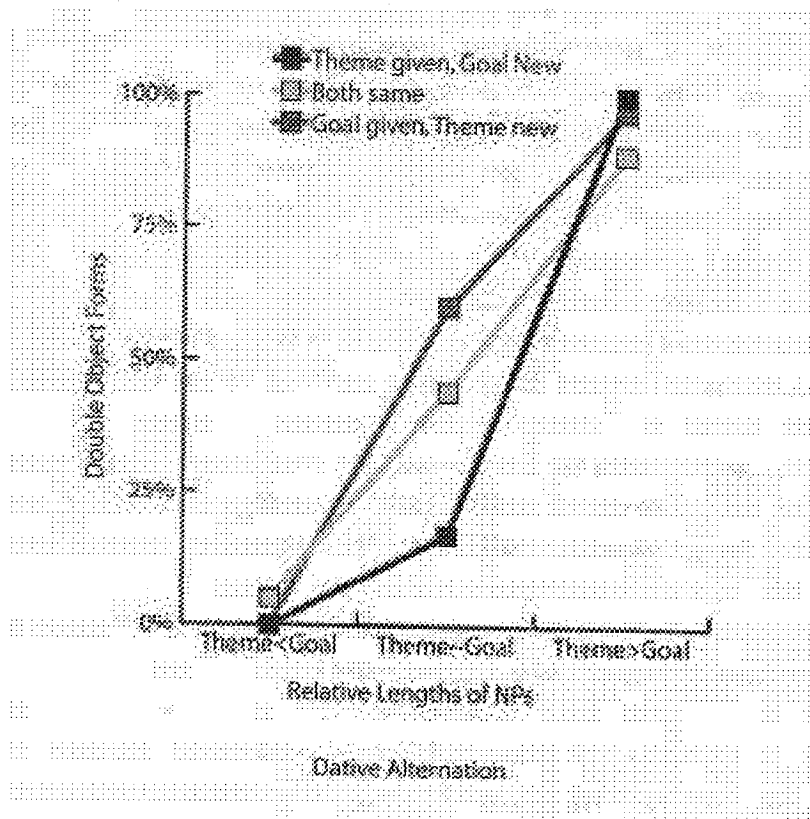


Figure 2.1: Results for Wasow's corpus study on the dative alternation

constructional choice.

Wasow (2002) does, however, not suggest that universally weight is the stronger predictor. He cites the research of Siewierska (1993) on Polish constituent ordering and the results of a production experiment run by himself and colleagues as evidence that information status can be more important than weight. In fact, he suggests that the corpus study's result may simply be due to the fact that the measure of newness might have been more coarse than that used for weight.

Since it will be of interest to compare Wasow's (2002) experimental findings to at least the findings based on a spoken corpus, I will now briefly summarize Wasow's (2002) production experiment and its results.

The task devised had pairs of participants talk to each other about the manip-

ulation of certain objects, with one participant (called Director) giving instructions and the other (called Actor) carrying them out. The manipulations mainly involved giving objects of various shapes and color to cardboard cutouts depicting animals. The objects varied in color and other physical characteristics so as to ensure that reference to some of them was easier because they were uniquely describable just by the head noun, whereas other items required the additional mention of a property of the item.

The Director received cue cards which indicated what they were supposed to tell the Actor to do. The cards had only one or two words on them so as not to prime use of one of the constructions of interest. Actors also received cue cards prompting them to ask questions of the Directors, thereby eliciting instructions. The instructions produced by the Directors make up the data set of 1684 sentences that was analyzed.

Each exchange was initiated by the Actor asking either about a set of three animals or a set of three objects indicated on their cue card. This was to prevent the Director from using pronouns in their responses, which were guided by matching cue cards showing how objects and animals should be matched up. Filler items were included between the test items in order to make sure the verb *give* was in fact used and to draw attention away from the core interest of the test, the dative alternation.

Wasow (2002) points out that the design leaves open which construction should be used by the director, although the word *give* appears on the cue card for the director. It also leaves open the choice of voice and any other dislocation construction compatible with the linking and the voice constructions. As a result, the recorded sentences were of eight different constructional patterns. The Actors' cue cards served to control the information status—either Theme or Recipient was always given in the Director's reply. The Director could also mention a referent before referring to it again in the instruction on which object goes to which animal; this produced cases where both referents were given. Overall, 808 sentences were collected in which the Theme was given, 849 sentences in which the goal was given, and 27 in which both were given. The Directors' cue cards served to control the length of Theme NPs (when the exchange was initiated with a question about the animals).

Wasow (2002) suggests that overall the cooperative nature of the task encourages

the participants to focus on the transmission of content, drawing on the same factors that are at work in spontaneous speech, including the desire to communicate the intended content clearly and efficiently.

The outcome of the experiment is summarized by Figure 2.2.⁶ The figure is organized differently from the one for the corpus study, Figure 2.1. In 2.2, the givenness statuses are mapped on the X-axis and the lines differentiate the relative-length categories, which is the opposite from 2.1. Figure 2.2 also shows a different distribution than for the corpus study. Notice that where the length values differ (i.e. *theme shorter than goal* or *goal shorter than theme*), we no longer find categorical behavior: not only the very light line but also the heavy black and the medium light line now have differing values across the three givenness categories. The reason for this likely is that in the corpus experiment the categories at the ends of the length difference scale were for cases where the absolute difference was 2 or more words, whereas in the experiment, the criterion is 1 or more. In other words, the extreme categories are not as extreme here as in the corpus analysis.

As in the case of the corpus study, it was found for the experimentally elicited sentences that both weight and discourse status were significant predictors of constituent ordering, at $p < .005$ and $p < .001$, respectively. Unlike in the corpus data, there also was a statistically significant interaction between the factors ($p < .001$).

2.2 Give in the BNC

The lexical unit that we will use to examine the properties of the double object and the caused-motion construction is *give*. The greatest advantage and disadvantages of this choice are its great frequency and its polysemy, respectively. In the basic sense that we will be interested here, *give* is used to refer to a transfer of possession such that the Giver possesses the Theme before the Transfer while the Recipient does not and that the Recipient possesses the Theme after the Transfer while the Giver does not. Prototypically, the transfer would involve physical motion of the Theme from the

⁶Wasow uses the term Goal where I speak of Recipient.

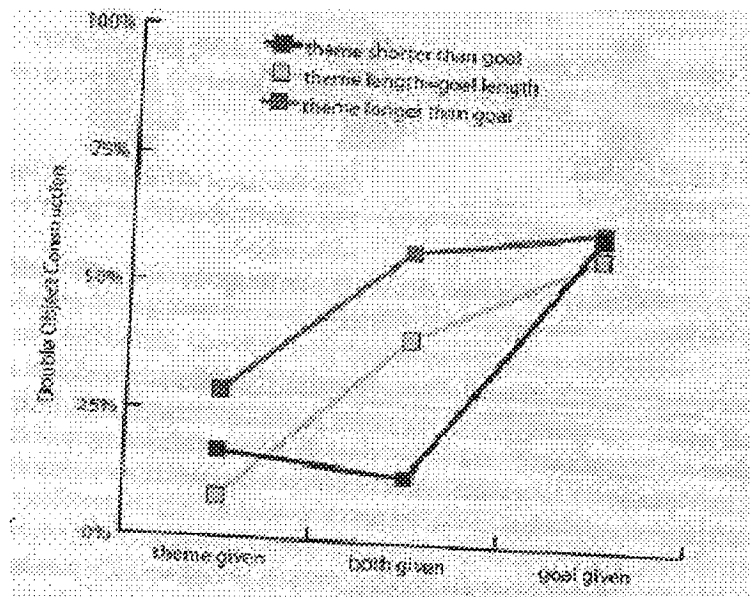


Figure 2.2: Results for Wasow's production experiment on the dative alternation

Recipient to the Giver. In my sample I have, however, also allowed for the transfer of immobile entities such as lands or estates. Excluded were metaphorical transfers of states (*The beer gave her a slight buzz*) and transfers of information (*Can you give me his name?*).

2.2.1 Possession transfer uses of *give*

Voice A χ^2 -test performed on the distribution of active and passive form sentences in the sample, which is shown in Table 2.2, yields a p-value of 0.9128, indicating no plausible difference between the constructions.

Frame element morphology Let us consider the morphology of the Recipient and Theme frame elements. Table 2.3 summarizes the forms found for the Theme frame element.

Table 2.3 shows the familiar picture: Themes are less likely to be definite in the double object construction than in the caused-motion construction. The difference is

	Active	Passive	Total
Double object	128	35	163
Caused-motion	35	10	45
Total	163	45	208

Table 2.2: Voice by construction

statistically significant according to a χ^2 -test, with $p=0.0002$

The data for the morphology of Recipient frame elements are presented in Table 2.4. They, too, hold no surprise. Recipients are inherently very likely to be definite. The difference between the constructions in percentages may not seem very large but it is statistically significant according to a χ^2 -test with $p=0.0006$.

Number of contextual evocations

⁷The reflexive forms are no mistake. Consider:

- (i) The I gives itself to the Thou , makes a complete act of dedication, links the two persons through the poetry.

Arguably, this sentence should not have been included in the data as it is not concrete. Since I had already performed a great many calculations before noticing that I had included this sentence I left it in the sample.

	Double object		caused-motion	
	n	%	n	%
Reflexive ⁷	0	0.0	2	4.7
3rd	5	3.1	17	37.8
Supplementary	0	0.0	1	2.2
relative				
This	2	1.2	0	0.0
Possessive	1	0.6	0	0.0
pronoun				
That	2	1.2	0	0.0
That NP	1	0.6	0	0.0
Proper Noun	0	0.0	1	2.2
Possessed NP	7	4.3	2	4.4
Definite NP	33	20.2	5	11.1
	51	31.3	28	62.2
Indefinite NP	62	38.0	5	11.1
Quantified	18	11.0	3	6.7
NP				
Interrogative	1	.6	1	2.2
pronoun				
Bare NP	16	9.8	4	8.9
Restrictive	2	.6	1	2.2
relative				
Reduced rela-	1	.6	0	0.0
tive				
What of head-	1	.6	0	0.0
less relative				
Negative de-	1	.6	0	0.0
terminer				
That-relative	9	5.5	3	6.7
Headless rela-	1	.6	0	0.0
tive				
Total	163	100.0	45	100.0

Table 2.3: Realizations of Themes

	Double object		Caused-motion	
	n	%	n	%
Definite zero	1	0.6	0	0.0
1st or 2nd	60	36.8	7	15.6
3rd	49	30.1	4	8.9
Proper Noun	15	9.2	13	28.9
Possessed NP	2	1.2	2	4.4
Non-restrictive relative	3	1.8	0	0.0
Definite NP	16	9.8	6	13.3
Reflexive	2	1.2	0	0.0
	148	90.8	32	71.1
Indefinite NP	5	3.1	2	4.4
Quantified NP	3	1.8	4	8.9
Bare NP	7	4.3	6	13.3
Headless relative	0	0.0	1	2.2
Total	163	100.0	45	100.0

Table 2.4: Realization of Recipients

	Recipients		Themes	
	Distance	Number	Distance	Number
N	163	163	163	163
Mean	3.0429	4.6442	10.1595	.1779
Median	1.0000	3.0000	11.0000	.0000
Mode	.00	.00	11.00	.00
Std. Deviation	4.27215	4.75145	2.54327	.59753
Skewness	1.264	1.067	-2.907	4.137
Std. Error of Skewness	.190	.190	.190	.190

Table 2.5: Prior mentions in the double object construction

	Recipients		Themes	
	Distance	Number	Distance	Number
N	45	45	45	45
Mean	7.3333	1.4667	5.8667	1.0444
Median	11.0000	.0000	8.0000	1.0000
Mode	11.00	.00	11.00	.00
Std. Deviation	4.69042	2.74359	5.21187	1.63701
Skewness	-.562	2.201	-.051	2.983
Std. Error of Skewness	.354	.354	.354	.354

Table 2.6: Prior mentions in the caused-motion construction

	Recipients			Themes		
	Distance	Number	Last	Distance	Number	Last
N	163	163	163	163	163	163
Mean	3.4233	4.6258	4.3436	7.9693	.5706	.7607
Median	1.0000	3.0000	4.0000	11.0000	.0000	.0000
Mode	1.00	.00	.00	11.00	.00	.00
Std. Deviation	4.44291	4.96169	4.12517	4.55342	1.30038	1.78414
Skewness	1.003	1.276	.247	-.895	5.189	3.021
Std. Error of Skewness	.190	.190	.190	.190	.190	.190

Table 2.7: Subsequent mentions in the double object construction

	Recipients			Themes		
	Distance	Number	Last	Distance	Number	Last
N	45	45	45	45	45	45
Mean	5.1556	2.3333	2.4222	8.6000	.4444	1.0222
Median	2.0000	1.0000	1.0000	11.0000	.0000	.0000
Mode	11.00	.00	.00	11.00	.00	.00
Std. Deviation	4.98158	3.33030	3.36755	4.24478	.94281	2.59798
Skewness	.273	2.078	1.279	-1.271	2.722	2.953
Std. Error of Skewness	.354	.354	.354	.354	.354	.354

Table 2.8: Subsequent mentions in the caused-motion construction

	Double object	Caused-motion	Total
brand-new	88	13	101
variable	12	4	16
unused	2	0	2
inferable	42	5	47
evoked	17	23	40
situationally given	2	0	2
Total	163	45	208

Table 2.9: Prior evocation of Themes

Prior evocation When we turn to prior evocation of the frame elements' referents, we also find differences across constructions.

Table 2.9 summarizes the data for Theme frame elements.

There is clearly a difference between constructions with respect to whether Theme referents are inferable or textually or situationally evoked, or not. If we collapse the tokens with 'unused' status with those of inferable or textually or situationally evoked status, and compare them to the group that consists of tokens of brand-new and variable status, a χ^2 -test yields a p-value of 0.0047, which is statistically significant. (Grouping the unused-status tokens with the brand-new and variable ones and contrasting this group to the group made up of tokens in the remaining statuses yields a lower p-value of 0.0029). Themes are more likely to have a prior evocation in the caused-motion construction than in the double object construction.

Table 2.10 shows the summary for the Recipient data. Repeating the same χ^2 -tests carried out for Themes, we get p-values of 0.006 for the case where the unused-status tokens are grouped with the inferable and evoked ones, and $p=0.0000$ for the case where the unused tokens are grouped with the tokens of brand-new status. Recipients in the double object are more likely to have a prior evocation.

Relative prior evocation We can now also look at whether differences in prior evocation might explain the choice of construction. Based on theoretical considera-

	Double object	Caused-motion	Total
brand new	9	10	19
unused	3	4	7
inferable	15	13	28
evoked	97	15	112
situationally given	39	3	42
Total	163	45	208

Table 2.10: Prior evocation of Recipients

tions we should expect, for instance, that in a situation where the Recipient referent has an explicit prior evocation whereas the Theme referent has none, the double object construction is used.

In order to make a comparison between the frame elements in terms of prior evocation manageable, I ordered the possible values on a scale in 3 subgroups as follows:

- Situationally evoked, Textually evoked > Inferable, Unused > Variable, Brand-new

Differences between immediately adjacent groups were counted as having magnitude 1, differences between the groups at the end as having magnitude 2. For the members of a group, I treated the first-mentioned one as having a slightly higher status than the other. On the basis of this treatment, Table 2.11 was generated for the difference in evocation status between Recipient and Theme referents.

The Table shows that there are only 39 cases in which no distinction in evocation status can be made. For the cases where the frame element referents are distinct with respect to prior evocation status, there is no categorical distinction. Sentences always appear in both constructions. It seems fairly clear that the greater the difference is in the positive direction, the more likely it is that the double object construction is used: for a +2 difference, there are only 3 cases out of 85 that violate the prediction that the double object construction ought to occur. For a +1 difference, the correlation is

	Double object	Caused-motion	Total
-2	0	[3]	3
-1	2	[9]	11
-.5	5	[1]	6
+.5	[4]	3	7
+1	[46]	11	57
+2	[82]	3	85
Tied	24	15	39
Total	163	45	208

Table 2.11: Relative prior evocation of Recipients

	Double object	Caused-motion	Total
Not evoked	37	18	55
Inferrable	4	3	7
Textually evoked	120	24	144
Situationally evoked	2	0	2
Total	163	45	208

Table 2.12: Subsequent evocation of Recipients

less strong but still there. With negative differences of -1 and -2, the constructional choice is also well predicted. The figures for the small positive (+.5) and negative (-.5) differences may suggest that the group-internal differentiation does not really hold up.

To get a sense of how well this relative measure of prior evocation status correlates with constructional choice, we will add up the correct predictions—enclosed in square brackets in Table 2.11—and divide them by the total: $145/208=0.6971$ which is 69.7%.

Subsequent evocation Looking at subsequent evocation of the referents of the frame elements, we find differences across constructions only for Recipients. Table 2.12 summarizes the data for that frame element.

	Double object	Caused-motion	Total
Not evoked	110	33	143
Inferrable	10	3	13
Evoked	42	9	51
Situationally evoked	1	0	1
Total	163	45	208

Table 2.13: Subsequent evocation of Themes

Grouping the inferable tokens with the evoked or the non-evoked cases and performing a χ^2 -test yields significant p-values of 0.0198 and 0.0052, respectively. Thus, we may conclude that Recipients are more likely to have a subsequent evocation in the double object construction than in the caused-motion construction.

Table 2.13 summarizes the data for Theme frame elements.

In the case of Theme frame elements, we get non-significant p-values of 0.4539 and 0.3815, respectively, for the case where the inferable tokens are grouped with evoked or the non-evoked tokens. Thus, for Themes there is no particular difference across construction in how likely a subsequent evocation is.

Relative subsequent evocation We can now also look at how differences in subsequent evocation correlate with the choice of construction. Based on theoretical considerations we should expect, for instance, that in a situation where the Recipient frame element's referent has an explicit subsequent evocation whereas the Theme frame element's referent has none, the double object construction is used.

For the purpose of comparing the frame elements in terms of their prior evocation status, I ordered the possible values on a scale in 3 subgroups as follows:

- Evoked > Inferrable > Variable

Differences between immediately adjacent statuses were counted as having magnitude 1, differences between the extreme statuses as having magnitude 2.

	Double object	Caused- motion	Total
-2.00	1	[3]	4
-1.00	1	[1]	2
.00	78	21	99
1.00	[11]	3	14
2.00	[72]	17	89
Total	163	45	208

Table 2.14: Relative subsequent evocation of Recipients

On the basis of this treatment, Table 2.14 was generated for the difference in subsequent evocation status between referents of Recipient and Theme frame elements.

The number of ties is rather large in the case of subsequent evocation (47.6%). For the positive differences the correlation is again reasonably strong: when the Recipient has a subsequent mention and the Theme does not, the construction used is mostly the double object construction. The number of negative differences is small so it is not quite conclusive whether they favor the caused-motion construction; they do not do so clearly in this sample.

If we calculate again how many sentence tokens the factor relative subsequent evocation can account for, we get a rather low rate of 41.8% ($208 - 99 - 22 = 87$). This is quite a bit lower than what we found for prior evocation.

Relative length by construction The two constructions also differ by the relative length of the constituents encoding the frame elements. Table 2.15 summarizes the data comparing relative length of frame elements by construction.

A χ^2 -test with six cells yields a χ^2 value of 53.488 and a p-value less than 0.001, indicating a highly significant result.

Around 46% of the χ^2 -value is contributed by the cell intersecting the caused-motion construction and the condition in which the Theme is shorter than the Recipient. For this combination, the observed value is a lot lower than the expected value of 9.08. Thus, there are more cases of a relatively shorter Theme than are to expected if the data are proportionally distributed. The cell contributing the next most sig-

	Double object		Caused-motion		Total	
	n	%	n	%	n	%
T < R	18	11.0	24	53.3	42	20.2
T=R	32	19.6	15	33.3	47	22.6
T > R	113	69.3	6	13.3	119	57.2
Total	163	100.0	45	100.0	208	100

Table 2.15: Relative length of frame elements by construction

nificant chunk to the overall χ^2 -value, 15.1, is the cell intersecting the caused-motion construction and the condition in which the Theme is longer than the Recipient. Not surprisingly, more tokens were expected (25.7) than were actually found (6).

So what we find is what we expected: the proportion of cases where the Theme is shorter than the Recipient is higher in the caused-motion construction than in the double object construction, and conversely, the proportion of cases where the Theme is longer than the Recipient is higher in the double object construction than in the caused-motion construction.

However, it is clearly the case that relative weight by itself predicts the wrong construction in a considerable number of cases. The 18 sentence tokens where a shorter Theme and a longer Recipient are paired in the double object construction contradict the prediction of relative weight as do the 6 sentence tokens where a longer Theme and shorter Recipient are paired in the caused-motion construction.

If we sum up the wrong predictions and the cases for which no prediction is made and subtract them from the total, we end up with $208-24-47=137$ cases, 65.9% of the total, where relative weight makes the right prediction. Note that this is slightly less than the percentage of cases that relative prior evocation status was able to predict correctly, namely 69.7% (cf. p. 53)

Relative definiteness The two constructions also differ by the relative morphological definiteness of the constituents encoding the frame elements. Table 2.16 summarizes the data comparing relative definiteness of frame elements by construction.

Performing a χ^2 -test on the data is not possible since two of the cells have expected

	Double object		Caused-motion		Total	
	n	%	n	%	n	%
both definite	50	30.7	24	53.3	74	35.6
both indefinite	13	8.0	8	17.8	21	10.1
Theme def., R. indef.	2	1.2	4	8.9	6	2.9
Theme indef., R. def.	98	60.1	9	20.0	107	51.4
Total	163	100.0	45	100.0	208	100.0

Table 2.16: Relative definiteness of frame elements by construction

	Double object		Caused-motion		Total	
	n	%	n	%	n	%
Same definiteness	63	38.7	32	71.1	95	45.7
Different definiteness	100	61.3	13	28.9	113	54.3
Total	163	100.0	45	100.0	208	100.0

Table 2.17: Relative definiteness II of frame elements by construction

values of less than 5. So we need to lump these cases in with other cases. One plausible grouping is to lump the cases that involve a matching definiteness value and those that involve different definiteness values. The resulting table is shown in 2.17. A χ^2 -value performed on this 2×2 table yields a χ^2 -value of 14.975 and a p-value of 0.0001, which we ought to consider statistically significant.

For the purpose of predicting the choice of either the double object or the caused-motion construction from the relative definiteness, a rate of 45% of cases in which the values match and no prediction is possible, as shown in Table 2.16, is not satisfactory. For the factor relative weight measured in the number of words, there were ties between the frame elements which allow no prediction in only 22.1% of the cases, as shown above in Table 2.15. However, it is possible to improve the predictive value of definiteness morphology by making finer distinctions. In particular, if we split up

	Double object	Caused-motion	Total
P>I	77	1	78
P>D	30	0	30
D>I	21	8	29
D<P	1	8	9
I<D	2	2	4
I<P	0	2	2
P=P	8	10	18
D=D	11	6	17
I=I	13	8	21
Total	163	45	208

Table 2.18: Relative definiteness III for Recipients by construction

the category definite into pronominal and simple definite by the criterion of whether the phrase contains a lexical head noun, then we can re-present the data from Table 2.16 as below in Table 2.18.

As a result of making a finer distinction there are now fewer cases where the morphological definiteness values match and no predication is possible. The bottom segment of Table 2.18 contains 56 cases, which is an improvement to the 95 cases in Table 2.16 for which no prediction is made.

Of course not all the cases for which a prediction is available are predicted correctly. In the top segment of the Table where the Recipient has a morphological form indicating greater definiteness, there are 9 cases which are incorrectly predicted to occur in the double object construction rather than in the caused-motion construction. Likewise, in the middle segment there are 3 cases which based on definiteness we would expect to occur in the caused-motion construction but which actually do occur in the double object construction.

If we sum up the wrong predictions and the cases for which no prediction is made and subtract them from the total, we end up with $208 - (56 + 12) = 140$ cases, 67.3% of the total, where relative definiteness defined as a three-way distinction makes the right prediction. This success rate is more or less the same as the one we saw for

	shorter	equal	longer	
	Double object			
P>I	56	14	*7*	77
P>D	29	1	0	30
D>I	11	4	*6*	21
D<P	0	0	%1%	1
I<D	*1*	0	%1%	2
P=P	1	(7)	0	8
D=D	5	(4)	*2*	11
I=I	10	(2)	*1*	13
Total	113	32	18	163
	Caused-motion			
P>I	0	*1*	0	1
D>I	%2%	0	*6*	8
D<P	0	2	6	8
I<D	0	2	0	2
I<P	0	0	2	2
P=P	0	(9)	1	10
D=D	*4*	(1)	1	6
I=I	0	(0)	8	6
Total	6	15	24	45

Table 2.19: Relative weight and relative definiteness for Recipients

what relative weight, namely 65.9

Relative weight by relative definiteness So far we have considered the factors in isolation. It is now time to see whether the factors taken together explain more of the data than they do individually. We will explore here the interaction of relative weight and relative definiteness, measured at three levels. The data is summarized in Table 2.19.

The markings in the table have the following meanings. Asterisk indicate constellations where the prediction of one of the factors is violated. For instance, in the first row, there are seven tokens that occur in the double object construction event though the Recipient is longer than the Theme. When a number is enclosed in per-

	shorter	equal	longer	
	Double object			
P>I	46	9	*5*	60
P>D	27	1	0	28
D>I	8	4	*5*	17
P=P	1	(7)	0	8
D=D	5	(4)	*1*	10
I=I	4	(1)	0	5
Total	91	26	11	128
	Caused-motion			
D>I	%1%	0	*5*	6
D<P	0	2	6	8
I<D	0	1	0	1
I<P	0	0	2	2
P=P	0	(9)	0	9
D=D	*4*	0	0	4
I=I	0	0	5	5
Total	5	12	18	35

Table 2.20: Relative weight and relative definiteness for Recipients in active voice sentences

centage signs, it means that both factors' predictions are violated. For instance, in row 4, there is one token that occurs in the double object construction even though the Recipient is longer than the Theme and only a lexical definite whereas the Theme is a pronominal definite.

If we proceed in the same way as before to calculate the rate of correctly predicted tokens, we find that $208-55=153$ or 73.6% of the tokens are handled correctly by the two factors combined, assuming that either both factors have to point in the same direction or that one is silent while the other makes a prediction. This is the highest number encountered so far. The increase over the individual factors is not dramatic but seems real enough. A regression analysis will be needed to ascertain this.

One factor that might interact with weight and definiteness is voice. Let us therefore consider how successfully weight and definiteness predicted the constructions once we separate the two voice constructions from each other. Let us begin by considering active voice sentences. The relevant data is summarized in Table 2.20.

	shorter	equal	longer	
	Double object			
P>I	10	5	*2*	17
P>D	2	0	0	2
D>I	3	0	*1*	4
D<P	0	0	%1%	1
I<D	*1*	0	%1%	2
D=D	0	0	*1*	1
I=I	6	(1)	*1*	8
Total	22	6	7	35
	Caused-motion			
P>I	0	*1*	0	1
D>I	%1%	0	*1*	2
I<D	0	1	0	1
P=P	0	0	1	1
D=D	0	(1)	1	2
I=I	0	0	3	3
Total	1	3	6	10

Table 2.21: Relative weight and relative definiteness for Recipients in passive voice sentences

When we calculate the number of successfully predicted tokens for this Table alone, we find that out of 163 tokens, 121 or 74.2% are predicted correctly, which is slightly better than in the case where we consider tokens of both voices together.

The data for passive tokens are summarized in Table 2.21.

When we calculate the number of successfully predicted tokens for this Table alone, we find that out of 45 passive tokens, 32 or 71.1% are predicted correctly, which is less than is the case where sentences in both active and passive voice were considered together.

2.2.2 Cases that contradict the predictions of weight and definiteness

To get a sense of what may cause certain tokens to violate one or both of the predictive factors of weight or definiteness, let us take a closer look at the sentences.

- Table 2.19 shows 6 cases where the caused-motion construction is used where the Recipient is longer than the Theme but also unexpectedly more definite. These are cases of Heavy NP shift, as shown below:

(2.37) The supervisor is required to keep accounts and records and file an abstract of them with the court at least once a year, **giving** copies to the debtor and to all creditors.

(2.38) Before the sale, Murphy and Cree deliberately **gave** some of their stake to their fellow directors and to an employee trust as a means of motivation.

In these 6 cases, the Recipient is on average 3.5 words longer than the Theme and it might be the magnitude of the difference that favors the caused-motion construction despite the fact that the Recipient is more definite.

- One token of the caused-motion construction involves a Recipient that is more definite than, but as long as Theme.

(2.39) It was here that items of the Emperor's clothing and the superb diamond necklace which had been **given** to him by the Princess Borghese [his sister, Pauline] as well as the Landau which had escaped the Moscow disaster in 1813 (sic) were taken.

This token we may explain just by the fact that both in the British and even more so in the American variety, passives of *give* and other ditransitive verbs which involve promoting the Theme to subject with double object-linking as a post-verbal NP for the Recipient, as in (2.40) are deprecated.

(2.40) It was here that items of the Emperor's clothing and the superb diamond necklace which had been **given** him by the Princess Borghese [his sister, Pauline] as well as the Landau which had escaped the Moscow disaster in 1813 (sic) were taken.

An active form of the relative clause would be more acceptable out of context (*which the Princess Borghese [his sister, Pauline] had given him*) but in this particular case the passive linking has the advantage that it preserves a non-transitive series of clauses through the main clause and both relative clauses.

- 2 cases involve the use of the caused-motion construction even though the Recipient is both more definite and shorter than Theme.

(2.41) Twice, though, Poiret **gives** a hostage to fortune by invoking the name of Feydeau.

(2.42) A grant of £ 950 will be **given** to Dramarama creative arts to part fund a theatre project in Middlesbrough for people who are physically handicapped or who have learning difficulties.

I coded example (4.102) as having a personified, culturally available unused referent “Fortune” for a Recipient. If that analysis were rejected, then at least the definiteness factor would no longer produce a violation. But that move does not seem warranted.

Now, it also seems to be the case that “give a hostage to fortune” is somewhat of a collocation: the expression with all possible inflected forms of *give* was matched 106 times in the caused-motion form on the internet by the Google search engine but no tokens of a double object counterpart could be found. We might thus be justified in classifying token (4.102) as an instance of an idiom with a fixed form.

The second example, (2.42), is amenable to a different explanation. Consider the sentence in its full context:

(2.43) GRANTS totalling more than £37,000 for a host of projects being run by organisations in Cleveland have been agreed by the county council. Respond, an inter-church organisation tackling unemployment, is to receive £8, 000 from the 1992 - 3 development fund.

The council is providing cash towards the cost of a major survey in Hemlington, Middlesbrough, where plans for a Community Action Plan are under way.

The county council and borough council are each providing £6,000 for a survey to find out what local people want in terms of estate development.

A grant of £950 will be **given** to Dramarama creative arts to part fund a theatre project in Middlesbrough for people who are physically handicapped or who have learning difficulties.

Other recipients are: Raybee Credit Union, Hartlepool, £800, for office furniture and equipment; Thornaby Impasse, £620, to buy equipment for the centre; Outreach Advice and Advocacy Service, Thornaby, £3,935, towards the cost of providing advice sessions.

...

The sentence is part of a news item about grants that were approved by a county council. Following the introductory sentence, the reader is presented with the information pairing the awardees (Recipients) and the amount of the grants made to them (Themes).

In cases of paired lists like this, it is somewhat arbitrary how one proceeds and in fact the order of presentation changes. The second sentence topicalizes the heretofore unmentioned organization Respond. The speaker then switches back to topicalizing the awarding council and linking the Theme to the object slot, putting off mention of the awardee or their cause till the end of the sentence. At the point where our sentence token occurs, the organization changes once again. Now the grant amount is topicalized in the passive construction and the awardee appears focally at the end of the sentence. In the next sentence the mode of presentation changes yet again to a simple comma-separated paired list without any overt linguistic relation expressed.

Thus, we might suggest that the sentence token in question might be the product

of a conscious editorial effort to vary the linguistic expression as the list of grants and awardees is presented, a stylistic consideration unlikely to be part of the basic factors that drive selection of constructional alternatives.

- Another group of unexpected tokens involves uses of the **caused-motion** construction where the Recipient is shorter than the Theme, though equally definite.

(2.44) I was unhappy, and afraid, and I was pregnant again. They took me to Lochleven Castle, and put me in a room like a prison. There, I did not eat for two weeks, and Bothwell's children—there were two babies—were born dead. I nearly died too—I was so angry and tired and ill. Then, one day after the babies died, Lord Lindsay gave me a letter. It said:

I, Mary, Queen of Scots, **give** the Kingdom of Scotland to my son, James. From today, James is the new King of Scots. But because he is a child, the Earl of Moray, my half-brother, can rule the country for him.

Because I was afraid, and tired, and ill, I wrote my name on the letter: Mary . But it is not important, James, it doesn't change anything. I am Queen of Scots, not you. That letter changes nothing.

(2.45) One of the world's greatest ethnographic museums. Founded in 1884 when General Pitt Rivers **gave** his collection of ethnology and prehistory to the University, it has now grown into an astonishingly diverse collection of the products of human art and ingenuity from all periods and cultures. Exhibition of Musical Instruments—to see and hear—and Hunter Gatherers, past and present, are in the nearby Balfour Building.

(2.46) Considering the two versions of *King Lear* , Andrew Gurr discusses the implications posed by the two endings of the play. The Quarto version **assigns** the play's final lines—"The weight of this sad time

we must obey; /Speak what we feel, not what we ought to say. / The oldest hath borne most; we that are young/ Shall never see so much nor live so long" (V, iii, 322 - 6) —to Albany, the Folio **gives** the lines to Edgar.

- (2.47) And thirdly, Paul argues that the way of self-commendation to God by merit is not the way of God's heroes in the Old Testament. The Jews of Paul's day could draw no credit for being sons of Abraham unless their attitude was like that of Abraham. God **gave** his inheritance to Abraham by promise, not as a reward for law-keeping (Gal. 3:18).

In (2.44), the line containing the target *give* needs to be imagined as part of a proclamation. Within this context of use, the used surface order reflects the fact that the referent James has not been mentioned. It's true that in the actually used order the object *the Kingdom of Scotland* also has not been mentioned before, however, it is much more inferable as the sentence begins with the self-reference to the queen as *Mary, Queen of Scots*, which evokes the idea of the reign, but not the idea of motherhood.

Example (2.45) is taken from a leaflet describing museums in Oxfordshire, England. One of the two major sections, the one where the sentence token above comes from, is concerned with the Museums that are part of Oxford University. The referent of noun phrase *the University* is clearly given in context; all references to the institution are just in that abbreviated form rather than with its full name. The collection of ethnology and prehistory, by contrast, has not been mentioned before and is anchored by the possessive in the referent General Pitt Rivers. One might thus expect to find the double object construction used.

One possible explanation why this is not the case relies on the comparison of the likelihood of subsequent mentions for Recipients between constructions. It was found that in the caused-motion construction, that likelihood is smaller. The use of the construction here may thus signal that the University itself is not an important local topic that is bound to be a sentence topic in following

clauses, which it indeed is not. While the University figures prominently in the background, it is not what the text segments are about: they are concerned with the holdings and exhibitions of the university-run museums.

Another explanation is that this token should have more appropriately been treated as a case of donation and excluded from this sample. The other lexical units that evoke the donation frame (donate, contribute) frame realize their arguments only in the caused-motion construction.

The third example, (2.46), clearly involves a case of contrastive focus between the referents of *Albany* and *Edgar*. Further, although the reader may be expected to recognize the two roles from *King Lear*—primed on the one hand by the mention of the play but also by the fact that the context of the sentence is about the reasonableness and the difficulty of trying to establish authoritative versions of Renaissance texts that exist in multiple versions—the two roles have not been recently mentioned at all.⁸ The lines whose attribution is at issue clearly have been mentioned and they have in fact been rendered fully previously. They are therefore more accessible than the characters of the play. Thus, from any consideration other than length, it makes sense that the caused-motion construction is used.

The third example much more clearly than the second also does not plausibly represent a case of concrete transfer and should have been excluded from this sample. On the other hand, it is not unreasonable to assume that the metaphorical cases of giving are sensitive to the same kinds of factors as the concrete ones: if we analyzed (2.46) in a sample of metaphorical sentences, it would still require an explanation.

The final example, (2.47), is a difficult one. The referent of *Abraham* is mentioned twice in the sentence before the sentence at issue, whereas the idea of inheritance has not been explicitly evoked; the reference to *sons of Abraham* may make the idea somewhat accessible, though. The sentence also involves an ambiguity in how the inheritance is referred to: it could be viewed as some-

⁸Not within the last 150 sentences, which is as far as I looked back at the prior context.

thing that God gives away or as something that Abraham receives. On the former reading, the pragmatics of the sentence is better since now the referent is anchored by the highly accessible subject God; on the latter reading, the pragmatic structure is worse since the pronoun is not anteceded by the closest possible antecedent, God.

Reordering the sentence in the double object construction would also allow two readings. But now the reading with Abraham as the antecedent of *his* is more prominent and plausible than in the caused-motion version: indirect/primary objects are known to be good secondary topics (Givón 1984:163). Given these paradigmatic considerations, it would sense for the caused-motion construction to be chosen if the intended reading is indeed that the inheritance is what God gives away and not what Abraham receives.

- The next set of unexpected tokens consists of sentences in the double object construction that have a Recipient that is equal in definiteness status to the Theme but longer.

(2.48) RC schools get extra cash for more classrooms

Two Catholic schools in Darlington will be among those to gain from new rules on extra classrooms, the town's MP said this week.

Mr Michael Fallon, who is also schools minister, said it was an historic change which he had been working on with Catholic bishops since November.

Canon Bob Spence, the senior Catholic priest in Darlington, said he was delighted with the news.

But Ms Hilary Armstrong, education spokesman for Labour and North West Durham MP, called it a desperate measure to save Mr Fallon's parliamentary seat.

The new move, announced by Mr Kenneth Clarke, Secretary of State for Education, means over-subscribed schools will be given money to expand even if empty places exist in nearby schools.

It will apply to state schools as well as those that are voluntary-aided. "But the first to benefit will be successful church schools because they might have been discriminated against in the past," said Mr Fallon.

- (2.49) A characteristic working practice which the companies brought with them covered the organization of the underground process of coal-winning. This was known as the "butty system", which has a long history in the area (Griffin, 1977, p. 26). "Butties"—the local term for subcontractors or chargemen—were employed on a contract basis to run much of the day-to-day operations of each pit. 593 Griffin (1971, p. 29) describes their role thus:

the company would sink a shaft, then invite tenders from the butties. The butty who quoted the lowest price per ton for getting and raising the coal was **given** the contract. The butty would then engage his men ... [and he] supervised and paid the men, and provided the working capital. Sometimes they were responsible only for work directly connected with production, but in some cases they were responsible also for development work, drainage and ventilation.

- (2.50) The obtrusive tone of the tuba makes it extremely useful in bringing important bass passages into prominence. It should be used somewhat sparingly, however, as it has a tendency, except in very full tuttis, to overload the bottom octaves of the orchestra.

Its everyday function is to provide a strong, firm bass to the trumpet-trombone group. In soft brass passages it provides a perfectly satisfactory bass by itself, but in loud ones the difference in quality between its "blunt" tones and the "sharp" utterance of the trumpets and trombones makes itself felt. It is therefore advisable to **give** the bass to the bass trombone and to double this at the octave below with the tuba.

In the case of (2.48), the chosen construction in my judgment is preferable to its alternative (... *means money will be given to over-subscribed schools to*

	Donor	Recipient	Theme	Unclear	Total	n/a
Caused-motion	3	4	0	2	9	5
Double object	7	45	4	1	57	29
Theme-only	16	2	0	1	19	10
Total	26	51	4	4	85	44

Table 2.22: Orientation of purpose clauses with *give money*

expand even if ...). One reason is that in the text the schools are more topical than the money. The sentences following the sentence of interest keep talking about how state versus church schools will be affected and what counts as an oversubscribed school. By contrast, no more mention is made of money.

Another reason why this particular sentence is more appropriate in its actual form than in its alternative formulation has to do with the presence of the purpose clause. It seems to be the case purpose clauses oriented towards the Recipient typically are found when that role is linked to primary object.

To test this hypothesis, I extracted all sentences from the BNC where a form of *give* was followed within 6 words to the right by the noun *money* and then within another five words by *to* and an infinitival verb form. (I picked *money* as a Theme because I expected it to be reasonably frequent as a Theme of *give*.) This query matched 142 sentences. In 13 of them the noun *money* was not an argument of *give*. In others, listed in the *n/a* column in the table, *give* occurs in one of the three usual constructions with *money* as a Theme but the infinitive phrase is not related to the verb *give*.

As Table 2.22 shows, the intuition about Recipient-oriented purpose clauses is largely true: if there is one, then it almost always occurs in the double object construction. The Table also shows that it's not impossible to have a Recipient-oriented purpose clause; it's just not as common.

We now turn to example (2.49), where both the referent of *contract* and the role *[T]he butty who quoted the lowest price per ton for getting and raising the coal* are inferable given the context. The chosen construction preserves topic

continuity with the following clause more readily than its alternative and the butty meeting the criteria is also more readily inferable from previous explicit references to butties than is the contract, which is only frame-inferable from the bidding process.

Finally, in example (2.50), we also have reasons to favor the chosen construction despite its being non-optimal with respect to weight. The example involves a contrast between the two referents tuba and bass trombone. The former is the explicit topic of the preceding several sentences, while the latter is only inferable. (The inference builds on the knowledge that the discussion is about which instrument should play bass passages and on the fact that the major trombone types are tenor and bass trombones, which the reader presumably knows.) By not linking the bass trombone to primary object in a double object construct, the speaker not only reflects the relative lesser topicality of the bass trombone but they also avoid misleading the reader about the topic status of the bass trombone in the following discourse: the expectation of subsequent mentions is lessened in the caused-motion construction, which is appropriate here as the next paragraph continues to talk about the tuba. Also note that the example under consideration is taken from a section on “The Tuba (in F, non-transposing)” from a book entitled *Orchestral technique: a manual for students*. This fits the impression that the tuba is the major topic of the passage.

- The next class of sentences where the **double object** construction is used unexpectedly involves cases where the Recipient is both less definite and longer than Theme. The sentences ostensibly violate the predictions both of weight and definiteness.

(2.51) Anyone working in the retail trade will often have to deal with customers who want to exchange or return goods. The item may be faulty, or the wrong size/colour, or the customer may have been **given** it as a present. The procedure for exchanging goods varies from shop to shop, but may include:

...

offering a replacement

...

- (2.52)
- We will complete the introduction of the National Curriculum offering 10 subjects at a nationally-defined standard—English, Mathematics, Science, History, Geography, Technology, Art, Music, PE and, in secondary schools, a foreign language.
 - Regular and straightforward tests will be in place for all 7, 11 and 14 year-olds by 1994.
 - GCSE at age 16 will be integrated into the National Curriculum, with a new A+ grade to test the most able. The majority of marks will come from a written exam.
 - We will continue to encourage the creation of nursery places. For the first time, over 50 per cent of three and four-year olds have places either in nursery or primary schools.
 - Full information will be published annually about the performance of all local schools in each area.
 - Independent inspection of schools will provide parents with straightforward reports on their child's school, together with an action plan from governors to remedy any weaknesses.
 - Popular schools which are over-subscribed will be **given** the resources to expand.
 - GM schools will be able to change their character if that is what parents clearly want and the change fits in with the wider needs of the local area.
 - The Technology Schools Initiative will be expanded across the country.
 - Existing schools which opt for GM status will be able to emulate City Technology Colleges and attract private technology sponsorship.

- We will maintain the Assisted Places scheme, which gives access to independent education to many families who could otherwise not afford it.
- We will ensure that the partnership between the state and the churches in education is maintained and strengthened.
- We will enable small schools to apply for GM status in groups.
- We will pay particular attention to raising educational standards in areas of deprivation in our cities.

The attested version of example (2.51) is not a completely unreasonable choice. In the preceding sentence, there is an overt reference to generic customers who want to exchange or return goods so the referent of customer is more accessible than just being inferable from the retail scenario.

Still, it seems to me that (2.51) could have been more felicitously formulated with a passive combining with the caused-motion construction (...or it may have been given to the customer as a present). It could also have been phrased without any overt realization of the Recipient (...or it may have **been** given as a present), given that the Recipient is just a generic customer and is never again mentioned overtly in the passage. (However, the customer has to be accessed in several clauses where the Recipient argument of *offer* is null instantiated, as exemplified above.) Both the caused-motion and the Theme-only constructions offer the benefit of topic continuity between the disjoined clauses, though they would leave it up to inference to establish that the person wishing to return the item has received it as a gift.

Actually, it seems the best way to express the facts with the customer as topic and the goods as object would be to simply use the lexical unit *receive*:

(2.53) The item may be faulty, or the wrong size/colour, or the customer may have received it as a present.

Here we need to remind ourselves that an attested sentence simply may not be

the optimal expression of its content in its context. In evaluating and generalizing over speaker choices between constructions, we should not fall into the trap of believing that every linguistic production is either perfect or completely ungrammatical/incomprehensible. Clearly, speakers are capable of articulating graduated judgements about the grammaticality and discourse appropriateness of at least some linguistic constructs. Matters of information structure generally fall into the class of linguistic phenomena where graduated judgments are made (as opposed to, say, the relative order of determiners and head nouns in NPs). The sentence at issue simply happens to be a case where a pragmatically non-optimal encoding of the content was chosen.

The example in (2.52) comes from a campaign-related publication of the British Conservative party ahead of the 1992 election. As shown above, the sentence token under discussion is part of a list of campaign promises about what the future situation for education will be.

In the early part of the list the subject switches back and forth between the governing party (*we*) and referents related to the schooling scenario (*GCSE, tests*). Our target sentence appears at the beginning of a short series of items where subclasses of schools and the changes affecting them are discussed (with the interruption of the sentence about *The Technology Schools Initiative*). Afterwards, the point of view shifts back to the governing party.

We may speculate that part of the reason why the target sentence has its form is the wish to create a parallelism between the items that discuss schools with particular statuses or characteristics. Another factor is that, if we disregard the particular grammatical place in which reference to the idea of school is made, the concept school is a great deal more prominent than the concept resource. The 10 sentences preceding the target sentence contain 8 tokens of word forms of the lexeme *school* and the 10 sentences following the target sentence contain another 4, whereas no other tokens of word forms of the lexeme resource occur anywhere in this 20 sentence window of context.

The concept of schools expanding is evoked before the target sentence. But

that reference is 10 sentences earlier and many different propositions intervene before the idea of expansion is re-evoked. The idea of “the resources (needed) to expand” is thus no longer accessible based on the initial mention. In fact, the definiteness of the NP here seems to rely on the fact that the the crowding of “over-subscribed” schools has been suggested by the subject NP. If one wanted to reverse the target sentence and make the resources-NP the subject (and topic), one would, in my opinion, preferably use a bare NP. Compare:

(2.54) Resources to expand will be **given** to popular schools which are over-subscribed.

(2.55) ?The resources to expand will be **given** to popular schools which are over-subscribed.

Examples like suggest (2.51) and (2.52) are, of course, cases where choice of sentence topic as a factor more or less pre-determines the choice of linking construction. And it seems that in those cases—where the two arguments do not occur adjacent to each other post-verbally—violations of the factors weight and definiteness may be more tolerable.

- Violations of the **double object** construction where the Recipient is more definite but longer than the Theme.

(2.56) She wore no jewellery, except a gold and platinum twisted ring that some friend had **given** her, and the heavy signet ring, perhaps it had been her father’s, that weighted one hand.

(2.57) I’ve bin rubbin’ ’im wiv that liniment ole Doctor Kelly **give** me, an’ I made ’im wear a sheet o’ brown paper under ’is vest.

(2.58) She lived on in them, in their memories, not in the ashes he knew he would soon be **given**.

(2.59) That’s the magic ointment Hepzibah **gave** me.

- (2.60) It was biting, but she was censoring herself instinctively now, because Florian's personality was truly one-dimensional, albeit in another way, and the only thing she had ever **given** him was the degree of tolerance his peculiar genius made his due.
- (2.61) The money Danica **gave** my mother was not enough to buy one of the better sorts of bicycle, but it was no good buying a cheap one as it would have to last me for years.
- (2.62) By the time he was seven he had "perfect pitch"; he could also improvise on the keyboard in various styles, play with the keys hidden under a cloth, sight-read anything he was **given**, add a bass line to a tune, and had "taught himself" the violin
- (2.63) One touching moment during a Love Feast in Edenderry was once when a visitor to the service, who had never attended a Love Feast before but who had felt he wished to share in it, stood and held up half the biscuit he had been **given**
- (2.64) I never knew how much Albert and Mary **gave** Mum and Dad, but they threw money around like water; as Mum had said, she was going to give Mary the best turn-out in the district.
- (2.65) When I took over the club from Connelly I agreed to **give** old Giuseppe there a job, he explained.
- (2.66) Take three thrillers
Our Foreign Secretary, Douglas Hurd, was **given** three thrillers to review by the Daily Telegraph: Geoffrey (not to be confused with Lord) Archer's Eagletrap (Century), Clive Egleton's Hostile Intent (Hodder) and J K Mayo's A Shred of Honour (Harvill). It was Mr Hurd's first paragraph that drew me to the rest of the review.
- (2.67) "Well, I love you for a start—and there's Mum and Dad and the kids. I told you Jenny wanted to get married, didn't I?"
It reminded Harry of his plan to **give** Jenny and her boy some money

towards the shop they were saving to buy. Alice thought it a wonderful gesture, and for the remainder of the journey they discussed their joint wish to make Will and Martha Pritchett financially secure for the rest of their lives.

- (2.68) The last prize to be given was £1,000 in cash. It was nice to hear everyone give the chap a big cheer as he came forward, obviously overcome with excitement. It was also nice to see him give the young girl who had drawn his winning number a £20.00 note as a way of saying thank you.

Among this last group of unexpected constructional uses, there is a significant subgroup consisting of examples (2.56-2.63) which is characterized by the fact that the Theme is relativized on. Since I treated cases like these as involving a variable and as having length zero for the Theme these tokens necessarily come out as violations. Note that this problem would not be fully solved by changing the analysis of *that* from that of a marker of relative clauses to an analysis as a relative pronoun. Only one of the relative clauses is a *that*-relative, the other six have no overt pronoun or marker.)

Obviously, it is a reasonable and not at all unusual discourse move for a speaker to choose to anchor a Theme referent in an event of giving. We should thus not count these cases as simple violations of what weight and definiteness predict but as situations where an overriding interest comes into play.

Let us now turn to the remaining examples in the group.

In example (2.64), the Theme argument is questioned. In this context it is clearly focal and indefinite, which properties go with the double object construction that is used. Use of the caused-motion construction would imply that there is also some focal status to be given to the Recipient, which would be misleading. Pragmatic considerations here strongly favor the double object construction regardless of weight. As noted before, it may be the case that relative weight is a lesser factor in situations where the two arguments are not

realized contiguously in post-verbal position.

In example (2.65) we find a (length difference between Theme (*a job*) and Recipient (*old Giuseppe there*), which should favor the caused-motion construction. That the double object construction is the form actually chosen is not surprising. First, the Recipient is mentioned 7 times in the space of the preceding 10 lines and is topical whereas the idea of a job or employment is not explicitly evoked (in nominal form) at all in the context. Second, the referent is given in the speech context (*old Giuseppe there*), making him even more accessible.

Example (2.66) is similar to the previous example. First, the Recipient is very prominent in the community due to his status as a government official. The Theme referents are presumably a lot less well known: none of the given titles seem familiar. Second, the Recipient re-appears as the possessive in the subject NP of the next sentence (*Mr Hurd's first paragraph*). Having a subsequent mention is more common for Recipients in the double object construction than in the caused-motion construction. The chosen construction also allows for closer proximity between the NP *three thrillers*, which introduces the compound referent and the list of novels that is given after the colon.

Example (2.67) belongs to this set only if we consider the towards-PP as not being part of the NP *some money*. Otherwise, the Theme would be longer than the Recipient and both definiteness and weight would predict the occurring construction. I do, however, think that the towards-PP is more like a purpose infinitive that is dependent on the verb *give*. Under this view, the chosen construction is more appropriate than the caused-motion alternative because the Recipient has been evoked in the last conversational turn by Harry's interlocutor and because the Recipient referents figure in the following sentence (*Will and Martha Pritchett, their*). By contrast, the NP *some money* has a non-specific referent due to the irrealis nature of the clause and there is no further reference to money which is to be given to the Recipient.

In example (2.68) neither the Recipient nor the Theme has a prior or subsequent mention. The Recipient, the prize-drawing girl, is clearly anchored in the sce-

nario, though, and is frequently part of prize drawing events. Definiteness and animacy thus weigh in favor of the chosen construction. Another consideration that may be important in this particular sentence is ambiguity avoidance: if the caused-motion construction were used, as in (2.69), the modifier *as a way of saying thank you* may be interpreted relative to the drawing event, thereby giving the impression of impropriety.

(2.69) It was also nice to see him give a £20.00 note to the young girl who had drawn his winning number as a way of saying thank you.

- The final type of violation involves a single use of the **double object** construction where the Recipient is less definite and shorter than the Theme.

(2.70) In the meantime the police had been called, but they stayed in the background throughout the proceedings. When Mr Szuluk and his brother emerged 2 hours later he still hadn't got his clothing grant.

He says that he hasn't been offered a grant, and the case has been referred to higher officials. He's fed up.

The DSS says every claim is decided upon its merits and the amount of cash it has available in the social fund.

He says there's only limited funds in the kitty—Mr Zuluc was offered a loan but refused this.

The prisoners welfare group Nacro, says every prisoner leaving jail should be **given** at least the basics to make a fresh start.

He says that every prisoner leaving jail should at least have enough clothing to make a fresh start, the danger is that prisoners who don't have enough funds can return to crime.

Mr Szuluk is appealing against the DSS's decision. In the meantime, it looks as though he's back to borrowing his brothers clothes again.

This last example, (2.70), comes from the transcript of a TV news broadcast. This origin is somewhat of a problem for the analysis in that the broadcast is a

composition of the speech of different speakers whose utterances were not meant to be related to each other but are juxtaposed to appear as if they formed a common discursive thread together with the announcer's words. Assuming for the sake of the exercise that the editor has done their work well, we can ask how the utterance fits in with what has previously been said.

Clearly the utterance represents a generic characterization of how released prisoners should be treated. The set of all released prisoners is inferable given the particular case that is reported on in the first part of the news item. Likewise, the concept of "the basics to make a fresh start" is also accessible given the problem described of not having enough money upon release to clothe oneself.

As in other cases we have seen, there basically is no choice of linking construction here that would be independent of the choice of sentence topic. Several considerations makes *every prisoner* a good sentence topic. First, the next utterance elaborates on the present one and presents the same subject referent. This sequence is more coherent due to the shared topic (though it is also repetitive). Further, despite the original coding of the quantified NP as being indefinite, we might want to treat the quantifier *every* here as a kind of honorary definiteness marker as it involves considering the set of entities of the specified type exhaustively. Finally, in the typically non-focal position that the Recipient occupies in the double object construction, whether in the active or passive voice, there is no sense of contrast between the asserted exhaustiveness and a possibility that the relation might be true for only some instances of the type. This kind of contrast does seem to be suggested when an NP quantified by *every* occurs as the oblique Recipient in the caused-motion. For instance, example (2.71), which is adapted from (2.70), seems to suggest that there is a position opposed to that of Nacro according to which only some prisoners should be given the basics to make a fresh start.

- (2.71) The prisoners welfare group Nacro, says at least the basics to make a fresh start should be **given** to every prisoner leaving jail.

The suggested focality of the Recipient in example (2.71) also clashes with the focal status of the Theme which is cued by *at least*, which is a so-called partial restrictive focusing modifier.

2.2.3 Cases where no prediction is made

In addition to the 32 cases in which either one or both of the factors' prediction was violated, there are 23 cases in which the two factors both result in ties and no prediction is made.

Two lexical NPs We will begin with the cases where two lexical NPs of equal length are involved. The first to consider involves two indefinite NPs appearing in the double object construction in (2.72).

(2.72) Looney balloon

This is a game which can be played in two ways depending whether you have younger or older children. Choose two teams and **give** each one a balloon. Set an obstacle, such as a chair, a distance away from the teams.

Example (2.72) is a case where the Recipient is inferable as a subset of the two teams mentioned in the preceding clause whereas the Theme referent has no prior mention within the text, though it is mentioned in the heading preceding the description of the game. However, there is no reason to assume that the title reference to the concept balloon ought to count as an antecedent in the text. For one, the reference to the balloon is indefinite rather than definite, suggesting that the hearer is not expected to have a balloon referent in their discourse model. Also, the reference in the title is part of a name and not of a mention of an actual referents. Finally, titles and intermediate headings are not places where readers are supposed to look for new referents being introduced.⁹

The next case again involves relativization.

⁹Consider a typical example from the New York Times (Thursday, August 26, 2004):

- (i) Law Banning Type of Abortion Is Unconstitutional, Judge Says By JULIA PRESTON
A federal judge in New York ruled today that a federal law that banned a form of abortion

- (2.73) ITT Sheraton : an inch thick manual of good practice; initiatives ranging from tree planting to voluntary donations from customers matched by the hotels and **given** to local conservationists.

We now turn to sentences with two definite NPs appearing in the double object construction.

- (2.74) John Hill had queried Miller's description and had published his views in Eden, or a Complete Body of Gardening (1759) and Justice retaliated in his British Gardener's Calendar (1759): "I thought it my indispensable duty to **give** Mr. Miller his due both from a love of truth and from a sense of gratitude to him for his public labours, as well as private friendship."

- (2.75) The one major drawback with this proposal to "**give** Channel Four its freedom" is that it would make it compete with the existing ITV companies.

- (2.76) "Come in!" Captain Seru called. She was a little woman with a fat face. "Come in! Would you like a drink?"
 "No, thank you," Kiah answered. "I'd like you to look at these numbers."
 Kiah **gave** Captain Seru his book.

- (2.77) In today's Property issue referring to flats in North Meggetland in Edinburgh, we are told, first, that the site was part of the Burgh Muir, whereas the nearest part of the muir is at Tipperlinn, almost a mile away.

Secondly, we are told that David I **gave** the town the muir in 1128, whereas to the best of our knowledge he did nothing of the sort. What is likely is that the royal burgh and its muir were set up by David's brother, Alexander

is unconstitutional because it does not include an exception for cases where the procedure might be necessary to protect a woman's health. The law, the Partial Birth Abortion Ban Act, enacted in November, makes it a crime for doctors to perform any "overt act" to "kill the partially delivered living fetus."

In (i), the first sentence of the article following the title re-introduces the judge-referent in indefinite form rather than continuing with either a name or a personal pronoun, indicating that the reader is not expected to be familiar with the referent.

I, in about 1120, but what is certain is that when David founded Holyrood Abbey in 1128, or soon after, his charter makes it clear that Edinburgh was by then a going concern.

The first two examples, (2.74) and (2.75) seem to be explainable in the same way: the Theme referent is anchored by a possessive pronoun whose antecedent is the Recipient.

Example (2.76) superficially looks like the previous two cases but here the possessive pronoun anchoring the Theme referent is anteceded by the subject. In this context, neither Theme nor Recipient contrast with anything. What favors the chosen construction are that it seems to be the more frequent pattern anyway, and the fact that the Theme is relatively less prominent. It is mentioned before in the text, but at a distance of 20 sentences, compared to the Recipient referent being referred to in the previous sentence. Moreover, the book is not referred to again in the remaining 70 sentences of the chapter, whereas the Recipient continues to feature prominently in the narrative.

Example (2.77) does not involve a comparable difference in prominence in terms of the number of contextual mentions. The Theme, the muir, is mentioned twice before our target sentence whereas the Recipient, Edinburgh, is mentioned only once before. The muir has one subsequent mention as does the town—unless we assume that the burgh metonymically stands for the whole settlement. The simple double object construction is of course favored by sheer frequency. Possibly there is also a sense that since Recipient and Theme stand in a spatial relationship together, the Recipient as Ground should be treated as more topical relative to the Figure, the Theme. But this is speculation; it would not seem inappropriate here for the caused-motion construction to have been used.

There also is one sentence where two definite NPs occur in the caused-motion construction.

(2.78) It was clear to Alexei that if the Merkuts did not wish to see Jehana **given** in marriage to Artai, there were reasons which had nothing to do with the suitability of the match.

Given that the Theme is chosen as a subject, example (2.78) is preferable over (2.79). Combinations of double object linking and Theme subject are rare and for the most part occur only when the Recipient is pronominal.

(2.79) ...see Jehana **given** Artai in marriage ...

But more importantly, the chosen construction has the right focal alignment: Artai here is contrasted with the person that Jehana has actually been given in marriage to, namely Alexei. This fact also makes the occurring form preferable over an alignment with Artai as the subject in a double object passive, as in (2.80), where we would naturally expect a non-focal subject.

(2.80) ...see Artai **given** Jehana in marriage ...

Two personal pronouns We now turn to cases where the arguments are non-lexical definites.

The majority of sentences in the sample involving two pronouns appear in the caused-motion construction, (2.81-2.78).

(2.81) And if I got summat Annie wanted—you know, like a little necklace or a whistle or summat, I'd just **give** it to her, and Mum would say what a polite boy I was, and Dad would smile at me, and we'd all be happy.

(2.82) If it is then I **give** it to you gladly.

(2.83) One of the band **gave** it to him earlier.

(2.84) **Give** it to me and I'll hand out the stuff.

(2.85) In fact you might as well sign the lot and **give** them to me.

(2.86) Barbie said she'd rather have a cigarette anyway and **gave** them back to us, but we couldn't break any of them off, so we kept passing this lump backwards and forwards, sucking it and grinding little chunks off it, till we got fed up with it and chucked

- (2.87) Gramps **gave** them to me.
- (2.88) The Pope had **given** it to him in honour of his piety and when he caught Philippa to himself the design left an imprint on her cheek, Defender of the Faith, carved in Latin.
- (2.89) If a woman poses for you, she **gives** herself to you.

In American English, examples (2.81-2.89) are standard and there is hardly any variation. When both Theme and Recipient are pronominal, then the caused-motion construction has to be used.

In British English, the use of the caused-motion construction in the case of two pronominal objects seems to be less entrenched. Example (2.90) was found in our BNC sample.

- (2.90) Sometimes our neighbours want the things we have, or have the things we want, so we both fight until they take ours or **give** us theirs.

For example (2.90) there is one very good reason why this particular sentence should appear in its chosen form, namely the rhetorical parallelism and contrast between the final two conjoined clauses, where the Theme appears clause-finally in each case.

Such a clear explanation is not available for all cases where two pronouns appear in the double object construction. Consider sentence (2.91).

- (2.91) I'll **give** you it.

The particular token in the sample is from work of fiction but the BNC contains a total of 18 sentences with the string *give you it*, and additional tokens with other personal pronouns for the Recipient. These tokens overwhelmingly belong either to the spoken part of the BNC or are part of quotations in works of fiction, which suggests that this a regular pattern in some British varieties, which was also confirmed by a native speaker of British English. The texts in the BNC contain information about the place of residence of the author of a work of fiction or of the interlocutors

for recorded speech but inspecting this information for the sentences containing *give you/her/him/us/me/them* it doesn't show any clear regional bias.

In addition to cases of the double object construction with the normal linear order, we also find some cases of reversed linear order, as in sentences (2.92) and (2.93).

(2.92) Francie appeared, wearing his harp tie-pin and carrying his fiddlecase.

"I like your harp," said Melanie because she loved him.

"They **gave** it me St Patrick's night," he said. "At the Dagenham Irish Club."

They were all ready, spruce and clean as for going to church, Sunday trim.

(2.93) "No!" she gasped. "That's not true—"

"Look here!" exclaimed Peter at the same moment.

Marc fished in his pocket and held something out to them.

"So?" blustered Peter. "Sure I **gave** them her. I thought they'd look pretty—"

Marc was holding out the Regency earrings. His dark eyes swept Sarella's face with contempt. "Pretty?" he demanded. "Thirty-thousand pounds' worth of prettiness?"

It is not clear what to make of these tokens. Both (2.92) and (2.93) are taken from novels so they are unlikely to be simple speech errors. Example (2.92) in particular stems from *The Magic Toyshop* by Angela Carter, a novel featuring speakers of Irish English. Unfortunately, sentence (2.92) is the only case of *give* with two pronominal objects in the part of the text that is incorporated into the BNC so it cannot be established whether the sentence is part of an attempt to render Irish English. In the case of (2.93) there are no clues at all pointing, for instance, to dialectal speech.

To get a better sense of how the patterns are distributed, I looked at several other corpora, the Frown, Brown, FLOB, and LOB corpora. The FLOB-corpus (Freiburg-LOB corpus) is a 1991 parallel corpus to the 1961 Lancaster-Oslo-Bergen (LOB)

corpus of British English, which is itself a parallel corpus to its contemporary, the American Brown corpus, which in turn has the 1999 Frown corpus¹⁰ as a match. All 4 corpora are about 1 million words in size and all 4 were constructed as balanced corpora, representing a variety of written genres.

In the BROWN corpus, there are 4 tokens combining the Theme *it* or *them* with a personal pronoun as a Recipient in the caused-motion construction. In the double object construction there is no attestation of either *them* or *it* as a Theme combining with a personal pronoun as a Recipient.

In the FROWN corpus, the combination of any of *me, you, him, her, them, us, it* as Recipient and either *it* or *them* as Theme never occurs in the double object construction. The caused-motion construction with *it* as Theme and a personal pronoun as Recipient occurs 9 times.

In the FLOB corpus, the combination of a personal pronoun as Recipient and either *it* or *them* as Theme never occurs in the double object construction. There are 5 tokens combining the Theme *it* with a personal pronoun as a Recipient in the caused-motion construction and one where *them* as a Theme is paired with *me* as a Recipient.

Finally, in the LOB corpus, there is a single instance of the double object construction with either *it* or *them* as Theme and a personal pronoun as Recipient (*I'm going to give you it straight*). There are four instance of the caused-motion construction with the Theme *it* and a personal pronoun as Recipient.

The numbers involved are too small to say anything conclusive. If anything, they confirm the clear predominance of the caused-motion form in the written language.

Personal pronoun Recipient, demonstrative Theme The situation that we observed for combinations of personal pronouns has somewhat of a parallel in cases where the Recipients are personal pronouns and the Themes demonstrative pronouns, except that the most frequent pattern in this case is the double object construction with arguments aligned in standard order.

¹⁰The name indicates that this corpus like the FLOB has Freiburg roots.

- (2.94) He **gave** her that for Christmas, just before he ran away.
- (2.95) Did they **give** you that too?
- (2.96) A friend seeing him off from New York on Sunday **gave** him this to relieve the tension he might face here.

As in the case of two personal pronouns co-occurring, an alternative construction to the dominant type is also used, namely the caused-motion construction, as exemplified in (2.97).¹¹

- (2.97) Then he blew over another spoonful and **gave** that to her.

Due to the luck of the draw no example like (2.97) was actually part of my 2000-sentence sample for *give* from the BNC. This reflects the fact that in the full BNC such sentences occur very rarely. Consider that there are 132945 tokens of potential verbal inflectional forms of *give* in the BNC¹³ Among those, there are 1624 tokens of the sequence lexeme *give* + word form *that* but a large number of the tokens belong to the idiomatic sequence *given that*, which functions as a subordinator, as in (2.98).

- (2.98) **Given** that Mexico is tightly linked with the US, Bergsten devoted his remarks to answering Lay's first point regarding the effects of the US economic slow-down

There are 41 tokens with the sequence *give that* , of which 10 which are continued by *to+* pronoun and belong to the caused-motion construction.

- (2.99) You can **give** that to him .

¹¹What does not seem to occur is use of the double object construction with simple re-ordering of the arguments as in example (i).

- (i) The client should be able to say: "I want circulation status, and if you stupid server can't **give** that me, then you return me an error message"

Example (i) is one of 2 matches of a Google-search for the string "give that {him,me,her}" with the intended reading the pronouns as Recipients.¹² For my purposes here, I will assume that the reverse-order double object pattern involving demonstrative-pronoun Themes and pronominal Recipients is not an established pattern of any variety of English.

¹³This excludes four uses of *gimme* and 11 tokens of *giveth*, which are lemmatized separately.

If we are looking for sequences consisting of a word form of *give* and the word form *this*, we find 604 matches. Expanded by *to* we find only 30 matches and of these only 1 is an instance of the caused-motion construction with a pronominal Recipient:

(2.100) **Give** this to him.

Turning now to uses of the double object construction, we find that the sequence *give* + Pronoun + *that* occurs 518 times; the token of *that* is a demonstrative pronoun in only 88 instances. 36 tokens have a concessive meaning, typically, modalized with *will* (*You don't miss a trick, Roy, I'll give you that*). Another 12 tokens are idiomatic uses of the negated imperative *Don't give me that!*, which serve to indicate the speaker's displeasure with a communicative act (speech or gesture) of their interlocutor. This leaves 40 tokens.¹⁴ The combination of Theme encoded by *that* and pronominal Recipient thus occurs about 4 times more often in the double object construction than in the caused-motion pattern.

There are 241 instances of *give* + Pronoun + *this*, of which 31 are instances of the double object construction, with *this* a demonstrative pronoun. Thus, as in the case of *that*, the double object construction is significantly more frequent than the caused-motion construction.

The large difference in frequency between the double object and the caused-motion construction that obtains in the BNC does for combinations of demonstrative *that* as Theme and a pronominal Recipient, however, not seem to be true of other varieties of English. In one simple experiment, I looked for the strings "giving her that" and "giving that to her" in the Google news group archive. The former pattern resulted in 1820 hits, of which I then inspected the first 300, while the latter resulted in 156 matches, all of which I inspected. Assuming for the larger sample that the first three hundred tokens are representative and that there is no significant amount of repetition among the matches, we would guess that there ought to be ca. 180 hits for *giving her that* where *that* is a simple demonstrative pronoun expressing the Theme argument.

¹⁴Included among them is one token which has *that* post-modified by *there*: *But he gave us that there*.

Weeding out duplicates and errors from the matches for the caused-motion pattern *giving that to her* left 125 sentences.

This was confirmed by repeating the effort for the pair *give her this* and *give this to her*. The double object pattern was reported to have 3628 instances and I sampled the first three hundred, finding 28 valid instances. Projecting from that to the corpus as a whole, we should expect around 340. For the caused-motion pattern, 456 hits were reported, all of which I inspected. There were 78 cases that I could easily rule out as cases where *her* was a possessive pronoun or as duplicates of the same token (usually due to re-postings of news group messages). There may have been more duplicates that I did not notice because they were not ordered next to each other. Even if we err on the side of caution and assume that the real number of inappropriate uses is twice what I counted, we still get around 300 cases of the caused-motion pattern, which is very close to what we may expect for the double object pattern.

Thus both exploratory queries suggests that there is a frequency difference between the constructions in general English, which possibly is more American-influenced, but of a much smaller magnitude than suggested by the BNC alone.¹⁵

2.2.4 Sentences with two personal pronouns

We saw earlier (p. 84) that the double object construction is sometimes used when both the Recipient and the Theme are personal pronouns.

In order to get a sense whether the cases in which simple double object order is

¹⁵I also looked for these patterns in the Brown corpus. There were no tokens of the form *give + Pronoun + that/this*. The caused-motion construction was represented by one instance of *give that to me*. It's not too surprising that there are so few tokens of these patterns as the corpus contains only about 1 million words, compared to 100 million words in the BNC.

I repeated the query in the Frown corpus. Neither the double object nor the caused-motion pattern was attested either with *this* or *that*.

In the LOB corpus, there are no uses of the caused-motion pattern with either *that* or *this* as Theme and a pronominal Recipient. For the DO pattern there are no uses with *that* as Theme and a pronominal Recipient. With *this* as Theme, there was a single match for *him* (*Nip round to Sammy and give him this*).

The FLOB corpus does not contain any instances of *give + Pronoun + that/this* either. The caused-motion pattern is not attested either with *that* or *this*.

used in the BNC with two pronouns are different in any way from the cases where the caused-motion construction occurs, I extracted all active form sentences where a Recipient personal pronoun and the Theme pronoun *it* occur post-verbally and coded the tokens with respect to the number and distance of contextual mentions.

There were a total of 90 cases. 17 cases had to be thrown out because the context was not clear enough to allow the tracing of anaphoric relations.¹⁶ There were also four cases where two pronouns followed each other spuriously as in (2.101). These were also discarded.

(2.101) I don't like the smell of and all that I, I, mean we had erm, we went out for the day and they **give** us it was sandwiches.

Of the remaining 69 tokens, all clearly exhibited the double object order except six tokens which exhibited two ungoverned bare pronouns in the caused-motion order of Theme before Recipient.

(2.102) "It would be interesting to know how they got it," he said. "The information is good."

"So good," said Garvin, "that—"

"I must have **given** it them?" Owen smiled. "I might have," he said, "if I'd thought of it. But even if I did, we're still left with the same question: What are we going to do about it?"

The properties of the contextual mentions for both frame elements are summarized in Tables 2.23 and 2.24. The tables mostly paint a familiar picture. Recipients have more prior and subsequent mentions than Themes. For subsequent mentions it seems that Recipients have their nearest mention closer than Themes, but for prior mentions the Themes may have their nearest mention closer to the target instance, which would

¹⁶That the number of unusable cases is so high reflects the fact that these tokens predominantly come from the spoken part of the BNC. Recall that unlike the Switchboard corpus, the conversations in the BNC were recorded with speakers that knew each other, were co-located and typically in familiar surroundings. There also often were more than two speakers, resulting in more multi-threaded and parallel conversations, which makes the data harder to use.

	Recipients		Themes	
	Distance	Number	Distance	Number
N	63	63	63	63
Mean	4.02	3.00	3.16	1.84
Median	3.00	2.00	2.00	1.00
Mode	0	2	1	1
Std. Deviation	4.006	3.637	3.730	1.648
Skewness	.792	3.437	1.278	1.602
Std. Error of Skewness	.302	.302	.302	.302

Table 2.23: Prior mentions for double object tokens

	Recipients			Themes		
	Distance	Number	Last	Distance	Number	Last
N	63	63	63	63	63	63
Mean	4.43	2.60	4.46	5.71	1.14	1.76
Median	3.00	2.00	4.00	4.00	1.00	.00
Mode	11	1	0	11	0	0
Std. Deviation	4.106	2.981	3.876	4.900	1.446	2.804
Skewness	.672	3.005	.110	.082	1.427	1.558
Std. Error of Skewness	.302	.302	.302	.302	.302	.302

Table 2.24: Subsequent mentions for double object tokens

be a difference to the overall situation for the double object construction, as shown in Table 2.5 on p. 49 above. Finally, Recipients seem to have their last subsequent mention farther on in the text than Themes, indicating greater persistence as topics of conversation.

In order to have a comparison for the double object cases, I also extracted 100 sentence tokens where *it* co-occurs with a Recipient pronoun in the expected caused-motion construction. 12 of these were discarded. Most of these were from the spoken part of the corpus and did not allow tracing of anaphoric relations. Other tokens involved idiomatic uses such as *Let me give it to you straight* or *Now, you Get back out there and give it to her but good*.

	Recipients		Themes	
	Distance	Number	Distance	Number
N	88	88	88	88
Mean	3.05	4.17	1.84	2.05
Median	2.00	3.50	1.00	1.00
Mode	0	2	0	1
Std. Deviation	3.500	3.093	2.626	1.748
Skewness	1.233	.607	2.208	1.673
Std. Error of Skewness	.257	.257	.257	.257

Table 2.25: Prior mentions for caused-motion tokens

	Recipients			Themes		
	Distance	Number	Last	Distance	Number	Last
N	88	88	88	88	88	88
Mean	4.30	3.56	4.92	7.40	1.00	1.31
Median	2.00	2.50	5.00	11.00	.00	.00
Mode	1	0	0	11	0	0
Std. Deviation	4.324	3.613	3.963	4.662	1.688	2.311
Skewness	.725	1.072	-.021	-.576	2.287	2.154
Std. Error of Skewness	.257	.257	.257	.257	.257	.257

Table 2.26: Subsequent mentions for caused-motion tokens

The contextual mention characteristics for these tokens are summarized in Tables 2.25 and 2.26. As in the case of the double object tokens, we find that Recipients have more prior and subsequent mentions than Themes. For subsequent mentions Recipients have their nearest mention closer than Themes, but for prior mentions the Themes seem to have their nearest mention closer to the instance under examination. Finally, Recipients have their last subsequent mention farther on in the text than Themes, indicating greater persistence as a topic of conversation. Overall, the pronominal tokens in the caused-motion construction behave in line with the overall characteristics of that Construction (cf. Tables 2.6 and 2.8).

Next, let us consider the contextual mention properties of sentences in which two

	Recipients		Themes	
	Distance	Number	Distance	Number
N	83	83	83	83
Mean	2.99	3.30	2.10	1.87
Median	1.00	2.00	1.00	1.00
Mode	0	2	0	1
Std. Deviation	3.900	3.067	3.095	1.368
Skewness	1.324	1.506	2.124	1.181
Std. Error of Skewness	.264	.264	.264	.264

Table 2.27: Prior mentions for reverse double object tokens

	Recipients		Themes	
	Distance	Number	Distance	Number
N	83	83	83	83
Mean	5.16	1.92	6.47	.86
Median	3.00	1.00	11.00	.00
Mode	11	0	11	0
Std. Deviation	4.530	2.548	4.920	1.201
Skewness	.371	3.060	-.202	1.714
Std. Error of Skewness	.264	.264	.264	.264

Table 2.28: Subsequent mentions for reverse double object tokens

bare pronouns that are not part of prepositional phrases are aligned in the caused-motion order, which constellation I shall refer to as the reverse double object construction. Tables 2.27 and 2.28 summarize the data.

As usual, we find that Recipients have more prior and subsequent mentions than Themes. As in the other two constructions, we also find that the prior mentions of the Themes seem to be closer than those of the Recipients and that the next subsequent mentions of the Themes are more distant than those of the Recipients.

We may wonder if there is really any clear pragmatic difference between the three constructions. To address this question, we will perform some statistical tests on the contextual mention properties. Let us begin with the measures for distance to

closest prior mention. These are spelled out for all three constructions and both frame elements in Tables 2.29-2.31

Since the data is not normally distributed—for all three constructions, the majority of tokens have distance values from 0 to 3—we need to use a non-parametric test to assess whether the constructions differ. We can use the Mann-Whitney U test again to compare the constructions pairwise. The Mann-Whitney results for the pairwise comparisons turned out as shown in Tables 2.32 and 2.33. The constructions are found to differ with respect to the distance of the closest prior mentions of Themes. For caused-motion and reverse double object tokens, prior mentions are closer than for double object tokens. It also seems that there might be a difference between the double object and the reverse double object constructions with respect to the distance to the closest prior mention for Recipients. However, a Kruskal-Wallis test comparing all three examples at once yields an insignificant result for distance to closest prior mention of Recipients ($p=.123$), while a Kruskal-Wallis test for distance to closest prior mention of Themes confirms the results of the Mann-Whitney test ($p=.047$).

The data for the number of prior mentions for both frame elements is given in greater detail in Tables 2.34 to 2.36. The Mann-Whitney results for the pairwise comparisons between the constructions are shown in Tables 2.37 and 2.38. The constructions are not found to differ when it comes to the number of prior mentions of Themes but they do differ with regard to the number of prior mentions of Recipients. In the caused-motion construction, Recipients have more prior mentions than in the double object or reverse double object constructions.

We now turn to the distance to the closest subsequent mention. The data for the different constructions are summarized in Tables 2.39-2.41. The statistics for this data are given in Tables 2.42 and 2.43. Mann-Whitney U tests indicate that there is no significant difference for Recipients. The comparison between the double object and the caused-motion construction produces a significant result but a Kruskal-Wallis test comparing all three samples at once turns up non-significant ($\chi^2=5.027$, $df=2$, $p=.081$). The Recipients in double object sentences may tend to have on average a shorter distance to the next subsequent mention of the Theme than those in reverse

Distance	Theme		Recipient	
	n	%	n	\$
0	14	22.2	15	23.8
1	16	25.4	6	9.5
2	12	19.0	10	15.9
3	4	6.3	8	12.7
4	2	3.2	3	4.8
5	1	1.6	1	1.6
6	2	3.2	3	4.8
7	1	1.6	4	6.3
8	1	1.6	1	1.6
9	1	1.6	0	0
10	1	1.6	0	0
11	8	12.7	12	19.0
Total	63	100.0	63	100.0

Table 2.29: Distance to prior mention in double object construction

Distance	Theme		Recipient	
	n	%	n	\$
0	26	31.3	25	30.1
1	25	30.1	20	24.1
2	18	21.7	14	16.9
3	1	1.2	3	3.6
4	3	3.6	3	3.6
5	1	1.2	0	0
6	0	0	2	2.4
7	0	0	0	0
8	1	1.2	3	3.6
9	1	1.2	0	0
10	1	1.2	0	0
11	6	7.2	13	15.7
Total	83	100.0	83	100.0

Table 2.30: Distance to prior mention in reverse double object construction

Distance	Theme		Recipient	
	n	%	n	\$
0	34	38.6	26	9.5
1	19	21.6	14	5.9
2	17	19.3	13	4.8
3	3	3.4	6	.8
4	4	4.5	8	.1
5	4	4.5	3	.4
6	2	2.3	4	.5
7	0	0	2	.3
8	1	1.1	2	.3
9	0	0	0	0
10	0	0	1	1.1
11	4	4.5	9	10.2
Total	88	100.0	88	100.0

Table 2.31: Distance to prior mention in caused-motion construction

	Themes		
	RDO-CM	DO-CM	DO-RDO
Mann-Whitney U	3524.500	2170.500	2128.000
Wilcoxon W	7440.500	6086.500	5614.000
Z	-.408	-2.332	-1.975
Asymp. Sig. (2-tailed)	.683	.020	.048
	Recipients		
	RDO-CM	DO-CM	DO-RDO
Mann-Whitney U	3453.500	2385.500	2119.500
Wilcoxon W	6939.500	6301.500	5605.500
Z	-.626	-1.481	-1.992
Asymp. Sig. (2-tailed)	.531	.139	.046

Table 2.32: Test statistics for distance to closest prior mention

	N	Mean Rank	Sum of Ranks
Themes			
Reverse double object	83	87.54	7265.50
Caused-motion	88	84.55	7440.50
Double object	63	85.55	5389.50
Caused-motion	88	69.16	6086.50
Double object	63	81.22	5117.00
Reverse double object	83	67.64	5614.00
Recipients			
Reverse double object	83	83.61	6939.50
Caused-motion	88	88.26	7766.50
Double object	63	82.13	5174.50
Caused-motion	88	71.61	6301.50
Double object	63	81.36	5125.50
Reverse double object	83	67.54	5605.50

Table 2.33: Ranks for distance to closest prior mention

Number of prior mentions	Theme		Recipient	
	n	%	n	%
0	4	4.5	9	10.2
1	48	54.5	10	11.4
2	11	12.5	13	14.8
3	9	10.2	12	13.6
4	6	6.8	7	8.0
5	6	6.8	11	12.5
6	1	1.1	5	5.7
7	2	2.3	6	6.8
8	0	0	5	5.7
9	1	1.1	5	5.7
10	0	0	1	1.1
> 10	0	0	4	4.5
Total	88	100.0	88	100.0

Table 2.34: Number of prior mentions in caused-motion construction

Number of prior mentions	Theme		Recipient	
	n	%	n	%
0	6	7.2	13	15.7
1	39	47.0	12	14.5
2	14	16.9	17	20.5
3	16	19.3	9	10.8
4	2	2.4	11	13.3
5	4	4.8	7	8.4
6	2	2.4	4	4.8
7	0	0	2	2.4
8	0	0	2	2.4
9	0	0	1	1.2
10	0	0	2	2.4
> 10	0	0	3	3.6
Total		100.0		100.0

Table 2.35: Number of prior mentions in reverse double object construction

Number of prior mentions	Theme		Recipient	
	n	%	n	%
0	8	12.7	12	19.0
1	28	44.4	11	17.5
2	13	20.6	14	22.2
3	4	6.3	9	14.3
4	5	7.9	3	4.8
5	2	3.2	3	4.8
6	2	3.2	4	6.3
7	0	0	3	4.8
8	1	1.6	2	3.2
9	0	0	0	0
10	0	0	0	0
> 10	0	0	2	3.2
Total	63	100.0	63	100.0

Table 2.36: Number of prior mentions in double object construction

	N	Mean Rank	Sum of Ranks
Themes			
Double object	63	71.16	4483.00
Reverse double object	83	75.28	6248.00
Double object	63	73.44	4626.50
Caused-motion	88	77.84	6849.50
Reverse double object	83	85.76	7118.00
Caused-motion	88	86.23	7588.00
Recipients			
Reverse double object	83	77.90	6465.50
Caused-motion	88	93.64	8240.50
Double object	63	63.93	4027.50
Caused-motion	88	84.64	7448.50
Double object	63	69.48	4377.50
Reverse double object	83	76.55	6353.50

Table 2.37: Ranks for number of prior mentions

Themes			
	RDO-CM	DO-CM	DO-RDO
Mann-Whitney U	3632.000	2610.500	2467.000
Wilcoxon W	7118.000	4626.500	4483.000
Z	-.067	-.655	-.617
Asymp. Sig. (2-tailed)	.947	.513	.538
Recipients			
	RDO-CM	DO-CM	DO-RDO
Mann-Whitney U	2979.500	2011.500	2361.500
Wilcoxon W	6465.500	4027.500	4377.500
Z	-2.093	-2.892	-1.011
Asymp. Sig. (2-tailed)	.036	.004	.312

Table 2.38: Test statistics for number of prior mentions of Theme

Number of subsequent mentions	Recipient		Theme	
	n	%	n	%
.00	10	15.9	11	17.5
1.00	9	14.3	11	17.5
2.00	10	15.9	6	9.5
3.00	7	11.1	1	1.6
4.00	5	7.9	4	6.3
5.00	3	4.8	2	3.2
6.00	0	0	0	0
7.00	1	1.6	0	0
8.00	3	4.8	0	0
9.00	0	0	0	0
10.00	2	3.2	0	0
11.00	13	20.6	28	44.4

Table 2.39: Distance to closest subsequent mention in the double object construction

double object sentences and those occurring in the caused-motion construction.

The final measure to consider is the number of subsequent mentions. The data for the different constructions are summarized in Tables 2.44-2.46.

Mann-Whitney U tests suggest that there is no significant difference between the constructions in the number of subsequent mentions that Themes have. The tests for Recipients do, however, suggest that there may be a significant difference between the reverse double object construction and the two other constructions: the comparison between the reverse double object and the caused-motion construction is very significant, that between the double object construction and the reverse double object construction is not significant by a small margin. A Kruskal-Wallis test comparing the data for all three constructions at one confirms this ($\chi^2=9.811$, $df=2$, $p=.007$). The Recipients in the reverse double object construction seem to have fewer subsequent mentions than those in the other two constructions.

Number of subsequent mentions	Recipient		Theme	
	n	%	n	%
.00	8	9.6	7	8.4
1.00	20	24.1	21	25.3
2.00	12	14.5	7	8.4
3.00	3	3.6	1	1.2
4.00	4	4.8	1	1.2
5.00	2	2.4	1	1.2
6.00	4	4.8	1	1.2
7.00	1	1.2	0	0
8.00	0	0	0	0
9.00	0	0	0	0
10.00	1	1.2	0	0
11.00	28	33.7	44	53.0

Table 2.40: Distance to closest subsequent mention in the reverse double object construction

Number of subsequent mentions	Recipient		Theme	
	n	%	n	%
.00	13	14.8	4	4.5
1.00	22	25.0	18	20.5
2.00	13	14.8	5	5.7
3.00	7	8.0	4	4.5
4.00	3	3.4	0	0
5.00	3	3.4	2	2.3
6.00	1	1.1	0	0
7.00	0	0	1	1.1
8.00	3	3.4	0	0
9.00	0	0	0	0
10.00	1	1.1	0	0
11.00	22	25.0	54	61.4

Table 2.41: Distance to closest subsequent mention in the caused-motion construction

	Themes		
	RDO-CM	DO-CM	DO-RDO
Mann-Whitney U	3267.500	2237.000	2384.500
Wilcoxon W	6753.500	4253.000	4400.500
Z	-1.329	-2.214	-.976
Asymp. Sig. (2-tailed)	.184	.027	.329
	Recipients		
	RDO-CM	DO-CM	DO-RDO
Mann-Whitney U	3245.000	2642.000	2423.000
Wilcoxon W	7161.000	6558.000	4439.000
Z	-1.287	-.498	-.771
Asymp. Sig. (2-tailed)	.198	.619	.441

Table 2.42: Test statistics for distance to next subsequent mention

	N	Mean Rank	Sum of Ranks
Themes			
Ranks Double object	63	67.51	4253.00
Caused-motion	88	82.08	7223.00
Double object	63	69.85	4400.50
Reverse double object	83	76.27	6330.50
Reverse double object	83	81.37	6753.50
Caused-motion	88	90.37	7952.50
Recipients			
Double object	63	78.06	4918.00
Caused-motion	88	74.52	6558.00
Double object	63	70.46	4439.00
Reverse double object motion	83	75.81	6292.00
Reverse double object	83	90.90	7545.00
Caused-motion	88	81.38	7161.00

Table 2.43: Ranks for distance to closest subsequent mention

Number of subsequent mentions	Recipient		Theme	
	n	%	n	%
.00	13	20.6	28	44.4
1.00	16	25.4	17	27.0
2.00	7	11.1	9	14.3
3.00	10	15.9	3	4.8
4.00	4	6.3	2	3.2
5.00	7	11.1	4	6.3
6.00	4	6.3	0	0
7.00	0	0	0	0
8.00	0	0	0	0
9.00	0	0	0	0
10.00	1	1.6	0	0
>10.00	1	1.6	0	0

Table 2.44: Number of subsequent mentions in the double object construction

Number of subsequent mentions	Recipient		Theme	
	n	%	n	%
.00	28	33.7	43	51.8
1.00	17	20.5	24	28.9
2.00	14	16.9	7	8.4
3.00	8	9.6	5	6.0
4.00	10	12.0	2	2.4
5.00	3	3.6	2	2.4
6.00	1	1.2	0	0
7.00	0	0	0	0
8.00	0	0	0	0
9.00	0	0	0	0
10.00	0	0	0	0
>10.00	2	2.4	0	0

Table 2.45: Number of subsequent mentions in the reverse double object construction

Number of subsequent mentions	Recipient		Theme	
	n	%	n	%
.00	22	25.0	54	61.4
1.00	12	13.6	11	12.5
2.00	10	11.4	8	9.1
3.00	8	9.1	9	10.2
4.00	8	9.1	2	2.3
5.00	3	3.4	2	2.3
6.00	7	8.0	0	0
7.00	6	6.8	1	1.1
8.00	3	3.4	0	0
9.00	2	2.3	1	1.1
10.00	1	1.1	0	0
>10.00	6	6.7	0	0

Table 2.46: Number of subsequent mentions in the reverse double object construction

	Themes		
	RDO-CM	DO-CM	DO-RDO
Mann-Whitney U	3508.500	2417.000	2340.000
Wilcoxon W	7424.500	6333.000	5826.000
Z	-.493	-1.469	-1.169
Asymp. Sig. (2-tailed)	.622	.142	.243
	Recipients		
	RDO-CM	DO-CM	DO-RDO
Mann-Whitney U	2699.500	2434.000	2145.500
Wilcoxon W	6185.500	4450.000	5631.500
Z	-2.996	-1.291	-1.891
Asymp. Sig. (2-tailed)	.003	.197	.059

Table 2.47: Test statistics for number of subsequent mentions

	N	Mean Rank	Sum of Ranks
Themes			
Double object	63	77.86	4905.00
Reverse double object	83	70.19	5826.00
Double object	63	81.63	5143.00
Caused-motion	88	71.97	6333.00
Reverse double object	83	87.73	7281.50
Caused-motion	88	84.37	7424.50
Recipients			
Double object	63	80.94	5099.50
Reverse double object	83	67.85	5631.50
Double object	63	70.63	4450.00
Caused-motion	88	79.84	7026.00
Reverse double object	83	74.52	6185.50
Caused-motion	88	96.82	8520.50

Table 2.48: Ranks for number of subsequent mentions

In sum, the statistical procedures we have carried out suggest the following:

- Themes in the caused-motion construction seem to have closer prior mentions than those in the other two construction.
- Recipients in the double object construction seem to have fewer prior mentions than those in the other two constructions.
- Recipients in the reverse double object construction seem to have fewer subsequent mentions than those in the other two constructions.

That Themes should have closer prior mentions in the caused-motion construction than in the other two constructions would seem to fall in line with the observation that the caused-motion construction is used when the Theme is locally topical, which for Themes only means recently mentioned rather than having a large number of prior and subsequent mentions. Where the recent mention criterion for theme-first alignment is not met as clearly, the likelihood for use of the double object construction may be increased. Recall that the difference between the double object construction and the reverse double object construction was also significant—the Mann Whitney test produced a p-value of 0.048. Thus, the use of reverse double object construction too may be motivated by the fact that it places the Theme before the Recipient, indicating greater local topicality of the Theme.

To help us make sense of the second finding let us consider the distribution of Recipient pronouns, which is shown in Table 2.49.

A χ^2 -test shows that there is a significant difference between the caused-motion and the double object constructions in terms of how many 1st/2nd versus 3rd person pronouns appear as Recipients ($\chi^2=5.676$, $df=1$, $p=0.01720$). If we repeat the test for the pair of double object and reverse double object construction, we get a result that is not significant by a slight margin with $\chi^2=3.313$, $df=1$, $p=0.0687$. A final χ^2 -test performed for the caused-motion and reverse double object construction shows that there is no significant difference with regard to the proportion of 1st/2nd versus 3rd person pronouns ($\chi^2=.364$, $df=1$, $p=0.5463$).

Recipient	Double object		Caused-motion		Reverse double object	
	n	%	n	%	n	%
you	16	25.4	18	20.5	18	21.7
us	8	12.7	2	2.3	5	6.0
me	23	36.5	29	33.0	27	32.5
	47	74.6	49	55.7	50	60.2
him	9	14.3	20	22.7	16	19.3
her	5	7.9	19	21.6	10	12.0
them	2	3.2	0	0	7	8.4
Total	63	100.0	88	100.0	83	100

Table 2.49: Distribution of Recipient pronouns across the two constructions

The referents of the 1st and 2nd person pronoun forms are speech act participants (or groups including speech act participants), they in principle do not need any textual evocation. By contrast, 3rd person pronouns require textual evocation much more strongly than 1st and 2nd person pronouns. The person asymmetry may explain the first of the three observations above. But of course it remains unexplained why the double object construction would be used at all. One explanation may be that it tends to occur as a default: 1st and 2nd person pronouns are relatively more frequent in the double object construction than in the double object construction and the double object construction itself is significantly more frequent than the caused-motion construction. If this explanation is on the right track, then we might expect to find that there is no difference between the 1st/2nd person Recipient referents across the constructions otherwise.

That is, however, not the case. When the above computations are repeated for only the tokens where the Recipient is a first or second person pronoun we still find differences. The third finding above, that Recipients in the reverse double object construction seem to have fewer subsequent mentions than those in the other two constructions, is also true of just the sentences with 1st or 2nd person pronouns Recipients. Pair-wise Mann-whitney U tests show this as does a Kruskal-Wallis test comparing all three constructions ($\chi^2=8.057$, $df=2$, $p=.018$). The first finding is also weakly confirmed. Themes in the caused-motion construction seem to have closer

prior mentions than those in the other two constructions, although the difference between the caused-motion and the reverse double object construction is no longer significant and the three-way comparison between all constructions in a Kruskal-Wallis test also has become non-significant by a slim margin ($\chi^2=5.779$, $df=2$, $p=.056$). The evidence for the second finding is weaker. Recipients in the double object construction were found again to have fewer prior mentions than those in the caused-motion construction but the difference between the double object and the reverse double object construction appear clearly insignificant. The three-way comparison also fails to register a significant difference ($\chi^2 = 3.833$, $df = 2$, $p = .147$). Overall, the comparison of only those tokens that have 1st or 2nd person pronouns as Recipients suggests that the differences between the constructions are not just due to differences in the relative share of 1st and 2nd person pronouns among the Recipients.

Finding an explanation for why recipients in the reverse double object construction should have fewer subsequent mentions than those in the other two constructions is tricky. Certainly, when we consider the constructions overall with all NP types, not just pronouns, it is the case that the Recipients in the double object construction have more subsequent mentions than those in the caused-motion construction. To the degree that the reverse double object construction is similar to the caused-motion construction, we would expect it to behave similarly to it where number of subsequent mentions is concerned: we might expect fewer subsequent mentions of Recipients. This explanation, however, assumes that speakers allowing for more than one construction with two pronominally coded referents make the distinction in terms of the criteria that apply to non-pronominally coded referents. Recall that when both Theme and Recipient are pronouns, Recipients in the caused-motion construction seem to have slightly more subsequent mentions than those in the other two constructions. But of course this is very likely just due to the two facts that most speakers represented in the British National corpus allow only the caused-motion construction when both Theme and Recipient are pronominal and that most Recipients are very topical.

A fuller explanation of the difference between the constructions would need to use data coming only from speakers that distinguish between more than one of these

constructions. The initial exploration undertaken here does however suggest that there might be genuine pragmatic differences between the constructions in addition to any, for instance, sociolinguistic considerations that may influence the use of the non-canonical double object and reverse double object constructions.

2.2.5 Sentences with a personal pronoun Recipient and a demonstrative Theme

We noticed earlier that in the case of a pronominal Recipient and a demonstrative Theme, the relative frequency of the two constructions is the opposite from that for the combination of two personal pronouns. The frequency difference can be gleaned from Tables 2.50 and 2.51, which further show the combinations of Recipient pronouns and singular demonstrative Theme that were found for the double object and caused-motion constructions. It is noticeable that the 1st and 2nd person pronouns account for a large part of the tokens. A large number of the tokens come from the spoken part of the corpus.

	that	this	Total
her	10	2	12
him	12	5	17
it	2	0	2
me	41	21	62
them	8	0	8
us	10	3	13
you	37	23	60
Total	120	54	174

Table 2.50: Recipients occurring in double object construction with singular demonstrative Theme

	that	this	Total
him	2	1	3
me	4	0	4
you	4	0	4
Total	10	1	11

Table 2.51: Recipients occurring in caused-motion construction with singular demonstrative Theme

We now turn to the contextual mention properties of the arguments in sentences with a demonstrative Theme and a pronominal Recipient in the double object con-

struction. Table 2.52 shows the number and distance of prior mentions for Recipients and Themes. With either demonstrative as Theme, we find that Recipients have significantly more prior mentions than Themes and that these mentions tend to be closer to the current instance. This difference goes in the same direction as that observed for Recipients and Themes in general within the double object construction (cf. Table 2.5, p. 49).

Table 2.53 shows the number and distance of subsequent mentions for Recipients and Themes in the double object construction. We find that with either demonstrative as, Theme Recipients have more and closer subsequent mentions than Themes. This is also the pattern found for Recipients and Themes in general in the double object construction (cf. Table 2.6, p. 50).

For the purpose of looking at the properties of contextual mentions of the arguments in sentences exhibiting the caused-motion construction with a demonstrative Theme and a pronominal Recipient, I will simply pool the one sentence that has *this* as Theme with the 10 sentences that have *that* as Theme. Table 2.54 shows the summary for the number and distance of prior mentions.

Recipients have more prior mentions than Themes and these mentions may also be a bit closer but compared to the Recipients in the double object construction the Recipients in the caused-motion construction appear less topical: their prior mentions are farther away and fewer (cf. 2.52). A similar picture emerges when we look at subsequent mentions. Recipients do have more and closer subsequent mentions than Themes but they fare less well when compared to their counterparts in the double object construction (cf. 2.53): they are again fewer in number and farther away. The frame elements in caused-motion sentences with a demonstrative Theme and a pronominal Recipient thus behave the same way with respect to prior and subsequent mentions as frame elements in the construction in general (cf. Tables 2.7 and 2.8 on p. 50).

Because the number of tokens that exhibit the caused-motion pattern is very small in the BNC we cannot come to firm conclusions about why combinations of demonstrative Theme and pronominal Recipient are aligned the way they are. The occurring tokens suggest that this combination behaves like Recipients and Themes

	That			
	Number		Distance	
	Recipient	Theme	Recipient	Theme
N	120	120	120	120
Mean	2.4833	.7833	4.3750	5.8750
Median	2.0000	1.0000	2.0000	4.5000
Mode	.00	.00	1.00	11.00
Std. Deviation	2.61856	.93650	4.17890	4.85688
Skewness	1.454	1.509	.701	.034
Std. Error of Skewness	.221	.221	.221	.221
	This			
	Number		Distance	
	Recipient	Theme	Recipient	Theme
N	54	54	54	54
Mean	3.0000	.5185	4.4630	9.2407
Median	2.0000	.0000	3.0000	11.0000
Mode	.00	.00	11.00	11.00
Std. Deviation	2.95245	1.22460	4.15598	3.51780
Skewness	1.032	2.676	.626	-1.620
Std. Error of Skewness	.325	.325	.325	.325

Table 2.52: Prior mentions for the double object construction

	That			
	Number		Distance	
	Recipient	Theme	Recipient	Theme
N	120	120	120	120
Mean	2.2083	.8667	4.9750	7.3167
Median	2.0000	.0000	3.0000	11.0000
Mode	.00	.00	11.00	11.00
Std. Deviation	2.42153	1.51704	4.51415	4.79668
Skewness	1.402	2.212	.402	-.580
Std. Error of Skewness	.221	.221	.221	.221
	This			
	Number		Distance	
	Recipient	Theme	Recipient	Theme
N	54	54	54	54
Mean	3.1667	.8148	5.0000	7.1296
Median	2.0000	.0000	3.5000	11.0000
Mode	.00	.00	11.00	11.00
Std. Deviation	3.28375	1.34689	4.49318	4.78222
Skewness	1.025	2.131	.324	-.469
Std. Error of Skewness	.325	.325	.325	.325

Table 2.53: Subsequent mentions for the double object construction

	This & That			
	Number		Distance	
	Recipient	Theme	Recipient	Theme
N	11	11	11	11
Mean	1.55	.64	4.91	5.00
Median	1.00	1.00	4.00	3.00
Mode	1	1	0	11
Std. Deviation	1.293	.505	4.369	4.899
Skewness	.726	-.661	.303	.468
Std. Error of Skewness	.661	.661	.661	.661

Table 2.54: Prior mentions for the caused-motion construction

	This & That			
	Number		Distance	
	Recipient	Theme	Recipient	Theme
N	11	11	11	11
Mean	2.18	.18	5.36	9.18
Median	2.00	.00	5.00	11.00
Mode	0	0	11	11
Std. Deviation	2.089	.405	4.296	4.045
Skewness	.672	1.923	.268	-1.923
Std. Error of Skewness	.661	.661	.661	.661

Table 2.55: Subsequent mentions for the caused-motion construction

overall. The great frequency difference between the constructions—174 instances of the double object construction, 11 of the caused-motion construction—and the fact that it does not seem to exist to the same degree in English in general, as suggested by queries with the Google search engine, need not necessarily be taken as a sign that the double object construction has the status of a default in the BNC.

One reason is that demonstrative *that* strongly tends to be used when its referent is not linguistically given but only contextually available. When the context is physical, this results in core deictic uses of demonstratives, as in *Hey, gimme that!*. Note that although the speech act participants are also situationally given, they are presumably more accessible to each other than entities in the physical context. The cases where *that* and *this* pick up on linguistic entities are similar:

(2.103) We're definitely not vacationing in Greece, {*that*/**it*'s} for sure.

In (2.103), *that* couldn't readily be replaced by *it*; the proposition *We're definitely not vacationing in Greece* does not seem to have a discourse referent that could be referred to pronominally simply as a result of its utterance.

Also note that *that* tends to be the first or second reference to an entity rather than occur later in a referential chain. At least it seems that the same speaker cannot easily refer repeatedly to an entity using *that*. Consider (2.104).

- (2.104) All I cared about is that *we made finals*. And *that* was a surprise to me, as *it/*that* was to a lot of others out there.

Further evidence that *it* and *that* are in near-complementary distribution comes from the fact that the former cannot be topicalized or focused unlike the latter:

- (2.105) Topicalization

- a. Now that I like!
- b. *Now it I like!

- (2.106) Focusing

- a. A: What's making that noise? B: (Pointing at object) That!
- b. A: What's making that noise? B: (Pointing at object) *It!

The differences in referent accessibility between *that* and *it* thus seem to account well for the high frequency of the double object pattern. But in that case, we should be looking more closely at the tokens in the caused-motion construction as they are unexpected. Tables 2.54 and 2.55, compared to Tables 2.52 and 2.53, suggest that the Recipients are less topical in the caused-motion patterns with fewer and more distant contextual mentions on either side. Three representative examples of the caused-motion pattern are given in (2.107-2.109).

- (2.107) "We must help him," said Endill.

"I know," said Matron. "Already he is looking weak. He has a cough too but the school has no medicine. All I get supplied are empty bottles and hundreds of boxes of bandages."

"I have some cough mixture from home," replied Endill. "You can **give** that to him."

- (2.108) "See that scar?" Gallagher cried, pointing down. "I **gave** that to him with a sleán !"

(2.109) “But what if he’s afraid? What if he doesn’t want to go back?”

Karr nodded, then reached into his tunic pocket yet again. “**Give** this to him. He’ll understand.”

In the first two cases, there is an antecedent for the Theme referent in the preceding utterance of the quoted dialog. In both cases, the speaker has just drawn the hearer’s attention to the referent. In the last case, though there is no linguistic antecedent, we might imagine that both the protagonist and the character Karr are closely attending to what Karr is getting out of his tunic pocket, causing this to be very highly accessible. This is not as clear for the three other cases without overt linguistic antecedent: the transcripts of the spoken conversations are not rich enough to let the analyst know what all the interlocutors are attending to.

However, as mentioned in section 2.2.3 (cf. p. 87), there are many additional instances of the caused-motion patterns available on the internet and these have the advantage that the communication happens between distant interlocutors who cannot assume sharing a physical environment. To test the hypothesis that the caused-motion pattern is used when the Theme is highly accessible, in particular by there being a prior mention in the same or the preceding sentence, I searched for the string *gave that to her* on Google and inspected 40 of the first 200 matches. (There were reported to be a total of 1670 matches.) The approximation of random sampling in this process consisted of checking only every 5th match: 5, 10, 15, etc. 34 of the cases had a previous mention of the Theme in the prior sentence. A typical example is (2.110).

(2.110) He ate more himself. Then he blew over another spoonful and **gave** that to her. Harold poured milk into a glass and she leaned forward over the table and drank .

In addition to the clear cases, there were another three cases where the demonstrative picks up on a saliently inferable referent. In (2.111), the feeling of safety is presumably what Max ‘gives’ the referent of her; in (2.112), the speaker is an author who tells the audience that he gives the character Rei the quality of being emotive; in (2.113), what is given (as a gift) is the speaker’s effort to learn how to sew.

- (2.111) He thought she was afraid to tell him what she wanted! She had a need to feel safe and Max never **gave** that to her! She swore she would never come to him for anything, ever again!
- (2.112) It was also here that I felt that Rei needed to become a bit more “normal”. Emotive. So I **gave** that to her, too. And it just kept going from there.
- (2.113) today of all days, i celebrate her by learning to sew. its something she wanted and today i **gave** that to her. ...

One of the three remaining examples is a comment under a cartoon entitled “Conan the librarian”:

- (2.114) Xena’s Mom is a librarian, and I **gave** that to her the day I saw it. I love that one!

It is clear from the context that in (2.114) the speaker is referring to the cartoon. Though not verbally, the criterion of high salience is fulfilled.

The two remaining examples involve cases where the antecedent of *that* is a little farther away.

- (2.115) my vet gave me uraprim(or however you spell it) i gave the last packet today.. I hope she quits soon. she hasnt coughed a lot after I **gave** that to her but if shes running in the pasture or something she will cough. one of my friends ponies has heaves really bad. they have to give him antihistamines. but shes going to get some of that cough free stuff and I think I will. Do you give it to them all the time??
- (2.116) Q. Who brought that to her?
 A. I don’t know.
 Q. Who **gave** that to her, if you know?
 A. I don’t know. I think I might have given her some.

In (2.116), a court transcript, the referent is clearly prominent and it is explicitly evoked only by the questioner. In (2.115), the antecedent is two sentences away and the Recipient is mentioned between the antecedent and the current mention of the Theme.

Regardless of whether we count any of the 6 cases where there is no overt antecedent in the previous sentence/utterance, the data seems to suggest that the use of the caused-motion construction with demonstrative is specifically motivated by a closeby prior antecedent of the Theme. Very often, the Recipient is not mentioned in the preceding sentence. (2.110) is a good example and many more like it were part of the matches found on the internet:

(2.117) Actually, I wanted to get Banana Republic's W perfume because my brother gave that to her once a long time ago and I know my mom likes it a lot.

When both Recipient and Theme occur in the same preceding clause and the caused-motion construction is chosen, then it is often the case that the Theme occupies a more prominent grammatical position in the prior clause. For instance, in (2.118), the antecedent of the Recipient pronoun is the possessive of the subject NP, *her first husband, my dad's father*.

(2.118) "Adam, how did you afford this ring?" "It was my grandmother's. Her first husband, my dad's father, gave that to her when they became engaged."

There are, however, cases like (2.119) and (2.120), where both Recipient and Theme have antecedents in the prior clause and the Recipient has the greater grammatical prominence but still the caused-motion construction is chosen.

(2.119) Jane also asked me for the phone # of her college roommate! So I gave that to her. It felt really cool .

(2.120) It was time for my wife to get an updated car, and she liked my Ford Escape so I gave that to her and we bought me the new car.

I would predict that in cases such as (2.119) and (2.120) speakers may find both versions acceptable as there is no question of strict grammaticality but just one of local topic coherence. However, it is conceivable that (certain) American speakers treat the demonstrative pronouns just like the personal pronouns and therefore prefer the caused-motion construction as a default. This possibility suggested itself during consultations with a few American speakers, in which I showed them two versions of sentences like (2.119) and (2.120), one original, one modified, in the original context.

Finally, note that although for instance (2.110) may suggest the relevance of contrast between Recipients, this is clearly not a requirement and there is no potential contrast present in the vast majority of cases. Example (2.119), for instance, informs the hearer simply that the inferable transfer of information did indeed happen; there is no question as to the identity of the (type of) information, the information seeker and the information provider.

We can tentatively conclude that what is relevant for the use of the caused-motion construction when the Theme is a demonstrative and the Recipient is a pronoun is the high accessibility of the Theme in the immediate prior context, with concurrent lower local prominence of the Recipient, i.e., the Recipient should not occur in the prior sentence or occupy a grammatical position of lower prominence. When this is not the case, the double object construction tends to get used: in the BNC corpus, among the *give* + Pronoun + demonstrative tokens, there were 37 which had a prior mention of the Theme in the preceding clause, but two thirds of them also had a mention of the Recipient in that same preceding clause.

The fact that the two constructions seem to be distributed in such different frequencies in the corpus as compared to the internet seems in part to be explained by differences between the two bodies of text. The vast majority of tokens found on the internet do not involve either actual transcription of spoken conversation or fictional rendering of such conversation. Rather, the texts are monologic and no shared context with the hearer is assumed; the speaker explicitly introduces most of the referents. By contrast, in the BNC, for 96 of the 174 sentences with a demonstrative Theme and a pronominal Recipient, there was no prior mention of the Theme, while only 40 sentence lacked a prior mention of the Recipient. At least in the case

of Themes, this lack of prior linguistic evocation is possible only when physical context is shared. Another possible contributing factor seems to be a difference between the varieties with regard to the importance of the Recipient. British English speakers seem to use the caused-motion construction in a situation with a demonstrative Theme and a pronominal Recipient only when both the Theme is locally prominent in the prior linguistic context and the Recipient is not prominent. For American speakers the local prominence of the Theme may by itself be reason enough to use the caused-motion form. This difference leads British speakers to use the double object construction when both arguments have antecedent mentions in the same or the immediately prior sentence, whereas American speakers may favor the caused-motion construction. Another factor that cannot be fully distinguished from the foregoing one may be a difference in the status of demonstrative pronouns. To the degree that American speakers grammatically treat them just like personal pronouns, they will prefer to use the caused-motion construction.

2.2.6 An unattested pattern: Theme *anything* plus Recipient *anybody*

A look back at Tables 2.3 and 2.4 shows that negatively determined NPs (e.g., *no man*, *(not) any books*) and negative pronouns (*(not) anybody*, *nobody*) are very rare as fillers of the frame element roles Recipient and Theme. We may wonder whether the distribution of these NP types which don't refer is influenced by the same kinds of factors that we found to be relevant for more typically referential NPs. To get a sense of this, I extracted 20 instances each of the alternate word form sequences *give anything to anybody* and *give anybody anything* from the Internet via the Google search engine. (The BNC corpus contained fewer than 3 instances of the two patterns).

Since *anybody* and *anything* do not normally refer to an identifiable and trackable entity in discourse, the data reported here on the contextual mentions is relative to the inference triggers. Negative *anything* and *anybody* are very often related by inference to sets of similar referents that are contextually mentioned. In example

(2.121), for instance, *anything* is related to the amounts of money that are mentioned before while *anybody* here clearly is not a free-choice use of the word but is used against the backdrop of the set of known 2004 Democratic presidential candidates.

(2.121) West Wing creator Aaron Sorkin gave \$2,000 each to Edwards, Clark and Gephardt, and \$1,000 to Dean. West Wing president Martin Sheen didn't **give** anything to anybody.

There are, however, also cases in which there is no specific contextually available set of persons or things mentioned. But even in those cases, uses of *anybody* and *anything* often seem to pair up with parallel uses in close-by clauses or sentences. An example of this kind is (2.122).

(2.122) You don't have to give anybody anything to come into this county, and you don't have to **give** anybody anything if they threaten to leave.

While examples (2.121) and (2.122) may suggest that rhetorical parallelism is all that is involved, there are many other cases where such parallelism does not exist and where instead discourse-pragmatic considerations seem to come into play again. In (2.123), for instance, *they* is clearly linked to *anybody*. Arguably this is a conditional construction which introduces a referent in a hypothetical mental space (or world) and then makes a predication about that referent. The link is thus one of co-reference. Note that *charity* too may be linked inferentially to the notion of giving something for free. However, even if one treated both *they* and *charity* as being inferentially linked to *anybody* and *anything*, it is clearly the case that the person-evoking NPs occupy the more grammatically prominent positions associated with higher topicality: *anybody* occurs as a first object in the first clause and as subject in the second clause while *anything* and *charity* occur as second object and direct object, respectively.

(2.123) You **give** anybody anything for free and they get lazy and pretty soon start expecting charity.

If we now consider the characteristics of the sample as a whole, then we find further evidence suggesting that pragmatic influences such as the closeness and number of

	Recipient				
	Dist. prior	Num. prior	Dist. sub.	Num. sub.	Last sub.
N	20	20	20	20	20
Mean	3.85	1.25	5.80	1.15	.50
Median	2.00	1.00	6.50	.50	-.50
Mode	1	1	11	0	-1
Std. Deviation	4.356	1.517	5.376	1.663	2.606
Skewness	1.059	3.044	-.047	1.720	2.321
Std. Error of Skewness	.512	.512	.512	.512	.512
	Theme				
N	20	20	20	20	20
Mean	5.40	.65	8.35	.25	-.65
Median	3.50	1.00	11.00	.00	-1.00
Mode	11	1	11	0	-1
Std. Deviation	4.795	.587	4.727	.444	.745
Skewness	.314	.212	-1.277	1.251	2.695
Std. Error of Skewness	.512	.512	.512	.512	.512

Table 2.56: Properties of link referents in the double object construction (co-referential mentions or inference triggers)

contextual mentions of inference triggers (including co-referential mentions such as in (2.123)) seem to line up in the same way as with the majority of referential NPs. Table 2.56 shows that in the double object construction, *anybody* seems to be linked to the surrounding discourse by expressions that are closer and which have more mentions within the window of context.

For the caused-motion construction, shown in Table 2.57, we find that the link expressions for Recipients now have slightly poorer measures compared to their counterparts in the double object construction. The measures for expressions linking the Theme *anybody* to the surrounding discourse are better than those of their counterparts in the caused-motion construction. As seen in various places above, in the caused-motion construction the measures for the Theme referents' contextual mentions are

	Recipient				
	Dist. prior	Num. prior	Dist. sub.	Num. sub.	Last sub.
N	20	20	20	20	20
Mean	4.75	1.45	6.00	1.15	.50
Median	2.00	1.00	7.00	.50	-.50
Mode	1	1	11	0	-1
Std. Deviation	4.789	1.877	5.181	2.254	2.439
Skewness	.574	1.651	-.061	3.499	2.492
Std. Error of Skewness	.512	.512	.512	.512	.512
	Theme				
N	20	20	20	20	20
Mean	4.10	.95	3.70	.80	.20
Median	1.00	1.00	1.00	1.00	.00
Mode	1	1	0	1	0
Std. Deviation	4.701	.887	4.932	.616	1.056
Skewness	.859	1.110	.905	.120	.453
Std. Error of Skewness	.512	.512	.512	.512	.512

Table 2.57: Properties of link referents in the caused-motion construction (co-referential mentions or inference triggers)

not necessarily always better in absolute terms than those of the Recipients' contextual mentions. In the case of the *anybody+anything* combination with *give*, Themes' link expressions fare better only with respect to the Distance between the current mention of the Theme and the link to prior discourse ($4.10 < 4.75$) and the Distance between the current mention of the Theme and its link to subsequent discourse ($3.70 < 6.00$).

The data presented in this sub-section suggest that the general pragmatic notions that were found to be relevant for characterizing the distinction of the bulk of the data also are relevant to rarer patterns that at first glance may seem to be outside their purview. Another relatively rare pattern that merits further investigation in the way this was done here for negative pronouns consists of cases in which the Recipient

is coded as an indefinite NP. Based on the data discussed here for *anybody+anything*, we should expect that in those cases, too, indefinite Recipients in the double object construction are inferentially linked to surrounding discourse more strongly than indefinite Recipients in the caused-motion construction.¹⁷ In conclusion, it is also worth pointing out that the findings of this section are in line with the observations about other reference phenomena and constructions, for instance by Birner 1997, that inferables and inference triggers often behave as if they were co-referential mentions.

2.2.7 Regression analysis

So far we have considered how predictive various factors are taken by themselves and combined in pairs, in the latter case focusing on weight and definiteness. We have also inspected all the cases that could not be accounted for in terms of weight and definiteness in order to see how the analysis needs to be enriched to give a better account of the occurring tokens. The prior evocation status, in particular, emerged as a factor that ought to be taken into account explicitly. We have also seen that it is important to know if an argument has been relativized on. For instance, treating the Theme as a variable with no definite referent of length zero in cases where it is relativized on results in conflicting predictions of weight and definiteness. But if the model was aware of relativization, the correct form could be chosen.

Proceeding with an analysis based on tabulation becomes impractical at this point where we want to take additional factors into account. We therefore turn to logistic regression as an analytic tool. Regression analysis attempts to fit a predictive model to a set of data and to use that model to predict values of a dependent variable from one or more independent variables. If there is only one independent variable or predictor, we speak of simple linear regression, if there are more we speak of multiple linear regression. Logistic regression is like multiple regression but with an outcome variable that is a categorical dichotomy, and predictor variables that are either continuous or categorical. What is being predicted in logistic regression is not the value of a quantitative variable but category membership.

¹⁷This additional work could not be carried out as part of this thesis for lack of time.

In order to generate binary outcomes, logistic regression uses a filter that modifies the predicted value for the dependent variable, the logit transformation. The logit transformation Y of a probability p of an event is defined as the logarithm of the ratio between the probability that the event occurs and the probability that the event does not occur, $Y = \log(p/(1-p))$.

There are different ways to run a regression analysis depending on what the purpose of the analysis is. The forced entry method of including all the predictors into the regression model at the same time is recommended when an existing hypothesis is being tested. There are also two other, so-called step-wise methods of entering predictors into a regression model. In the forward method, the analysis begins with only a constant, which represents the relative frequency of the more frequent outcome, and then adds single predictors into the model based on a specific criterion, the value of the so-called score statistic. Analysis proceeds until no predictor variable with a significant score statistic remains. In the backward method, analysis begins with all predictors included. Subsequently, predictors are removed if their removal does not result in a substantial reduction of the goodness of fit between model and data. The step-wise methods are recommended for exploratory work.

We will begin by testing the pre-existing model based on weight and definiteness rather than trying to explore all of the factors. Consequently, we will use the forced entry method of including all the predictors into the regression model at the same time.

The constant for the model, that is the relative frequency of the more frequent construction, is $163/208 = 0.784$. That is, without knowing anything about the factors discussed above, guessing that the form chosen is the double object construction results in 78.4% right guesses. Predictors can only be added to the model if they improve on that baseline score.

The two predictors that are used in the model are relative weight and relative definiteness. Both have categorical values in my coding: relative weight has three levels (Recipient is shorter, equal, or longer than Theme), and relative definiteness has eight levels for the different attested combinations of Pronoun, Definite lexical, and Indefinite lexical. For carrying out the regression analysis, the different levels

	Frequency	(1)	(2)
Theme shorter	42	1.000	.000
Theme equal	47	.000	1.000
Theme longer	119	.000	.000

Table 2.58: Parameter coding

of the two variables are encoded separately as so-called dummy variables. Thus, there are in the model a binary variable that records whether a case did or did not involve a pronominal Recipient and an indefinite Theme; a variable that records whether a sentence has a pronominal Recipient and a lexical definite NP as Theme; etc. This coding is useful as it allows us to check specifically how influential specific combinations of definiteness values of Recipients and Themes are. (We would, for instance, expect the case where both arguments are pronouns to be a good predictor as many of these tokens will simply be pairings of two personal pronouns appearing in the caused-motion construction.)

We will begin by doing a minimal regression analysis which includes only one of the variables. Let us first consider the variable relative length. The regression starts with an initial -2 Log likelihood of 217.254. This is a measure of how much unexplained information there is in the model.

After the single step of regression we end with a value of 163.770, a significant reduction. We also get a p-value of 1.000 for the Hosmer and Lemeshow Goodness of Fit test. The test tests the hypothesis that the observed data are significantly different from the predicted values of the model. In other words, the high p-value that we found indicates that this is not the case and that in fact the model does match the observed data well.

The Classification table of the model at the end of our only model step, Step 1, is given in Table 2.59. The model with the given factor relative length handles roughly 3% more of the cases than simple guessing the more frequent double object construction would get right.

As shown in Table 2.60, the two dummy variables that represent the three length

Observed			Predicted		
			Construction		Percentage Correct
			do	cm	
Step 1	Construction	Double object	145	18	89.0
		Caused-motion	21	24	53.3
	Overall Percentage				81.3

Table 2.59: Classification table for model with relative length as only factor

								95.0% C.I. for EXP (B)	
		B	S.E.	Wald	df	Sig.	Exp (B)	Lower	Upper
Step 1	LGTHDIFF			38.161	2	.000			
	LGTHDIFF(1)	3.223	.522	38.094	1	.000	25.111	9.023	69.888
	LGTHDIFF(2)	2.178	.523	17.348	1	.000	8.828	3.168	24.602
	Constant	-	.419	49.100	1	.000	.053		
		2.936							

Table 2.60: Variables in the equation after the single regression step for the parameter relative length

difference levels in the regression are all reported to be significant in predicting the chosen construction, $p=.000$ in each case.

Table 2.60 also shows the lower and upper bounds for $\exp \beta$. $\exp \beta$ is an indicator of the change in odds resulting from a unit of change in the predictor variable. A value greater 1 for $\exp \beta$ indicates that as the predictor increases the odds of the outcome occurring—in our case, SPSS picks the caused-motion construction occurring as the event to be predicted is—increase, conversely a value less than 1 indicates that as the predictor increases, the odds of the outcome occurring decrease. It is important that for both dummy variables the upper and lower bounds of the confidence interval for $\exp \beta$ are on the same side of 1. This means that we can be fairly confident that the

Observed			Predicted		
			Construction		Percentage Correct
			do	cm	
Step 1	Construction	Double object	152	11	93.3
		Caused-motion	23	22	48.9
	Overall	Per-			83.7
	centage				

Table 2.61: Classification table for model with relative definiteness as only factor

population value of $\exp \beta$ does indeed lie on the side of > 1 .

Inspection of the standardized residuals indicates 6 cases (2.9%) where the model fits badly: they all have a value of $z=4.340$. In an average normally distributed sample less than 1% of the values should have a z-score more extreme than ± 2.5 . The 6 cases are all ones in which the prediction of length is violated and a shorter recipient occurs in the caused-motion construction. Since there is no good reason to exclude these tokens—they are not ungrammatical in any sense—we need to acknowledge that the model as it is fits these cases poorly.

We will now repeat this simple regression analysis with the variable relative definiteness as predictor. The event to be predicted is the occurrence of the caused-motion construction. The model begins as before with a -2 Log likelihood value of 217.254. After the 1-step regression with eight dummy variables encoding the 9 observed relative definiteness combinations, we get a value of 131.40, which is lower than what we got for the regression with only the predictor variable relative length. The Hosmer and Lemeshow Goodness of Fit test comes out as $p=1.000$, which indicates that our model does fit the data. The resulting classification table is given in 2.61.

The model is better than that based on relative length alone; it correctly classified 83.7% of the cases. Of the 8 dummy variables that are used, only 2 come up as significant. They are the variable that uniquely encodes the combination of pronominal Theme and definite Recipient ($p=.026$) and the the variable that encodes the combination of indefinite theme and pronominal Recipient ($p=.000$). These two also

have $\exp \beta$ values that do not cross the value 1: the variable for pronominal theme and definite recipient has bounds of 1.360 and 124.297 and the variable for indefinite theme and pronominal recipient has bounds 0.002 and 0.183. These boundary values indicate that in the first group of cases the odds of the predicted event occurring—namely use of the caused-motion construction—are increased, while those for the second group of cases indicate that the odds of the predicted event occurring are decreased, all of which is in line with our previous discussions. The other dummy variables, and thus the other relative definiteness combinations, do not contribute significantly to the predictiveness of the model. Note that they also have $\exp \beta$ values which cross 1, indicating that it is not clear whether they increase or decrease the odds of the event occurring. Inspection of the standardized residuals indicates that with relative definiteness as a predictor, we have only 2 cases (.96%) that have a z-score more extreme than 2.5, which is acceptable. Each of these cases violates one of our two significant dummy variables. One case is a passive in the double object construction with a definite Recipient subject and a pronoun Theme (cf. (2.51)), the other case involves a passive of the caused-motion construction where the Theme has been relativized on and therefore counts as indefinite (cf. (2.39)).

We now perform a regression analysis combining the two factors relative length and relative definiteness. After the single step of regression we end with a -2 Log likelihood value of 117.859, a significant improvement over the initial value of 217.254. We also get a p-value of .764 for the Hosmer and Lemeshow Goodness of Fit test, indicating that our model approximates the data well.

The Classification table of the model at the end of our only model step, Step 1, is given in Table 2.62. The model with the two given factors handles roughly 8% more of the cases than simple guessing would get right.

Inspection of the standardized residuals indicates that with our two predictors, we have 4 cases (1.9%) that have a z-score more extreme than 2.5. Ideally, one would want to have 1% or less such cases. Two of the cases we saw before in the regression analysis for relative definiteness, namely (2.51) and (2.39). Another case is (4.102), which is actually kind of idiomatic but was coded as having a longer and less definite Theme in the caused-motion construction. Finally, there is a passive caused-motion

Observed			Predicted		
			Construction		Percentage Correct
			do	cm	
Step 1	Construction	Double object	151	12	92.6
		Caused-motion	16	29	64.4
	Overall	Per-			86.5
	centage				

Table 2.62: Classification table for model with two factors relative length and relative definiteness

construction sentence with a longer and less definite Theme as subject (cf. (2.42)).

A look at the variables in the equation shows the composite representations of relative definiteness and relative length as significant, with $p < .05$ and $p < .01$ respectively. Of the individual levels of the relative definiteness variable, only the one encoding the combination of indefinite Theme and pronominal Recipient is reported as significant. The upper and lower bounds for its $\exp \beta$ are also both less than 1 so we can be sure that membership in the category is associated with decreased odds of appearing in the caused-motion construction.

Since we have two variables in the model we also need to look into the possibility that they are highly correlated and therefore should not be used together. An inspection of several collinearity diagnostics indicates that multi-collinearity does not seem to be a problem. Both the tolerance and the VIF values are 1.000, neither of which is in the region raising concern ($< .1$ for tolerance, >10 for VIF). The condition index values range from 1.000 to 8.319. This is still be acceptable, if go with the cutoff value of 15 suggested by the SPSS package. There also is no component associated with a high condition index which would contribute substantially to the variance of two or more variables.

Note that this last regression analysis did not include a variable for the interaction of relative length and relative definiteness. If we do this and re-run the analysis, none

of the individual variables is reported as significant.¹⁸ Nonetheless, running the same regression with a step-wise method and removal criterion will result in all variables being kept. As has been pointed out to me, it is a limitation of logistic regression as statistical procedure that it may not deal well with interactions as these introduce a multiplicative component into the regression equation. What this means for my investigations here is that I will not try to determine exactly how different variables overlap.

In the regression analyses so far, we used a three level categorization for relative length which included only cases where both Theme and Recipient were exactly equally long in the category for equally long, while Wasow 2002 included cases of one word differences in the about equally long class. To see if using Wasow's measure would produce better results, I recoded the data and re-ran the analysis. The overall outcome is basically the same, though the classification plot is slightly different now: 86.5% of the cases are predicted correctly and the same variables and dummy variables as before are reported as significant.

Another variation on our basic regression analysis consists of making a four- rather than a three-way distinction in terms of definiteness and separating out the variable cases, i.e. cases where an argument is relativized on, into a class of their own. Combined with Wasow's measure of relative length this results in the model handling an additional 1 % of cases for a total of 87.5%. (With our treatment of length, the model does not improve.) We also now have 5 outliers rather than 4. 4 of the 5 outliers have a high Cook's distance value, indicating that they do have great influence on the model coefficients. Excluding them would change the model significantly.

We can also try to use the absolute length values rather than the relative length information. As the actual length values carry more information than the relative variable, this move might improve the model. It turns out, however, that that does not improve the model. We again get 87.5% of the cases right.

What does improve the model significantly is the addition of a new variable, Subject choice. The subject of a clause headed by a ditransitive verb can be either

¹⁸This was pointed out to me by Sridhar Narayanan of ICSI.

Observed			Predicted		
			Construction		Percentage Correct
			do	cm	
Step 1	Construction	Double object	154	9	94.5
		Caused-motion	7	38	84.4
	Overall Percentage				92.3

Table 2.63: Classification table for model with four factors (absolute lengths of frame elements, relative definiteness, subject choice)

the Donor (in active voice) or the Theme or the Recipient. The passives show a strong correlation between chosen subject and construction: choosing the Recipient as subject normally means that the double object construction is used, and choosing the Theme as subject normally means that the caused-motion construction is used.

If we add a categorical variable for subject choice to the regression model in a third block, we get the classification table in 2.63 as a result, showing that now 92.3% of the sentences can be assigned the right construction.

The Hosmer and Lemeshow tests at each step indicate a good match between data and model. The initial -2 Log likelihood of 217.254 is reduced to 76.189 at the end of the regression. Inspection of multi-collinearity diagnostics suggests that there is no problem: VIF values as well as Tolerance values are around 1.000. The values for the Condition index range from 1.000 to 5.097, well below the mark of 15. No component associated with a high condition index contributes substantially to the variance of two or more variables. All variables entered into the backward stepwise regression model are kept when a removal test is performed.

There are three outliers reported for the standardized residuals (1.4%), which is just a bit more than is acceptable. One of them is (2.69) above, which involves a strong violation of length while relative definiteness is adhered to. Another case that appears is the idiomatic *give a hostage to fortune* seen above in (4.102). The last of the three cases is (2.73), which involves a strong violation of relative length. 2 of

these cases also show a Cook's distance value of >1 .

The next attempt was to see whether information about contextual mentions would account for additional data. I added two categorical variables to the third block of the stepwise backward regression. One variable represents the prior evocation status and the other the subsequent evocation status of the Theme. The model does not keep either of the variables. Neither does it keep categorical variables for relative prior or subsequent evocation status.

We could continue trying to re-code the variables or use different entry methods etc. However, that would just be going on a fishing expedition. So we will end the analysis here, having found that we can reproduce Wasow's (2002) finding that weight is a significant factor. However, we have not directly attempted to replicate his finding regarding prior evocation status since we relied on variables representing definiteness morphology. A final run of the regression analysis suggests that replacing the relative definiteness variable with the prior evocation variable leaves the total number of correctly classified cases exactly the same. Which cases are predicted correctly does change, however: the model gets slightly better at predicting the double object tokens and worse at predicting the caused-motion tokens. There is clearly a relation between definiteness morphology and relative prior evocation but it seems with the data at hand and the statistical techniques available it is not possible to pick one over the other.

One factor that Wasow (2002) did not mention but which does improve the model significantly is Subject choice, which for main clauses approximates choice of sentence topic. This variable is significant since it more or less automatically determines the construction choice in the passive voice. It is not clear whether the corpus study on the dative alternation that Wasow (2002) reports on included or eliminated passive voice uses. If it did include such tokens, it is curious that there is no discussion of the interaction between passive voice and choice of linking construction.

	n	%
administer	5	1.7
concrete	94	32.1
donate	1	.3
metaphorical	193	65.9
Total	293	100.0

Table 2.64: Uses of *give* in the Switchboard corpus

	n	%	n	%
Caused-motion	30	31.9	43	14.7
Double object	53	56.4	199	67.9
Theme-only	11	11.7	51	17.4
Total	94	100.0	293	100.0

Table 2.65: Constructions for all and only concrete uses of *give*

2.3 *Give* in the Switchboard corpus

In order to have a comparison to the mainly written data from the British National corpus, I also examined uses of *give* that were sampled from the Switchboard corpus. I extracted a random sample of 400 sentences coded them for the features construction used, voice, theme and recipient morphology, theme and recipient length, relative definiteness and relative length, presence of particle, concrete, metaphorical or idiomatic meaning.

After eliminating the cases involving particle verb uses such as *give away*, uses of *given that*, and tokens that were incomplete or incoherent, there remained in the sample 293 tokens. They belonged to the uses shown in Table 2.64

Note that the tokens coded metaphorical also include cases where information is given (*give your name and address*). As usual, the double object construction is more frequent than the caused-motion construction: 82.2% of all tokens that express a Recipient overtly belong to the double object construction.

The distribution of tokens across constructions is compared in Table 2.65.

The table shows that the concrete tokens include a relatively higher proportion of

	Double object	Caused-motion
BNC (concrete)	163	45
Switchboard (all)	199	43
Switchboard (concrete)	53	30

Table 2.66: Distribution of constructions for different corpora

	Caused-motion	Double object	Theme-only	Total
a	36	184	48	268
p	7	15	3	25
Total	43	199	51	293

Table 2.67: Voice by construction for all tokens

caused-motion sentences than all tokens do. The difference is statistically significant ($\chi^2=14.112$, $p=0.0009$).

Table 2.66 offers a comparison of the distribution of the two constructions to what we found for concrete possession transfer uses in the BNC.

Pairwise χ^2 -tests suggest the following:

- The tokens of *give* in the BNC with a concrete meaning are different from the concrete tokens in Switchboard with regard to their distribution across the constructions. In Switchboard, sentences are more likely to occur in the Caused motion construction than in the BNC ($\chi^2=6.529$, $p=0.0106$).
- The Switchboard overall and the concrete BNC tokens do not show a difference ($\chi^2=1.063$, $p=0.30253$).
- The *give* tokens from the Switchboard overall are differently distributed than the concrete tokens alone ($\chi^2=11.983$, $p=0.0005$). The difference is, naturally, even greater when concrete and non-concrete tokens are compared ($\chi^2=29.195$, $p < 0.0001$).

Voice At 8.5%, the rate of passives for the 293 tokens is significantly lower, according to a χ^2 -test yielding $p=0.0000$, than for the sample of concrete uses of *give* from

	Caused-motion	Double object	Theme-only	Total
a	26	51	11	88
p	4	2	0	6
Total	30	53	11	94

Table 2.68: Voice by construction for concrete tokens

the BNC where 21.6% of the tokens were passive (45/208).

If we examine only the tokens with concrete meanings, we find the results in Table 2.68. Here too the rate of passives is, at 6.3%, significantly lower, according to a χ^2 -test yielding $p=0.0011$, than in the sample from the BNC. Within the Switchboard sample, there is no difference in the rate of passives between the concrete and the non-concrete uses ($\chi^2=.819$, .3655).

Frame element morphology The morphological realizations of the Theme and Recipient arguments are summarized in various tables below. What we find is not surprising: Recipients overall tend to be definite and Themes indefinite (cf. Table 2.69). Recipients are significantly more likely to be definite in the double object construction than in the caused-motion construction, while Themes are significantly more likely to be indefinite in the double object construction than in the caused-motion construction (cf. Tables 2.71 and 2.72). For instance, χ^2 -tests performed for all tokens yield a χ^2 -value of 15.307 and a p-value of 0.000 for Themes and a χ^2 -value of 53.922 and a p-value of 0.000 for Recipients.

Comparing Tables 2.69 and 2.70 might suggest that the number of pronominal Recipients is even higher among non-concrete uses than among concrete uses of *give*, which we can infer from the relatively higher number of personal pronoun Recipients in Table tab:choteau. However, this difference is not statistically significant; a χ^2 -test yields a χ^2 -value of 3.279 and a p-value of 0.07017.

Relative length of the Recipient Table 2.73 shows the relative length of the Recipient across the various constructions. In the double object construction, the Recipient is shorter in 83.4% of the cases, compared to 16.3% of the cases in the

	Recipients		Themes	
	n	%	n	%
1st or 2nd	114	47.1	0	0
3rd	76	31.4	17	7.0
This	1	.4	1	.4
This NP	2	.8	3	1.2
That NP	0	0	3	1.2
Proper Noun	2	.8	0	0
Possessed NP	8	3.3	5	2.1
Definite NP	15	6.2	33	13.6
	218	90.1	62	25.6
Indefinite NP	3	1.2	67	27.7
Quantified NP	10	4.1	43	17.8
Bare NP	5	2.1	44	18.2
Indefinite pronoun	4	1.7	5	2.1
Indefinite zero	1	.4	1	.4
Interrogative pronoun	0	0	2	.8
That-relative	1	.4	9	3.7
Reduced relative clause	0	0	2	.8
What of headless relative	0	0	2	.8
Embedded <i>wh</i> -question	0	0	2	.8
Negative determiner	0	0	2	.8
Negative pronoun	0	0	1	.4
Total	242	100.0	242	100.0

Table 2.69: Realization of Recipients and Themes for all tokens

	Theme		Recipient	
	n	%	n	%
3rd	16	19.3	21	25.3
1st or 2nd	0	0	36	43.4
This	1	1.2	0	0
This NP	1	1.2	2	2.4
Possessed NP	3	3.6	5	6.0
Definite NP	2	2.4	5	6.0
	23	27.7	69	83.1
Indefinite NP	16	19.3	2	2.4
Quantified NP	20	24.1	5	6.0
Bare NP	14	16.9	2	2.4
Indefinite pronoun	1	1.2	3	3.6
Indefinite zero	0	0	1	1.2
Reduced relative clause	1	1.2	0	0
What of headless relative	4	4.8	0	0
Negative determiner	1	1.2	0	0
That-relative	3	3.6	0	0
Total	83	100.0	83	100.0

Table 2.70: Realization of Recipients and Themes for concrete tokens

	Caused-motion	Double object	Total
3rd	17	0	17
This NP	0	3	3
That NP	0	3	3
Possessed NP	0	5	5
Definite NP	2	31	33
This	1	0	1
	20 (46.5%)	37 (18.5%)	
Indefinite NP	4	63	67
Quantified NP	6	37	43
Bare NP	6	38	44
Indefinite pronoun	1	4	5
Interrogative pronoun	0	2	2
Reduced relative clause	1	1	2
What of headless relative	2	0	2
Embedded wh-question	1	1	2
Negative determiner	0	2	2
Negative pronoun	0	1	1
Indefinite zero	1	0	1
That-relative	1	8	9
Total	43	199	242

Table 2.71: Realization of Theme frame elements across constructions

	Caused-motion	Double object	Total
1st or 2nd	5	109	114
3rd	10	66	76
This	0	1	1
This NP	1	1	2
Proper Noun	2	0	2
Possessed NP	4	4	8
	24 (55.8%)	181 (91.0%)	
Indefinite NP	1	2	3
Quantified NP	7	3	10
Bare NP	3	2	5
Indefinite pronoun	2	2	4
Indefinite zero	0	1	1
That-relative	0	1	1
Total	43	199	242

Table 2.72: Realization of Recipient frame elements across constructions

	Caused-motion	Double object	Total
shorter	7	[167]	174
equal	14	18	32
longer	[22]	14	36
Total	43	199	242

Table 2.73: Relative length by construction for all tokens

caused-motion construction.

If we consider the cases where the Recipient is shorter in the double object construction and longer in the caused-motion construction to be predicted by weight, we find that 78.1% (189/242) of the cases are predicted correctly.

We can repeat this calculation for the concrete cases. Their distribution is given in Table 2.74.

Among the concrete tokens, the value of relative length as a predictor of constructional choice is lower than for all tokens: only 63.9% of the cases (53/83) are predicted.

	Caused-motion	Double object	Total
shorter	2	[37]	39
equal	12	8	20
longer	[16]	8	24
Total	30	53	83

Table 2.74: Relative length by construction for concrete tokens

	Caused-motion	Double object	Total
I<P	[6]	0	6
I<D	0	2	2
D<P	[3]	0	3
I=I	7	9	16
D=D	1	3	4
P=P	9	0	9
D>I	11	[9]	20
P>D	1	[37]	38
P>I	5	[139]	144
Total	43	199	242

Table 2.75: Relative definiteness by construction for all tokens

Relative definiteness of the Recipient Table 2.75 shows the relative definiteness of the Recipient across the various constructions. The cells in which values are enclosed in brackets are the ones that are directly predicted by considerations of definiteness. In the double object constructions, the recipient should be higher on our three point definiteness scale than the Theme and the reverse should be true in the caused-motion construction.

By itself, the consideration of relative definiteness handles 80.2% of the cases (194/242).

We can repeat this calculation for the concrete cases. Their distribution is given in Table 2.76.

Among the concrete tokens, the value of relative definiteness as a predictor of constructional choice is lower than for all tokens: only 67.5% of the cases (56/83) are predicted.

	Caused-motion	Double object	Total
I<P	[5]	0	5
D<P	[3]	0	3
P=P	9	0	9
D=D	1	0	1
I=I	3	5	8
D>I	6	[3]	9
P>D	0	[5]	5
P>I	3	[40]	43
Total	30	53	83

Table 2.76: Relative definiteness by construction for concrete tokens

Relative definiteness by relative weight With the concrete uses of *give*, the combination of relative weight and relative definiteness handles 67.5% of the cases (56/83) correctly. There is thus no improvement over relative definiteness alone.

Considering all uses of *give*, the combination of relative weight and relative definiteness handles 81.4% of the cases (197/242) correctly. This represents a very slight improvement over what relative definiteness predicts by itself (80.2%) but is still worse than always guessing the double object construction, which would be right 82.2% of the time for our sample.

Cases where two personal pronouns co-occur constitute a rather large set of the unpredicted cases, namely 9. In our Switchboard sample, there basically was no variation in the ordering of two personal pronouns. If we added a third hard constraint to that effect as a factor, our accuracy for concrete cases would go up to 78.3% (65/83) from 67.5%. For the complete data set, the improvement would be a bit less pronounced, going from 81.0% to 84.7%.

Prior evocation Table 2.79 summarizes the distribution of Recipient tokens over prior evocation statuses for both the double object and the caused-motion constructions.

In the caused-motion construction only 18.6% of the Recipients (8) belong to the Variable and Brand-new categories. In the case of the Double object construction, the

	Shorter	Equal	Longer	
Caused-motion				
I<P	0	0	5	5
D<P	0	0	3	3
P=P	0	(9)	0	9
D=D	*1*	0	0	1
I=I	0	0	3	3
D>I	%1%	*1*	*4*	6
P>I	0	*2*	*1*	3
Total	2	12	16	30
Double object				
I=I	2	0	*3*	5
D>I	0	1	*2*	3
P>D	5	0	0	5
P>I	30	7	*3*	40
Total	37	8	8	53

Table 2.77: Relative definiteness by relative length by construction for concrete tokens

	Shorter	Equal	Longer	
Caused-motion				
I<P	0	0	6	6
D<P	0	0	3	3
I=I	*1*	0	6	7
D=D	*1*	0	0	1
P=P	0	(9)	0	9
D>I	%3%	*3*	*5*	11
P>D	%1%	0	0	1
P>I	%1%	*2*	*2*	5
Total	7	14	22	43
Double object				
I<D	*2*	0	0	2
I=I	4	(1)	*4*	9
D=D	3	0	0	3
D>I	5	2	*2*	9
P>D	37	0	0	37
P>I	116	15	*8*	139
Total	167	18	14	199

Table 2.78: Relative definiteness by relative length by construction for all tokens

	Caused-motion	Double object	Total
Variable	1	1	2
Brand-new	7	3	10
Unused	1	2	3
Inferrable	15	21	36
Generic <i>you</i>	0	20	20
Evoked	14	105	119
Situationally given	5	47	52
Total	43	199	242

Table 2.79: Prior evocation status of Recipients across constructions

percentage is even lower at 2.0%. This difference is statistically significant according to a χ^2 -test with $\chi^2=20.661$ and a p-value of 0.000.¹⁹

As expected, the percentage of Recipients that are explicitly textually evoked or situationally evoked (i.e. the referent is a speech act participant) is significantly higher in the double object construction (76.4%) than in the caused-motion construction (44.2%).

Note that in Table 2.79 I exceptionally identified the number of generic *you*'s occurring in the sample because their number seemed to be strikingly high. By generic *you* I mean uses of the pronoun as in (2.124), where the speaker is not referring specifically to their interlocutor but to whoever participates in the relevant event.

(2.124) It **gives** you no relaxation.

I only assigned to generic *you* those cases that occurred without a prior mention of the generic referent; cases with prior mentions were treated as evoked rather than situationally evoked. Ideally, I should have separated out the issue of genericity from the issue of prior evocation. The understanding of generic *you* and its status with

¹⁹Note though that the expected value for 1 out of 4 cells is under 5, which argues against performing the test. Since the tool that I am using for performing a Fisher's exact test, which would be appropriate in this situation, cannot work with cell values as large as the ones here, I decided to report the χ^2 -value anyway.

	Caused-motion	Double object	Total
Variable	3	12	15
Brand-new	12	133	145
Unused	0	1	1
Inferable	8	44	52
Evoked	20	9	29
Total	43	199	242

Table 2.80: Prior evocation status of Themes across constructions

respect to evocation is not a simple matter (see Rubba 1996). For my own purposes, I have usually treated these cases in the class of situationally given referents rather than that of inferable referents.

Turning now to Themes, we notice that the situation is very different. Unsurprisingly, there are no situationally given referents: the corpus consists of phone conversations and they don't share a physical context; the interlocutors themselves also do not figure as Themes. More importantly, there are a great number of Theme referents with either Variable or Brand-new status in the caused-motion construction (34.9%) and the double object construction (80.4%). Conversely, there are very few explicitly evoked Themes, and this is true in the caused-motion construction (18.6%) and even more so in the double object construction (4.5%).

Relative prior evocation status If we cross-tabulate the evocation statuses of Recipients and Themes in the two constructions, we get the results shown in Table 2.81.

The caused-motion construction seems to be less well predicted by relative evocation status than the double object construction. In the double object construction, there is 1 case where the Recipient is inferable and the Theme evoked, where we might have expected the caused-motion construction. There are 17 cases where both arguments have the same status and no prediction is made. All other 181 cases (91.0%) are correctly predicted.

In the caused-motion construction, only 12 cases (27.9%) are predicted outright,

	Theme status					
	Caused-motion					
	v	b	u	i	e	Total
Variable	0	1	0	0	0	1
Brand new	0	(2)	0	1	4	7
Unused	0	*1*	0	0	0	0
Inferrable	0	*4*	0	(5)	6	15
Evoked	*2*	*4*	0	*2*	(6)	14
Situationally given	*1*	0	0	0	(4)	5
Total	3	12	0	8	20	43
	Double object					
	v	b	u	i	e	Total
	Variable	0	0	0	1	0
Brand new	(1)	(2)	0	0	0	3
Unused	0	2	0	0	0	2
Inferrable	1	13	0	(6)	*1*	21
Generic you	0	19	0	1	0	20
Evoked	7	65	1	27	(5)	105
Situationally given	3	32	0	9	(3)	47
Total	12	133	1	44	9	199

Table 2.81: Relative prior evocation status for all tokens

in 17 cases there is a tie, and 14 cases are wrongly predicted. Clearly, this is a much worse correlation than we'd want in order to be able to use relative evocation status as the sole predictor of constructional choice. On the other hand, appearances are a bit deceptive here. Cases like the 10 sentence (6+4) where the Theme is evoked and the Recipient textually or situationally evoked are, in the American English to which the Switchboard corpus belongs, pretty much categorically assigned to the caused-motion construction and thus there really is no need to predict them.

Overall, 193 cases (79.8%) are predicted correctly by the factor relative prior evocation status. This is a higher rate than we found for the concrete tokens of the BNC sample, where the rate was 69.7% (cf. p. 53). The difference is statistically significant ($\chi^2 = 6.032, p = 0.0140$).

The cases with a Variable-Theme and evoked Recipients in the caused-motion construction involve the relative clauses, embedded questions, and headless relatives in (2.125-2.127).

- (2.125) So the advice we **gave** to them was, number one, visit the college that you're thinking about.
- (2.126) All my income goes for things like luxuries, I mean, it really does, except for what we **give** to my daughter, to, you know, put her through college and stuff.
- (2.127) s [This is, + this is, part of giving back a portion of what He's **given** to us.

In two of the cases, there is a notion of contrast between the Recipient and somebody else. In example (2.127), *us* refers to humans and they contrast with God as the Recipients of giving.

Example (2.125), comes from a context where *them*, the speaker's children, contrast with a child of the Addressee's, whose reasons for selecting a particular college were just discussed.

In the case of (2.126) one factor might be that the prior evocation of the referent has occurred very much earlier whereas the use of the double object construction usually goes with recent mention of the Recipient. Note, too, that for merely inferable

		Theme status				
		Caused-motion				
	v	b	i	e	Total	
variable	0	(1)	0	0	1	
brand-new	0	(1)	1	4	6	
inferable	0	*2*	(4)	5	11	
evoked	*1*	0	*2*	(4)	7	
situationally given	*1*	0	0	(4)	5	
Total	2	4	7	17	30	
		Double object				
	v	b	i	e	Total	
brand-new	(1)	(1)	0	0	2	
generic you	0	4	0	0	4	
inferable	1	2	(3)	1	7	
evoked	3	20	3	(2)	28	
situationally given	0	10	1	(1)	12	
Total	5	37	7	4	53	

Table 2.82: Prior evocation status of Themes by prior evocation status of Recipient

Recipient referents the double object construction is not as pre-dominant as in the case of evoked Recipients.

Table 2.82 repeats the cross-tabulation for the concrete data only. 53 cases (63.9%) are predicted correctly. Of the 30 cases that are not predicted, 6 involve the violation of a prediction, 24 are due to tied evocation status.

Subsequent evocation The next factor we will look at is the subsequent evocation of referents. The data on the number of subsequent referents is summarized in Table 2.83.

There is no surprise in the fact that Recipients on average have more subsequent mentions than Themes. It is also the case that of the Recipients more (52) have a subsequent mention than don't (31), whereas the opposite is the case for Themes where only 19 have a subsequent mention and 64 do not.

	Concrete uses		All uses	
	Recipients	Themes	Recipients	Themes
N	83	83	242	242
Mean	1.40	.52	1.59	.25
Median	1.00	.00	1.00	.00
Mode	00	00	0	0
Std. Deviation	1.623	1.2133	1.961	.788
Skewness	1.417	3.0633	1.640	4.704
Std. Error of Skewness	.264	.264	.156	.156

Table 2.83: Number of subsequent mentions

	Theme status		
	Caused-motion construction		Total
	no	yes	
no	(10)	[2]	12
yes	6	(12)	18
Total	16	14	30
	Double object construction		
	no	yes	Total
no	19	0	19
yes	[29]	(5)	34
Total	48	5	53

Table 2.84: Relative subsequent mentions for concrete tokens

If we cross-tabulate the subsequent mentions of Recipients and Themes for the concrete tokens we get the Table in 2.84.

It is clear that subsequent mention as a single predictor is not useful, if we predict the double object construction only when the Recipient has a subsequent mentions and the Themes does not and conversely predict the caused-motion construction only when the Theme has a subsequent mention while the Recipient does not, we get only 31 out of 83 cases right (37.3%). Guessing the double object construction whenever the Recipient has a subsequent mention would do better, at 34 cases (40.9%) but guessing the caused-motion construction only when the Theme has a subsequent mention does even better: we would get 62 cases (74.7%) right.

	Theme status		
	Caused-motion construction		Total
	no	yes	
no	(14)	[5]	19
yes	8	(15)	24
Total	23	20	43
	Double object construction		Total
	no	yes	
	no	74	0
yes	[110]	(15)	125
Total	184	15	199

Table 2.85: Relative subsequent mention for all tokens

Observed			Predicted		
			Construction		Percentage Correct
			do	cm	
Step 1	Construction	Caused-motion	38	5	88.4
		Double object	4	195	98.0
	Overall Percentage				96.3

Table 2.86: Classification table for model with six factors

The situation is different when we look at all Switchboard tokens at once. On the strict criterion, we would get only 115 (110+5) cases right (47.5%). Guessing based on the Recipient gives us the correct prediction in 125 cases (51.1%). But guessing the caused-motion construction whenever the Theme has a subsequent mention and the double object construction when it doesn't gives us the best result: we get only 15 wrong and 227 right, which amounts to 93.8%.

2.3.1 Regression analysis

In this section I will summarize a regression analysis I performed on the data sampled from the Switchboard corpus of American English.

The best model that could be built for the data was able to classify 96.3% of the data, as shown in Table 2.86.

The model was generated by a hierarchical regression with three blocks using the backward stepwise method of entry with likelihood ratio as a removal criterion. The first block contained recipient length and theme length; the second recipient morphology, coded with 4 levels and theme morphology, also coded with four levels. The third block contains two categorical variables, Subject choice and Recipient focus. All of the variables except the last two should be familiar.

Subject choice basically record which of Donor, Theme, or Recipient is chosen as clausal subject, which is typically equivalent to clausal topic. This factor by itself is not very predictive as the sentences with Donor as subject can take either the caused-motion or the double object form. However, when either Recipient or Theme are chosen as subject, the construction that appears is, in the Switchboard corpus and probably in American English at large, almost completely predictable: Theme subject clauses require use of the caused-motion construction while Recipient subject clauses. (The one exception seem to be passives with a pronominal Recipient, especially in reduced relative clauses.)

The second factor that is new is Recipient focus. This factor was added to record with a binary variable those cases where a Recipient is clearly focal either because a focusing marker such as *only* or *just* modifies it, as in (2.128), or because it appears in an explicitly contrastive context, as in (2.129).

- (2.128) A.93 utt3: D like our band gave money away. /
 A.93 utt4: We're a band booster club. /
 A.93 utt5: [We gave, + we **give**,] F uh, two five hundred dollar scholarships just to kids, F uh, who we think were worthy, D you know /
 A.93 utt6: D so,
- (2.129) A.65 utt18: Use a lot of discernment. /
 A.65 utt19: Don't **give** money to every nation. /
 A.65 utt20: You know, what I'm saying, /
 A.65 utt21: I think we're loaning money to too many nations, /
 A.65 utt22: C but I think if we **gave** to a few select ones, that

really needed it, that it would work out better for us and for them. /

In the first step of the regression, 88.8% of the cases are classified correctly. The initial -2 Log Likelihood of 226.447 is reduced to 151.677. The Hosmer and Lemeshow Test of Goodness of fit comes up significant at .003 ($\chi^2=21.846$, $df=7$), indicating that the model so far does not fit the data well. Nonetheless, the model does not remove the two length variables as that would make the model significantly worse; in each case $p < .000$.

The regression shows that increasing length of the Recipient reduces the odds for the double object construction, which is the event to be predicted in this regression. For Recipient length, $\text{Exp}(\beta) = .241$ and the lower and upper bounds of the confidence interval are both less than 1, namely .124 and .468. For Theme length, $\text{Exp}(\beta)$ is 1.662, with lower and upper bounds 1.278 and 2.162. Longer themes thus are found to increase the odds of the double object construction occurring. Both these findings are in accord with our previous considerations.

The second step of regression, where the categorical variables for Theme and Recipient morphology are introduced, reduces the -2 Log likelihood to 91.132. The Hosmer and Lemeshow test at this step is insignificant at .595 with $df=8$ and $\chi^2=6.467$. 93.0% of the data are classified correctly by the model. The removal test shows both variables to make a significant contribution to the model. The p-value for the Recipient morphology removal test is .002, that for Theme morphology .000.

In the final step, Recipient focus and subject choice are introduced into the model. This reduces the -2 Log likelihood to 34.067. The Hosmer and Lemeshow test for goodness of fit now gives a significance value of 1.000 ($\chi^2=.307$, $df=8$), indicating very good fit between model and data. Both variables are retained in the model as the removal test gives significance values of .000 for both variables.

Inspection of the Cook's distance values shows that no value is greater than 1. There are two cases with large leverage values close to 1 but these have perfectly acceptable standardized residuals and small Cook's distance values. As Field (2000) points out, cases with large leverage values need not necessarily have a large influence

on the regression coefficients because they are measured on the outcome variables rather than the predictors. I will thus assume that the two cases of high leverage values do not actually show undue influence.

Inspection of standardized residuals for outliers yields one case with a z-score outside of the ± 2 range, which is acceptable for a data set of 242 tokens. The relevant token is shown in (2.130). The context is a discussion of teachers that are drunk at school. The example violates the prediction of weight in the absence of a prediction by relative definiteness. Subject choice has nothing to say about the case and I coded the case as not involving a focal Recipient.

- (2.130) A.97 utt1: Isn't that amazing, /
 B.98 utt1: C And, D you know, my children brought it to my attention. /
 B.98 utt2: I never—
 A.99 utt1: [[If it, + if it's —] +
 B.100 utt1: — was there to see it. /
 A.101 utt1: — if it was] true there, I'm sure it's true with just
 about probably to some degree every school there is —/
 B.102 utt1: Uh-huh. /
 A.103 utt1: — D you know, C because they represent a pretty good, -/
 B.104 utt1: C And that's not **giving** a good example to students.
- A.105 utt1: No, /
 A.105 utt2: certainly not, /
 A.105 utt3: certainly not, /
 A.105 utt4: they don't need that example /
 A.105 utt5: C but, -/

Inspection of DFBeta values for the various predictor variables finds 2 cases that seem to affect the regression parameters unduly. In the case of (2.131), the DFBeta value of subject choice registers as unusual and the same is true for (2.132). It is not clear to me why these values should be so unusual. Certainly, (2.131) is odd in

that the subject choice causes a definite Recipient to be coded in the caused-motion construction. Moreover, relative length also would have favored the Use of the double object construction. The only thing that seems clearly unusual about (2.132) is the fact that the Recipient is indefinite. Relative length favors the use of the double object construction as does the absence of focus on the Recipient.

(2.131) A.79 utt1: C Because, [there's no, + the answer is not] being **given** to them, in the court system /

(2.132) A.77 utt4: [C So even before, + C so even when] things were good in the Middle East, Americans weren't really **given** visas to go into Baghdad. /

The final check on the model consists of looking at multi-collinearity diagnostics to see which factors are correlated with each other. The Tolerance values for all six variables are all above 0.2, which is sometimes suggested as a cutoff below which there is reason for concern. The VIF statistics are all just about 1 so there is no great reason for concern.

The values for the Condition index range from 1.000 to 10.378, which is below the value of 15 that the SPSS package gives as a threshold for being concerned about collinearity.

The variance proportions suggest that, not surprisingly, Recipient morphology and length, and Theme morphology and length are somewhat correlated. In the former case, the same dimension accounts for 50% of the Recipient length variance and for 61% of the Recipient morphology variance. In the Theme case, the percentages are 49% and 41%, respectively.

Overall, it seems that the model is doing an acceptable job. We are left with only 9 out of 242 cases that are not predicted correctly. This involves 5 tokens in the caused-motion construction that are wrongly assigned to the double object construction and 4 double object construction that are assigned to the caused-motion construction.

Example (2.130) belongs to the former group, as do the following examples:

(2.133) B.64 utt3: Ronald Reagan basically said, D you know, I'm going to give this much of a tax cut to the tax payers [and, + and,] created [this, +

this] ludicrous budget that he knew, [[that, + that,] + F um, that] the Congress

- (2.134) A.22 utt2: C but if it were ((told to us)) that we would find the time somewhere between our seventeenth and, F uh, twenty-sixth birthday to **give** a year or two to the, to the country in the form of maybe building better roads or the parks
- (2.135) B.111 utt3: These people have **given** so much of their lives, D you know, [to, + to] America —
- (2.136) B.108 utt1: F Uh, we have fifty acres. /
 A.109 utt1: F Oh, my, /
 A.109 utt2: D well, then. /
 B.110 utt1: F Oh, pardon me, /
 B.110 utt2: we have forty, /
 B.110 utt3: we **gave** ten to one of our children <laughter>. /

Beginning with the last one, (2.136), we can notice that it would be handled correctly if the model treated close proximity of the inference trigger of a Theme referent in the same way as close prior antecedent. The *ten* clearly refers to part of the original fifty acres and is highly inferable.

Examples (2.134) and (2.135) involve disfluencies at the point where the Recipient ought to be produced. We might take the position that these tokens ought to be eliminated completely as the speaker did not have the utterance fully planned or did not the utterance in the way they originally planned it.

Example (2.133), finally, involves a quotation, which may not be evaluated fairly relative to its embedding context.

We now turn to unaccounted for tokens where the double object construction occurs but the caused-motion construction is predicted.

- (2.137) A.165 utt1: I really and truly did. /
 A.165 utt2: E I mean, I have absolutely no problems with them winning

seven Oscars. /

A.165 utt3: None at all. /

B.166 utt1: Right. /

A.167 utt1: Normally I'm one of these people that I don't like [one, + one] movie taking all the honors. /

B.168 utt1: Everythin-, - /

B.168 utt2: yeah, /

B.168 utt3: yeah. /

A.169 utt1: I wanted him to win best actor on top of it, /

A.169 utt2: he was, -/

B.170 utt1: Yeah, F oh. /

A.171 utt1: Like I said, I wouldn't **given** the supporting actresses either one. /

In (2.137), the interlocutors are talking about the Academy awards and the fact that the movie *Dances with Wolves* won 7 Oscars. The example involves a violation of length: the Theme is shorter than the Recipient. The chosen form seems nonetheless appropriate as the Recipient is definite and very much inferable in the context. (Although speaker A prefaces the last utterance with *Like I said*, what they say is no repetition. It seems that one is supposed to take what they say as somehow following from the earlier statement that they don't like one movie to sweep the awards. This conflict of course with the earlier assertion that speaker A is fine with *Dances with Wolves* winning 7 Oscars. Be that as it me, it also is the case that *either one* seems to require a focal interpretation, despite the fact that that is not set up explicitly in the context. Focus on the Theme should also favor the double object construction.

Example (2.138) involves a length violation. The difference is small though and the Recipient referent, the generic homeless, has been mentioned before. Speaker B has also just used the same construction a few utterances before.

(2.138) B.96 utt7: I'm not saying just give them money - /

B.96 utt8: C but, E I mean, C because they have to be taught to be, -/

A.97 utt1: That's the last thing you want,

B.98 utt1: Yeah <laughter>. /

A.99 utt1: to do. /

B.100 utt1: E I mean, I know that's what my father-in-law was talking about, D you know, when you **give** these people money /

B.100 utt2: C or, <breathing>, what was it /

B.100 utt3: they were bringing up points asking as far as whether they need to give, F um, more money to these people, /

B.100 utt4: E I mean, C and that's not really the answer, ((E I mean)) , because if they're alcoholics they'll just go out and spend,

Example (2.139) seems not to have been assigned to the double object construction by the model because there is no length difference. Definiteness certainly would point to the double object construction.

(2.139) B.20 utt5: C and now they're wanting to go back to the old days where, D you know, women stayed home with the kids, D you know, and try to **give** the kids quality time. /

Example (2.140) involves a clear violation of length.

(2.140) A.115 utt1: [It, + it] astound, /

A.115 utt2: you realize that the Kuwaiti government **gives** every Kuwaiti man, woman, and child money. /

B.116 utt1: F Uh, no. /

B.116 utt2: [I wasn't, + I [didn't, + wasn't]] aware of that, /

Both referents and the proposition are somewhat inferable in the context—a discussion of the fact that American soldiers are working on the reconstruction of Kuwait in the aftermath of the First Gulf War while Kuwaitis do not seem to want to participate. Still, it's not clear if one is supposed to consider one of the referents an argument focus.

Focus	86.0
Subject choice	85.1
Theme length	82.2
Recipient length	86.0
Recipient morphology	84.7
Theme morphology	89.7

It seems we might be able to improve the model by rejecting some tokens as part of what should be accounted, notably ones involving disfluencies and quotes. Maybe a way to capture the relation between inferred referents and their inference triggers, at least for sub-set type relations, might improve the model somewhat, too. Likewise, encoding clear cases of Theme focus might also make the model better, as might the use of more data to allow for some cases of small length violations in the double object construction.

But clearly, the bulk of the data is very well described by the six variables we have looked at. What is missing now is a better analysis of the interrelationships between the different variables. Table 2.3.1 shows how many data points can be successfully classified by regression models with only one factor.

All of these factors by themselves offer an improvement over guessing the more frequent construction. The sets of data that each can help classify must overlap given that there are only 100% variance to account for. We already saw earlier on p. 155 that length and definiteness status of the arguments are somewhat correlated. Unfortunately, statistical techniques available to me at this time do not allow me to resolve the issue of which factors explain how much of the data. I will leave this open for future research.

2.4 Non-concrete Themes with *give*

In this section we will take a look at uses of *give* where no transfer of possession occurs. One substantial subclass of uses involves combinations in which the verb **give** is not the (sole) semantic head of the clause but mostly serves to project a clausal

structure. The event reported on in (2.141) is not giving but thanking.

(2.141) We **give** thanks to God.

Another large subset of non-concrete nouns occurring with *give* has to do with states of affairs that hold—that the Recipient metaphorically possesses—due to the action of the metaphorical Donor, as in (2.142).

(2.142) In some cultures sons are **given** priority, others give daughters priority, following the McIver principle “when you educate a man you educate an individual: when you educate a woman you educate a whole family”.

The sample contains very few idiomatic combinations of *give* and noun such as *give somebody the bird*.

Length Table 2.87 shows that overall there are very pronounced differences between the constructions when it comes to argument length.

The average absolute length difference between the two arguments is 5.05 words if both construction are considered together. It is slightly smaller at 4.66 words in the double object construction and slightly larger at 6.16 for the caused-motion construction.

If we compare the relative length tabulation in Table 2.88 to the one we saw earlier for the concrete tokens in the BNC (cf. 2.15 on p. 56), we can notice that length by itself predicts the right construction more often with support verbs, namely in 173 cases (79.7%), than with concrete tokens (65.4%). This difference is statistically significant according to a χ^2 -test ($\chi^2=10.334$, $df=1$, $p=0.001$).

Definiteness Non-concrete uses also differ significantly from concrete ones in terms of the definiteness of their arguments. As Table 2.89 shows, the definiteness rate for Themes is very low; Recipients by contrast are very likely to be definite.

We can compare the situation to that for the arguments that occur with the concrete uses of *give* in the BNC. The information from the earlier Tables 2.3 and 2.4 is summarized in compressed form in Table 2.90.

	Themes	Recipients
	Double object	
N	160	160
Mean	5.8125	1.5938
Median	4.0000	1.0000
Mode	3.00	1.00
Std. Deviation	5.36491	1.44195
Skewness	1.954	3.851
Std. Error of Skewness	.192	.192
	Caused-motion	
N	57	57
Mean	1.7193	7.2456
Median	1.0000	3.0000
Mode	1.00	2.00
Std. Deviation	1.53244	9.83122
Skewness	1.604	2.510
Std. Error of Skewness	.316	.316

Table 2.87: Length of arguments in support verb uses of *give*

	Caused-motion		Double object	
R < T	8	14.0	134	83.8
R = T	10	17.5	13	8.1
R > T	39	68.4	13	8.1
Total	57	100.0	160	100.0

Table 2.88: Relative length of arguments in support verb uses

	Recipient				Theme			
	Double object		Caused-motion		Double object		Caused-motion	
	n	%	n	%	n	%	n	%
Supplementary relative	1	.6	0	0	0	0	0	0
Reflexive	3	1.9	0	0	0	0	0	0
1st or 2nd	35	21.9	1	1.8	0	0	0	0
3rd	48	30.0	1	1.8	0	0	1	1.8
Possessive pronoun	0	0	0	0	1	.6	0	0
That	1	.6	0	0	1	.6	0	0
This	0	0	1	1.8	0	0	0	0
That NP	1	.6	0	0	1	.6	0	0
This NP	0	0	0	0	2	1.3	0	0
Possessed NP	6	3.8	2	3.5	7	4.4	5	8.8
Proper Noun	17	10.6	7	12.3	0	0	0	0
Headless relative	0	0	1	1.8	1	.6	0	0
Definite NP	24	15.0	18	31.6	22	13.8	1	1.8
	136	85.0	31	54.4	35	21.9	7	12.3
Quantified NP	5	3.1	0	0	13	8.1	4	7.0
Bare NP	12	7.5	16	28.1	43	26.9	27	47.4
Indefinite zero	1	.6	0	0	0	0	0	0
Indefinite NP	3	1.9	8	14.0	61	38.1	7	12.3
Negative pronoun	0	0	0	0	2	1.3	0	0
Negative determiner	0	0	0	0	1	.6	2	3.5
Restrictive relative	1	.6	1	1.8	0	0	1	1.8
That-relative	0	0	0	0	3	1.9	3	5.3
Reduced relative	2	1.3	0	0	0	0	6	10.5
embedded wh-question	0	0	1	1.8	0	0	0	0
what of headless relative	0	0	0	0	1	.6	0	0
Other	0	0	0	0	1	.6	0	0
Total	160	100.0	57	100.0	160	100.0	57	100.0

Table 2.89: Definiteness of frame elements for support verb uses

	Recipient		Theme	
	Double object	Caused-motion	Double object	Caused-motion
definite	90.8	71.1	31.3	62.2
indefinite	9.2	28.9	68.7	37.8

Table 2.90: Percentages of definite and indefinite arguments with concrete uses of *give* in the BNC

	both def.	both indef.	R def., T indef.	R indef., T def.	Total
Double object	(31)	(20)	105	[4]	160
Caused-motion	(6)	(25)	[25]	1	57
Total	37	45	130	5	217

Table 2.91: Construction by frame element morphology for support verb uses of *give*

The comparison clearly shows that with support verbs Themes are not as often definite as they are with concrete tokens; the difference is especially large in the caused-motion construction. This suggests that definiteness cannot play as large a role in choosing between the double object and caused-motion constructions as it does with concrete tokens.

Relative definiteness Table 2.91 shows the combinations of definiteness values for the Theme and the Recipient frame elements by construction. To the extent that differences in definiteness are predictive of constructional choice, we would expect to find that when the Recipient is definite and the Theme is not, the double object construction is used, and that when the situation is reversed, the caused-motion construction is used.

No particular prediction can be made for the 82 cases—enclosed in parentheses in the Table—where both frame elements have the same definiteness value.

Of the remaining 135 cases, 105 have a definite Recipient and an indefinite Theme in the double object construction, 1 has a definite Theme and an indefinite Recipient in the caused-motion construction, and the remaining 29 cases (21.5%) constitute

	R < T	R = T	R > T	Total
Double object	134	(13)	[13]	160
Caused-motion	[8]	(10)	39	57
Total	142	23	52	217

Table 2.92: Construction by relative length for support verb uses of *give*

mismatches. (As above they are enclosed in square bracket.) Most of these 29 cases specifically involve uses of the caused-motion construction where the Recipient is definite and the Theme is indefinite. The number of tokens that can be predicted correctly only by relative definiteness is a low 106 (48.8%).

Weight Table 2.92 shows the combinations of relative weight of the Recipient frame element and construction for our sample of 217 support verb uses of *give*. No prediction is made for 23 cases. Of the remaining 194 tokens, all are predicted by weight considerations except 8 cases where the caused-motion construction occurs even though the Recipient is shorter than the Theme, and 13 cases where the Recipient is realized as primary object in the double object construction even though it is longer than the Theme. (In the table these 21 cases are enclosed in square brackets.) Thus, 173 of the 217 tokens (79.7%) are predicted directly by weight.

Weight and definiteness Again we may ask whether weight and definiteness account for more of the data when taken together than they do separately. This is addressed by the tabulation in Table 2.93.

In Table 2.93, the cells that simply contain a number are those where either both factors, weight and relative definiteness predict the construction correctly or where one factor makes no prediction and the other predicts the correct construction (155/217=71.4%). The cells in which the number is contained in parentheses are cases where both factors make no prediction as to which construction should occur (12/217=5.5%). The cells in which the number is enclosed in square brackets contain cases in which one factor predicts the actually occurring construction, while the other incorrectly predicts the alternative construction, or where one factor makes no predic-

	both definite	both indefinite	Theme def., Recipient indef.	Theme indef., Recipient def.	Total
R > T T=R R<T	Double object				
	[1] (3)	[2] (3)	*.00*	[10] 7	13
	27	15	[4]	88	134
	31	20	4	105	160
	Caused-motion				
R > T R=T R<T	1	18	1	[19]	39
	(2)	(4)	.00	4	10
	[3]	[3]	[.00]	*2*	8
	6	25	1	25	57

Table 2.93: Relative length by relative definiteness

tion while the other factor's prediction is not borne out (42/217=19.4). Consider, for instance, in the top half of the table, under the heading "double object", the cell that intersects "R>T" and "Theme indef., Recipient def.". The 10 cases that are found there are in accord with what definiteness predicts—when the Recipient is definite and the Theme is indefinite, the double object construction should be used—but they are not in accord with what weight considerations predict given that the Recipient is longer than the Theme, which should give rise to the caused-motion construction. The final category is that which involves violations of both factors' predictions; these cases are enclosed by asterisks (2/217=.9%).

Clearly, the combination of both factors covers more ground than the factors do individually: when both are combined, there are only 12 (5.5%) out of the 217 for which no prediction can be made, whereas this was the case for 23 sentences (10.6%) when weight alone was considered, and for 82 cases (37.8%) where definiteness alone was taken into account.

However, it seems that the combination of both factors leads to fewer total correct predictions (161/217=74.2%) than weight alone is able to provide (173/217=79.7%), though it is of course better than definiteness alone (106/217=48.8%).

	Double object			Total
	T<R	T=R	T>R	
D>I	0	0	[4]	4
D<P	0	1	17	18
I<D	[7]	1	29	37
I<P	[3]	6	59	68
P=P	0	(2)	0	2
D=D	1	0	10	11
I=I	2	(3)	15	20
Total	13	13	134	160
	Caused-motion			Total
	T<R	T=R	T>R	
D>I	1	0	0	1
D<P	0	0	*1*	1
I<D	[19]	[3]	*2*	24
I<P	0	[1]	0	1
P=P	0	[1]	0	1
D=D	1	(1)	[2]	4
I=I	18	(4)	[3]	25
Total	39	10	8	57

Table 2.94: Relative length by relative definiteness measured at three levels

Maybe the measure of definiteness was not fine-grained enough since pronouns are treated just like definite nouns. A more nuanced treatment of definiteness may show a clear effect. Table 2.94 repeats the calculations for a three level coding of definiteness, with definite pronouns separated from definites with a lexical head.

The number of cases for which no prediction is made is reduced to 10 (4.6%), 161 (74.2%) cases are predicted correctly and 46 (21.2%) cases wrongly. There is basically no improvement over the tabulation with a binary definite-indefinite distinction.

Why might this be? Table 2.94 suggests that 19 of the violations in the caused-motion construction consist of cases such as (2.143), where weight outweighs definiteness.

(2.143) It is much larger than previously, **giving** more room to the business of sending out carpet samples to hungry customers.

The double object construction by contrast has 10 violations where weight is not

	Active	Passive	Total
Double object	132	28	160
Caused-motion	41	16	57
Total	173	44	217

Table 2.95: Voice by construction for support verb uses of *give*

honored. These split into two subgroups. Half the cases involve passives like (2.144), the other half are actives such as (2.145).

(2.144) The only way the trust can be **given** effect is by construing it as charged on the intestate heir in favour of some other person.

(2.145) The ecological disaster created by the diversion of water from the Aral Sea for the sake of a better cotton crop has **given** that sort of thinking a severe knock.

If one discounts cases like (2.144), where the arguments are not contiguous, then it seems that definiteness is ignored in favor of weight more often than the other way around.

Voice Table 2.95 shows the distribution of voice by construction for support verb uses of *give*.

Overall, the active voice uses outnumber the passive ones. The magnitude of the difference is 2.56 for the caused-motion construction and 6.40 for the Theme-only construction. According to a χ^2 -test this is by a small margin not statistically significant, with $p=0.07$.

It is noteworthy that of the 16 passive voice cases in the caused-motion construction, 7 involve relativization on the Theme as in (2.147) but 0 cases involve relativization on the Recipient. 6 of the 7 cases of relativization on the Theme in the caused-motion construction involve reduced relative clauses like (2.147), whereas this type of relative is used only for 2 of the 4 cases of relativization on the Recipient in the double object construction.

Relative clauses also show up among the 28 passive voice sentences in the double object construction. 1 sentence involves relativization on the Theme, namely (2.146), and 4 cases involve relativization on the Recipient (cf. (2.148)).

- (2.146) Recall praise and criticism you've been **given**.
- (2.147) All these twists **given** to human character by the conditions of camp life were noticeable in varying degrees throughout the years.
- (2.148) When it comes to such matters as the protection of creditors and third parties dealing with a company, the obvious way forward is for the Community to lay down in its Directives detailed rules which are to be **given** similar effect under national law in all member states.

The numbers are small but they are in line with what we found earlier in our examination of various relative clause types. If a Theme needs to be anchored by a Recipient in a passive voice clause, then the caused-motion construction is to be much preferred over the double object construction. Recipients are typically relativized on with passive double object clauses.

Lexical bias Certain combinations of non-concrete noun and *give* are more strongly associated with one construction than another. To illustrate this, let us consider the combination *give emphasis*. A random sample of 150 sentences containing both words included 75 cases where they were in construction with each other. In 2 cases no Recipient was expressed. Of the 73 other cases, 59 sentences exhibited caused-motion syntax and 14 double object syntax.

For the caused-motion sentences there is a greater difference in length between the arguments than in double object sentences. In the former case, Themes are on average 1.73 words long and Recipients 8.24; in the latter case, Themes have an average length of 2.14 and Recipients one of 2.79. The prediction of length alone is, not surprisingly, better for caused-motion sentences than for double object sentences. Of the 59 caused-motion sentences, 50 have a longer Recipient. By contrast, of the

	Double object				Caused-motion			
	Recipient		Theme		Recipient		Theme	
3rd	2	14.3	0	0	0	0	0	0
Supplementary relative	1	7.1	0	0	0	0	0	0
Definite NP	2	14.3	0	0	21	35.6	0	0
That NP	0	0	0	0	0	0	1	1.7
This NP	1	7.1	0	0	0	0	0	0
Possessed NP	2	14.3	1	7.1	5	8.5	0	0
Proper Noun	1	7.1	0	0	2	3.4	0	0
	9	64.2	1	7.1	28	47.5	1	1.7
Quantified NP	3	21.4	0	0	4	6.8	3	5.1
Indefinite NP	0	0	3	21.4	1	1.7	3	5.1
Bare NP	2	14.3	10	71.4	21	35.6	38	64.4
Vping	0	0	0	0	5	8.5	0	0
Restrictive relative	0	0	0	0	0	0	1	1.7
Reduced relative	0	0	0	0	0	0	9	15.3
That-relative	0	0	0	0	0	0	3	5.1
Interrogative NP	0	0	0	0	0	0	1	1.7
Total	14	100.0	14	100.0	59	100.0	59	100.0

Table 2.96: Argument morphology for *give emphasis*

14 double object sentences, only 6 have a shorter Recipient. Overall length predicts 56 out of 73 cases (76.7%) of sentences correctly.

Table 2.96 shows that there is a great asymmetry in terms of definiteness as with non-concrete uses in general. The Table also suggests that Pronominal recipients only occur in the double object construction and that when Themes are relativized on, the caused-motion construction is used.

Relative definiteness does not work particularly well as a predictor of constructional form. There are 24 cases where definiteness is contradicted by the chosen form: a definite Recipient and an indefinite Theme occur in the caused-motion construction.

- (2.149) The examples used should be realistic and relevant to the pupils concerned. Emphasis should also be **given** to the way in which percentages are used both for comparative purposes in many everyday situations and also as a numerical measure.

There are 26 cases where the definiteness values match. 20 of these involve paired indefinites in the caused-motion construction.

(2.150) It is true that the paper still **gave** considerable emphasis to conventional forces .

Overall, relative definiteness accounts for only 23 of 73 cases (31.5%). This a low figure when compared for instance to the concrete tokens in the BNC (cf. Table 2.18 on p. 58), 67.3% of which can be accounted for directly by relative definiteness.

Note, too, that metaphorical Themes headed by *emphasis* hardly ever have subsequent mentions: the rate of .04 subsequent mentions is significantly lower than that for concrete Themes, which is .52. (cf. 2.83).

The situation is slightly different for *give chance*. A sample of 100 sentences that contain both lexical units included 78 sentences where they were in construction with each other. 1 case included no Recipient. The other 77 all exhibited the double object construction, 11 in the passive voice and 66 in the active voice. The Recipient is shorter in 73 cases, as long as the Theme in 3, and only in 1 case is it longer (cf. (2.151)). In other words, 94.8% of the cases are directly accounted for by relative length.

(2.151) We did it to **give** some of the great youngsters a chance.

The Recipients are pronouns or lexical definites in 62 of the 77 sentences (80.5%); 14 Recipients are indefinite; 1 is relativized on. Of the Themes, 50 are indefinites and 27 lexical definites. The definiteness value is tied in 12 cases. There is one violation but it involves relativization on the Recipient. Thus, 83.1% of the data conform to the predictions of relative definiteness.

Caused-motion uses do occur but they are exceedingly rare. Example (2.152) was retrieved from the internet. It strongly seems to be motivated by weight.

(2.152) “Part of our mission as an upstart company is to **give** a chance to somebody who’s so gifted or extraordinary, but because of luck or exposure, hasn’t had a chance,” Paula Alprin said.

As in the earlier cases we have seen, the metaphorical Theme is basically inactive as a discourse referent. There is only 1 case in which there was a subsequent mention of the Theme referent.

For *give lie* the situation is basically the opposite from *give chance*. *Give lie* occurred 50 times in the corpus. 47 sentence tokens exhibit the caused-motion construction, 2 the double object construction (cf. (2.153-2.154)).

- (2.153) He heard the calm voice , but could feel the trembling body which gave it the lie.
- (2.154) It is true that today's speakers, not knowing the origins of generic he , may regard it as just a feature of grammar. But historical scholarship gives this notion the lie: since the form was originally prescribed for sexist reasons, feminists who find it sexist are hardly projecting some novel and bizarre interpretation onto an innocent and neutral rule.

The sequence *the lie* appears in 46 of the 50 sentences, in 3 cases the bare noun *lie* is used. In one use, the NP is indefinite (*Does that not give an absolute lie to the belief that the Opposition are united on Europe ?*).

The Recipient noun phrase is typically longer than the Theme NP. It is as long as the Theme in 5 cases (4 in the caused-motion construction; 1 in the double object construction). It is shorter in 3 cases (2 in the caused-motion construction; 1 in the double object construction, namely (2.154)). On average, the large length difference between Recipient and Theme for the caused-motion sentences is -5.86 words.

Relative definiteness is not a good predictor for *give lie*. In 31 out of 50 cases, both arguments are lexical definites. In 12 cases the Recipient is an indefinite but all 12 sentences exhibit the double object construction. The Recipient is a pronoun in 3 cases but 2 of these cases appear in the caused-motion construction anyway. Notice, too, that the Theme *lie* never had a subsequent mention in any of its 50 occurrences, indicating its pragmatic inertness.

Finally consider the combination *give boost*. A random sample of 200 sentences contained 103 double object sentences, 53 caused-motion sentences, and 3 Theme-only sentences, where no Recipient was expressed.

	Theme	Recipient
	Double object	
N	103	103
Mean	2.94	3.03
Median	3.00	2.00
Mode	3	2
Std. Deviation	1.385	3.315
Skewness	3.290	3.068
Std. Error of Skewness	.238	.238
	Caused-motion	
N	53	53
Mean	2.55	4.19
Median	3.00	3.00
Mode	3 2	
Std. Deviation	1.136	3.453
Skewness	-.898	2.242
Std. Error of Skewness	.327	.327

Table 2.97: Length of arguments

The Theme *boost* had a subsequent mention in only 6 of 156 sentences (3.8%). In all cases there was just a single subsequent mention. In 5 cases this mention was in the next sentence, in 1 it was in the same sentence. Thus the metaphorical Theme *boost* on average has 0.038 subsequent mentions, which is clearly lower than the rate for concrete uses, shown in Table 2.83 above to be .52.

Table 2.97 summarizes the argument length data. It shows that in the double object construction, the length difference between the arguments is on average less than .1. There is a more pronounced difference in the caused-motion construction, with Recipients being longer than the Themes.

Table 2.98 shows the relative length of arguments for *give boost*. Note that the double arrows mean that the difference is 2 or greater.

The table indicates that there are a significant number of cases in which no prediction is made or in which it is violated. 75 out of a total 156 cases (48.1%) are not directly predicted by weight.

However, as Table 2.99 shows, Relative definiteness does not look like a much

	Caused-motion		Double object	
	n	%	n	%
T << R	22	41.5	[18]	17.5
T < R	7	13.2	[7]	6.8
T=R equal	(8)	15.1	(26)	25.2
T>R	[10]	18.9	30	29.1
T>>R	[6]	11.3	22	21.4
Total	53	100.0	103	100.0

Table 2.98: Relative length of arguments for *give boost*

	caused-motion	double object	Total
I=I	(20)	(21)	41
D=D	0	(2)	2
I>V	[1]	0	1
D>I	[26]	60	86
D>V	[6]	0	6
P>D	0	1	1
P>I	0	18	18
P>V	0	1	1
Total	53	103	156

Table 2.99: Definiteness of Recipient relative to Theme for *give boost*

better predictor. No prediction in terms of definiteness is made in 43 cases and the prediction of definiteness is violated in 33 cases. Thus, only $156-76=80$ cases (51.3%) are predicted correctly by relative definiteness alone.

Table 2.100 shows a cross-tabulation of relative length and relative definiteness. Note that the row labels for relative definiteness mention Theme before Recipient. The Table shows that there are quite a few cases where one or both factors' predictions are violated. In fact, there are 80 cases with either a violation or no prediction so that less than half of the tokens (48.7%) are predicted correctly.

Proportionally, the combination of factors works less well for the caused-motion construction than for the double object construction. Only 4 of 53 tokens (7.5%) have both a shorter and more definite Theme or have one of the two characteristics

	T <<R	T<R	T=R	T>R	T>>R	
Caused-motion						
I=I	3	1	(2)	[8]	[6]	20
V<I	[1]	0	0	0	0	1
I<D	[12]	[6]	[6]	*2*	0	26
V<D	[6]	0	0	0	0	6
Total	22	7	8	10	6	53
Double object						
I=I	[6]	0	(4)	7	4	21
D=D	0	0	0	1	1	2
I<D	[12]	[6]	22	15	5	60
D<P	0	0	0	0	1	1
I<P	0	0	0	7	11	18
V<P	0	[1]	0	0	0	1
Total	18	7	26	30	22	103

Table 2.100: Relative length versus relative definiteness

with the other feature's value being equal to that of the Recipient. In the case of the double object construction, there is no prediction for 4 tokens, and 25 tokens have a violation. None of the 25 has two violations and all violations are of the weight factor: the double object construction is used in spite of the Theme being shorter than the Recipient. By contrast, in the caused-motion construction, the violations in the left two columns (where the Theme is shorter than the Recipient) all honor weight over definiteness. Notice in particular the cells which combine the definiteness constellation I<D (where the Theme is indefinite and the Recipient definite) with either a Theme being much shorter or just shorter than the Recipient. It appears that basically, the alignment can go either way, as the two pairs of cells in question each account for 18 tokens. The table also shows 14 violations of weight where the caused-motion construction is used even though the Theme is longer and both noun phrases are indefinite.

We may speculate that the violations that occur in the caused-motion construction all serve to produce contiguity between *give* and the NP headed by *boost*. This would reflect the lexicalization of the sequence as a unitary concept. The violations in the

double object construction, by contrast, favor the alignment that definiteness predicts and that is more frequent overall in the language.

Of the four combinations that we have considered—*give lie*, *give boost*, *give chance*, and *give emphasis*—two seemed to be influenced mostly by weight, namely *give lie* and *give emphasis*. The patterns for *give chance* are accounted for well both by relative definiteness and by relative weight. The uses of *give boost*, by contrast, are not predicted particularly well by either factor. Interestingly, none of these four combinations is predicted well only by relative definiteness and I cannot think of any others for which that would be the case.

Modality So far we have looked at quite a bit of evidence arguing that support verb uses are different from concrete uses. This seems to conflict with the fact that we were able to build a quite good model for predicting constructional choice for the Switchboard data, which consisted of 242 tokens of which 159 were non-concrete. Is there reason to think that the non-concrete verb uses in the Switchboard differ significantly from those in the BNC? That does indeed seem to be the case.

We can note that 146 (91.8%) of the 159 non-concrete cases occur in the double object construction. This is a greater bias than occurs in the case of the non-concrete tokens occurring in the BNC. There, only 160 (73.7%) of 217 sentences exhibit the double object construction.

The non-concrete tokens in the Switchboard are clearly different from the ones in the BNC that were considered above when it comes to weight and definiteness as predictors of constructional choice. (cf. 164f.) The non-concrete uses in the Switchboard corpus are very well predicted by length. Table 2.101 shows that 136 (85.5%) out of 159 sentences are predicted directly by length. The non-concrete uses in the Switchboard corpus are also well predicted by relative definiteness. 138 (86.8%) of 159 cases follow the prediction of relative definiteness as shown by Table 2.102. And the combination of relative length and relative definiteness yields an even better prediction of constructional choice: for 143 (89.9%) of 159 sentences, the choice of linking construction is predicted correctly.

The central difference between the BNC and the Switchboard data seems to be

	cm	do	
R < T	[5]	130	135
R = T	(2)	(10)	12
R > T	6	[6]	12
Total	13	146	159

Table 2.101: Relative length for non-concrete uses in the Switchboard corpus

	cm	do	
I < P	1	0	1
I < D	0	[2]	2
I = I	(4)	(4)	8
D = D	0	(3)	3
D > I	(5)	6	11
P > D	(1)	32	33
P > I	(2)	99	101
Total	13	146	159

Table 2.102: Relative definiteness for non-concrete uses in the Switchboard corpus

this: the Switchboard corpus has a lot more pronominal Recipients than the BNC. In the former 134 (84.3%) of 159 sentences with non-concrete Themes have pronominal Recipients, while in the latter only 86 (39.6%) of 217 sentences have pronominal Recipients. There are thus many fewer sentences in the Switchboard, in which length and definiteness have a chance to conflict: pronouns are minimally short and high on the definiteness scale. Thus, the distinction between concrete and non-concrete tokens is significantly levelled in conversation as the support verb uses that are produced also favor the double object construction and do not produce conflicts between weight and definiteness.

Summary The non-concrete uses considered here are significantly different from concrete uses. Many of them occur a lot more frequently in one of the two linking constructions than the other. They basically also lack certain patterns that are attested with concrete tokens. While this could not be shown with the samples in the

	Caused-motion			
	R < T	R=T	R>T	
I<P	0	0	1	1
I=I	[1]	0	[3]	4
D>I	[2]	[2]	[1]	5
P>D	[1]	0	0	1
P>I	[1]	0	[1]	2
Total	5	2	6	13
	Double object			
	R < T	R=T	R>T	
I<D	[2]	0	0	2
I=I	2	(1)	[1]	4
D=D	3	0	0	3
D>I	5	1	0	6
P>D	32	0	0	32
P>I	86	8	5	99
Total	130	10	6	146

Table 2.103: Relative length by relative definiteness for non-concrete uses in the Switchboard

Lexical bias section, it is true for the 219 token sample of non-concrete uses taken from the British National corpus, that there was no pronominal Theme attested. In other words, although possible, one does not actually find cases like (2.155), in which the metaphorical Theme is a pronoun.

(2.155) Our team needed a boost and you gave it to them.

Some idiomatic combinations of *give* plus noun occur more frequently in the caused-motion construction than in the double object construction, while the situation is the reverse for others. Of the ones that we considered, the combinations that favor the caused-motion construction (*give lie*, *give emphasis*) seem to be motivated mostly by weight. Of the other two, *give chance* behaves more or less like the concrete uses; the linking construction used for its tokens typically matches both the predictions of weight and definiteness. With *give boost*, neither factor works very well. In some respects, there seems to be a trend to favor contiguity between the pieces of the expression but there are also cases where definiteness and/or high absolute frequency motivate the use of the double object construction.

The extent to which semantics influences the relative frequencies of the constructions is an interesting question. One might say that *give lie* and *give emphasis* are less readily conceived of as giving than *give boost* and *give chance*, which more often have sentient Recipients. The Recipients in these latter cases also can be said to have ‘gotten’ the relevant metaphorical Themes, which is not the case for *give lie* and *give emphasis*. One possible hypothesis then is that if the combination has a weak metaphorical connection to the core sense of giving it is more apt to be conceived of as a single unit, with contiguous pieces appearing in the caused-motion construction. But, clearly, more work is needed here and further discussion is beyond the scope of this work. For our local purposes, the main point is that non-concrete uses are significantly different from concrete uses and that mixed up with them they might obscure the relatively strong correlations between length, definiteness, and constructional choice. An exception of course are conversational contexts such as in the Switchboard corpus, where non-concrete uses frequently involve pronominal Recipients and thus rarely produce the conflicts between weight and definiteness that can be observed in writing.

2.5 Other lexical units: *deny* and *donate* in the BNC

In this section we will consider the lexical unit *deny*, which can also occur in the double object and caused-motion constructions. The goal is to see whether for this lexical unit, too, the choice between constructions is explained by the same kinds of factors as for *give*. We will also look at the lexical unit *donate*, which appears only in the caused-motion construction, to ask how well its uses in that construction match the uses of the construction that were observed for *give*, where it alternates with the double object construction.

	Frequency	Percent
deny_statement	609	76.1
deny_benefit	167	20.9
deny_score	22	2.8
other	2	.3
Total	800	100.0

Table 2.104: Uses of *deny*

2.5.1 Deny

The data for *deny* come from a random sample of 800 sentences taken from the British National corpus. The distribution of senses is shown in Table 2.104.

The sense that we will be interested in here is the one I call *deny_benefit*. It is illustrated in examples (2.156) and (2.157).

(2.156) Non-hipster **denied** entrance to ‘Lost in Translation’

(2.157) Members of the Election Commission in the polling station first **denied** the man the ballots for his wife.

The other two clearly identifiable senses occurring in the corpus, denying the truth of a statement and the idiomatic use of *deny* to talk about the thwarting of a scoring chance in sports, are illustrated in (2.158) and (2.159).

(2.158) He **denies** the allegation, and said he wanted an investigation to “clear” his name.

(2.159) Makaay almost sealed the points on 54 minutes but Weidenfeller **denied** him with a save which rivalled the earlier one from Kahn.

The scoring chance use of *deny* may not really be different from the benefit use, which may also cover the semi-idiomatic sequence *would not be denied* and active sentences such as (2.160).

	Double object	Caused-motion	
active	66	21	87
passive	51	11	62
Total	117	32	149

Table 2.105: Voice across constructions for *deny*

- (2.160) Ms. Simon-Harris maintains that he only requested visitation three times and that she **denied** him once, and only because she had made plans for Earen to see other relatives.

Note that these uses are special, though, in that they involve definite null instantiation of the (metaphorical) Theme.

We begin by noticing that overall the double object construction occurs 3.66 times more often than the caused-motion construction, which is about the same quotient as was found for *give*, where 163 uses of the double object construction and 45 of the caused-motion construction were found in the sample.

Let us now consider how active and passive voice pair with the double object and caused-motion constructions. Table 2.105 shows that the rate of passives overall is high: 62 (41.6%) of all uses of *deny* are in the passive voice. For the concrete uses of *give* in the BNC, we found a rate of 21.6% (45/208). With *deny*, passives are more common with the double object construction (43.6%) than the caused-motion construction (34.4%).

Closely related to voice is of course the factor subject choice, which is tabulated in 2.106. As expected, we find that choice of an argument other than the Donor as subject strongly predicts the linking construction used. Choosing the Theme as subject basically means use of the caused-motion construction, while choosing the Recipient means using double object linking. The three exceptional sentences in which the Theme is subject but the Recipient is realized ungoverned post-verbally all involve personal pronouns as Recipients.

When we consider the morphological realization of Recipients in the two constructions, we find a clear asymmetry: Recipients in the double object construction are

	Double object	Caused-motion	Total
Donor	66	21	87
Recipient	48	0	48
Theme	3	11	14
Total	117	32	149

Table 2.106: Subject choice for *deny*

more often definite than in the caused-motion construction. A similar difference exists with *give* but it is smaller: 90.8% of all Recipients in the double object constructions were found to be definite and in the caused-motion construction the percentage was still 71.1%.

The morphological realization of Theme arguments with *deny* also seems to be a bit different from that of *give*. While *give* has 62.2% definite Themes in the caused-motion construction—most of them 3rd person pronouns—and 31.3% definite Themes in the double object construction, *deny* has only 37.5% definite Themes in the caused-motion construction—few of these are 3rd person pronouns—and, surprisingly, more definite Themes, namely 52.1%, in the double object construction.

If we consider the length of Theme and Recipient NPs we find a clear difference in the double object construction where Themes tend to be clearly longer than Recipients. In the caused-motion construction, by comparison, the difference is very small.

If we tabulate the relative length combinations for the frame elements Theme and Recipient we see that in the double object construction the expected constellation of shorter Recipient is found more often than the expected constellation of shorter Theme is found in the caused-motion construction. Overall, we can predict the linking construction on the basis of relative length correctly for 99 (66.4%) of 149 sentences. This rate is comparable to what we found for concerted tokens of *give* (65.9%).

Next we will consider how well relative definiteness predicts the choice of linking construction. As Table 2.113 shows, only 73 (49.0%) of 149 cases are predicted correctly, which is significantly less than what we found for concrete uses of *give* (67.3%).

	Frequency	Percent
Possessed Noun	2	6.3
1st or 2nd	1	3.1
3rd	1	3.1
That	1	3.1
Definite NP	8	25.0
	13	40.6
Indefinite NP	1	3.1
Proper Noun	3	9.4
Quantified NP	4	12.5
Bare Noun	10	31.3
Restrictive relative	1	3.1
Total	32	100.0

Table 2.107: Recipient morphology with caused-motion uses of *deny*

	Frequency	Percent
Definite Zero	1	.9
Appositive zero	1	.9
Reflexive	7	6.0
1st or 2nd	10	8.5
3rd	24	20.5
That	1	.9
This NP	1	.9
That NP	1	.9
Proper Noun	15	12.8
Nonrestrictive relative	2	1.7
Possessed Noun	7	6.0
Definite NP	23	19.7
	93	79.5
Indefinite NP	5	4.3
Quantified NP	5	4.3
Bare Noun	11	9.4
Restrictive relative	1	.9
Reduced relative	1	.9
Negative determiner	1	.9
Total	117	100.0

Table 2.108: Recipient morphology with double object uses of *deny*

	Frequency	Percent
3rd	2	6.3
This NP	2	6.3
That NP	1	3.1
Possessed Noun	2	6.3
Definite NP	5	15.6
	12	37.5
Indefinite NP	2	6.3
Quantified NP	1	3.1
Bare Noun	13	40.6
Restrictive relative	1	3.1
Reduced relative	1	3.1
That-relative	2	6.3
Total	32	100.0

Table 2.109: Theme morphology with caused-motion uses of *deny*

	Frequency	Percent
3rd	2	1.7
That	1	.9
This NP	3	2.6
That NP	5	4.3
Proper Noun	1	.9
Possessed Noun	16	13.7
Definite NP	33	28.2
	61	52.1
Indefinite NP	19	16.2
Quantified NP	11	9.4
Bare Noun	16	13.7
Interrogative pronoun	1	.9
Restrictive relative	2	1.7
Reduced relative	1	.9
Negative determiner	1	.9
That-relative	5	4.3
Total	117	100.0

Table 2.110: Theme morphology with double object uses of *deny*

	double object		caused-motion	
	Recipient	Theme	Recipient	Theme
N Mean	1.89744	4.82906	2.96875	3.18750
Median	1.00000	4.00000	2.00000	1.50000
Mode	1.000	2.000	2.000	1.000
Std. Deviation	2.073580	4.196243	3.237725	4.775188

Table 2.111: Length of Themes and Recipients with *deny*

	double object		caused-motion	
	n	%	n	%
R>T	[16]	13.7	16	50.0
R=T	(18)	15.4	(7)	21.9
R<T	83	70.9	[9]	28.1
Total	117	100.0	32	100.0

Table 2.112: Longer frame element by construction for *deny*

	Double object	Caused-motion	Total
V<D	0	1	1
V<I	[2]	0	2
I<P	[1]	1	2
I<D	[8]	5	13
I=I	(11)	(7)	18
D=D	(28)	(4)	32
P=P	(1)	(1)	2
I>V	0	[2]	2
D>I	18	[9]	27
D>V	4	[1]	5
P>D	23	0	23
P>I	17	0	17
P>V	4	[1]	5
Total	117	32	149

Table 2.113: Relative definiteness for *deny*

	Double object			
	R<T	R=T	R>T	
V<I	[2]	0	0	2
I<P	0	[1]	0	1
I<D	[7]	[1]	0	8
I=I	7	(1)	3	11
D=D	20	(5)	3	28
P=P	0	(1)	0	1
D>I	12	3	[3]	18
D>V	0	2	[2]	4
P>D	21	1	[1]	23
P>I	14	3	0	17
P>V	0	0	[4]	4
Total	83	18	16	117
	Caused-motion			
V<D	[1]	0	0	1
I<P	0	0	1	1
I<D	[2]	1	2	5
I=I	[1]	(3)	3	7
D=D	[2]	(1)	1	4
P=P	0	(1)	0	1
I>V	0	0	[2]	2
D>I	*3*	[1]	[5]	9
D>V	0	0	[1]	1
P>V	0	0	[1]	1
Total	9	7	16	32

Table 2.114: Relative length by relative definiteness for *deny*

Next, we can tabulate relative length and relative definiteness to see if the combination predicts the correct linking construction for more antecedences than either factor does taken by itself. Table 2.114 shows that no improvement results over the number of cases that relative length by itself can handle correctly. The two factors together predict the correct linking construction for only 97 (65.1%) of 149 sentences.

Now that we have looked individually at the basic set of factors that we employed when looking at *give*, we can also perform a logistic regression on the *deny* data to evaluate how well all the factors combine to predict the choice of linking construction.

Observed			Predicted		
			Construction		Percentage Correct
			do	cm	
Step 1	Construction	Double object	115	2	98.3
		Caused-motion	12	20	62.5
	Overall Percentage				90.6

Table 2.115: Classification table for *deny* for model with relative length, definiteness status, and subject choice as factors

Since I did not code for Recipient focus for the *deny* data, that variable cannot be included.

The regression is a stepwise backward regression with likelihood ratio as a removal criterion. It consists of three blocks. The first contains the two length variables; the second the two variables for definiteness status, coded at four levels; the last contains only the categorical variable for subject choice.

The model predicts 90.6% of the data correctly. As in all cases we've seen before, the model makes more mistakes on the caused-motion construction than on the double object construction.

The initial -2 Log likelihood for the model is 155.020, the final value is 75.502. The Hosmer and Lemeshow goodness of fit tests come out as non-significant at each stage, indicating good fit between model and data. All variables that we have input into the model pass the regression test—with the exception of the variable for the definiteness status of the Theme: it can be removed without significantly changing the model's success rate at predicting the linking construction. While, Theme morphology seems to be less predictive in the case of *deny* than with *give*, we can note that the role of subject choice is much larger for *deny*: repeating the regression without the subject choice factor results in a drop of correct predictions to 81.9%, with predictions of caused-motion sentences being only correct in 25% of the cases. Subject choice thus raises the percentage of correct predictions by 8.7% in the case of *deny* while the

improvement was only 4.8% in the case of the concrete uses of *give*.

At all stages of the regression The boundaries for the 95% confidence interval for the $\exp(\beta)$ value for Recipient length are both greater than 1 indicating that longer Recipients increase the odds that the caused-motion construction is used. For Theme length the boundaries cross 1 at all stages of the regression which indicates that the relationship between Theme length and the use of the caused-motion construction found for this sample may not be true of the population as a whole. For Recipient definiteness, we find that lesser definiteness status than pronominal reliably increases the odds that the caused-motion construction occurs. However, this does not hold for the Recipients that are variables, that is, relativized on. (As we have seen variously, relativized-on Recipients frequently appear as subjects in passive voice clauses with double object linking.) Finally, choice of Theme as subject is found to reliably increase the odds of the caused-motion construction being used.

The model has two cases with high Cook's distance values and high standardized residuals. The first one, shown in (2.161) is due to the fact that weight is violated quite strongly. In its fuller context, the Recipient has several prior mentions whereas the Theme, as might be expected from its great length, does not. It seems to me that the use of the double object form would have been acceptable here.

(2.161) It misses the significance of the points: first, that the occasional engagement of trade unionism in the promotion of industrial co-operatives had been tactical and opportunistic, rather than principled; and second, that Co-operation identified as the Consumers' Movement had opposed any significant role for the industrial co-operative, properly so-called. The significance is that, together, the two operated to remove an essential feature from the full description of Co-operation; and in so doing to **deny** to the movement that function which it was uniquely able to discharge and which was, therefore, most likely to guarantee its growth and survival in perpetuity. That function was not the provision of wholesome food at fair prices. It was, and remains, authentic industrial democracy.

The second problem case involves two pronouns. Since the model had discarded

the variable for Theme definiteness—it did overall not contribute significantly to the model—it predicted here only based on the information that the Recipient was a pronoun, which in the grand scheme of things would favor the use of the double object construction. A greater amount of data in the model might prevent the discarding of the Theme definiteness variable from happening

(2.162) What could be more arrogant than to **deny** it to them?

Some representative examples of other cases that were not predicted correctly include the ones shown in (2.163) to (2.168).

(2.163) Enraptured by the celluloid screen, we revelled in the romance that had been **denied** to us, or had been put in cold storage “for the duration”.

(2.164) Many speakers highlighted the apparent paradox of a party which gave its blessing to separate internal organizations for both women and young people, but which **denied** the same privilege to its black members.

(2.165) If fully independent castellans such as those of Bemy and the Mâconnais remained a rarity, nevertheless the princes’ claim to control fortifications was at best spasmodically enforced, at worst absent in practice if recognized in theory. For the concession of hereditary tenure, though made piecemeal over a long period of time, was universal by the end of the century. Only rebellion could **deny** a castellan his inheritance; then, the punishment might be harsh.

(2.166) He [Eden; JR] was also partly, though not wholly, responsible for the “non-intervention” policy which Britain and France pursued with regard to the Spanish Civil War, even though it was being blatantly ignored by Germany, Italy and Russia, and though it **denied** a legitimate Republican government the right of access to the world market for arms with which to defend itself.

(2.167) There is no single way to measure creativity or industrial innovation; they are complicated things. One man’s breakthrough is often another’s trifling

adaptation. Spending is not an infallible guide to results or intentions. Some investment does not bear fruit. Other investment—in patents and licences, for example—may be intended to **deny** ideas to rivals rather than to search for useful products. Nevertheless, Dr Mowery and Dr Teece have varied sources that make their picture of Japan convincing.

- (2.168) Following the general election of March 22 (which aimed to return the country to civilian rule) a loose coalition of five pro-military parties headed by the air force-backed Samakkhi Tham had attempted to form a government [see p. 38816]. However, this coalition's proposed premier, the Samakkhi Tham leader Narong Wongwan, had then lost credibility after the publication of US State Department information about his suspected involvement in drug trafficking.

On March 26 the US State Department confirmed that it had in July 1991 **denied** a visa to Narong because of suspicions over his role in heroin trading. Initially Narong continued to receive the support of the five coalition members, but by April 3 reports from Bangkok claimed that the military had withdrawn their support.

- (2.169) Charles III's minister, Floridablanca, believed good government would solve the colonial problem; this was a profound misconception, inherited by Spanish liberalism, for it was not *mal gobierno* that the creoles rejected but government *peninsulares* (peninsular Spaniards) as such. The efficiency of government could only be increased at the expense of its popularity; the more corrupt and less efficient the government, the more acceptable it was to creole merchants and landowners. Charles III's reforms improved the quality of the imperial civil service while **denying** to the creoles a share in the system—perhaps because, as Floridablanca maintained, creoles were too enmeshed in local graft to be trusted.

Example (2.163) seems to have predicted as a double object construction because the sample contained 3 other cases where the Theme was relativized on in a passive

clause with double object linking. While this combination also occurred with *give*, it was not the more frequent combination relative to the combination of pronominal recipient and caused-motion construction. For instance, in the study on postnominal reduced relatives we found that of 38 cases where a Recipient is expressed, 37 appeared in the caused-motion construction and 1 in the double object construction. 6 of the 37 caused-motion cases had a pronominal Recipient and the sole double object case, of course, had a pronominal Recipient, too.

Example (2.164) involves focus on the Recipient, which is a feature not coded for *deny*. Clearly, it is a useful predictor, even if it comes into play rarely.

Example (2.165) illustrates two factors seen at play earlier. One is that if the Theme is anchored by the Recipient referent via a co-referential possessive, then the double object construction is to be preferred. The other is that although the Recipient is formally indefinite it is by no means a brand new referent. The context is a generic one and a representative generic castellan is clearly accessible in the context; castellans are also the topic of the paragraph as a whole. Clearly, inferable status would need to be represented in the model to allow for such cases to be handled correctly.

Example (2.166) involves a violation of definiteness in favor of weight. The example is also a bit similar to (2.165) in that the Recipient antecedes a reflexive pronoun contained in a postnominal modifier of the Theme. Further, the clause headed by *deny* constitutes a generic assessment of what the policy of non-intervention in the Spanish Civil War amounted to: the generic referent *a legitimate Republican government* is highly accessible given the immediately preceding discussion of the actual world state.

In example (2.167), both NPs are indefinite and equally long. It is probably a combination of the shortness of the Recipient—recall that Theme morphology was eliminated from the model—that led the model to guess use of the double object construction. However, in the particular case the actual construction seems quite appropriate. The Theme *ideas* seems slightly more accessible in this paragraph as prior reference is made to innovation, creativity, and breakthroughs. The idea of business rivals by contrast has not been recently evoked. To capture cases like this as part of the model would require a sophisticated representation of inferables, based

on a lot of data.

Example (2.168) constitutes a violation both of length and of definiteness. is a case that I think could also have had the opposite alignment, although in that case it might be preferable to move the Time phrase *in 1991* towards the end of the VP, as in example (2.170).

(2.170) On March 26 the US State Department confirmed that it had **denied** Narong a visa in July 1991 because of suspicions over his role in heroin trading.

One possible consideration that may have motivated the use of the caused-motion construction is the fact that the sentence occurs paragraph-initially: there may be a kind of contextual re-set effect that erases the referent familiarity built up in the previous paragraph. On the other hand, that familiarity is not so great to begin with: in the previous paragraph the referent Narong is introduced only in an appositive phrase and the reader might not even expect to hear more about the referent given the actual content of the paragraph: Narong is described as having lost credibility, which may lead the reader to infer that he may play no role in the formation of the Thai government.

The final example (2.169) is particularly interesting. Here, Heavy NP shift combines with the caused-motion construction to produces the linear arrangement of the simple double object construction. Thus, in the linear arrangement, both relative weight and relative definiteness are honored. However, from the point of view of the linking construction alone, relative definiteness is violated as we should expect the Theme to be equally or more definite than the Recipient in the caused-motion construction.

One plausible explanation for why this set of constructions might be combined the way it is would be focus on the Recipient, which would favor use of the caused-motion construction. That motivation does, however, not seem to apply here. There is a weak contextual presupposition that the reform will deny a share to somebody since there is explicit mention that the reform will involve increasing efficiency at the cost of popularity. But at the same time it is highly inferable that the group with

whom the reform will be least popular are the creole landowners and merchants. As that group also does not contrast with another, there is no good informational reason for presenting the referent *the creoles* as focal. And, to me at least, the use of the double object construction here would have been acceptable, too.

Although the overall success rate for predicting linking constructions was slightly lower for *deny* than for *give*, the examination of data for *deny* suggests that basically the same factors are at work with that lexical unit. The relevant factors may however have different weights in predicting constructional choice. One example of this is that subject choice was found to more important as the rate of passives was higher for *deny* than for *give*. This particular difference would not have been obvious from comparing the proportions of caused-motion and double object constructions as *give* and *grant* both have about a 3.6:1 ratio of double object to caused-motion sentences. Inspection of sentences for which the linking construction was predicted incorrectly, also suggests that the same kinds of additional considerations that we noticed as missing from our other models also would improve the one for *deny*: encoding cases of focus on the Recipient; having a representation of inferables, especially partitive/subset inferables; having more data so that certain idiomatic subregularities—for instance, the possibility of combining a Theme subject in the passive with double object linking as long as the Recipient is a pronoun—could be learned. Finally, we also saw cases with *deny* in which the occurring token does not seem to be clearly better than its alternative.

2.5.2 Donate

Finally, let us briefly consider a lexical unit that does not alternate but only combines with the caused-motion construction. The interest here is to see whether the occurring tokens are similar to the caused-motion tokens of alternating verbs. The lexical unit we are considering here is the verb *donate*. Our sample contains 185 randomly selected tokens, which were vetted so they exclude cases of organ donation.

The semantics of *donate* is different from that of the concrete uses of *give* in that charitable donation does not require a Recipient; rather than a specific individual or organization, the Cause that they pursue may be referred to. Recipients may also

be definitely null instantiated, which is not possible for concrete giving. Most of the Recipients are organizations, including countries, (80.5%) rather than individuals.

In our sample, 26 sentences had an overt Cause as second argument in addition to a Donor and a Theme. Of the remaining 159 sentences, 79 lacked an overt Recipient. In 77 cases, the Recipient was omitted under definite null instantiation; in 2 cases, the Recipient NP was zero in a generic sentence and had no definite referent to retrieve. If we treat the cases of null instantiation as part of the caused-motion construction, we find a clearly unusual picture: 89.9% of Recipients are definite. In the case of concrete uses of *give*, the rate was 71.1%, which is significantly lower.

If we consider only those cases as part of the caused-motion construction in which there is an overt Recipient, as in Table 2.117, we still get a very high rate of definite Recipients, 82.5%.

Tables 2.118 and 2.119 show that the length asymmetry that we expect for Recipients and Themes in the caused-motion construction only holds when we exclude the cases of zero Recipients. With zero recipients included, Recipients would on average be shorter than Themes.

The same difference also shows up when we tabulate the relative length of the Recipients and Themes. When zero Recipients are included, we find that they typically are shorter or equally long as Themes, as shown in Table 2.120). When zero Recipients are excluded, Recipients are by a slim margin more likely than not to be longer than Themes, which can be seen from Table 2.121. The rate of 50.6% is not much lower than what we found for Recipients with *give* (53.3%). With concrete uses of *give*, however, there were fewer cases where the Recipient was actually shorter than the Theme (13.3%) than in the case of *donate* (40.5%). Length is thus violated more often with *donate* than with *give*.

In Table 2.116 above, we can also note that the rate of definite Themes is much lower for *donate* than for concrete uses of *give*: 37.8% versus 62.2%. The reason seems to be that donated Themes are typically less well individuated referents, especially sums of money. Themes are not brand-new or variables in 92 (49.7%) of 185 cases

	Theme		Recipient		Cause	
	n	%	n	%	n	%
Definite Zero	5	2.7	77	48.4	12	46.2
Supplementary relative	13	7.0	1	.6	0	0.0
Non-restrictive relative	2	1.1	0	0.0	0	0.0
This	1	.5	0	0.0	0	0.0
This NP	2	1.1	2	1.3	2	7.7
That NP	1	.5	0	0.0	0	0.0
3rd	6	3.2	2	1.3	1	3.8
Possessed NP	9	4.9	5	3.1	2	7.7
Proper Noun	1	.5	37	23.3	0	0.0
Definite NP	30	16.2	19	11.9	3	11.5
	70	37.8	143	89.9	20	76.9
Quantified NP	42	22.7	2	1.3	0	0.0
Bare Noun	22	11.9	7	4.4	1	3.8
Indefinite NP	12	6.5	4	2.5	1	3.8
Restrictive relative	2	1.1	0	0.0	0	0.0
Reduced relative	30	16.2	1	.6	0	0.0
Negative pronoun	1	.5	0	0.0	0	0.0
Indefinite Zero	2	1.1	2	1.3	0	0.0
That-relative	4	2.2	0	0.0	0	0.0
VP _{to}	0	0.0	0	0.0	3	11.5
Adverbial clause	0	0.0	0	0.0	1	3.8
Total	185	100.0	159	100.0	26	100.0

Table 2.116: Morphology of frame elements with *donate*

	Frequency	Percent
Supplementary relative	1	1.3
3rd	2	2.5
This NP	2	2.5
Possessed NP	5	6.3
Proper Name	37	46.3
Definite NP	19	23.8
	66	82.5%
Indefinite NP	4	5.0
Quantified NP	2	2.5
Bare NP	7	8.8
Reduced relative	1	1.3
Total	80	100.0

Table 2.117: Morphological realization of non-zero Recipients with *donate*

N	Theme	Cause
	185	26
Mean	2.6162	2.9231
Median	2.0000	2.0000
Mode	.00	.00
Std. Deviation	3.02862	3.93876
Skewness	1.884	1.347
Std. Error of Skewness	.179	.456

Table 2.118: Length for Causes and Theme arguments with *donate*

	All Recipients	Non-zero Recipients
N	159	79
Mean	2.0818	4.1899
Median	.0000	3.0000
Mode	.00	2.00
Std. Deviation	3.38644	3.77942
Skewness	2.851	2.488
Std. Error of Skewness	.192	.271

Table 2.119: Recipient length with *donate*

	Theme		Recipient		Cause	
	Shorter	48	25.9	77	48.4	11
Equal	49	26.5	42	26.4	7	26.9
Longer	88	47.6	40	25.2	8	30.8
Total	185	100.0	159	100.0	26	100.0

Table 2.120: Relative lengths for frame elements of *donate*

	n	%
Shorter	32	40.5
Equal	7	8.9
Longer	40	50.6
Total	79	100.0

Table 2.121: Relative length for non-zero Recipients for *donate*

	Theme		Recipient		Cause	
	n	%	n	%	n	%
Variable	38	20.5	0	0.0	0	0.0
Brand new	55	29.7	13	8.2	7	26.9
Unused	1	.5	13	8.2	0	0.0
Inferrable	52	28.1	31	19.5	8	30.8
Evoked	36	19.5	97	61.0	10	38.5
Situationally given	3	1.6	5	3.1	1	3.8
Total	185	100.0	159	100	26	100.0

Table 2.122: Prior evocation status with *donate*

with *donate*; with *give*, that is the case in 28 (62.2%) of 45 cases.²⁰ This difference is not statistically significant according to a χ^2 -test ($\chi^2=2.462$, $df=1$, $p=0.1166$). Recipients overall are not brand-new or variables in 91.8% of all cases. When only non-zero Recipients are considered, the rate is a bit lower at 85.0% (68/80) but still higher than the rate for Recipients in the caused-motion construction for *give*, where 31 out of 45 Recipients are not brand-new or variables (68.8%). This latter difference is statistically significant according to a χ^2 -test ($\chi^2=4.538$, $df=1$, $p=0.0332$).

For subsequent evocations, we find that about 4 out of 10 Recipients have a subsequent evocation while that is true for only 1 out of 8 Themes. In the case of caused-motion uses of *give*, 24 of 45 (53.3%) Recipients had a subsequent evocation and 9 of 45 Themes (20.0%). For neither frame element is the difference significant according to a χ^2 -test.

Finally, we turn to the contextual mention properties of the arguments of *donate*. Table 2.124 shows that with *donate*, too, Themes have fewer prior mentions than Recipients. The mean for Recipients is a little bit higher than for Recipients in the caused-motion construction with *give* (1.4667) and the mean for Themes is a little bit lower than for Themes with *give* in the caused-motion construction (1.0444).

With *donate*, the average distance to the closest prior mention for all Recipients is lower than with *give* in the caused-motion construction (7.3333) but for Themes it

²⁰Recall that here the term 'variable' refers to being relativized on.

	Theme		Recipient		Cause	
	n	%	n	%	n	%
Not evoked	150	81.1	92	57.9	21	80.8
Variable	1	.5	0	0.0	0	0.0
Inferrable	11	5.9	4	2.5	1	15.4
Evoked	23	12.4	63	39.6	4	3.8
Total	185	100.0	159	100	26	100.0

Table 2.123: Subsequent evocation status with *donate*

	Theme	Recipient	Cause
N	185	159	26
Mean	.4432	1.8679	.6538
Median	.0000	1.0000	.0000
Mode	.00	.00	.00
Std. Deviation	1.51380	2.99602	.97744
Skewness	5.498	2.675	1.897
Std. Error of Skewness	.179	.192	.456

Table 2.124: Number of prior mentions for *donate*

	Theme	Recipient	Cause
N	185	159	26
Mean	9.0216	5.4654	7.2692
Median	11.0000	4.0000	11.0000
Mode	11.00	11.00	11.00
Std. Deviation	4.08996	4.81924	4.78797
Skewness	-1.623	.166	-.609
Std. Error of Skewness	.179	.192	.456

Table 2.125: Distance to closest prior mention for *donate*

is higher than for *give* in the caused-motion construction (5.8667).

The properties of subsequent mentions are summarized in Table 2.126. Here we see that for Recipients the number of subsequent mentions is lower than for Recipients with *give* in the caused-motion construction (2.333). Themes have very few subsequent mentions just as in the case of *give* used in the caused-motion construction (0.4444). For both Themes and Recipients, the distance to the next subsequent mention is greater than is the case with caused-motion uses of *give*, where we find 8.6000 and 5.1556, respectively. The last subsequent mention for both Themes and Recipients is closer for *donate* than for caused-motion uses of *give* where we find 2.4222 sentences for Recipients and 1.0222 for Themes.

Based on the data on *donate* that we have considered here we can make the following observations. There seem to occur more violations of relative length with *donate* than with *give*. With *donate* the Recipients may be slightly more accessible from prior context than those found with *give*. In terms of most measures for subsequent mentions, both Recipients and Themes seem a bit less topical than their counterparts with *give*. The most interesting observation is that overt pronominal Recipients hardly occur. That no 1st and 2nd person Recipients occur is no surprise : donations are mostly made to organizations; the corpus consists largely of writing; and the basic and more frequent lexical unit *give* also has a *donate*-sense, in which it also allows definite null instantiation of Recipients. What is surprising is that so few overt 3rd person pronouns appear as Recipients. There seems to be a real preference against

	Cause	Recipient	Theme
	Number of subsequent mentions		
N	26	159	185
Mean	.8077	1.3585	.3351
Median	.0000	.0000	.0000
Mode	.00	.00	.00
Std. Deviation	2.19124	2.67024	.84449
Skewness	3.428	3.103	3.130
Std. Error of Skewness	.456	.192	.179
	Distance to next subsequent mention		
N	26	159	185
Mean	9.0769	7.2075	9.1189
Median	11.0000	11.0000	11.0000
Mode	11.00	11.00	11.00
Std. Deviation	4.02912	4.67971	3.94761
Skewness	-1.679	-.518	-1.662
Std. Error of Skewness	.456	.192	.179
	Distance last mention		
N	26	159	185
Mean	1.0385	1.8805	.3297
Median	.0000	.0000	.0000
Mode	.00	.00	.00
Std. Deviation	2.55313	3.20870	1.03967
Skewness	2.594	1.587	4.437
Std. Error of Skewness	.456	.192	.179

Table 2.126: Properties of subsequent mentions for *donate*

expressing given Recipient referents pronominally in the caused-motion construction. It is true that pronouns do occur as Recipients in the caused-motion construction with *give* but with that lexical unit (in its non-charity uses) there is no possibility for null instantiation of the Recipient and a lot of the cases also involve a pronominal Theme, triggering the caused-motion argument alignment as a default. The preference against pronominal Recipients in the caused-motion construction seems to be quite general. Other lexical units data for which I drew and analyzed random samples (*provide*, *supply*, and *bequeath*) confirm this trend. As an example, the data for *bequeath* are given in 2.127. The asymmetry observed with *bequeath* in the share of pronominal referents across the constructions is significant according to a χ^2 -test ($\chi^2=15.249$, $df=1$, $p\leq 0.000$).

Thus, it seems that for lexical units that participate in alternating constructions speakers avoid the use of pronominal Recipients in the caused-motion construction and use an alternative instead. Note that the alternative is not always the double object construction. With *provide* the main alternative to the caused-motion construction is the *NP PP_{with}* pattern. (In addition *provide* also allows Recipient definite null instantiation.)

It is not clear what motivates the preference against expressing pronominal Recipients in the caused-motion construction. It may be that given Recipients should only exceptionally be displaced from their primary object slot—as when the Theme is also pronominal or when the Recipients is focused on. An alternative explanation would be that this is part of a wider generalization involving caused-motion independently of whether there is a Goal that can act as a sentient Recipient. Some suggestive evidence in this direction comes from simple web searches. For instance, while the word sequence *put it in* is matched 1,610 million times by the Google search engine, the sequence *put it in it* occurs only 7,300 times. Difference of similar magnitude can also be found for other verb-preposition combinations such as *stick it*, *sprinkle it*, *lay it*. I will leave this question open for further research.

	Theme+Cause		caused-motion		double object	
	n	%	n	%	n	%
Definite zero	3	27.3	1	1.1	0	0
1st or 2nd	0	0.0	1	1.1	7	43.8
3rd	0	0.0	7	7.9	4	25.0
Non-restrictive relative	0	0.0	1	1.1	0	0
That NP	0	0.0	1	1.1	0	0
Possessed NP	1	9.1	12	13.5	1	6.3
Proper Noun	0	0.0	36	40.4	2	12.5
Definite NP	0	0.0	22	24.7	1	6.3
	4	36.4	81	91.0	15	93.8
Quantified NP	0	0.0	1	1.1	0	0
Bare NP	1	9.1	2	2.2	0	0
Indefinite NP	0	0.0	3	3.4	0	0
Interrogative pronoun	0	0.0	0	0.0	1	6.3
Restrictive relative	0	0.0	2	2.2	0	0
Indefinite zero	0	0.0	0	0.0	0	0
To-marked infinitive	6	54.5	0	0.0	0	0.0
Total	11	100.0	89	100.0	16	100.00

Table 2.127: Form of Recipient/Cause frame elements for *bequeath*

2.6 Summary

In this chapter, we have seen that the distribution of the linking constructions alternating in the dative alternation can reasonably well be approximated in terms of weight, discourse status (either as cued by morphology or as cued by contextual mentions), and subject choice. The latter two are basically pragmatic in nature while the first seems to have to do with processing. A clearly pragmatic factor that was also found to be relevant in the Switchboard corpus, even though it was coded only in the clearest cases, was focus. Importantly, it was argued that focus needs to be distinguished as a factor from evocation status. Both are independently needed, in general and to talk about the dative alternation data.

We found that the data in the Switchboard corpus of telephone conversations were modelled slightly better (96.3%) than the mostly written data from the BNC (92.3%). However, for the BNC data I had not coded for Recipient focus so the comparison is not quite fair. Having that additional variable may well improve the model. All the models that we built also illustrated that it is generally easier to get the prediction for the use of the double object construction right than that for the use of the caused-motion construction. In terms of frequency and the alignment of pragmatic properties, the double object construction is unmarked. The caused-motion construction seems to be used under special circumstances such as exceptionally high local topicality of the Theme; focus on the Recipient; a significant weight asymmetry.

Both corpora do contain uses that a well-edited text may not and the failure to get a greater percentage right is to a small degree explained by the fact that rare but stark violations of otherwise predictive factors cannot be incorporated into the logistic regression model, unless we find additional factors that motivate the exceptions.

We also considered at some length the properties of non-concrete uses of *give*. The basic finding was that relative definiteness is less predictive in *give* + Noun combinations where either the noun is the semantic head (*give approval, give boost*) or which are fully idiomatic (*give lie*). Much of the choice between the constructions is governed by weight. With some combinations like *give boost*, there may be a considerable number of tokens in which either the prediction of weight or definiteness

is violated.

Although the models built here do describe over 90% of the data, they do not tell the whole story. Some combinations of form and discourse status are too rare in the corpus to have been found at all or in significant number in our samples. Recall the discussion of demonstrative Themes combining with pronominal Recipient. Likewise, we have not been able to adequately investigate the use of the double object and reverse double object constructions that are possible when both Theme and Recipient are personal pronouns. While the constructional choice for combinations of demonstrative Theme and pronominal Recipient seems to be governed mainly by the recency of mention of the Theme, it remains open whether the two rarer constellations that occur with two pronominal arguments are similarly governed by the pragmatic factors observed for NPs of all form types.

Although it is remarkable that the model works as well as it does just based on a morphology-based treatment of discourse status, we have also seen in places that an explicit treatment of inferables and their inference triggers might improve the model further. Missing from the model, too, are variables for differences between individual lexical units and between individual speakers. In the case of differences between lexical units, however, the examination of the *deny* data advises caution: comparing only the relative frequencies of the two alternative constructions without controlling for other factors may lead one to believe that no differences exist where the differences lie in the relative importance of the various factors. Conversely, though this could not be illustrated here, one may think that a difference in idiosyncratic preference for one construction over the other exists for a lexical unit which simply happens to participate in types of discourses that naturally result in the more frequent use of one of the alternatives. In other words, it is important to make sure that all else is really equal.

We also briefly looked at two additional lexical units, *deny* and *donate*, to see how well the observations that we made about the double object and the caused-motion construction hold up. For *deny* we found that the same kinds of factors seem to apply as to *give*. We further found that the factor subject choice plays a greater role with *deny* because more clauses appear in the passive. It also seemed to be the case that

in passive clauses, where there is no question of contiguous post-verbal alignment, the constraints on relative weight and definiteness may be more violable. The data for *donate* suggested that at least one of the properties that *give* exhibits in the caused-motion construction also seemed to carry over to the non-alternating *donate*. While Themes were less often definite than with *give* in the caused-motion construction, speakers do seem to avoid use of personal pronoun Recipients with *donate* by using definite null instantiation instead when the Recipient referents are highly given. It was not resolved whether this preference against pronominal Recipients is a local property of Recipients or whether it is part of a broader generalization encompassing Goal phrases.

2.7 Application of pragmatics to syntax

In the introduction I indicated that the discourse-pragmatic research on the dative alternation could be used to inform the syntactic and semantic analysis of that construction. Specifically, I claimed that the discourse-pragmatic characteristics of the variably linked arguments provide evidence that Basilico's (1998) analysis of the difference between the alternates in terms of a difference between syntactically layered focus structures does not account for all the data and should be re-cast in pragmatic terms other than thethetic-categorical distinction. The argument will be spelled out in greater length in this section.

Basilico's analysis ties together two lines of generative syntactic research, the VP-internal subject hypothesis and the VP-external object hypothesis. The first hypothesis changed the traditional analysis of subjects by proposing that some subjects are generated in a VP-internal position and then move to their surface subject position from there. The second hypothesis challenged the traditional analysis of objects by claiming that some objects are generated outside of the VP. More specifically, Basilico's analysis for objects assumes that finite clauses have two VPs, an inner one that is lower down in the tree and an upper one that is higher up. Some objects are generated in the inner VP and move higher to the specifier of a TransP (transitivity phrase) functional projection located between the two VPs; other objects are said to

be generated directly in the [spec, TransP] position.

Now the connection between subjects and objects is that Basilico assumes that objects are subjects of inner predications, where the inner predicate is 'defined by the verb and any indirect complements' (1998:542). The inner predications are said to come in two types paralleling the *thetic-categorical* distinction on the sentence-level, which is further identified with the stage-level/individual-level distinction. 'When an object occupies an internal position [internal to the VP; JR], this corresponds to a 'thetic predication' within the inner predication in which an assertion is being made as to the existence of an object or of an event involving the object. When an object occupies an external position, this corresponds to a 'categorical predication' within the inner predication, in which a property is being ascribed to the object' (ibid.). The syntactic structures that go along with the two predication types are presented in Figures 2.3 and 2.4 on page 207.

The semantic analysis in terms of different predication types is taken to be strengthened by the fact that certain syntactic and semantic distinctions between the alternates can be analyzed in terms of the two different object positions that were posited. Four kinds of asymmetries can be observed between the members of the dative alternation (and also between the alternates of the locative spray/load alternation and between the alternates of the creation/transformation alternation). In the double object construction, which is taken to involve categorical inner predication, the (categorical) first object's quantifier needs to take scope over any quantifier of the second object (2.171a); there is no such requirement for the direct object of the thetic caused-motion construction relative to the accompanying prepositional phrase (2.171b). If a frequency adverb is placed between the two post-verbal arguments, the first object in the categorical double object construction needs to take scope over the frequency adverb whereas in the thetic caused-motion construction the (direct) object need not outscope the frequency adverb (2.172a-b). *There*-existential sentences combining with the categorical double object construction allow only the interpretation that the existence of the post-copular NP's referent is asserted (2.173a), whereas *there*-existential sentences combining with the thetic caused-motion construction allow an eventive interpretation also (2.173b). Finally, inversion is acceptable only with the

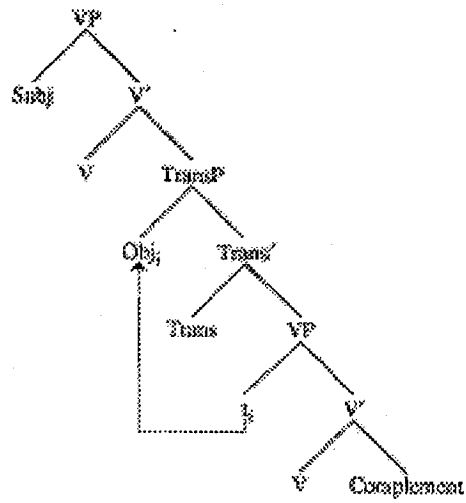


Figure 2.3: Thetic inner predication

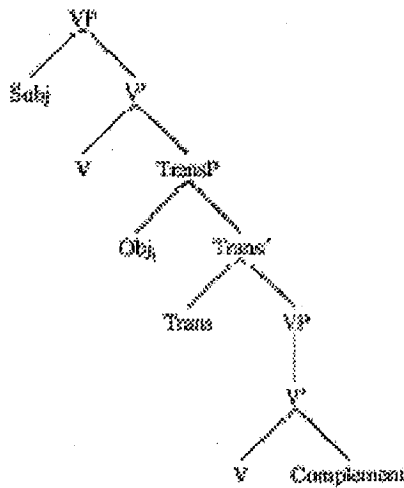


Figure 2.4: Categorical inner predication

thetic caused-motion construction (2.174a) but not with the categorical double object construction (2.174b).²¹

(2.171) Quantifier scope

- a. She **gave** a child every gift.
- b. She **gave** a gift to every child.

(2.172) Frequency adverb scope

- a. While on patrol, the policeman **gave** a speeding motorist occasionally a ticket.
- b. While on patrol, the policeman **gave** a ticket occasionally to a speeding motorist.

(2.173) *There*-insertion

- a. There were winners **given** awards.
- b. There were awards **given** to the winners.

(2.174) Inversion

- a. To those less fortunate than ourselves will be **given** a holiday basket and a free meal.
- b. *In the church will be **given** those less fortunate than ourselves a holiday basket.

Basilico discusses and explicitly rejects an account of the asymmetries between the alternates in terms of specificity of the referents of arguments. In such an account, certain constructional alternates would constrain certain of their argument positions by requiring NPs in those positions to denote specific entities. This kind of account

²¹For the compatibility test between the double object and the inversion constructions we have to assume that adjunct PPs can participate in locative inversion. Otherwise we couldn't apply this test.

could successfully handle the fact that the double-object sentence in (2.171a) is unambiguous with regard to the interpretation of the NP *a child* (one child only) while the caused-motion sentence in (2.171b) is ambiguous with regard to the relative scopes of the existential and universal quantifiers. However, as Basilico notes, it seems that objects in what are categorical constructional alternates according to the tests above may also be interpreted non-specifically.

(2.175) The sculptor **carved** some/sm slabs or marble into statues.

(2.176) The farmer **loaded** some/sm trucks with bales of hay.

(2.177) The angry voters **sent** some/sm congressmen a letter.

That categorical constructions can have nonspecific objects is unexpected under Basilico's account: exact parallelism between subjects and objects would predict that there should only be specifically interpreted VP-external objects in the categorical construction just as VP-external subjects of individual-level predicates require specific interpretation. Faced with this problem, Basilico argues that the unexpected availability of non-specific interpretations is due to other primitive predicates higher up in the tree—in many cases this an inchoative BECOME predicate—that may provide existential closure. The reason given for adopting this solution is the following. When one compares the transitive and the intransitive locative alternations in (2.178) and (2.179), it can be observed that both of them exhibit the four syntactic-semantic asymmetries illustrated above in (2.171)-(2.174). (This exercise is left to the reader.)

(2.178) Transitive locative alternation

- a. The farmer **loaded** a truck with every bale of hay. (categorical)
- b. The farmer **loaded** a bale of hay onto every truck. (thetic)

(2.179) Intransitive locative alternation

- a. The garden **swarmed** with bees. (categorical)
- b. Bees **swarmed** in the garden. (thetic)

Given that both alternations exhibit the set of asymmetries that is taken to be indicative of the difference betweenthetic versus categorical inner predications, the reason for the specificity difference must lie somewhere else. The prime candidate for the locus of difference is the difference in semantic structure: the transitive verbs include a BECOME-predicate that the intransitive verbs lack. Basilico suggests '[t]he presence of BECOME licenses and additional unselective existential operator above and beyond that sanctioned by the VP structure itself, and this allows for NPs within its scope to undergo existential closure' (p. 587).

On a technical level, Basilico's account for the unexpected non-specific readings of objects occurring with verbs that have a categorical inner predication seems to work. However, it does have problems. It is actually incomplete since it is not spelled out when the existential closure applies at the level of LF and when it does not.

Further, the augmentation of the basic account by appealing to additional existential binders in the lexical representations of verbs with categorical inner predications introduces an unforeseen problem. There are other categorical/individual-level intransitive predicates such as *turn* that one would expect to include a BECOME-operator within their lexical representation but which do not allow non-specific readings of their subjects.

(2.180) A frog **turned** into a prince.

The lack of a non-specific reading for the indefinite subject NP of (2.180) is not accounted for by Basilico.

Another argument against adopting athetic-categorical inner predication analysis for the dative alternation is the following: if the caused-motion variant is indeed thethetic member of the dative alternation, then it is at least somewhat suspicious that that construction would have to be used whenever the Theme is a pronoun. It is not by itself a problem for a topical NP to occur in athetic predication as is shown by sentences like *It's ME* and *That's YOU*; the possibility for such combinations indeed follows from Lambrecht's separation of the topic role from a particular activation status. However, what is unexpected and cannot be derived at all from Basilico's account is that pronominal Themes cannot occur in the categorical double object

construction. No pragmatic theory I am aware of would prohibit topical non-subject arguments in categorical predications.

Another problem of the account given is that the thetic construction can occur when there is clearly no issue of introducing a referent, or an unanalyzed event apprehended whole:

(2.181) Does Michael have my copy of the New Yorker? Did you **give** it to him?

Also, the thetic-categorical distinction does not constitute the complete set of focus structure types. Argument focus sentences such as B's answer in (2.182) are neither thetic nor categorical. Basilico's account does not explain how the constructions should behave in argument focus contexts.

(2.182) A: Who did you give *The unequalled self* to?—B: I **gave** that book to Sue.

Finally, Basilico argues that specificity is too strong an assumption for the objects occurring in categorical constructions. To the degree that speakers share the judgement that sentence (2.177) is just as acceptable on the reading with weak *sm* as sentences (2.175) and (2.176), that claim is justified. To me, however, it seems that the example of the double object construction, (2.177), is actually worse than the others. There is also evidence from comparing the sequences *give anything to anybody* and *give anybody anything* that even when both arguments seem to be non-specific, the choice between the constructions is still sensitive to such pragmatic factors as the closeness and number of expressions that link the non-specific expressions to context (cf. section 2.2.6). Consider examples (2.183)-(2.186).

(2.183) You don't have to give anybody anything to come into this county, and you don't have to **give** anybody anything if they threaten to leave.

(2.184) A millionaire can donate thousands. How can a have-not **give** anything to anybody?

(2.185) You **give** *anybody* anything for free and *they* get lazy and pretty soon start expecting charity.

- (2.186) We are going to instill in the minds of the people as far as possible that statement by Brigham Young. . . to the effect that it was his policy not to **give anybody anything** unless *he* earned *it*; that people must do something to earn that which they receive.

While in cases such as (2.183) it may appear that only rhetorical parallelism is involved, examples like (2.184), (2.185), and (2.186) are less readily explained away in that way. In the case of (2.184), the relevant prior clause has no overt Recipient. Clearer still are (2.185) and (2.186). I would contend that in each case the double object construction is chosen because it sets up the correct expectation(s) of (relative) topicality and topic continuity.

Overall, the corpus data cast considerable doubt on Basilico's treatment of the difference between the two members of the dative alternation in terms of focus structure. It is at least incomplete and it might be possible to replace it altogether by an account in terms of a constraint that the Recipient has to bear the topic pragmatic relation to the proposition. Notice that under the Lambrechtian view of information structure adopted here a sentence can have multiple topics so the subject of a double object sentence could still also be a topic. Further, since pragmatic relations are not tied to activation status, cases like (2.185) are no problem for the theory: in such cases, the use of the double object construction topicalizes the referent and sets up an expectation for topic continuity. An account in terms of assuming topic-status for Recipients in the double object construction would also handle the observed asymmetries rather well if we assume that topics that are pre-specified as such by argument structure constructions cannot be out-scoped either by quantifiers or frequency adverbs. Further, topics in general do not fit into presentational *there* or inversion constructions. Finally, an analysis in terms of attributing topic status to Recipients in the double object construction also predicts their strong dispreference for, if not outright incompatibility with being focal.

Now it might also be possible to cast the analysis in terms of a specificity requirement on Recipients in the double object construction insofar as the examples involving weak *some* in (2.175-2.177) do not involve nonspecific arguments after all.

In other words, on one understanding of specificity, being weakly quantified is not the same as being nonspecific. Nonspecific arguments do not refer to individuals, that is, non-specific arguments do not have wide scope of the existential quantifier. However, the arguments in (2.175-2.177) clearly do have wide scope and are thus specific. (The 'marble' case is a possible exception but of course it refers to a mass type.)

It is difficult to evaluate whether an analysis in terms of specificity or one in terms of topic status is what is needed since topics are specific. One of the problems in trying to resolve this question is that the inversion construction does not seem to ever be possible with the double object construction. We can, however, note that at least some inversion constructions seem to allow specific referents.

(2.187) Yes, this is no ordinary general election.

'Evans is a Democrat; Daley is a Democrat. Different Democrats have different points of view about the city of Chicago and its politics,' Jackson notes. 'The war between forces within the party continues, and within our coalition.'

Standing in the middle of it all is Jesse Jackson. (Birner & Ward 1998:164-165, (210); emphasis in original)

The referent *Jesse Jackson* is specific in example (2.187). It is not an established topic at that point in the discourse following the discussion of various Democratic actors in Chicago politics.

A second problem in comparing a possible account in terms of topic status and one in terms of specificity is that in *there*-existential sentences combining with the double object construction what follows the copula is a single NP containing a reduced relative. There is thus no act of reference made that would just be based on the head of the NP and any pre-modifiers and determiners. The NP as a whole is specific in the sense spelled out above of having wide scope of the existential quantifier.

I will leave it open to future research which is the more appropriate account. This should also involve careful investigation of the other alternations that Basilico has examined. The three sets of alternations seem to not in fact be as parallel as Basilico's analysis suggested. Recall that in the dative alternation pronominal Themes

were found to require the use of the presumablythetic caused-motion construction. Now when we consider the creation/transformation alternation, we find that for the categorical member, the transformation-construction involving *into*, there exists also a very strong tendency for the second or non-object post-verbal complement to not be pronominal. However, though rare, there are exceptions such as *She usually wasn't the rambling type, but he had **turned** her into it*. Corpus study of the other constructions might turn up additional differences that need to be taken into account before deciding how much parallelism to assume between the constructions and how to account for shared behaviors.

Chapter 3

Raising constructions

In this chapter I explore the question whether raising alternations exhibit a correlation between the choice of construction and the discourse pragmatic properties of the referents and whether they might in fact be fully conditioned by discourse properties rather than semantics. I consider three different raising constructions in this chapter. The first is the well known one involving such verbs as *seem* and *appear*, which I will call epistemic raising alternation. The second is the body part possessor raising alternation involving verbs such as *smack*, *touch*, and *hit*. The final type of raising alternation I consider is the change property alternation involving predicates such as *fall*, *rise*, and *drop*.

3.1 Epistemic raising

The central raising alternation involves two valence patterns where one valence includes a propositional, clausal element, frequently extraposed, and the other valence includes two elements, corresponding to the subject and predicate of the alternative pattern. The former pattern, which I will call the unraised pattern, is exemplified by (3.1a) and the latter pattern, which I will call the raised pattern, by (3.1b).

- (3.1) a. It is **likely** that John will leave
 b. John is **likely** to leave

In the generative tradition the interest in the epistemic raising alternation came strictly from the syntactic side, in particular case theory. The difference between the two constructions was never adequately addressed. It was tacitly assumed that raising predicates had the same meaning in both constructions and that the sentences were truth-conditionally equivalent, though Postal (1974) allowed that certain raising uses involved assumptions that 'are not part of the core meanings of sentences' (as cited by Langacker 1995: 6). Later generative analyses were only concerned with generating the right surface forms and keeping the analyses consistent with their theories of case and movement. The question why one construction was used instead of the other was not addressed.

3.1.1 Langacker's active zone analysis

Langacker (1995) offers a Cognitive Grammar analysis of raising constructions:

- (3.2)
- a. She **expected** a specialist to examine her mother (subject-to-object raising)
 - b. David is **likely** to criticize this plan (subject-to-subject raising)
 - c. This plan is **easy** for David to criticize (object-to-subject raising; *tough*-movement)

Under Langacker's analysis, raised and unraised sentences 'are not derived from the same underlying structure, nor one from the other. They instantiate separate and parallel constructions, each representing its own way of construing and symbolizing situations that may in some cases be the same' (Langacker 1995:36). Following Borkin (1973), Langacker presents the following set of sentences as evidence for the non-equivalence of raised and unraised sentences.

- (3.3)
- a. I **find** that this chair is uncomfortable.
 - b. I **find** this chair to be uncomfortable.
 - c. I **find** this chair uncomfortable. (= (15a-c), p. 5)

The sentence in (3.3a) can be used on the basis of indirect evidence for the finding. Sentence (3.3b) seems to require more direct evidence and sentence (3.3c) seems to imply that the speaker has actually sat in the said chair and derives their finding from first-hand experience.

According to Langacker, '[T]he semantic non-equivalence of raising sentences and their counterparts is attributable to three factors: the meanings of the grammatical elements they contain (e.g. *that, to, will, -ing*); the reference-point function of the "raised" NP; and the highlighting effects of focal prominence (trajector/landmark status)' (Langacker 1995:36).

One of the crucial elements of Langacker's analysis, and the one that is involved in all three of the raising types illustrated in (3.2), is the concept of *reference point*, which refers to 'our basic cognitive ability to invoke one entity as a conceptual reference point for purposes of establishing mental contact with another' (Langacker 1995:27). The use of reference points is said to be independently motivated and, indeed, pervasive in grammar. Metonymy is said to be an instance of the concept of reference points also (e.g., **That car** *doesn't know where he's going*). A subtype of metonymy are *active zones*. For instance, as Langacker points out, *the cigarette in her mouth* does not refer to a situation where a cigarette is completely inside a person's mouth. Rather, it involves a part of the cigarette (the non-burning end) and a part of the mouth (the lips). The cigarette butt and the lips are the active zones of the referents of the arguments of *in*. With an awareness of active zones, raising constructions are characterized as follows.

In raising constructions a raised NP is a reference point that facilitates mental contact with the active zone of the main predicate, the event or process whose status as real or likely is assessed. For instance, in (3.2b), *David* is the reference point that allows the hearer to access the process denoted by (*criticize this plan*) more easily. Importantly, Langacker argues that '[T]he "raised" NP can be thought of as a kind of local topic, i.e. as a topic for purposes of ascertaining the actual (or direct) participant in the profiled main-clause relationship ...' (Langacker 1995:38). This is said to be true for all the raising constructions. Note that for Langacker the fact that *it* and *there* (as well as certain idiom chunks) can appear as subjects in

the raised construction is no problem. While for many linguists, *it* and *there* are non-referring expressions, so-called expletives, Langacker says “they should probably be thought of as abstract presentational frames serving to announce the intention of subsequently introducing something in those frames” and as settings they should be natural reference points for the elements in the presentational frames they introduce (1995:42).

3.1.2 French raising and topicality

Achard (2000) studied raised and unraised sentences in a corpus of sentences containing the French raising verb *sembler* ‘seem’. He argues that “each construction’s meaning strongly predicts its distribution in text, and that the results of the corpus analysis confirm that prediction” (p. 1).

The basic semantics of *seem* is epistemic: the verb is said to be one that “evaluates the possible occurrence of a given event or proposition in reality” (p. 1-2). In particular, Achard claims that while control verbs encode situations where “certain facts can only be discovered if people make the conceptual effort to seek them out”, verbs like *seem* encode situations, where facts “become available for anyone to see with no particular effort on any conceptualizer’s part” (p. 2).

Achard calls the unraised construction a setting-subject construction, in which the abstract setting is given focal prominence and thus marked as the subject, and the event or proposition as a whole (including its main participant) is “viewed as the secondary figure and marked as the landmark” (p. 2). By contrast, in the raised construction, the main participant in the event to be located in reality is taken as the “trajector of the main relation due to its focal prominence” (p. 2), and the process/eventuality that the main participant participates in is seen as the landmark of the relation. Further, “[u]nlike the case with the unraised construction, reality remains an unprofiled part of the base” (p. 2). The raised subject thus functions as a reference point for accessing the ‘logical’ subject of the raising verb, the process or event denoting verb phrase. Raising is thus treated as similar to possessor raising as found in pairs like ‘She hit my arm’ and ‘She hit me on the arm’.

The French raising construction that Achard sets out to test Langacker's (1995) prediction on is very similar to the English one. In sentences with raising such as (3.4a), the raising verb takes an NP subject and an infinitival complement. In unraised sentences such as (3.4b), a pronominal form occupies the subject position of the raising verb and the process or event which is located relative to reality is encoded in a finite subordinate *que*-clause.

- (3.4) a. Jean **semble** enfin comprendre.
 John seems to finally understand.
- b. Il **semble** que Jean comprend enfin.
 It seems that John finally understands.

In his study of 300 sentences containing *sembler*, 205 raised, 95 unraised, Achard classified referents in terms of "cognitive availability". The three different availability statuses that Achard recognizes are topical, inferable, and non-topical, following Givón 1995.

Construction	Topical		Inferable		Non-topical		Relative		Total
	n	%	n	%	n	%	n	%	
Raised	116	56.58	39	19.02	25	12.19	25	12.19	205
Unraised	37	38.94	8	8.42	49	51.57	1	1.05	95
Total	153	51.00	47	15.67	74	24.67	26	8.67	300

Table 3.1: Achard's results for the referents of *sembler*

Achard finds that, as predicted, for the most part ($\approx 75\%$) non-expletive subjects of *sembler* in the raising construction are either topical or inferable. He takes this as good evidence for the correctness of the prediction with respect to the raising construction. He also examines more closely the tokens that were not topical or inferable and suggests that about half of them can be accounted for by specific stylistic and rhetorical patterns that serve to enhance the cognitive availability of the referent. One of these patterns is a list structure with elements that by themselves may be relatively unfamiliar and Achard notes that in personal communication Ellen Prince

suggested treating these cases as inferables also. Another pattern consists of negative quantified expressions, which are also typically related to a contextually salient set, that is, they pick out an empty subset. Thus, these elements too may be included in the inferable class. The last pattern that Achard identifies consists of uses of a less readily available referent in order to delay revealing the intended referent's identity for rhetorical purposes.¹

In the unraised construction, Achard finds 47.36% of subjects to be topical or inferable. He argues, however, that in many of these cases the speaker uses the unraised construction on purpose in order to increase the "conceptual distance" between the conceptualizer and the conceptualized scene. Achard characterizes the notion of conceptual distance this way:

The choice of the raised variant involves reduced conceptual distance between the conceptualizer and her object of conceptualization because her initial contact with the complement process is made directly with a participant in that process, or an entity directly inferable from the context. In discourse, this characteristic of the construction represents a useful way of maintaining textual continuity by introducing new information through familiar entities used as reference points. Conversely, with the unraised variant, the complement process can only be located through the consideration of the more abstract notion of reality. The distance between the conceptualizer and her conceptualization is thus increased. (p. 6)

The choice to increase conceptual distance is said to be made in response to specific features of the situation, or simply for strategic reasons. Moreover, Achard says, the speaker's intention to increase the conceptual distance is clearly visible independently of the presence of the unraised construction. Achard mentions the following sub-cases: the referent of the complement clause is reintroduced as a topic as part of a topic shift; a difference of opinion is highlighted by introducing the second opinion in an unraised construction, which is said to bring out the difference between the previous and the current position; the unraised construction serves the narrative purposes of the author better in that it allows them to re-present the referent as different or altered in some way from what it was like in an earlier episode or, more generally, earlier in the text.

¹Given that only one example is given it is not clear how consistent this class is.

Achard's analysis has some problems. For instance, from the fact that 75% of the raised sentences have topical or inferable subject it does not automatically follow that this is good evidence for his prediction that raised sentences have more topical subjects than unraised subjects. The reason is that Achard has no baseline data on what subjects in general are like. It could be the case that the subjects of raising sentences are no different than the average subject and that the unraised construction represents a departure from the normal case when it comes to topicality of subjects. More generally, it seems that the analysis is too much predetermined by an a priori assumption that raised and unraised constructions ought to be maximally different: raised sentences have topical or inferable subject referents while unraised constructions have non-topical subject referents. This expectation also leads Achard to consider the high rate of 47% topical or inferable referents in unraised sentences as problematic. But again we might not have to be so troubled by this number. After all, these subjects of finite complement clauses are still subjects and thus should be rather likely to be topical or inferable. But no evidence is presented about what percentage of topical or inferable subjects we ought to expect for subordinate clauses. Finally, the contrast between the two constructions—if it is real—may just be a matter of degree and consist of very subtle effects that are not readily captured by the broad category 'topical' (or the categories 'topical' and 'inferable'). Employing additional measures might have provided relevant clues and preempted the need for the more subjective and qualitative analyses of exceptional cases in terms of narrative and rhetorical strategies.

Still, Achard's study is important. It is the only one that examines sentence pairs which do not involve an alternation between linking as an oblique dependent and linking as a core argument since in the unraised condition the 'raising' argument is not a direct argument of the raising predicate at all. Further, Achard's study suggests that maybe all of Langacker's (1995, 2000) reference point constructions involve the predicted kind of topicality difference in the construal of one of the participants, a point to which we will return in sections 3.2 and 3.3. And finally, the study also makes the methodological point that it is worthwhile to study in detail the data points that deviate from the general pattern suggesting that factors other than topicality also

play a role in the choice of one construction over the other.

3.1.3 English data

Given Achard's analysis of French data, it is reasonable to also explore the relevant discourse-pragmatic properties for the English pair of constructions. However, the focus here will be somewhat broader. We will investigate several different raising predicates (*seem*, *appear*, *likely*) as well as explore additional formal, semantic, and pragmatic properties of the referents occurring in the constructions.

Before proceeding to the results of various statistical analyses, it is necessary to elaborate some more on which constructions I take to participate in the raising alternation for the verbs *seem* and *appear*. These two verb pose a problem for the analyst because different from *likely* they occur in five valence patterns that allow for similar meanings to be expressed, as is shown in (3.5).

- (3.5)
- a. It **seems** that Peter is a nice guy.
 - b. Peter **seems** to be a nice guy.
 - c. It **seems** as {if/like} Peter is a nice guy.
 - d. Peter **seems** like a nice guy.
 - e. Peter **seems** a nice guy.
 - f. Peter **seems** like he's a nice guy.

Do (3.5c-f) participate in the raising alternation? Are they maybe just stylistic alternatives to (3.5a-b) but semantically equivalent?

I will assume that *seem* is at least two-way polysemous. More specifically, I will treat sentences such as (3.5c) as belonging to an epistemic sense of *seem* while treating (3.5d-f) as belonging to a separate perception-based sense of the verb. Perception *seem* is tied to a Perceiver having some kind of immediate access to a perceived entity, whereas the epistemic *seem* is just concerned with the certainty of one's knowledge about a state of affairs regardless of what kind of evidence it is based on. The

differences between the raising and the bare predicate constructions with *seem* are illustrated by the following examples:²

- (3.6) a. It **seems** that three of our members are sick today.
 b. Three of our members **seem** to be sick today
 c. Three of our members **seem** sick today
- (3.7) a. Well, actually, they're just late.
 b. What makes you think that?

Whereas examples (3.6a-b) can be followed felicitously both by (3.7a) and (3.7b), (3.6c) can only be followed by (3.7b) but not by (3.7a). The reason is that (3.6c) is necessarily perception-based and therefore incompatible with a statement that the presumed objects of perception are not even present. (3.6a) and (3.6b), by contrast, are compatible with, but do not require a perceptual basis for their assertion.

In the analysis in this section, I will focus on the uses of *seem* that belong to the epistemic raising sense, commenting on perceptual uses only occasionally. While I do consider the pattern in (3.5c) to exemplify the epistemic raising sense, I will not include tokens of this pattern in my analysis but just focus on the two English patterns that most closely match Achard's French data.

The data analyzed here consisted of sentences sampled randomly from the BNC. For *appear* and *seem*, I collected data separately from the spoken and the written part of the BNC, while the data for *likely* are derived only from the written texts in the BNC. Tables 3.2 and 3.3 show the number of tokens extracted from the corpus for each of the verbal predicates and the distribution of the tokens across the relevant constructions. The uses of *likely* are summarized separately in Table 3.4 because as an adjective it participates in several constructions not available to the verbal predicates.

Several points are noteworthy. One of them is the different percentages of tokens occurring in deficient contexts, which caused them to be excluded from analysis. Not surprisingly, spoken data included a great many more tokens with deficient contexts.

²The set of examples in (3.6) was brought to my attention by Charles Fillmore. They are parallel to Borkin's set of sentences in (3.3).

Contexts were deemed deficient for several reasons. One kind of deficiency consisted of personal names being blacked out for privacy protection. In multi-speaker contexts such as city council meetings, this often prevented co-reference being established. Another reason for excluding tokens was the speaker's abandoning their utterance altogether, or their repairing or changing it so that the target construction of interest was not fully instantiated. Still other tokens could not be used because the transcriptional record was incomplete due to poor recording quality, interfering noise obscuring utterances, etc.

predicate	def. context		other senses		prenominal		Total
	n	%	n	%	n	%	n
appear (spoken)	40	10.0	187	46.8	1	0.3	400
appear (written)	7	2.3	126	42.0	0	0	300
seem (spoken)	27	13.5	0	0	0	0	200
seem (written)	9	4.5	0	0	0	0	200

Table 3.2: Non-raising uses of four verbal predicates

	appear (spoken)		appear (written)		seem (spoken)		seem (written)		likely (written)	
	n	%	n	%	n	%	n	%	n	%
unraised	21	12.2	17	10.2	19	11.0	20	10.5	22	13.2
raised	125	72.7	111	66.5	102	59.0	89	46.6	136	81.4
predicative	23	13.4	33	19.8	45	26.0	79	41.3	0	0
other constructions	3	1.7	6	3.6	7	4.0	3	1.6	9	5.4
Total	172	100.0	167	100.0	173	100.0	191	100.0	167	100.0

Table 3.3: Raising uses of four verbal predicates

As a comparison between the Totals in Tables 3.2 and 3.3 shows, *appear* is a lot more polysemous than *seem* so more sentences had to be inspected. Many of the non-epistemic uses of *appear* have to do with scenes in which a person shows up or arrives at a place. Closely related further senses are 'appearing in court', 'participating in

Construction Type	n	%
unraised	22	11.0
raised	136	68
NP subject	9	4.5
prenominal	14	7.0
nominalizations	3	1.5
adverbial	9	4.5
deficient context	7	3.5
Total	200	100.0

Table 3.4: Uses of *likely*

a performance', 'being published', etc. Interestingly, these non-epistemic senses of *appear* were more frequent in the spoken language than in the written language.

The predicates differ in how frequently they occur in the different kinds of constructions. Particularly noteworthy is the large percentage of uses of *seem* in written contexts in predicative sentences such as *He seems nice*. Predicative uses are not available with *likely* and they are much less common with *appear* and with *seem* in spoken texts. Note that the 'other' constructions mainly include cases where *seem* or *appear* occur with a *like*-PP (e.g. *They seem like nice people*); where the same two verbs occur in a *there*-construction with just a plain noun phrase following (e.g. *There seemed no problem*); and cases where *likely* takes an event-denoting noun subject (e.g. *A takeover by a foreign bank is likely*).

With regard to the two constructions that we are centrally interested in it, we can observe that the raised construction is much more frequent than the unraised one. The ratio of raised to unraised construction is smallest for written *seem* at 4.45:1 and biggest for written *appear* at about 6.5 to 1.

The frequency difference observed here is not just an artifact of using the BNC as a data source. A look at some of the same lexical units in the Wall Street Journal Corpus suggests that in writing in general the raised construction is more frequent. Table 3.5 shows the relevant data.

The data for *appear* show the constructions in the same relative order. With written *seem* the relative frequency order of constructions is the same as in the BNC except

	seem		likely		appear	
	n	%	n	%	n	%
unraised	7	3.5	4	2.0	10	7.0
raised	95	47.5	139	69.5	99	69.7
predicative	81	40.5	0	0.0	31	21.8
np subject	n/a	n/a	8	4.0	n/a	n/a
prenominal	n/a	n/a	15	7.5	n/a	n/a
other constructions	17	8.5	34	17.0	2	1.4
Total	200	100.0	200	100.0	142	100.0

Table 3.5: Raising predicates in the Wall Street Journal Corpus

that the unraised and the other constructions have swapped rank. With *likely*, the raised construction is again the most frequent one but, unlike in the BNC, the unraised construction is less likely than the NP-subject construction.³

Equally, it should be noted that in the sample from the BNC the ratio of raised to unraised sentences is far greater than in Achard's French data, which were collected from newspaper articles. If we pool the tokens of the different lexical units, then we have 563 raised sentences and 99 unraised sentences; in Achard's sample, there were 205 raised sentences and 95 unraised sentences. This is a statistically very significant difference: $p \leq 0.001$ for a χ^2 -test.

It is not clear what the nature of the frequency difference is. It could be real and obtain between the two languages or just across genres, dialects, or modalities. Alternatively, the sample itself might not be random and thus not reflective of the relative frequencies. That might have happened if, as in English, the unraised construction was a lot less frequent in French than the raised one. If the analyst then felt he needed more tokens to be able to say something meaningful about the unraised construction, he might have collected additional tokens only of that construction.

Unfortunately, Achard does not describe the composition of his sample in detail. For instance, no mention is made of the predicative uses that *sembler* also occurs in (e.g. *Il semble difficile d'établir une liste exhaustive des actes à risque de transmis-*

³In the Wall Street Journal there were no nominalizing constructions involving *likely* such as *Fairfax is the most likely to win the nomination* and no deficient contexts.

sion de ATNC), which he ignored in his analysis, as I will for mine.

What is important for our concerns here is that, if there is a real significant frequency difference in the use of the two constructions between French and English, then we might reasonably expect that the usage conditions are not the same ones.

3.1.4 Formal properties of raised and unraised constructions

Before we begin looking at the discourse properties of the referents in the raised and unraised constructions, we will consider certain formal properties of the sentences in which the two constructions occur.

One such property concerns the nature of the embedded proposition and the syntax of its encoding. In many of the sentences, both raised and unraised ones, one finds that the embedded proposition expresses a categorization or a property attribution, both of which are encoded syntactically in a copular construction as in (3.8).

- (3.8) a. Kim **seems to be** quite nice.
 b. It **seems that Pat is** a teacher by profession.

The relative frequencies of copular subordinate predications with the different raising predicates are presented in Table 3.6. *Appear* clearly occurs most frequently with such predications, followed by *seem* and *likely*. However, when one compares how often unraised and raised constructions occur with a copular subordinate predication, it turns out that there is no statistically significant difference for any of the lexical units or for the pooled data. Table 3.7 shows the frequencies and reports the result of a χ^2 test of independence (including the degrees of freedom), which in every case suggest the absence of a significant difference between the construction with respect to the likelihood of occurring with a copular subordinate predicate.

Another factor that might play a role in the choice of construction is genre. For the purposes of my investigation, I classified the tokens taken from written data into two very broad categories, fiction and non-fiction. This distinction corresponds to the British National Corpus's distinction between imaginative and non-imaginative writing. The results for each of three lexical units and the pooled data are presented

	Copular			
	yes		no	
	n	%	n	%
appear (spoken)	71	47.7	78	52.3
appear (written)	59	44.0	75	56.0
seem (spoken)	44	34.4	84	65.6
seem (written)	23	20.5	89	79.5
likely (written)	42	25.1	125	74.9

Table 3.6: Copular subordinate predication

	Copular	un-raised	raised	Total	χ^2 -statistic	df	p-value
seem (spoken)	n	10	69	79			
	y	9	33	42			
	Total	19	102	121	1.593	1	0.207
seem (written)	n	16	70	86			
	y	4	19	23			
	Total	20	89	109	.018	1	0.894
appear (spoken)	n	13	64	77			
	y	8	61	69			
	Total	21	125	146	.827	1	0.363
appear (written)	n	12	57	69			
	y	5	54	59			
	Total	17	111	128	2.196	1	0.138
likely (written)	n	17	99	116			
	y	5	37	42			
	Total	22	136	158	.195	1	0.659
pooled	n	68	359	427			
	y	31	204	235			
	Total	99	563	662	0.8906	1	0.345

Table 3.7: Presence of copula by construction

	Genre	unraised	raised	Total	p-value of Fisher's exact test
appear (written)	fiction	1	13	14	
	non-fiction	16	98	114	
	Total	17	111	128	0.6916
likely (written)	fiction	2	6	8	
	non-fiction	20	130	150	
	Total	22	136	158	0.3084
seem (written)	fiction	7	23	30	
	non-fiction	13	66	79	
	Total	20	89	109	0.4164
pooled	fiction	10	42	52	
	non-fiction	49	294	343	
	Total	59	336	395	*0.7053

Table 3.8: Construction by genre

in Table 3.8. The results of the Fisher's exact tests undertaken for each lexical unit and a χ^2 -test for the pooled data show no significant difference between the two super-categories.⁴ A more fine-grained analysis of text genres might yield a different result, however.

The final non-pragmatic factor we will consider here is the tense of the raising predicate. It is the case for all raising predicates that they occur more frequently, and in most cases much more frequently, in the present tense than in the past tense. This is shown in Table 3.9. The difference is greatest for *likely* and smallest for *seem* in written contexts.

When we consider whether the tenses are distributed proportionally across the constructions, we find that in some cases we get a significant χ^2 test result, as shown in Table 3.10.

⁴The reason for using a χ^2 -test rather than a Fisher's exact test for the pooled data reason is that most software packages are not equipped to perform the Fisher's exact test once the cell values have become very large for some of the cells. However, with large samples the difference in precision between the Fisher's exact and the χ^2 -test becomes less important. Note that I mark those statistical measures that derive from a different procedure than that used for the majority of the measures with an asterisk '*'. That is, in Table 3.8 the asterisk marks the result of the χ^2 -test.

	Present	Past	Future	Dependent	Imperative
appear (spoken)	131	32	1	8	0
appear (written)	110	54	3	0	0
seem (spoken)	128	42	0	3	0
seem (written)	102	86	1	2	0
likely (written)	123	27	1	16	0

Table 3.9: Tense of raising predicates

According to the χ^2 -tests, there is no statistically significant skewing for the written data used for *appear*, *seem*, and *likely* but there is such skewing within the spoken data for both *seem* and *appear* and when all data is pooled. However, a closer look at the calculations reveals that in most cases the major contribution to the χ^2 value comes from the category 'other', which consists mainly of un-tensed uses. For instance, in the case of *appear* from spoken data, almost 5 of the χ^2 -value of 7.964 is contributed by the cells for the 'other' category. With *seem* in spoken data, the large χ^2 value is due to the low number of non-present tense sentences in the unraised construction (3.1445) as well as the large number of non-present tense sentences among the other constructions (6.6358). In the case of the pooled data, 2/3 of the χ^2 -value is contributed by the 'other' category.

Moreover, it is the case for all lexical units and the pooled data that testing the distribution of tenses across only the raised and unraised constructions for skewing yields no significant results. With spoken *seem* and *appear*, a χ^2 -test for only the raised and unraised constructions is not possible because too many cells would have expected frequencies lower than 5. Therefore, I report the results of Fisher's exact test for the five individual lexical units.⁵ Since there is enough data in all cells for the pooled data, I report the result of a χ^2 -test in that case. The finding is that in all cases, there is no significant difference, as shown in Table 3.11.

I also looked at the temporal relation of the dependent predicate to the raising predicate. The resulting tabulation for the instances of *seem* taken from the written

⁵The Fisher's exact test allows cells to have values less than 5. The test is actually more precise than the χ^2 -test but is harder to compute and the difference between the tests gets smaller the larger the sample and the cell values.

	Construction	Present	Non-Present	Total	χ^2	df	p-value
appear (spoken)	unraised	14	7	21			
	raised	102	23	125			
	other	15	11	26			
	Total	131	41	172	7.964	2	0.018 < 0.05
appear (written)	unraised	12	5	17			
	raised	76	35	111			
	other	22	17	39			
	Total	110	57	167	2.054	2	0.358
seem (spoken)	unraised	18	1	19			
	raised	81	21	102			
	other	29	23	52			
	Total	128	45	173	14.778	2	0.006 < 0.05
seem (written)	unraised	10	10	20			
	raised	54	35	89			
	other	38	44	82			
	Total	102	89	191	3.627	2	0.163
likely (written)	unraised	14	8	22			
	raised	104	32	136			
	other	5	4	9			
	Total	123	44	167	3.2128	2	0.201
pooled	unraised	68	31	99			
	raised	417	146	563			
	other	109	99	208			
	Total	594	276	870	32.918	2	< 0.001

Table 3.10: Tense by construction

part of the BNC is presented in Table 3.12. As the results of Fisher's exact tests reported in the last column of Table 3.12 indicate, there is no asymmetry within the constructions as far as the combinations of tense of the raising predicate and relative event time for the embedded predicate are concerned. But we can notice that in the raised construction, it is much more likely that the embedded event will be simultaneous with the time indicated by the tense of the raising predicate. If we perform

	p-value
seem (spoken)	0.1440
seem (written)	1.0000
appear (spoken)	0.1920
appear (written)	0.4540
likely (written)	0.1985
pooled	*0.2647

Table 3.11: Results of two-tailed Fisher's exact and χ^2 -tests for tense by construction for raised and unraised constructions only

another Fisher's Exact test comparing the Total rows in Table 3.12, we find that there is a significant difference between the constructions in the relative proportion of event time relations, with $p=0.0271$ (2-tailed). The unraised construction seems to be used disproportionately often when the embedded event is temporally anterior to the time indexed by the raising predicate.

Construction		ante- rior	simul- taneous	Total	p-value of Fisher's Exact test
raised	past	5	33	38	
	present	9	42	51	
	Total	14	75	89	
unraised	past	3	7	10	
	present	5	5	10	
	Total	8	12	20	

Table 3.12: Temporal relationship between embedded predicate and raising predicate for *seem* in the written part of the BNC

I repeated the tabulation for the instances of *likely* extracted from the written part of the BNC. A Fisher's exact test cannot be performed on the data as given in Table 3.12 because there are too many columns and rows. A χ^2 -test is not possible either because of low cell values. A natural remedy for this situation is to collapse some of the contrasts. I decided to contrast the simultaneous category event all others for the time of the embedded event and to contrast present with any non-present

forms for the tense value of the raising predicate. The resulting tabulation is shown in Table 3.14. Fisher's Exact tests for the difference within constructions between the frequency of the four possible combinations yield non-significant result; $p=1.000$ (two-tailed) for the raised construction; $p=0.6244$ (two-tailed) for the unraised construction. However, if we perform a Fisher's exact test for the Totals rows, we get a highly significant result with $p \leq 0.0001$. The unraised construction occurs much more often with events whose relative time is explicitly coded as not simultaneous with the time indicated by the tense of the copular clause containing *likely*. The rather convoluted formulation in the previous sentence is necessary because events embedded under *likely* are often quite naturally understood as future-oriented. However, *likely* does not demand such a temporal relation. All that is required is that the verification of the state of affairs lie in the future relative to speech time, as shown by 3.9.

- (3.9) 'I hardly think anyone is **likely** to have survived that fire,' snorted Robertson.

		anterior	simul- taneous	futurate	hypothe- tical	Total
raised	past	1	23	0	0	24
	present	3	98	2	1	104
	simul- taneous	0	8	0	0	8
	Total	4	129	2	1	136
unraised	past	0	1	0	0	1
	present	4	3	6	1	14
	simul- taneous	2	2	2	0	6
	future	0	0	1	0	1
	Total	6	6	9	1	22

Table 3.13: Temporal relationship between embedded predicate and raising predicate for *likely* in the written part of the BNC

The final predicate to be considered here is *appear* as used in the spoken part of the BNC. The relevant tabulation is shown in Table 3.15. The results of Fisher's

		simul- taneous	non-simul- taneous	total	p-value of Fisher's exact test
raised	present	98	6	104	1.000
	non-present	31	1	32	
	Total	129	7	136	
unraised	present	3	11	14	.6244
	non-present	3	5	8	
	Total	6	16	22	

Table 3.14: Temporal relationship between embedded predicate and raising predicate for *likely* recoded with fewer contrasts

exact tests are reported in the last column of Table 3.15. Note that in the calculation for the raised construction the non-present tenses values past and simultaneous were merged in order to arrive at a 2x2 table; this is marked by the symbol †. The results indicate that there is no statistically significant asymmetry within the constructions as far as the combinations of tense of the raising predicate and relative event time for the embedded predicate are concerned. As before, we notice that in the raised construction, it is much more likely that the embedded event will be simultaneous with the time indicated by the tense of the raising predicate. If we perform another Fisher's Exact test comparing the Total rows in Table 3.12, we find that there is a significant difference between the constructions in the relative proportion of event time relations, with $p = 0.0045$ (2-tailed). The unraised construction is used disproportionately often when the embedded event is temporally anterior to the time indexed by the raising predicate.

To sum up: there is no evidence of a statistically significant interaction between either presence of a copular predication, genre, or tense on the one hand and choice of construction on the other. However, we did find a reliable difference between the constructions in terms of the temporal characteristics of the embedded predicate: the unraised construction is used disproportionately often in cases when the event of the embedded predicate is temporally located before the time indicated by the tense of the raising predicate's clause. We will return to the significance of this finding later.

		anterior	simultaneous	Total	p-value of Fisher's exact test
raised	past	1	19	20	†0.19130
	present	17	85	103	
	simultaneous	0	3	3	
	Total	18	107	125	
unraised	past	5	2	7	0.1588
	present	4	10	14	
	Total	9	12	21	

Table 3.15: Temporal relationship between embedded predicate and raising predicate for *appear* in the spoken part of the BNC

	Animate		Total
	n	%	
appear (spoken)	52	30.2	172
appear (written)	58	34.7	167
seem (spoken)	78	45.1	173
seem (written)	80	41.9	191
likely (written)	60	35.9	167

Table 3.16: Animacy of subject referents

3.1.5 Pragmatic properties of referents in raised and unraised constructions

Animacy We will now turn to discourse-pragmatic properties of the referents that are realized as subjects of the raising predicate in the raised construction or as subjects of the dependent clause of the raising predicate in the unraised construction. The first property we will consider is animacy. The percentage of animate referents ranges from 30.2% with *appear* in spoken data to 45.1% but it is always less than 50%.

When we analyze the distribution of animate and inanimate referents across the three construction types with a χ^2 -test, we find that there is usually no significant difference in terms of the proportions of referents that are either animate or inanimate, though the test comes close to significance for several lexical units. This is shown by

Table 3.17.

		un-raised	raised	other	Total	χ^2	df	p-value
seem (spoken)	animate	11	50	17	78			
	inanimate	8	52	35	95			
	Total	19	102	52	173	5.123	2	0.0772
seem (written)	animate	13	41	26	80			
	inanimate	7	48	56	111			
	Total	20	89	82	191	8.519	2	0.0141 < 0.05
appear (spoken)	inanimate	10	36	6	52			
	animate	11	89	20	120			
	Total	21	125	26	172	3.762	2	0.1524
appear (written)	inanimate	8	77	24	109			
	animate	9	34	15	58			
	Total	17	111	39	167	3.550	2	0.1695
likely (written)	inanimate	12	86	9	107			
	animate	10	50	0	60			
	Total	22	136	9	167	5.955	2	0.0509
pooled	inanimate	54	290	82	426			
	animate	45	273	126	444	10.271	2	0.0059 < 0.05

Table 3.17: Animacy by construction

Closer inspection of the cases with low p-values shows the following. Though *seem* in the written data shows a significant result for the χ^2 -test, the factors that are most responsible for this largely have to do with the 'other' category: the high number of animate referents in the unraised construction contributes 2.551 to the total χ^2 -value; the low number of animate referents in the 'other' constructions contributes 2.028; and the low number of inanimate referents in the unraised construction 1.839. For spoken *seem*, the statistic is very close to 0.05 but most of the χ^2 -value is contributed by the 'other' constructions: $1.772 + 1.455 = 3.227$. The reason is that fewer than expected animate referents were found in those constructions. For *likely*, too, the statistic is very close to 0.05 but once again the cells that contribute to the high χ^2 value have nothing to do with the raised and unraised constructions: the uneven distribution of

animate referents across the 'other' constructions contributes $3.234+1.813 = 5.047$ out of the total χ^2 value of 5.955. When we pool the data for all lexical units and perform a χ^2 -test on the combined data, we also get a significant result, $p = 0.0059$. However, as with the individual lexical units the major portion of the χ^2 value is contributed by the cells for the 'other' category: 3.868 derives from there being fewer animate referents than expected in the 'other' constructions and 3.711 derives from there being too many inanimate referents in these constructions.

Fisher's exact tests for the distribution of animacy values only between the raised and unraised constructions yield an insignificant result for every lexical unit. And a χ^2 test for the animacy values in the raised and unraised constructions for the pooled data also yields a non-significant result, with $p=0.5771$. We thus have no good reason to believe that the raised and unraised constructions differ significantly in terms of the animacy of the subject referent.

Subject length We now consider the length of the subject NPs. In general, it should be the case that more readily identifiable, and likely more topical, referents are encoded by shorter linguistic descriptions such as pronouns or simple definite NPs rather than by clauses or by complex NPs with relative clauses or modifiers. For the five lexical units under consideration here the subject length data is as presented in Table 3.18.

The length data does not have a normal distribution for the five lexical units nor for all the data combined. To assess whether there is a significant difference in terms of subject length between the two constructions, we therefore have to use the non-parametric Wilcoxon-Mann-Whitney test. This test works by (1) ranking the scores in increasing order while retaining the information on where each score came from—for tied scores each of the tied observations gets assigned the average of the ranks they would have if no tied had occurred; (2) summing the ranks for each group; (3) determining the average rank for both groups. We would expect the average ranks in each of the sample groups to be about equal if they came from the same population. The results of the test for subject length are shown in Table 3.19.

With four of the five lexical units, the test comes up insignificant with $p > .05$.

	appear (spoken)	appear (written)	likely (written)	seem (spoken)	seem (written)	pooled
	Unraised Construction					
N	21	17	22	19	20	99
Mean	2.8095	3.0000	4.8182	2.8947	2.6000	3.2626
Median	2.0000	1.0000	2.0000	1.0000	2.0000	2.0000
Mode	1.00	1.00	2.00	1.00	1.00	1.00
Std. Dev.	2.80391	3.55317	6.13838	2.57972	2.83586	3.90339
Variance	7.86190	12.62500	37.67965	6.65497	8.04211	15.23645
	Raised Construction					
N	125	111	136	102	89	563
Mean	2.3600	2.9279	3.7868	1.5784	2.4382	2.6874
Median	1.0000	2.0000	2.0000	1.0000	1.0000	1.0000
Mode	1.00	1.00	1.00	1.00	1.00	1.00
Std. Dev.	4.31128	4.15104	4.54347	2.09401	3.46395	3.95729
Variance	18.58710	17.23112	20.64308	4.38488	11.99898	15.66011

Table 3.18: Subject length

Only with spoken *seem* do the subjects of the raised construction have a significantly smaller mean rank. Specifically, they rank higher (that is, closer to rank 1, 2, 3 than to rank 144, 145, 146) because their subjects are shorter than those of the unraised construction. With spoken *seem* the situation is thus as one should expect if the raised construction is associated with more topical subject referents. The influence of spoken *seem* also causes the pooled data to show a significant difference between the constructions with respect to subject length.

Morphology We now turn to the morphology of the subject referents. The data for the different lexical units are summarized in Tables 3.23-3.27 below. The first thing to notice is that for almost all combinations of lexical unit and construction, the majority of subjects is definite, which is no great surprise. The one exception is uses of written *likely* in the category ‘other’, where indefinite NPs, bare/plural NPs and negatively quantified NPs predominate over definite uses 5 to 4. An example of this is given in (3.10):

	appear (spoken)	appear (written)	likely (written)	seem (spoken)	seem (written)	pooled
Group size	21; 125	17; 111	22; 136	19; 102	20; 89	99; 563
Mann-Whitney U	1092.50	908.50	1332.00	633.00	763.50	23809.00
Wilcoxon W	8967.50	1061.50	10648.00	5886.00	4768.50	182575.00
Z	-1.363	-.262	-.850	-3.086	-1.102	-2.521
p (2-tailed)	.173	.793	.395	.002	.270	.012

Table 3.19: Wilcoxon-Mann-Whitney tests for subject length between constructions

(3.10) So on the rare occasions they did report themselves in a State of Readiness they knew an Agile Blade [military exercise] was **likely**.

But this construction is specific to *likely* and the numbers involved are also quite small. Overall, then, it is the case that with these predicates subjects are mostly definite.

The first question we may ask is whether one of the constructions is associated more strongly with definite subjects than another. To address this question I performed two sets of tests: two-tailed χ^2 -tests for the three categories raised, unraised, and other; and two-tailed Fisher's exact tests only for the categories raised and unraised. The counts that go into the tests are given in Table 3.20, which is derived from Tables 3.23-3.27 below, which display the distribution of NP form types in greater detail.

As Table 3.21 shows, none of the tests yielded a significant result. The one result that came very close to statistical significance is the χ^2 -test for *appear* (spoken), $p = 0.0578$. The χ^2 value in that case is 5.70 and the two largest contributors are the cells for the category 'other', with values of 1.74 and 3.08 respectively. The value of 1.0 for the Fisher's exact test for just the raised and unraised constructions confirms the impression that only the numbers for the category 'other' are significantly out of proportion. So far, we have no reason to believe that either the raised or the unraised construction has significantly more definite subjects than the other.

	appear (written)			appear (spoken)		
	Raised	Unraised	Other	Raised	Unraised	Other
Definite	68	12	27	75	13	22
Indefinite	43	5	12	50	8	4
	seem (written)			seem (spoken)		
	Raised	Unraised	Other	Raised	Unraised	Other
Definite	64	15	55	75	13	42
Indefinite	25	5	27	27	6	10
	likely (written)					
	Raised	Unraised	Other			
Definite	78	14	4			
Indefinite	58	8	5			

Table 3.20: Definite and indefinite subjects in the raised and unraised constructions

	p for χ^2 - test for 3 categories	p for Fisher's exact test for 2 categories
appear (spoken)	0.0578	1.0
appear (written)	0.5641	0.5938
likely (written)	0.6163	0.6467
seem (spoken)	0.4764	0.7794
seem (written)	0.6952	1.0

Table 3.21: Tests for definiteness by construction

We may however calculate one-tailed probabilities for the Fisher's exact tests in line with Achard's directional hypothesis that the subjects in the raised construction are more topical. The Fisher's exact test computes the probability, given the observed marginal frequencies, of obtaining exactly the frequencies observed and any configuration more extreme.⁶ "More extreme" is understood to mean any configuration (given observed marginals) with a smaller probability of occurrence in one direction (one-tailed) or in both directions (two-tailed). More extreme configurations in the 'same' direction are identified by locating the smallest frequency in the table, subtracting 1, and then computing the remaining items given the observed marginal

⁶Sheskin provides a very good discussion of Fisher's exact test (2004: 505-511).

frequencies. More extreme configurations in the opposite directions then simply lie on the other side of the tail that is said to go in the same direction. Now consider the 2 x 2 frequency sub-table for the raised and unraised constructions for *likely* in Table 3.20. For Achard's hypothesis to be confirmed and the null hypothesis that there is no difference to be rejected, we need to take into account, in addition to the attested configuration, all additional configurations in which the count in the cell definite+raised is higher than 78 and conversely the count for indefinite+raised is smaller than 58, which in turn entails that the cell for indefinite+unraised would have to be greater than 8 so as to keep the marginal frequencies constant. We thus need to consider the tail that goes in the opposite direction from the one in which the smallest cell frequency, 8, keeps decreasing. The results calculated in this way are as follows: $p=.5349$ for *appear* (spoken); $p=.8434$ for *appear* (written); $p=.7831$ for *likely*; $p=.7039$ for *seem* (written); $p=.7741$. Thus, even when using the directional hypothesis, which should allow easier rejection of the null hypothesis that there is no difference between the constructions, we do not obtain a significant result.

Though the constructions do not seem to differ overall in terms of definite versus indefinite NPs, it might still be the case that more topical definite NPs such as personal pronouns tend to be associated more strongly with one of the two constructions. To test this I repeated the battery of tests above for the contrast between personal pronouns, first, second, or third person, and all other definite NPs. (Note that I use the term 'personal pronoun' to designate a form type. Pronouns do not have to refer to a person or an animate being to be counted as personal pronouns for my purposes. Expletives are, however, excluded.) The results are presented in Table 3.22. Note that a χ^2 value for *likely* could not be calculated because the number of cells with values smaller than 5 was too high.

As for the contrast between definites and indefinites we find that one χ^2 -test comes up significant. But as in the earlier case, a substantial part of the χ^2 value of 6.40 is contributed by the 'other' category. In particular, the number of personal pronouns in the 'other' construction is much higher than expected, adding 2.23 to the total χ^2 -value. As above, if we leave aside the category 'other' and perform a Fisher's exact test only on the raised and unraised constructions, we find no significant difference

	p for χ^2 – test for 3 categories	p for Fisher’s exact test for 2 categories
appear (spoken)	0.99	1.0
appear (written)	0.0407	0.1438
likely (written)	n/a	0.5390
seem (spoken)	0.3593	0.1917
seem (written)	0.4743	0.5717

Table 3.22: Tests for the contrast between highly topical and less topical definites by construction

in the number of personal pronouns versus other definites. It appears that there is no sub-regularity within the class of definite NPs that associates more highly topical NPs with one of the constructions.

	other		raised		unraised		Total	
	n	%	n	%	n	%	n	%
Def. Zero	0	0	2	1.6	0	0.0	2	1.2
1st/2nd Person Pronoun	2	7.7	4	3.2	1	4.8	7	4.1
3rd Person Pronoun	6	23.1	25	20.0	4	19.0	35	20.3
non-restrictive relative pronoun	0	0.0	5	4.0	0	0.0	5	2.9
This NP	0	0.0	2	1.6	0	0.0	2	1.2
That NP	0	0.0	1	0.8	0	0.0	1	0.6
This	1	3.8	2	1.6	0	0.0	3	1.7
That	2	7.7	5	4.0	0	0.0	7	4.1
Possessed NP	3	11.5	1	0.8	0	0.0	4	2.3
Proper Noun	1	3.8	10	8.0	3	14.3	14	8.1
Definite NP	7	26.9	18	14.4	5	23.8	30	17.4
Intermediate Total	22	84.6	75	60.0	13	61.9	110	64.0
Indefinite NP	0	0	2	1.6	0	0.0	2	1.2
Quantified NP	0	0	10	8.0	2	9.5	12	7.0
Bare/Plural NP	1	3.8	6	4.8	2	9.5	9	5.2
restrictive relative pronoun	0	0	9	7.2	0	0.0	9	5.2
pronoun of free relative	0	0	7	5.6	0	0.0	7	4.1
negative determiner	0	0	0	0.0	1	4.8	1	0.6
negative pronoun	0	0	1	0.8	0	0.0	1	0.6
indefinite zero	1	3.8	0	0.0	0	0.0	1	0.6
zero in integrative <i>that</i> -relative clauses	0	0.0	3	2.4	0	0.0	3	1.7
finite clause (including <i>that</i> -clauses)	1	3.8	2	1.6	0	0.0	3	1.7
adverbial clause	1	3.8	0	0.0	0	0.0	1	0.6
expletive	0	0.0	10	8.0	3	14.3	13	7.6
Total	26	100.0	125	100.0	21	100.0	172	100.0

Table 3.23: Subject form for *appear* in spoken part of the BNC

	other		raised		unraised		Total	
	n	%	n	%	n	%	n	%
Def. Zero	0	0.0	0	0.0	0	0.0	0	0.0
1st/2nd Person Pronoun	0	0.0	1	0.9	1	5.9	2	1.2
3rd Person Pronoun	12	30.8	13	11.7	4	23.5	29	17.4
non-restrictive relative pronoun	0	0.0	5	4.5	0	0.0	5	3.0
This NP	0	0.0	3	2.7	0	0.0	3	1.8
That NP	0	0.0	0	0.0	0	0.0	0	0.0
This	0	0.0	1	0.9	0	0.0	1	0.6
That	0	0.0	2	1.8	0	0.0	2	1.2
Possessed NP	1	2.6	2	1.8	0	0.0	3	1.8
Proper Noun	2	5.1	16	14.4	2	11.8	20	12.0
Definite NP	12	30.8	25	22.5	5	29.4	42	25.1
Intermediate Total	27	69.2	68	61.3	12	70.6	107	64.1
Indefinite NP	0	0.0	2	1.8	1	5.9	3	1.8
Quantified NP	1	2.6	7	6.3	0	0.0	8	4.8
Bare/Plural NP	5	12.8	14	12.6	1	5.9	20	12.0
restrictive relative pronoun	2	5.1	6	5.4	0	0.0	8	4.8
pronoun of free relative	0	0.0	5	4.5	0	0.0	5	3.0
negative determiner	0	0.0	1	0.9	0	0.0	1	0.6
negative pronoun	1	2.6	0	0.0	1	5.9	2	1.2
indefinite zero	0	0.0	0	0.0	0	0.0	0	0.0
zero in integrative <i>that</i> -relative clauses	1	2.6	3	2.7	0	0.0	4	2.4
finite clause (including <i>that</i> -clauses)	0	0.0	0	0.0	0	0.0	0	0.0
VP _{to}	1	2.6	0	0.0	0	0.0	1	0.6
adverbial clause	0	0.0	0	0.0	0	0.0	0	0.0
for-to clause	1	2.6	0	0.0	0	0.0	1	0.6
expletive	0	0.0	5	4.5	2	11.8	7	4.2
Total	39	100	111	100.0	17	100.0	167	100.0

Table 3.24: Subject form for *appear* in written part of the BNC

	other		raised		unraised		Total	
	n	%	n	%	n	%	n	%
Def. Zero	0	0.0	0	0.0	0	0.0	0	0.0
1st/2nd Person Pronoun	0	0.0	7	5.1	3	13.6	10	6.0
3rd Person Pronoun	1	11.1	18	13.2	0	0.0	19	11.4
non-restrictive relative pronoun	0	0.0	1	0.7	0	0.0	1	0.6
This NP	0	0.0	3	2.2	2	9.1	5	3.0
That NP	0	0.0	1	0.7	1	4.5	2	1.2
This	0	0.0	5	3.7	0	0.0	5	3.0
That	2	22.2	2	1.5	1	4.5	5	3.0
Possessed NP	0	0.0	4	2.9	1	4.5	5	3.0
Proper Noun	0	0.0	4	2.9	0	0.0	4	2.4
Definite NP	1	11.1	33	24.3	6	27.3	40	24.0
Intermediate Total	4	44.4	78	57.4	14	63.4	96	57.5
Indefinite NP	1	11.1	7	5.1	0	0.0	8	4.8
Quantified NP	0	0.0	10	7.4	3	13.6	13	7.8
Bare/Plural NP	3	33.3	26	19.1	5	22.7	34	20.4
interrogative pronoun	0	0.0	1	0.7	0	0.0	1	0.6
restrictive relative pronoun	0	0.0	6	4.4	0	0.0	6	3.6
reduced relative pronoun of free relative	0	0.0	2	1.5	0	0.0	2	1.2
negative determiner	1	11.1	0	0.0	0	0.0	1	0.6
negative pronoun	0	0.0	0	0.0	0	0.0	0	0.0
indefinite zero	0	0.0	0	0.0	0	0.0	0	0.0
zero in integrative <i>that</i> -relative clauses	0	0.0	1	0.7	0	0.0	1	0.6
finite clause (including <i>that</i> -clauses)	0	0.0	0	0.0	0	0.0	0	0.0
VP _{to}	0	0.0	0	0.0	0	0.0	0	0.0
adverbial clause	0	0.0	1	0.7	0	0.0	1	0.6
for-to clause	0	0.0	0	0.0	0	0.0	0	0.0
expletive	0	0.0	4	2.9	0	0.0	4	2.4
Total	9	100.0	136	100	22	0.0	167	100.0

Table 3.25: Subject form for *likely* in the written part of the BNC

	other		raised		unraised		Total	
	n	%	n	%	n	%	n	%
Def. Zero	0	0.0	0	0.0	0	0.0	0	0.0
1st/2nd Person Pronoun	2	2.4	8	9.0	2	10.0	12	6.3
3rd Person Pronoun	28	34.1	20	22.5	6	30.0	54	28.3
non-restrictive relative pronoun	1	1.2	0	0.0	0	0.0	1	0.5
This NP	1	1.2	4	4.5	1	5.0	6	3.1
That NP	1	1.2	0	0.0	0	0.0	1	0.5
This	0	0.0	3	3.4	0	0.0	3	1.6
That	2	2.4	2	2.2	0	0.0	4	2.1
Possessed NP	2	2.4	4	4.5	0	0.0	6	3.1
Proper Noun	6	7.3	10	11.2	2	10.0	18	9.4
Definite NP	12	14.6	13	14.6	4	20.0	29	15.2
Intermediate Total	55	61.8	64	71.9	15	75.0	134	70.2
Indefinite NP	1	1.2	1	1.1	1	5.0	3	1.6
Quantified NP	5	6.1	4	4.5	2	10.0	11	5.8
Bare/Plural NP	4	4.9	11	12.4	2	10.0	17	8.9
Interrogative pronoun	0	0.0	0	0.0	0	0.0	0	0.0
restrictive relative pronoun	3	3.7	2	2.2	0	0.0	5	2.6
reduced relative pronoun of free relative	0	0.0	0	0.0	0	0.0	0	0.0
negative determiner	1	1.2	2	2.2	0	0.0	3	1.6
negative pronoun	0	0.0	0	0.0	0	0.0	0	0.0
indefinite zero	0	0.0	0	0.0	0	0.0	0	0.0
zero in integrative <i>that</i> -relative clauses	0	0.0	1	1.1	0	0.0	1	.5
finite clause (including <i>that</i> -clauses)	9	11.0	0	0.0	0	0.0	9	4.7
VP _{to}	2	2.4	1	1.1	0	0.0	3	1.6
adverbial clause	1	1.2	0	0.0	0	0.0	1	0.5
for-to clause	0	0.0	0	0.0	0	0.0	0	0.0
Expletive	1	1.2	3	3.4	0	0.0	4	2.1
Total	82	100.0	89	100.0	20	100.0	191	100.0

Table 3.26: Subject form for *seem* in the written part of the BNC

	other		raised		unraised		Total	
	n	%	n	%	n	%	n	%
Def. Zero	1	1.9	3	2.9	0	0.0	4	2.3
1st/2nd Person Pronoun	8	15.4	16	15.7	5	26.3	29	16.8
3rd Person Pronoun	22	42.3	39	38.2	2	10.5	63	36.4
non-restrictive relative pronoun	1	1.9	2	2.0	0	0.0	3	1.7
This NP	0	0.0	0	0.0	2	10.5	2	1.2
That NP	1	1.9	0	0.0	0	0.0	1	0.6
This	0	0.0	1	1.0	0	0.0	1	0.6
That	1	1.9	3	2.9	0	0.0	4	2.3
Possessed NP	1	1.9	1	1.0	0	0.0	2	1.2
Proper Noun	1	1.9	5	4.9	0	0.0	6	3.5
headless <i>wh</i> -clause	0	0.0	0	0.0	1	5.3	1	0.6
Definite NP	6	11.5	5	4.9	3	15.8	14	8.1
Intermediate Total	42	80.8	75	73.5	13	68.4	130	75.1
Indefinite NP	1	1.9	2	2.0	1	5.3	4	2.3
Quantified NP	1	1.9	4	3.9	1	5.3	6	3.5
Bare/Plural NP	0	0.0	3	2.9	1	5.3	4	2.3
Interrogative pronoun	0	0.0	0	0.0	0	0.0	0	0.0
restrictive relative pronoun	0	0.0	2	2.0	0	0.0	2	1.2
reduced relative pronoun of free relative	0	0.0	0	0.0	0	0.0	0	0.0
negative determiner	0	0.0	0	0.0	0	0.0	0	0
negative pronoun	0	0.0	3	2.9	1	5.3	4	2.3
indefinite zero	0	0.0	0	0.0	0	0.0	0	0.0
zero in integrative <i>that</i> -relative clauses	0	0.0	3	2.9	0	0.0	3	1.7
finite clause (including <i>that</i> -clauses)	3	5.8	1	1.0	0	0.0	4	2.3
VP _{to} adverbial clause	4	7.7	0	0.0	1	5.3	5	2.9
for-to clause	0	0.0	0	0.0	0	0.0	0	0.0
Expletive	1	1.9	7	6.9	1	5.3	9	5.2
Total	52	100.0	102	100.0	19	100.0	173	100.0

Table 3.27: Subject form for *seem* in the spoken part of the BNC

Evocation status We now turn to the question whether the constructions differ in how their subjects are connected to prior discourse, and in particular, whether the subjects have been explicitly mentioned. Table 3.28 summarizes the relevant data for the raised and unraised constructions.

	brand new	un- used	infer- rable	text. evoked	sit. evoked	var- i- able	n/a	Total
appear (spoken)	7	2	36	67	2	19	13	146
appear (written)	7	1	48	50	1	14	7	128
seem (spoken)	6	7	25	61	7	7	8	121
seem (written)	3	4	38	55	1	5	3	109
likely (written)	8	0	81	54	1	10	4	158
pooled	31	14	228	287	12	55	35	662

Table 3.28: Connection of subject to prior discourse

Not all predicates behave alike: with *likely* it seems to be the case that usually the subject referent has not been evoked before, whereas in the case of *seem* in the written data, prior textual evocation is just as likely as a lack of prior evocation. But regardless of the overall differences between the lexical units, there doesn't seem to exist a difference between the uses of a given lexical unit in the raised and unraised constructions. Table 3.29 shows the input and results of Fisher's exact tests for all five lexical units.⁷ The p-values are greater than 0.05 for all the lexical units and pooling the data for all the lexical units actually results in the highest p-value of all. As we did earlier for the definiteness-by-construction data, we can perform one-tailed directional Fisher's exact tests on all the data in Table 3.29 with the exception of the pooled data since the numbers involved there are too large to be calculated by software available to me. Following Achard, we would state the alternative hypothesis as saying that in the raised construction subjects will be more likely to have an explicit prior evocation than in the unraised construction. The results found for the five lexical units are as follows: $p=.2953$ for *seem* (spoken); $p=.8836$ for *seem* (written); $p=.8111$

⁷Note that the non-referring, expletive *there* subjects that are sometimes found with *seem* are included in the 'not evoked' category.

		Evocation		Total	p-value of Fisher's exact test
		evoked	not evoked		
seem (spoken)	unraised	8	11	19	
	raised	53	49	102	
		61	60	121	0.4639
seem (written)	unraised	12	8	20	
	raised	43	46	89	
		55	54	109	0.4590
appear (spoken)	unraised	11	10	21	
	raised	56	69	125	
		67	79	146	0.6373
appear (written)	unraised	8	9	17	
	raised	42	69	111	
		50	78	128	0.5946
likely (written)	unraised	4	18	22	
	raised	50	86	136	
		54	104	158	0.0968
pooled	unraised	43	56	99	
	raised	244	319	563	
		287	375	662	*1.0

Table 3.29: Evocation by construction

for *appear* (spoken); $p=.3196$ for *appear* (written); and $p=.0679$ for *likely*.

While evocation versus lack of evocation does not yield a significant difference, we might find one when we lump together the class of evoked and of inferable referents and compare it to all other evocation statuses. Table 3.30 presents the relevant tables and test statistics. Note that for the pooled data, I again report the p-value of a χ^2 -test (cf. footnote 4 on p. 229).

		Evocation		Total	p-value of Fisher's exact test
		evoked, unused, or inferable	other		
seem (spoken)	unraised	16	3	19	
	raised	84	18	102	
		100	21	121	1.0
seem (written)	unraised	20	0	20	
	raised	78	11	89	
		98	11	109	0.2104
appear (spoken)	unraised	16	5	21	
	raised	91	34	125	
		107	39	146	1.0
appear (written)	unraised	15	2	17	
	raised	85	26	111	
		100	28	128	0.3600
likely (written)	unraised	21	1	22	
	raised	114	22	136	
		135	23	148	0.2030
pooled	unraised	88	11	99	
	raised	453	110	563	
		541	121	662	*0.0454

Table 3.30: Evocation/inferability by construction

The tests do not yield statistically significant results for the lexical units. In the case of the pooled data, though, the asymmetries for the individual lexical units have combined to bring down the p-value below 0.05 and to statistical significance.

Statistically, this drastic reduction in the p-value of the pooled data compared to the individual lexical units makes good sense: the more data points one has, the more information one has used and the more confident one can be that a discrepancy is real rather than just due to chance. Adding more data points to the analysis of individual lexemes—lexical idiosyncrasies notwithstanding—should produce a similar result.

The finding about the pooled data is of great interest since Achard in analyzing French data for written *sembler* also collapses the class of inferables with that of evoked referents and then finds a significant result. However, a closer look at the χ^2 -table for the pooled data shows that we found a statistically significant result in the opposite direction from Achard's. Achard finds more evoked referents in the raised construction and fewer non-evoked ones than could be expected. Likewise, he finds fewer evoked referents in the unraised construction and more non-evoked referents than could be expected. In the pooled data, however, the tendency is the opposite: there are fewer evoked referents in the raised construction and more non-evoked ones than should be expected if the distribution was proportional. Also, we find more evoked referents in the unraised construction and fewer non-evoked ones than should be expected.

It seems clear that we cannot reproduce Achard's result for evocation status of subjects of French *sembler* with our data. We therefore have to look towards other factors that might exhibit differences between raised and unraised sentences that suggest that the raised construction is associated with greater topicality of the subject. One factor that suggests itself at this point is subsequent evocation, that is, whether the subjects of the two constructions are equally likely to be mentioned again in later discourse. The findings for subsequent evocation in raised, unraised, and other constructions are summarized in Table 3.31.

Note that the category inferable here refers to something other than in the case of prior evocation. Other referents may be inferred on the basis of the subject referent in the raised or unraised target sentence. But the current referent, since it has been mentioned, cannot be inferred again, at least within the window of context. Thus, for the purposes of analyzing the subsequent evocation status of the subject referents, I will collapse the inferables appearing in Table 3.31 with the non-evoked class. The

	inferred	evoked	none	Total
appear (spoken)	7	69	96	172
appear (written)	5	75	87	167
seem (spoken)	2	76	95	173
seem (written)	6	88	97	191
likely (written)	8	41	118	167

Table 3.31: Subsequent evocation

analysis for subsequent evocation of the subject is presented in Table 3.32.

As with prior evocation, we do not get a statistically significant result: raised and unraised constructions are not associated with an unusual likelihood for subsequent evocation of the subject referent. All the p-values are considerably greater than 0.05, both for individual lexical units and the pooled data. Note that the p-value for the pooled data again comes from a χ^2 rather than a Fisher's exact test.

As we did in the case of prior evocation, we can also perform directional Fisher's exact tests for the data on the five lexical units. The alternative hypothesis following from Achard's work has to be that subjects in the raised construction will have more subsequent evocations than subjects in the unraised construction. The results are as follows: $p=.4418$ for *appear* (spoken); $p=.6103$ for *appear* (written); $p=.4720$ for *likely* (written); $p=.4933$ for *seem* (spoken); and $p=.8836$ for *seem* (written). As with the two-tailed tests, there are no statistically significant results to record.

		Evocation		Total	p-value of Fisher's exact test
		evoked	non-evoked		
appear (spoken)	unraised	10	11	21	
	raised	54	71	125	
	Total	64	82	146	.8132
appear (written)	unraised	8	9	17	
	raised	52	59	111	
	Total	60	68	128	1.000
likely (written)	unraised	6	16	22	
	raised	33	103	136	
	Total	39	119	158	0.7917
seem (spoken)	unraised	9	10	19	
	raised	45	57	102	
	Total	54	67	121	0.8068
seem (written)	unraised	12	8	20	
	raised	43	46	89	
	Total	55	54	109	0.4590
pooled	unraised	45	54	99	
	raised	227	336	563	
	Total	272	390	662	*0.3383

Table 3.32: Subsequent evocation (including inferable) by construction

Number of contextual mentions So far we have looked mostly at categorical variables: presence or absence of some feature, such as a animacy, prior and subsequent mention. But it might be the case that there is a continuous effect on the choice of construction. This is easiest to explore for the number of prior and subsequent mentions, where a greater number of prior or subsequent mentions might correlate with a greater likelihood that one of the two constructions is chosen over the other.

We first look at the number of prior mentions that the five lexical units have. This data is summarized in Table 3.33.

	appear (spo- ken)	appear (writ- ten)	likely (writ- ten)	seem (spo- ken)	seem (writ- ten)	pooled
Unraised Construction						
N	21	17	22	19	20	99
Mean	2.05	2.06	.86	3.95	2.25	2.19
Median	1.00	.00	.00	.00	1.00	.00
Mode	0	0	0	0	0	0
Std. Dev.	3.309	3.152	2.376	7.619	3.626	4.391
Variance	10.948	9.934	5.647	58.053	13.145	19.279
Raised Construction						
N	125	111	136	102	89	563
Mean	1.49	1.28	1.26	1.80	2.61	1.63
Median	.00	.00	.00	1.00	.00	.00
Mode	0	0	0	0	0	0
Std. Dev.	2.418	2.382	2.776	3.083	4.435	3.043
Variance	5.849	5.676	7.704	9.506	19.673	9.260

Table 3.33: Number of prior mentions

Tests of normality show that the distribution of prior mentions is not normal for any of the lexical units (details not displayed here). The Kolmogorov-Smirnov and the Shapiro-Wilk test yield .000 for all lexical units. In order to compare the two constructions in terms of the number of prior mentions we thus again have to use the Wilcoxon-Mann-Whitney test. As Table 3.34 shows, there are no significant differences between the constructions for any of the lexical units, with p clearly greater

than .05 for all of them.

	appear (spoken)	appear (written)	likely (written)
Group size	21; 125	17; 111	22; 136
Mann-Whitney U	1181.000	828.000	1244.500
Wilcoxon W	9056.000	7044.000	1497.500
Z	-.802	-.925	-1.498
p (2-tailed)	.423	.355	.134
	seem (spoken)	seem (written)	pooled
Group size	19; 102	20; 89	99; 563
Mann-Whitney U	969.000	838.000	27133.500
Wilcoxon W	1159.000	4843.000	185899.50
Z	.000	-.434	-.464
p (2-tailed)	1.000	.664	.643

Table 3.34: Prior mentions

We can also evaluate the comparison of the constructions in terms of the number of prior mentions for the directional alternative hypothesis that follows from Achard's work: in the raised construction, subjects should have more prior mentions. Because of the way the Wilcoxon-Mann-Whitney test works—the smallest value gets the lowest rank—we should expect the mean rank for the raised construction to be greater than that for the unraised construction. Further, the Z-score should exceed the critical value at our pre-specified .05 level of significance. Consider as an example the lexical unit *seem* (written). The mean ranks (which were not given in Table 3.34) are 57.60 for the unraised construction 54.42 for the raised construction.⁸ This by itself indicates that a difference of the kind Achard's work would predict does not in fact

⁸The reader may wonder how it is possible that the mean rank of the raised construction is lower than that of the unraised construction given the numbers in Table 3.33. The answer is that a data set can have a higher average but a lower mean rank if the scores it contains are more extreme than those in the other data set. The reader may verify this with two constructed data sets A and B. For A {0, 0, 0, 0, 0, 10, 10, 10, 10, 10} and B {1, 2, 2, 3, 3, 3, 5, 7, 8, 10} we find that A has an average score of 5.0 and B has an average score of 4.4. The mean ranks for A is 10.25 and for B 10.75 (assuming tied ranks are averaged).

obtain: the mean rank in the raised construction is lower, rather than greater, than in the unraised construction. Moreover, it is the case that for *seem* (written), the calculated z-score is $-.434$, which means that its absolute value, $.434$, does not exceed the critical value of 1.65 for a one-tailed test at the $.05$ significance level. Inspection of the table shows that the z-scores calculated for the other lexical units also do not exceed the critical value of 1.65 . Thus, even when using the directional alternative hypothesis, we find no clear evidence that would allow us to reject the null hypothesis that the two constructions do not differ significantly with regards to the number of prior mentions of their subjects.

We now turn to the number of subsequent mentions. The data are summarized in Table 3.35. Again, according to Kolmogorov-Smirnov and Shapiro-Wilk tests (results not shown), the distribution is not normal for any of the lexical units in any of the constructions. As for prior mentions we find that the differences between the constructions suggested by their different mean values as shown in Table 3.35 are not significant according to Wilcoxon-Mann-Whitney tests, with p again clearly greater than $.05$ for all lexical units, as shown by Table 3.36. The z-scores calculated for the lexical units also all have absolute values smaller than the critical value of 1.65 which would have to be exceeded to support the directional alternative hypothesis suggested by Achard's findings for French that the subjects of the raised constructions ought to be more topical and have more subsequent mentions.

Overall, we find that analyzing the number of contextual mentions leads to the same conclusion as the analysis in the previous section of the binary opposition between prior or subsequent evocation and lack thereof: no significant differences between the constructions emerge.

	appear (spo- ken)	appear (writ- ten)	likely (writ- ten)	seem (spo- ken)	seem (writ- ten)	pooled
Unraised Construction						
N	21	17	22	19	20	99
Mean	2.38	2.06	.91	2.95	2.20	2.07
Median	.00	.00	.00	.00	1.00	.00
Mode	0	0	0	0	0	.0
Std. Dev.	4.653	4.562	2.223	4.755	3.334	3.957
Variance	21.648	20.809	4.944	22.608	11.116	15.658
Raised Construction						
N	125	111	136	102	89	563
Mean	1.52	1.56	1.07	1.85	2.49	1.63
Median	.00	.00	.00	.00	.00	1.63
Mode	0	0	0	0	0	.00
Std. Dev.	2.465	3.041	3.080	3.868	4.408	3.372
Variance	6.074	9.249	9.484	14.958	19.435	11.368

Table 3.35: Number of subsequent mentions

	appear (spoken)	appear (written)	likely (written)
Group size	21; 125	17; 111	22; 136
Mann-Whitney U	1249.500	940.500	1455.000
Wilcoxon W	9124.50	7156.50	10771.00
Z	-.388	-.023	-.272
p (2-tailed)	.698	.982	.785
	seem (spoken)	seem (written)	pooled
Group size	19; 102	20; 89	99; 563
Mann-Whitney U	871.500	831.000	26260.500
Wilcoxon W	6124.50	4836.00	185026.50
Z	-.764	-.494	-1.029
p (2-tailed)	.445	.621	.303

Table 3.36: Subsequent mentions

Distance of contextual mentions While neither the binary distinction between prior or subsequent mention and lack thereof nor the number of contextual mentions provided any evidence for a difference between the constructions, it might be the case that the distance of the immediately prior or subsequent mentions or of the last mentions within the window of context differ to a significant extent between the constructions.

Table 3.37 summarizes the data for the distance of immediately prior mentions. Note that in terms of the mean values there is no clear trend across all the lexical units. For spoken and written *appear* and written *seem* the mean is greater in the raised construction but for *likely* and spoken *seem* the opposite is true.

	appear (spo- ken)	appear (writ- ten)	likely (writ- ten)	seem (spo- ken)	seem (writ- ten)	pooled
Unraised Construction						
N	21	17	22	19	20	99
Mean	6.29	6.47	8.82	7.05	5.45	6.86
Median	6.00	11.00	11.00	11.00	3.50	11.00
Mode	11	11	11	11	11	11
Std. Dev.	4.849	5.269	4.148	4.904	4.861	4.836
Variance	23.514	27.765	17.203	24.053	23.629	23.388
Raised Construction						
N	125	111	136	102	89	563
Mean	6.45	7.32	7.42	6.37	6.22	6.80
Median	11.00	11.00	11.00	8.50	11.00	11.00
Mode	11	11	11	11	11	11
Std. Dev.	5.130	4.893	4.822	4.793	5.029	4.942
Variance	26.314	23.945	23.253	22.969	25.290	24.424

Table 3.37: Distance of prior mentions

The distribution of the distance data is not normal for any of the lexical units in either construction. Kolmogorov-Smirnov and the Shapiro-Wilk tests yield $p = .000$ in all cases. Performing two-tailed Wilcoxon-Mann-Whitney tests for all five lexical units yields no statistically significant result, as is shown in Table 3.38. We may also look at the z-scores to evaluate the directional alternative hypothesis flowing from

	appear (spoken)	appear (written)	likely (written)	seem (spoken)	seem (written)	pooled
Group size	21; 125	17; 111	22; 136	19; 102	20; 89	99; 563
Mann-Whitney U	1307.500	839.000	1240.500	894.000	838.000	27554.000
Wilcoxon W	1538.500	992.000	10556.500	6147.000	1048.000	186320.000
Z	-.031	-.838	-1.499	-.572	-.435	-.198
p (2-tailed)	.976	.402	.134	.567	.663	.843

Table 3.38: Distance of prior mentions

Achard's work: the prior mentions of subjects in the raised construction should be closer than those in the unraised construction. However, we see that the z-scores' absolute values do not exceed 1.65, the critical value for a significant result for a one-tailed test.

We now turn to the distance data for subsequent mentions. They are summarized in Table 3.39. As in the case of prior mentions, the mean distance values show no common tendency: some are greater in the raised construction and some in the unraised construction. As with most variables we have seen, the data for the distance of subsequent mentions is not normally distributed according to Kolmogorov-Smirnov and Shapiro-Wilk tests (results not shown). Two-tailed Wilcoxon-Mann-Whitney tests for assessing the difference between the constructions in terms of the distance of the target to the subsequent mention do not yield statistically significant results, with p greater than .05 in all cases, as shown by Table 3.40. One tailed tests for a directional alternative hypothesis also fail to yield significant results: the z-scores all have absolute values below the critical value of 1.65 for a one-tailed test. The data for subsequent distance thus pattern like the data for subsequent evocation and number of subsequent mentions.

The final distance measure to analyze is the distance from the current mention to the last mention within the window of context, which is similar to Givón's persistence measure. More topical referents should be talked about longer and thus have a later final mention than more transient referents. The data for the five lexical units

	appear (spo- ken)	appear (writ- ten)	likely (writ- ten)	seem (spo- ken)	seem (writ- ten)	pooled
Unraised Construction						
N	21	17	22	19	20	99
Mean	6.81	6.82	8.14	6.37	5.20	6.70
Median	11.00	11.00	11.00	11.00	2.50	11.00
Mode	11	11	11	11	11	11
Std. Dev.	4.885	4.747	4.804	5.177	4.927	4.906
Variance	23.862	22.529	23.076	26.801	24.274	24.701
Raised Construction						
N	125	111	136	102	89	563
Mean	7.02	6.79	8.60	6.89	6.24	7.21
Median	11.00	11.00	11.00	11.00	11.00	11.0
Mode	11	11	11	11	11	11
Std. Dev.	4.833	4.783	4.304	4.809	5.066	4.789
Variance	23.355	22.875	18.524	23.127	25.660	22.930

Table 3.39: Distance of subsequent mentions

	appear (spo- ken)	appear (writ- ten)	likely (writ- ten)	seem (spo- ken)	seem (writ- ten)	pooled
Group size	21; 125	17; 111	22; 136	19; 102	20; 89	99; 563
Mann-Whitney U	1259.50	918.50	1421.00	890.50	838.00	26354.00
Wilcoxon W	1490.50	7134.50	1674.00	1080.50	1048.00	31304.00
Z	-.327	-.191	-.498	-.616	-.438	-.971
p (2-tailed)	.744	.849	.618	.538	.662	.332

Table 3.40: Distance of subsequent mentions

	appear (spo- ken)	appear (writ- ten)	likely (writ- ten)	seem (spo- ken)	seem (writ- ten)	pooled
Unraised Construction						
N	21	17	22	19	20	99
Mean	2.57	2.71	.55	2.79	2.70	2.21
Median	.00	.00	.00	.00	.50	.00
Mode	0	0	0	0	0	0
Std. Dev.	3.723	3.853	2.154	4.090	3.614	3.558
Variance	13.857	14.846	4.641	16.731	13.063	12.659
Raised Construction						
N	125	111	136	102	89	563
Mean	1.99	1.85	1.04	2.09	2.33	1.80
Median	.00	.00	.00	.00	.00	.00
Mode	0	0	0	0	0	0
Std. Dev.	3.224	3.117	2.665	3.270	3.592	3.170
Variance	10.395	9.713	7.102	10.695	12.904	10.052

Table 3.41: Distance of last mentions

considered here is presented in Table 3.41. With the exception of *likely*, the mean values show an unexpected pattern: the last mentions in the unraised construction seem to be farther away from the current mention than the last mentions in the raised construction.

The data for the distance to the last contextual mention is not normally distributed for any of the lexical units in either of the two constructions according to Kolmogorov-Smirnov and Shapiro-Wilk tests (results not shown). As with all the other number and distance measures, we find no significant difference between the constructions, with p-values for two-tailed Wilcoxon-Mann Whitney tests solidly greater than .05 in all cases, as shown in Table 3.42. The z-scores in that table also indicate that one-tailed tests based on a directional alternative hypothesis fail to yield significant results: none of their absolute values exceeds the critical value of 1.65.

Interim summary Given the evidence that we have considered in this section there is no clear reason to assume a pragmatic distinction between the subjects of

	appear (spo- ken)	appear (writ- ten)	likely (writ- ten)	seem (spo- ken)	seem (writ- ten)	pooled
Group size	21; 125	17; 111	22; 136	19; 102	20; 89	99; 563
Mann- Whitney U	1213.50	878.50	1349.00	907.50	827.50	26884.00
Wilcoxon W	9088.50	7094.50	1602.00	6160.50	4832.50	185650.00
Z	-.631	-.514	-1.110	-.496	-.539	-.659
p (2-tailed)	.528	.607	.267	.620	.590	.510

Table 3.42: Distance of last mentions

the two constructions. Neither animacy nor subject length, subject definiteness, evocation status, number of contextual mentions, or distance of contextual mentions have resulted in statistically significant results that would hold for even just a majority of the lexical units. In most cases, none of the lexical units shows a significant result.

The question of what explains the difference between Achard's data for French *ssembler* and my English data for *appear*, *seem*, and *likely* remains open. Clearly, if one takes seriously the cognitive grammar account that Langacker has suggested, one cannot simply dismiss the lack of evidence for a role of topicality or, more generally, discourse status in English raising as a matter of cross-linguistic variation. Both languages usually have topical subjects and speakers of both languages have to be assumed to be able to attend to the same kind of factors.

There are several plausible sources for the disparity. One is that the samples are very different from each other. It might be the case that the British National Corpus data represents too many different authors and genres: data from dialects or idiolects or genres that make a distinction along the lines that Achard has found for French might be cancelled out by others where the distinction does not exist or follows different boundaries. Given that Achard's texts and method of sampling are not known to me this is difficult to assess.

Another possible explanation is that the coding criteria used by Achard and me were not consistent. In particular, this would affect the coding of inferables. For

	raised	unraised	Total
other	26	5	31
inferable anchored by modifier	3	1	4
frame	51	10	61
kinship, possession, attribute	13	3	16
verbal proposition	34	5	39
co-subset, co-subtype	13	9	22
subset, subtype	32	6	38
super-set, super-type	15	2	17
unused	11	3	14
situationally given	11	1	12
Total	209	45	254

Table 3.43: Kinds of inferables

instance, Achard might have coded certain referential expressions as inferable that I did not consider to be inferable and which moreover occurred more frequently in the raised construction. Alternatively, I might have recognized more inferables than Achard did, with the result of increasing the number of inferables in the unraised constructions. Given that I have no access to the coding protocol that Achard used—Achard cites Givón 1995 and Birner 1997 as relevant to his coding but these works do not provide coding protocols either—this question cannot be answered either.

However, for my own data one way to address this issue is to repeat the evocation status analysis with different sets of inferables included in the inferable class. The set of tokens where the subject referent is either unused, situationally given or inferable consists of 254 sentences. They belonged to the types shown in Table 3.43.

The most reliable kinds of inferables to recognize are the ones in the middle part of Table 3.43: set relations (*four children ... the oldest*), omni-relevant or unused entities (*the Pope*), situationally given entities (*He sneezed and it got on my sleeve*), inferables referring to a closely preceding proposition (*this* in *This seemed to be the consensus view*), and referents anchored via kinship (*my brother*), attribute (*its size*), or body part possession (*her foot*). The first three kinds of inferables are more difficult and there might be less agreement between coders.

	evoked+inferable	non-evoked	Total
unraised	72	27	99
raised	373	190	563
Total	445	217	662

Table 3.44: χ^2 -table for reduced set of inferables

If we therefore collapse the first three kinds of inferables with the non-evoked referents, but keep the other inferables and group them with the evoked referents, then we get the distribution shown in Table 3.44 for the pooled data (cf. Tables 3.29 and 3.30).

A χ^2 -test yields an insignificant result with $\chi^2 = 1.602$ and $p=0.2056$ for $df=1$. A closer inspection of the cell values of χ^2 finds the same result as earlier: there are fewer evoked referents in the raised construction and more non-evoked ones than we would expect if the distribution was proportional. Also, there are more evoked referents in the unraised construction and fewer non-evoked ones than we expect for a perfectly proportional distribution. Omitting certain kinds of inferables that are harder to code thus has not led to a result more similar to Achard's for French.

A third possible explanation might be that the samples I collected just happened to fall outside of the normal range and that larger samples could still confirm Achard's findings. This view is contradicted by the findings for the pooled data, which did not result in confirmation of Achard's findings even though the greater amount of data might have established as reliable differences certain tendencies that did not result in statistically significant results for the individual lexical units. The argument would have force, however, on the assumption that the pooling of the data was not appropriate and that differences between the individual lexical items are to be expected.

Though it is not clear on what basis this could be asserted—in fact, work on verb classes seems to tacitly assume uniformity of the members both with regard to tokens and with regard to types—we can explore this idea by extending the sample for one of the lexical units. In particular, I propose to expand the sample of written *seem* by adding instances of the unraised construction. This is motivated by two

considerations. First, Achard worked on French *sembler* in newspaper data and written *seem* is the closest match for it. Second, since the data for the unraised construction with written *seem* is based on a lot fewer tokens than that for the raised construction (20 versus 89), it is more likely that the results for calculations of properties of the unraised construction might change if the number of tokens is increased.

The new extended sample I constructed includes all the original tokens for written *seem* as well as additional tokens of written *seem* in the unraised construction found among the first 800 randomized tokens of the original sample for the lexical unit. As a result of adding tokens in the unraised construction, I ended up with 89 tokens in the raised construction, the same number as in the original sample, and with 76 tokens in the unraised construction, as against 20 tokens in the original sample. Thus, extracting unraised sentences from a four times greater randomized sample (800:200) has yielded a circa four times greater number of tokens of the unraised construction (76:20).

The results of the analysis of the extended sample in terms of comparisons between the raised and unraised constructions can be summarized as follows:

- Animacy: The raised construction has 41 animate and 48 inanimate subjects, the unraised one has 45 animate subjects and 31 inanimate ones. According to a χ^2 -test this does not constitute a statistically significant difference ($\chi^2=2.838$, $df=1$, $p=0.0921$).
- Subject length: Subjects in the raised construction are shorter than subjects in the unraised construction. A Wilcoxon-Mann-Whitney test yields $p=0.14$, with $W = 68887.500$ and $Z = -2.464$. This result differs from what was found for the sample of spoken *seem* and the pooled data but it falls in line with the vast majority of results showing no significant differences between the raised and unraised constructions.
- Subject morphology: There is no significant difference between the constructions in terms of having a subject with definite or non-definite morphology, as

	raised		unraised	
	n	%	n	%
1st/2nd Pronoun	8	9.0	8	10.5
3rd Pronoun	20	22.5	18	23.7
that	2	2.2	0	0.0
this	3	3.4	0	0.0
This NP	4	4.5	3	3.9
Possessive NP	4	4.5	4	5.3
Proper Noun	10	11.3	6	7.9
headless wh-clause	0	0.0	2	2.6
Definite NP	13	14.6	18	23.7
Intermediate Total	64	71.9	59	77.6
Quantified NP	4	4.5	6	7.9
Bare/Plural NP	11	12.4	5	5.5
Indefinite NP	1	1.1	3	3.9
negative pronoun	0	0.0	1	1.3
restrictive relative pronoun	2	2.2	0	0.0
pronoun of free relative	2	2.2	0	0.0
expletive	3	3.4	1	1.3
zero in that-relative	1	1.1	0	0.0
VP _{to}	1	1.1	1	1.3
Total	89	100.0	76	100.0

Table 3.45: Subject morphology for extended written *seem* sample

can be seen from Table 3.45. In the raised construction 64 out of 89 sentences have a subject with definite morphology, and in the unraised construction 59 out of 76 do. A χ^2 -test yields $\chi^2=0.707$ and $p=0.4004$ at $df=1$. Distinguishing only between personal pronouns and all other types of XPs does not result in a significant skewing of the distribution either. With 28 personal pronouns in the raised construction and 26 in the unraised construction, a χ^2 -test yields $\chi^2=0.141$ and $p=0.7073$ at $df=1$.

- Prior Evocation: Of the tokens in the raised construction, 43 had a prior mention, 46 did not; for the tokens in the unraised construction, the relevant num-

bers are 38 and 38, respectively. There is no significant difference according to a χ^2 -test ($\chi^2=0.047$, $df=1$, $p=0.8284$). An analysis that collapses inferables and evoked referents and contrasts that class with all other referents also yields no significant results. The evoked+inferable class includes 76 out of a total 89 tokens in the case of raised sentences, and 64 out of a total 76 tokens in the case of unraised sentences. A χ^2 -test yields 0.045 and $p=0.8320$ at $df=1$.

- Subsequent evocation: There is no significant difference between the constructions. Of 89 tokens in the raised construction, 39 have a subsequent mention and 50 do not. In the unraised construction, 42 tokens have a subsequent mention and 34 do not. A χ^2 -test yields $\chi^2=2.148$ and $p=0.14275$ at $df=1$.
- Number and distance of contextual mentions: Wilcoxon-Mann-Whitney tests to detect differences between the constructions in the number and distance of prior and subsequent mentions as well as in the distance of last subsequent mentions yield insignificant results, as shown in Table 3.46

	Prior mentions		Subsequent mentions		
	number	distance	number	distance	distance of last
Mann-Whitney U	3248.000	3339.500	3291.000	3311.500	3194.500
Wilcoxon W	6174.000	7344.500	7296.000	6237.500	7199.500
Z	-.473	-.150	-.317	-.246	-.667
p (2-tailed)	.636	.881	.751	.806	.505

Table 3.46: Wilcoxon-Mann-Whitney tests for contextual mention measures of extended sample for written *seem*

Overall, it seems the expanded sample for written *seem* does not give a result that is closer to what Achard found. The various measures that we have compared across the two constructions differ from the results for the smaller sample in only one case, subject length with spoken *seem* and the pooled data.

The result for the extended sample in combination with the earlier result we got for analyzing the pooled data with respect to a modified set of inferred and evoked referents does not provide evidence that my original samples might have been too small, or that the distinction between different evocation statuses I adopted was different from Achard's. Both might still be the case but it is ultimately not possible to make that determination given that the construction of Achard's sample and the coding of the tokens is not sufficiently documented.

However, before we conclude that the distribution of the English constructions is not influenced by the discourse status of the participants, we can use one more analytical tool, logistic regression, to check if it might not be the case that combinations of the factors we have so far explored singly can explain the observed pattern.

Logistic regression The dichotomous dependent variable, that is the outcome we want to predict, is the choice between the raised and the unraised constructions, and the factors we can use to try to predict the construction are the ones we described individually above. In addition, we can create factors that take interactions between individual variables into account.

In our first attempt at finding a good regression model, we will apply logistic regression to the pooled data, using as predictor variables the ones that we ought to consider as relevant based on Achard's research: the binary evocation status, the number of prior mentions, and a variable representing the interaction between these first two, as there is a necessary correlation between statuses like inferable or unused and a count of 0 for the number of prior mentions.

Since we are testing a pre-existing model rather than trying to explore the data, we will use the forced entry method of including all the predictors into the regression model at the same time. There are also two other, so-called step-wise methods of entering predictors into a regression model. They can roughly be characterized as follows. In the forward method, the analysis begins with only a constant, which represents the relative frequency of the more frequent outcome, and then adds single predictors into the model based on a specific criterion, the value of the so-called score statistic. Analysis proceeds until no predictor variable with a significant score statis-

tic remains. In the backward method, analysis begins with all predictors included. Subsequently, predictors are removed if their removal does not result in a substantial reduction of the goodness of fit between model and data. The step-wise methods are recommended for exploratory work.

In the first step of the regression, the model uses only information about the relative frequency.⁹ The data in our model consists of 563 tokens of the raised construction and 99 tokens of the unraised construction. Therefore, based on frequency information the model should always bet on the raised construction. This by itself account for 563/662, or 85% of the tokens. At the end of step one, score statistics are calculated for the variables that have not been inserted into the model yet. These are shown in Table 3.47.¹⁰

		Score	df	p
Step 0	prior evocation status (binary)	4.003	1	.045
	number of prior mentions	2.514	1	.113
	interaction of above	2.514	1	.113

Table 3.47: Variables not in the equation

The table suggests that only one of the variables, the binary prior evocation status, will lead to an increase in the predictive power of the model since its p-value is smaller than 0.05. However, when the variable is actually inserted into the model, it turns out not to be a significant predictor of what construction occurs, as shown by Table 3.48. Note that the variable representing the interaction was eliminated as redundant and is not represented in the table.

Thus, the only factor that has a significant predictive value according to the model is the constant, which represents the relative frequency of the constructions. By conventional statistical criteria, prior evocation status is not a significant predictor, by a slight margin.

⁹The use of relative frequency in the regression is the reason that here we include only the original data for written *seem* rather than the data used in the extended sample analyzed above. The construction of that extended sample led to an over-representation of tokens of the unraised construction.

¹⁰Residual Chi-Squares were not computed because of redundancies.

	B	S.E.	Wald	df	p	Exp (B)	95.0% C.I. for Exp (B)	
							Lower	Upper
prior evocation status	-.591	.345	2.942	1	.086	.554	.282	1.088
number of prior mentions	-.033	.030	1.173	1	.279	.968	.912	1.027
Constant	2.303	.316	53.019	1	.000	10.000		

Table 3.48: Variables in the equation after step 1

Redoing this analysis with the number of prior mentions and an evocation variable that represents not just a binary distinction but a three-way distinction between evoked, inferable, and non-evoked leads to an even poorer model with worse significance scores for the variables than in the regression model above.

These findings are not too surprising given the results of the analyses for individual formal and pragmatic features. Since we did not find any of those features to differ significantly in a systematic fashion across constructions for each lexical unit or the pooled data with the exception of a single significant finding for subject length, the only features that we could reasonably expect as predictors were interactions anyway. But these do not seem to be predictive either. We thus have found no good, consistent evidence to support an analysis that associates the choice between the constructions in English with a difference in the discourse-status of the subject referents.

3.1.6 Rhetorical structure

A possible explanation for the disparity between Achard's and my findings might be this: the difference between the constructions does not primarily lie in the discourse status of the referents involved but in what we may call the rhetorical structure of the discourses they are used in. Recall from section 3.1.2 that Achard broadly speaks of conceptual distance as the factor that explains why topical subjects exceptionally appear in the unraised construction. It might be the case that conceptual distance is not just an explanation for the exceptions but the primary explanation for most of

the English data, with subject topicality having only a secondary status, if any.

In order to get a sense of this, I coded the tokens of the extended written *seem* sample with the following three additional features, which are inspired by Achard's characterization of his French data.

- (3.11) Rhetorical contrast: This feature codes whether or not the proposition in the scope of *seem* contradicts or otherwise conflicts with another proposition that is explicit or implicit in the context.
- (3.12) Closeness to paragraph boundary: For the most part this is a purely formal criterion, noting whether the sentence token appears at the beginning or end of a paragraph. In texts that were part of transcribed conversations I took a closer look at the content of the speech to see if an utterance constituted an episode boundary.
- (3.13) Point of view shift: This feature records whether the proposition in the scope of *seem* shifts the point of view, for example, from one story character to another, or from a cited source to the author of a newspaper article.

Not having any prior reason to weight them otherwise, I decided to combine these three factors by assigning the value 1 to the presence of a feature and 0 to its absence. The combined score could thus vary from 0, lack of contrast, to 3, maximal contrast. The cross-tabulation of constructions against scores for the extended sample of written *seem* is shown in Table 3.49. The table shows that the tokens in the unraised construction tend to involve greater contrast than the tokens in the raised construction. The average values for the two constructions confirm this: 0.494 for the raised construction, and 1.382 for the unraised construction. Nonetheless, it is clear that there is no categorical distinction involved: both constructions occur with scores of 0, 1, and 2; only the maximum contrast score of 3 is associated exclusively with one of the constructions, the unraised one.

If we perform a χ^2 -test on the distribution of the contrast scores across the two constructions in Table 3.49 above, we find $\chi^2 = 39.618$ at 3 degrees of freedom, which gives p less than or equal to 0.001. The distribution is significant. Let us look

Score	raised	unraised	Total
.00	56	14	70
1.00	22	29	51
2.00	11	23	34
3.00	0	10	10
Total	89	76	165

Table 3.49: Contrast scores for extended sample of written *seem*

more closely at how the χ^2 -value is built up. The biggest contributors are the cells for row 1, which corresponds to a contrast score of 0. Here the raised construction has many more tokens with this low contrast score than expected and the unraised construction has fewer than expected. The two cells contribute 8.8 and 10.3, which adds up to 19.1. The next biggest contributors are the cells for the highest score, 3. They contribute $5.4 + 6.3 = 11.7$ to the total. In particular, it is the case that the unraised construction has more tokens with such a high contrast score than expected while the raised construction has too few. The next highest contributors are the cells for a score of 2 and the smallest contribution is made by the cells for a contrast score of 1.

Assuming that the chosen way of scoring is acceptable, we can conclude that the two constructions differ as to the degree of contrast they are associated with. The question now becomes what the status of the rhetorical contrast is. In Achard's account, the semantic difference between the constructions is primary. Its main pragmatic correlate is said to be the topicality of the subject referent. Cases where this correlation seems to be violated are treated much like exploitations of Gricean maxims: for instance, using a topical referent in the unraised construction signals that the hearer should increase the conceptual distance over what the formal coding alone would suggest. The increased distance then leads to the perception of a contrast.

While this kind of account might still make sense for Achard's French data, where 75.6% of tokens in the raised construction have a topical subject and only 47.4% of tokens in the conflated construction, it does not seem to fit the English situation very well. For one, in English the relative shares of the discourse statuses among

the tokens are statistically not distinguishable between the two constructions. In fact, the percentage of topical referents was actually slightly higher in the unraised construction (88.8%) than in the raised construction (80.5%). This is far different from the distribution that Achard observed in his French sample: if we maintained the same analysis, then the vast majority of tokens in the unraised construction would constitute exploitations of the “basic” pattern, which consequently ought to be very difficult to learn. For this reason, we may want to analyze the English data without appealing to the discourse pragmatics of the subject referents. Instead of saying that the primary correlate of the semantic difference between the construction is the discourse status of the subject referent, we can say that the primary correlate is just a notion of rhetorical contrast, which is independent of identifiability or familiarity of any referents involved. Propositions may contrast either because they involve different logical subjects or different logical predicates.

3.1.7 Semantics

Now that we have explored the discourse pragmatics of raised and conflated sentences it is worthwhile turning to their semantics. According to Achard’s analysis, both constructions have to do with inserting a proposition into reality, though in different ways. While the unraised construction treats the proposition at issue as “embedded in a sub-section of reality that enables the speaker to identify it as a possible part of that reality” (p. 2), in the raised construction, “the main character in the located event is taken as the trajector of the main relation due to its focal prominence. The process she participates in is viewed as the landmark of that relation. Unlike the case with the unraised construction, reality remains an un-profiled part of the base” (ibid.). Crucially, the difference here is one of profiling: that *sembler/seem* is an epistemic verb that evaluates the possible occurrence of a given event or proposition in reality is true for both constructions.

The fact that the main difference between the constructions is cast as focal prominence agrees with the thrust of the discourse pragmatic argument Achard makes: focal prominence in scene construal seems to hook up quite naturally with topicality

in ongoing discourse.

However, at least for the English verbs *appear* and *seem* we can expand Achard's account and ask the question why focal prominence falls where it does in the two constructions. The answer is similar in spirit to Borkin's analysis of the complement types for verbs like *seem* and *appear* and others like *find*, *imagine* etc. The main thrust of her analysis is that the progression of that-clause; to-infinitive; and bare infinitive or predicative adjective/noun/prepositional complement parallels "a semantic movement from an empirically oriented or discourse given proposition toward a matter of personal experience, individual perception, or a conventionally determined state of affairs" (Borkin 1973: 44).

A difference of this kind among the constructions seems to be reflected by the presence of three semantic elements that preferentially occur with the unraised construction. These are (a) expressions denoting a Cognizer; (b) expressions denoting an Evidence frame element; and (c) the modal *would*. These three elements occur with the unraised construction because it is not associated with direct perception: (a) assertions that are not based on one's own perception more often need indications of what the Evidence is; (b) indications of whose reasoning goes into making the assertion; (c) and indications that they are made with less certainty than assertions based on perception. I will discuss the distribution of these three elements now in turn.

Presence of Cognizer The first element is the presence of an overt Cognizer, which I had coded for both the written and spoken *seem* and *appear* data. With *likely* and other predicates expressing notions related to likelihood or probability, this frame element does not occur because likelihood is presented as a kind of objective fact that does not depend on anybody's perception (**It is not likely to me that there will soon be another volcanic eruption as great as, say, Krakatoa in 1883*).

Note that the presence of a Cognizer is of interest to the rhetorical account sketched above in as far as explicit mention of a Cognizer might be made more often in contrastive situations than in non-contrastive ones. As Table 3.50 shows, the Cognizer role is usually not overtly instantiated in the text, even though there

		yes	no	Total	p-value of Fisher's exact
appear (spoken)	unraised	4	17	21	0.054
	raised	7	118	125	
	Total	11	135	146	
appear (written)	unraised	1	16	17	0.4386
	raised	3	108	111	
	Total	4	124	128	
seem (spoken)	unraised	12	7	19	0.000
	raised	6	96	102	
	Total	18	103	121	
seem (written)	unraised	3	17	20	0.0420
	raised	2	87	89	
	Total	5	104	109	
pooled	unraised	20	57	77	0.000
	raised	18	427	445	
	Total	38	484	522	

Table 3.50: Presence of Cognizer

necessarily is an Experiencer or Cognizer in the scene.

Though the number of explicit Cognizers is small per lexical unit and also in the pooled data, it is the case that they occur disproportionately often in the unraised construction. This is true for the individual lexical units except written *appear*. It holds more strongly for the pooled data.

Evidence Another feature of unraised *appear* sentences that struck me as noteworthy was the fact that, unlike the raised ones, they seemed to more often include Evidence frame elements, usually *from*-PPs as in (3.14).

(3.14) It would **appear** from the report of Hudson that the defendant had simply overbalanced in his excitement.

I decided to take two new random BNC samples for the verbs *seem* and *appear* where each sentence would contain one of the verbs together with a *from*-PP. Among

1000 inspected sentences for *seem* there were 29 where the *from*-PP was an Evidence-PP construed relative to the verb. Of these 29 sentences, 18 exhibited the unraised construction, 9 the predicative construction, and 2 the raised construction. In the case of *appear*, there were 56 Evidence-encoding *from*-PPs in a sample of 1000 sentences. 43 of the 56 sentences were in the unraised construction, 6 in the raised construction, 3 in the predicative construction, 2 in a parenthetical construction, and 2 sentences occurred in the type of existential construction shown in 3.15

- (3.15) The provision . . . may be amended or revoked . . . by the parties to the treaty without the consent of the State entitled to the right . . . , unless—(a) the parties to the treaty had entered into a specific agreement with the latter with regard to the creation of the right ; or (b) a contrary intention **appears** from the terms of the treaty , the circumstances of its conclusion or the statements of the parties .

Clearly, the distribution of constructions with uses that contain Evidence-denoting *from*-PPs is very different from what one finds when one just looks at *seem* and *appear* sentences in general.

We may compare Evidence expressions with expressions such as *at first*, *at first glance*, and *at first sight*. These are similar to *from*-PPs in drawing explicit attention to the act of perception or judgment but they differ in that they do not introduce any specific Evidence. Table 3.51 shows the distribution of sentences containing such an expression construed relative to the target lexical units *appear* and *seem*. Both data sets come from an initial sample of 500 randomly chosen and manually inspected sentences.

In the case of *appear*, there is a statistically significant difference between the distribution of constructions for *appear* generally and the distribution of *appear* when it co-occurs with *at first*. A χ^2 -test for Table 3.51 yields a χ^2 -value of 6.496 at $df=2$ and $p \leq .05$. However, closer inspection shows that the majority of the χ^2 -value is contributed by the cells representing the raised and predicative constructions. Only slightly greater than 1.06 is contributed to the total χ^2 -value by the cells for the unraised construction.

	appear at first (glance/sight)	appear	seem at first (glance/sight)	seem
unraised	11	17	10	20
raised	36	111	34	89
predicative	23	33	79	79
Total	70	161	123	188

Table 3.51: *At first (glance/sight) with seem and appear*

With *seem*, too, there is a statistically significant difference between the distribution of constructions for *seem* generally and the distribution of *seem* when it co-occurs with *at first*. A χ^2 -test for Table 3.51 yields a χ^2 -value of 15.00 at $df=2$ and $p \leq .001$. As with *appear*, closer inspection shows that the majority of the χ^2 -value is contributed by the cells representing the raised and predicative constructions. Only slightly greater than 0.48 out of the total of 15.00 is contributed to the total χ^2 -value by the cells for the unraised construction.

Thus, with both lexical units, *at first* occurs as often in the unraised construction as one might expect it to based on the probabilities of the unraised construction occurring and of *at first* occurring with *appear*. It is the relative frequencies of the raised and the predicative constructions that differ markedly between samples for the lexical units in general and samples for the lexical units where they co-occur with *at first*.

We can thus conclude that *at first* {*glance/sight*} has a different distribution across the raised and unraised constructions than *from*-PPs. What is associated with the unraised construction is specifically the notion of Evidence instantiated by *from*-PPs rather than just any kind of element that draws attention to the act of perception or judgment.

Would Another feature that is more difficult to make sense of is the presence of the modal *would*. Working with the data, I had gained the impression that the sequences *would seem* and *would appear* occurred disproportionately frequently with the unraised construction. A typical use is presented in (3.16).

	would seem		seem		would appear		appear	
	n	%	n	%	n	%	n	%
unraised	54	27.0	20	10.5	73	42.7	17	10.2
raised	91	45.5	89	46.6	79	46.2	111	66.5
predicative	41	20.5	79	41.3	12	7.0	33	19.8
other	14	7.0	3	1.6	7	4.1	6	3.6
Total	200	100.0	191	100.0	171	100.0	167	100.0

Table 3.52: Distribution of constructions for the sequences *would seem* and *would appear*

(3.16) It would **seem** that great pains were taken to ensure that the potential division among Incomer/Shetlander lives would not be spoken about , and that if it was spoken about in a public forum , the arrival of incomers would be made to look less ominous. This can be gauged from the Dunrossness District Plan, published prior to the Structure Plan.

I decided to explore this question in a separate sample from the main data. For both lexical units, I extracted all sentences from the British National Corpus that contained the sequence *would*+verb, randomized them, sampled 200 from the randomized set, and classified them according to the construction that they participated in. The results are summarized in Table 3.52, which also repeats the distributions for the full written *seem* and *appear* data from Table 3.3 above.

χ^2 -tests confirm the impression that the *would seem* sequence is significantly different compared to all uses of *seem* regardless of modalization by *would*. A comparison just involving the raised and unraised constructions yields a χ^2 -value of 10.758 with $p = 0.0010$ at $df=1$ for *seem*, and $\chi^2=38.459$ with $p= 0.000$ at $df=1$ for *appear*. χ^2 -tests involving all four construction types result in significant results as well.

In order to ascertain whether *would* specifically rather than the presence of a modal was responsible for the greater frequency of the unraised construction, I decided to compare the distribution of constructions in a random sample for *may/might seem*. The data is summarized in Table 3.53.

The situation with *may/might seem* is very different from that with *would seem*. In the case of *may/might seem*, the predicative construction is the most frequent

	may/might seem		may/might appear	
	n	%	n	%
unraised	12	6.0	10	6.1
raised	26	13.0	65	39.4
predicative	150	75.0	84	50.9
other	12	6.0	6	3.6
Total	200	100.0	165	100.0

Table 3.53: Distribution of constructions for the sequences *may/might seem* and *may/might appear*

construction. The unraised and raised construction have the same relative order as in the case of *would seem* but both occur at much lower rates. *May/might seem* sentences are also different from *seem* sentences in general, where raised sentences are much more common than with *may/might seem*. With *may/might appear*, the situation is similar. The predicative construction is a lot more frequent than with either *would appear* or *appear* in general. The unraised construction is slightly less frequent for *may/might appear* than for plain *appear* and much less frequent than with *would appear*.

We can conclude that sequences involving the modals *may/might* and *would* have significantly different distributions than *seem* and *appear* do in general. *Would* sentences exhibit a higher number of unraised constructions, whereas *may/might* sentences show a drastically higher number of predicative sentences. Thus, it is specific modals that are associated with particular constructions rather than all modals being associated with one of the constructions qua modal.

It is not clear what the meaning contribution of *would* in sentences with *seem* and *appear* is. It clearly is not a habitual form since the verbs are stative. As I perceive it, the overall effect is to weaken the strength of the commitment the speaker makes to the truth of the proposition in the *that*-clause. This goes along with the fact that *would V* sentences occur at beginning or the end of a passage where evidence is considered that warrants coming to the conclusion that is expressed in the *that*-clause. This is the case, for instance, in example (3.16) above, where the last sentence of the

paragraph provides the transition to a new paragraph that presents detailed evidence for the claim in the *that*-clause complement of *seem*. Thus, *would V* sentences occur where one is concerned with reasoning or making inferences from evidence other than direct sensory perception.

Negation A fourth semantic difference between the constructions is that the raised construction occurs more readily with negation than the unraised one. A 400-sentence random sample containing the lexical unit *seem* included 202 sentences in the raised pattern and 43 in the unraised pattern. In the raised pattern 19 (10.4%) sentences had sentence-level negation, while there were no such sentences in the unraised pattern.

seem	Affirmative	Negated	Total
unraised	43	0	43
raised	183	19	202
Total	226	19	245
likely	Affirmative	Negated	Total
unraised	42	0	42
raised	355	13	368
Total	397	13	410

Table 3.54: Predicate affirmation and negation of *seem*

The difference between the construction with regards to negation on the level of the epistemic predicate is statistically significant ($\chi^2=4.385$, $p= \leq .05$). It is not the case that sentences of the type shown in (3.17) are ungrammatical, they just happen to be very rare: the sequence *not seem that* occurs only 14 times in the BNC, whereas the sequence *not seem to* occurs 953 times.

(3.17) It did **not seem** that I could survive such a volcanic attack from deep within my own system.

Table 3.54 also shows results for *likely* based on an initial random sample of 500 sentences taken from the BNC. Since the frequency asymmetries between the raised and unraised construction and between affirmative and negatives are as large as they

are, the data in the table do not give a significant χ^2 -result. The trend however seems to go clearly in the same direction as for *seem*: negation of the epistemic predicate is exceedingly rare when the complement is a finite clause in the unraised construction. Given a large enough sample size, one should find a significant difference in terms of negation between the constructions for *likely*, too.

When it comes to negation of the predicate in the post-verbal complement of *seem*, the preference seems to be the other way around.

seem	Affirmative	Negated	Total
unraised	39	4	43
raised	200	2	202
Total	239	6	245
likely	Affirmative	Negated	Total
unraised	41	1	42
raised	368	0	368
Total	409	1	410

Table 3.55: Affirmation and negation of the predicate in the clausal or verb phrase complement of *seem*

Negation is relatively more frequent for finite clause complements occurring with the unraised construction than for VP complements occurring with the raised construction. Although the numbers involved are small, according to a χ^2 -test, the difference is still significant for *seem* ($\chi^2 = 10.253$, $p \leq 0.01$). For *likely*, no valid test can be performed due to low cell values (< 5). The trend, however, seems to be same as for *seem*: there are fewer negations of the predicate of clausal complements in the unraised construction than of the verb phrase complement in the raised construction.

The distribution of negation is not surprising—introspection would probably lead most speakers to the same conclusions about its relative frequency with the two constructions. Nevertheless, it is not easy to know what to make of the asymmetry. One possible explanation may be that negation occurs most naturally on the predicates of topic-comment sentences, insofar as speakers tend to negate to correct beliefs of others about some agreed upon topic and some set of presuppositions.¹¹ Raising is a

¹¹The explanation offered here was suggested to me by Laura Michaelis in personal communication.

topic-comment construction despite the fact that the subject is not assigned a thematic role by the main verb, so it makes sense that one should find negation to be more prevalent in the raising pattern than in the extraposition pattern, whose post-verbal clause, its 'logical subject', is not a topic but a focus. Note, however, that the focal status of the proposition as a whole does not mean that the arguments of the predicate on which the proposition is built have to be non-topical. The focality of the proposition as a whole is still compatible with its predicate having a topical subject.

I will leave attempts at a more detailed explanation of the negation data for future research.

3.1.8 Summary

The suggested semantic difference between the raised and unraised constructions with *appear* and *seem* is subtle. But there are some features that are clearly and reliably oriented towards the unraised construction, namely the frame elements Evidence and Cognizer and the modalization with *would*. Appealing to Borkin's analysis, we can attribute these facts to the greater distance between the conceptualizer and the proposition: where an assertion is not based on perception, indications of Evidence, of the identity of the Cognizer, and hedging become more important. The difference between perception and reasoning may also be relevant for the fact, noted earlier in section 3.1.4, that the unraised construction is used disproportionately often when the embedded event is temporally anterior to the time indexed by the raising predicate: a simultaneous state of affairs is more readily open to perceptual apprehension than (the aftermath) of an event or state that is in the past. Such anterior states or events are, however, quite naturally objects over which a Cognizer may be reasoning.

An analysis in these terms also can also be related to Achard's account in terms of conceptual distance and focal prominence: where there is no direct perception, focal prominence on a participant in the scene is not appropriate.

Where does this leave the analysis of other raising predicates such as *likely* or *certain* that do not occur with expressions denoting Evidence or a Cognizer? For instance, for all cases where an Evidence *from*-PP occurs in a sentence with *likely* in

the BNC, we find that both the PP and *likely* are dependents of either *seem* or *look*. The adjectival raising predicates by themselves do not license Evidence expressions. However, even in those cases the notion of evidence necessarily plays a role: to assess probability, one needs some kind of knowledge or evidence of how the system behaves in different circumstances. Consequently, the choice between constructions may here still reflect how close the conceptualizer is to the basis of their reasoning, even if the notion of evidence is not given linguistic expression as syntactic dependent of the predicate.

The above characterization can serve at least for the prototypical cases, where speakers have a strong preference for one of the two constructions over the other. It may well be the case that the difference between the constructions is being eroded in English. Achard's numbers for French suggest 2 raised uses per 1 unraised use; in English the ratio seems to be about 6.5 to 1.

With respect to our original question, whether there is a good correlation between the discourse status of the subject referent and the choice of construction, we have come up with a negative finding both for the measure that Achard had used, prior evocation, and several others that are known to be relevant to other constructions. The only promising correlation that we have found had to do with the notion of rhetorical contrast. But it is not clear whether the main non-semantic correlate of the distinction between the construction in English just happens to be rhetorical rather than discourse pragmatic or whether this reflects a reanalysis resulting from the loss of a clear semantic contrast that used to be accompanied by the kind of discourse pragmatic difference that Achard found for his French data. These are questions for historical linguistics that cannot be resolved here. We will now, however, consider the properties of other raising-type constructions in order to assess whether the findings for the epistemic raising construction seem unusual. Maybe at least these other raising constructions exhibit the discourse pragmatics predicted by Langacker and Achard.

3.2 Possessor raising

Both Langacker and Achard suggest that the notions of reference point and active zone apply equally well to inter-clausal raising, that is subject-to-subject or subject-to-object raising, and to possessor raising constructions such as the body part possessor raising construction in (3.18).

- (3.18) a. Kim **slapped** Pat's face. (conflated)
 b. Kim **slapped** Pat across the face. (raised)

According to Langacker's analysis, in body part possessor raising sentences, the Possessor functions as a reference point facilitating access of the Body Part referent. In (3.18a), the object NP refers directly to the active zone of the verbal process, *Pat's face*. In (3.18b), by contrast, the Possessor is given the status of an independent clausal argument of the verb and the Body Part, which is still the active zone of the slapping process, is realized as an oblique argument. Furthermore, under the reference point account, the Possessor of the Body Part, which for brevity's sake I will henceforth call Possessor, should be a local topic in the possessor-raised construction where Possessor and Body Part are expressed separately but not in the simple construction, where the two are expressed in a single constituent.

3.2.1 Possessor/Attribute factoring alternations

Though it is probably the most prominent type, body part possessor raising is not the only type of possessor raising construction. Levin (1993) lists a group of so-called possessor-attribute factoring alternations following her entry for the body part possessor raising alternation. Although she does not make an explicit connection in her writing between body part possessor raising and this group of alternations, the juxtaposition and the semantic similarity centered on the broad notion of possession suggests that ultimately these alternations are similar.¹² Langacker provides no discussion of these other possessor-attribute factoring alternations but it seems to me

¹²In a 2003 talk to the Berkeley linguistics department, Levin explicitly grouped all these raising constructions together.

that these alternations would presumably also fall under the reference point-active zone analysis. The relevant alternations include:

- Judgement verbs, which allow separate or joint expression of an Evaluatee and a Reason, which may be an attribute of, or an action taken by the Evaluatee. When the Evaluatee and the Reason are expressed disjointly, the Evaluatee appears as the direct object.

- (3.19) a. I **admire** [you] [for your courage]
 b. I **admire** [your courage]

Levin distinguishes cases like (3.19a), which she calls Possessor object, from what we might call Reason object cases such as (3.20a). There, Evaluatee and Reason are referred to disjointly, with the Reason appearing as the direct object.

- (3.20) a. I **welcome** [ambition] [in new employees].
 b. I **welcome** [their ambition]

- Some verbs such as *admire* can participate in both types of construal. One fact that goes unmentioned by Levin is that in the case of joint expression the head noun does not have to be Attribute-denoting. An alternative way of conflating the Possessor and the Attribute consists of adding a modifier expressing the Attribute to the Possessor head noun as in (3.21)

- (3.21) a. We **welcome** [curious students]
 b. He seems too quick to **condemn** [people who criticize him]

- The alternations illustrated above for the object slot are also found in the subject slot with many transitive experiencer-object verbs, where a Protagonist and an Attribute/Act can be expressed jointly or disjointly. With these verbs, too, NPs that jointly express Protagonist and Attribute/Act may be headed by a noun of either of the two semantic types. When the two notions are expressed separately, it is the Protagonist that occupies the subject slot as shown in (3.22a), parallel to the Possessor object case seen above in (3.19a).

- (3.22) a. [Mark] **terrified** me [with his single-mindedness]
 b. [Mark's single-mindedness] **terrified** me
 c. [People of such single-mindedness] **terrify** me.

The only construction that seems to be able to produce a disjoint expression with the Attribute as subject consists of free-relative clauses, where the Possessor can be encoded in an about-PP:

(3.23) That is [what] **annoys** me [about him]!

(3.24) I haven't found [anything that] **annoys** me [about him].

(3.25) But [the one thing that] **annoys** me [about him] is that he talks S O DAMN MUCH.

- Another major semantic class allowing a possessor-attribute-type alternation consists of verbs denoting a change of state with respect to a scalar property and, mostly, adjectives expressing a ranking on a scale. With this kind of predicate we find the two participants expressed separately with the Possessor as subject as in (3.26a) and (3.26b); expressed separately with the Attribute as subject as in (3.26c); expressed as a single constituent as in (3.26d) and (3.26e).

- (3.26) a. Among the top 50 drugs sold to seniors, [more than half (27)] **rose** [in price] at least three times the rate of inflation from January 2002 to January 2003
- b. [Some cereals] are **high** [in fat], so check the nutrition information panel
- c. But, you say, [prices] are **rising** [for health care, college tuition and home heating].
- d. But this time, the deal seemed to work and [the price of oil] **rose** and rose
- e. Salmon or tuna is the best type of fish to consume due to its high omega 3 essential fatty acid and low toxicity, though [its fat level] is quite **high**.

However, these predicates do not allow joint expression with the Possessor as head of the NP.

(3.27) *(The) Oil of that price has been **rising** considerably.

- Another class of predicates that allows a kind of possessor raising involves adjectives that evaluate a person's mental properties on the basis of specific acts, or on the basis of their general behavior. This class is not covered by Levin since the predicates are not verbs. The predicates participating in the alternation have many of the same possibilities for syntactic expression of the semantic roles as are found with verbs. Consider the adjective *smart* in (3.28).

- (3.28) a. [You] were **smart** [to sell midday for the profits]
 b. [Selling midday for the profits] was **smart** [of you]
 c. [To sell midday for the profits] was **smart** [of you]
 d. [Your selling midday for the profits] was smart.
 e. [Those who sold midday for the profits] were **smart**

Example (3.28a) shows the disjoint expression with the Evaluatee as subject and the Action as post-predicate complement. Examples (3.28b-c) show a different kind of disjoint expression with the Action as subject and the Evaluatee as oblique complement. Finally, examples (3.28d-e) show the conflated presentation of Evaluatee and Action in a phrase headed by the Action and a phrase headed by the Evaluatee, respectively.

I decided to test Langacker's claim that raising and possessor raising are both examples of active zones and reference points for two of these constructions. The first was the body part possessor raising construction, the construction that for most people is a clear case of possessor raising. O'Connor (1996) in fact conjectures that an implicational universal exists whereby a language that has any kind of possessor raising also allows the use of the construction with the Possessors of Body Parts. The second construction I investigate is the change property factoring alternation illustrated in (3.26) above, which involves verbs such as *rise*, *drop*, *soar*.

3.2.2 Body part possessor raising

While we will be mainly concerned with the discourse-pragmatics involved in body part possessor raising, it is appropriate to briefly revisit semantic and syntactic properties for the sake of comparison with the other alternations just considered in section 3.2.1.

Unlike some of the other cases of possessor-attribute factoring cited above from Levin, body part possessor raising involves disjoint expression with the Possessor as object and the Body Part as oblique and joint expression in an NP headed by the Body Part. It is not possible to have disjoint reference with the Body Part as object or joint reference with the Possessor as head of the NP, as is illustrated in (3.29).

- (3.29) a. *I **smacked** John {of the arm/with his arm}.
 b. *I **smacked** the arm {in/with} John.

Another important concern is the kinds of verbs that may participate in the alternation. Levin (1993), like Fillmore (1970), points out that in English body part possessor raising is not possible with verbs of breaking: **The horse broke Penny in the shin*. Likewise, the alternation is said to be possible with verbs of cutting but not with verbs of carving such as *bore*, *bruise*, *carve*, *chip*, *chop* etc. The alternation is possible with verbs of impact, touch, poking and cutting. Though scenes of body part possessor raising are often associated with negative impact on the Possessor of the Body Part, this is not necessarily so. Most verbs of touching are at least neutral in this regard and there are other verbs such as *kiss* and *stroke* that usually occur in describing scenes involving a positive experience by the Possessor of the Body Part.

(3.30) My sister **touched/tapped/patted** me on the arm.

(3.31) Kim **kissed** Pat on the forehead

(3.32) Elena McKean came up to me and **stroked** me on the cheek.

It is not clear to what degree the restrictions on the English body part possessor raising construction also hold cross-linguistically. In German, for instance, the

constraints on the body part possessor raising construction seem to be very similar. Verbs of breaking and carving cannot occur in the construction which introduces the the Body Part Possessor as a separate argument in the accusative case:

- (3.33) *Peter hat mich in den Arm **gebrochen**
 Peter has me.ACC in the arm.ACC broken
 'Peter broke my arm'
- (3.34) *Peter hat mich in die Leber **durchtrennt**
 Peter has me.ACC into the liver.ACC perforated

The German accusative construction seems to be the equivalent of the English construction, which is also suggested by formal facts: the Possessors in the accusative construction can become the subjects of passives, unlike certain datives appearing in a broadly used ethical dative construction that do combine with the above verbs of contact, impact, etc.¹³

Still, the German accusative construction is somewhat different from the English body possessor raising construction in allowing at least one additional verb class to participate in the alternation: verbs of bodily experience such as *schmerzen* 'hurt', *jucken* 'itch' can either take the Possessor of the Body Part as subject while expressing the Body Part as an oblique, or they can conflate both in a subject NP headed by the Body Part:

- (3.35) [Mein Arm] **juckt** wie bekloppt
 My arm is itching like crazy

¹³Examples of the ethical dative construction with the equivalents of the verbs of breaking and carving are given in (i).

- (i) ia. Peter hat mir den Arm **gebrochen**
 Peter has me.DAT the arm.ACC broken
 'Peter broke my arm'
- ib. Peter hat mir die Leber **durchtrennt**
 Peter has me.DAT the liver.ACC perforated
 'Peter perforated my liver'

- (3.36) [Der Fuß] **juckt** [mich] gar nicht.
lit. The foot itches me not at all
- (3.37) Es **juckt** [mich] [am Ohr und in der Nase]
lit. It itches me on the ear and in the nose.

It is probably the case that not all languages have a construction that corresponds closely to the English kind of body part possessor raising construction. O'Connor (1996) reports for Northern Pomo that there is a construction that is used for the equivalent of the English body part possessor raising but which also occurs in cases where contact with a person's Body Part is not at issue.

- (3.38) a. mi? ?e:-nam k'edi phit'a
'Your hair looks nice' (Its color or something else about the hair is pretty) [conflated]
- b. mito ?e: k'edi phit'a
'You look nice with that hairstyle' [raised] (O'Connor 1996:136, 24a-b)

O'Connor rejects an analysis that is at the general level of AFFECTEDNESS because it is unclear and there are cases of compliments such as in (3.38) where the concept with its negative overtones actually makes no sense. Still, the Northern Pomo construction does not seem to be a general ethical experiencer construction rather than a body part possessor raising construction. At least, that is what is suggested by the absence of examples like *The car died on me* in O'Connor's data. There has to be a body part involved.

Closely related to the issue of which verb classes appear in the construction is another major semantic question surrounding body part possessor raising constructions, namely what if any semantic difference exists between the two alternates. O'Connor (1996) discusses several factors that favor the use of the raised construction over the unraised one in Northern Pomo.

- The more-well known the Possessor is to the speaker, the more likely the possessor raising construction is.

- When the predicate is one of high impact, the possessor raising construction is more likely
- High general significance of the Body Part for the Possessor can also make the construction more likely (e.g. with *face*, possessor raising is likely in Northern Pomo).

Some of the notions invoked are difficult to quantify: what counts as impact or high effect? It may vary in context and certainly examples like (3.38) are difficult to evaluate in this regard compared to verbs with meanings such as ‘smack’ and ‘touch’. Likewise, it is difficult to assess the felt, embodied significance of Body Parts such as *cheek*, *forehead* and *neck* compared to more prominent ones such as *face*, *mouth*, or *eyes*. In language as represented by a corpus, some body parts are clearly mentioned more frequently than others. For instance, the lemma *cheek* occurs 3250 times in the BNC, *nose* 4703 times, *leg* 11840 times, *face* 28740 times, and *eye* 39162 times.¹⁴ But it’s not clear that the face is in any readily demonstrable way roughly 9 times more prominent overall than the cheek.¹⁵

In the end, O’Connor suggests that all that can be reliably said about the possessor raising construction is that it foregrounds the experience of and the consequences for the Possessor of the Body Part:

So for any verb used with an argument denoting a body part, there will be an interpretation of the meaning of that verb predicated of that particular body part, and there will also be a set of implications or consequences of that meaning for the body part possessor. This set will include physical, social, functional, and cultural consequences of that verb-body part combination for the body part possessor. As such, these consequences will be highly context-specific. (1996:144).

Crucially, O’Connor’s account does not make any claims about a difference in discourse pragmatic status of the Possessor of the Body Part across the two constructions.

¹⁴Senses of these lemmas are not distinguished in these counts, unfortunately.

¹⁵In fact, the *face* should be even more prominent given that many uses of verbs like *kiss* and *slap* do not mention a Body Part but leave it to the hearer/reader to interpret *face* as the default location.

	kiss		slap		smack	
	n	%	n	%	n	%
Conflated	41	24.4	20	18.5	40	19.0
Own Body Part	1	0.0	13	12.0	7	3.3
Body Part	13	7.7	7	6.5	0	0.0
Possessor	94	56.0	25	23.1	125	59.5
Raised	19	12.3	36	33.3	37	17.6
Total	168	100.0	101	100.0	209	100

Table 3.56: Distribution of tokens across constructions

3.2.3 English body part possessor raising

I investigated 200 randomly sampled instances each for the verbs *kiss* and *slap* as well as all 476 attested uses of *smack* in the British National Corpus. The uses of the verbs are distributed as shown in Table 3.56.

The totals suggest that *smack* and *slap* have more uses than *kiss* in which expression of a Body Part is not readily possible: 267 out of 476 sentences for *smack* and 99 of the 200 sampled sentences for *slap* were not concerned with simple acts of hitting another person or one's own Body Part. In some cases, this represents polysemy, e.g. *slap on a fine*; *smack one's fist down on the table*; in others, homophony, *smack* of corruption; in still others we may have the right sense of the basic word but another construction used is incompatible with expression of a Body Part, e.g. *slap somebody around*.

By contrast, only 32 of the 200 sampled sentences for *kiss* were not concerned with simple acts of one person kissing another but had to do with e.g. kissing an object; reciprocal kissing; or kissing as part of greeting (*kiss goodbye, goodnight*). With these particular kinds of kissing, expression of a body part is highly unlikely, if not altogether excluded.

While Langacker's analysis is concerned only with the conflated and the raised constructions, it is important to note that the verbs also appear in additional constructions, as shown in (3.39).

- (3.39) a. Pat **smacked** Kim. (Possessor)
- b. Margaret **kissed** the side of his cheek and went to gather daisies in the grass and make a daisy chain to lead her new friend and spectacular pet through all the illuminated and flower strewn pages of Christian manuscripts. (Body Part)
- c. My aunts and uncles **slapped** their knees. (Own Body Part)

In (3.39a), only the Possessor is overtly expressed but not the Body Part, which for all three lexical units *slap*, *smack*, and *kiss* is presumably understood as the face by default. These default locations do not have to be expressed overtly in the context.¹⁶ In (3.39b), there is a single constituent that refers to a sub-region of a Body Part. However, in most cases like (3.39b) the Possessor is overtly expressed, though not as the syntactic possessor of the larger NP but only as the syntactic possessor of the Body Part-referring NP that is the complement of the dependent prepositional phrase (e.g., *kissed the side of his cheek*). With *smack* and *slap* but not with *kiss* one also finds cases where the Agent and the Possessor are the same person. Self-slapping as in (3.39c) is usually not punitive or offensive and under those circumstances the raised construction seems to be dispreferred.

Interestingly, the different constructions do not have the same relative ranking for the three verbs. In the case of *kiss* and *smack* the simple Possessor construction is the most common one (56.0%; 59.5%) ahead of the conflated construction (24.4%; 19.0%) and the raised construction (12.3%; 17.6%). With *slap*, however, the raised construction (33.3%) ranks ahead of the simple Possessor construction (23.1%) and the conflated construction (18.5%). Given the difference between *slap* and *smack* it is difficult to argue along the lines of O'Connor's analysis that a notion of affectedness or intensity is relevant to the choice of construction. *Smack* seems to be at least as affecting as *slap* yet the relative shares of the different constructions are as with *kiss*. Alternatively, one may speculate that with *slap* the active zone more often differs from the default, the face, than in the case of *kiss* and *smack* and therefore has to

¹⁶With *kiss* the default region may vary depending on who the participants are. Lovers are more likely to kiss each other on the mouth than more distant relatives or heads of state.

be mentioned overtly. However, one would need psycholinguistic experimentation to establish what the default locations involved are.

It is worthwhile taking a look at the Body Part referents that occur in the two constructions in my sample. They are shown in Table 3.57

	Conflated Construction	Raised Construction
slap	face, <u>hand</u> , shoulder, <u>wrist</u> , <u>side</u> , <u>cheek</u> , <u>limbs</u> , <u>rump</u>	<u>back</u> , <u>moustache</u> , face, <u>mouth</u> , <u>shoulder blades</u> , <u>arse</u> , <u>bottom</u> , <u>behind</u> , <u>chops</u> , <u>bottom</u> , shoulder
kiss	<u>brow</u> , lips, <u>cheekbone</u> , <u>hair</u> , hand, <u>mouth</u> , <u>neck</u> , <u>eyes</u> , <u>chest</u> , <u>forehead</u> , <u>nose</u> , <u>upper arm</u> , <u>face</u> , <u>arse</u> , <u>back</u> , <u>cheek</u>	cheek, <u>bald pate</u> , <u>head</u> , lips, hand
smack	bottom, <u>rump</u> , <u>arse</u> , <u>behind</u> , <u>bum</u> , face, back, <u>hands</u> , <u>botty</u> , head, <u>fingers</u> , head, <u>hind</u>	<u>chops</u> , face, <u>mouth</u> , <u>ear</u> , head, <u>eyes</u> , <u>nose</u> , <u>snout</u> , <u>haunches</u> , back, bottom, <u>knee</u>

Table 3.57: Types of Body Part referents occurring in the constructions

The underlined lexical units are attested in only one of the constructions in the sample. It might seem that there are specific lexical biases that favor one construction over the other. However, in all cases, the absence of certain types in one of the constructions is just an artifact of the sampling: most of the missing uses are attested in the full British National corpus and all of them are attested on the Internet.

3.2.4 Morphology

Before we begin to look at more immediately pragmatic notions such as the number of prior and subsequent mentions, let us consider what the morphology of the NPs denoting Possessors and Body Parts is. Table 3.58 shows the ways in which Possessors are encoded.

Assuming that more attenuated morphological expression is correlated at least with greater identifiability, and maybe even with greater topicality, we may divide up the formal types into lexical and pronominal forms and perform a Fisher's exact-test to assess whether there is a significant difference between the constructions along

these lines. The relevant data is shown in Table 3.59. Note that for the purposes of this comparison I treat the possessive pronouns occurring as part of the object NP in the conflated constructions the same as the independent personal pronouns that occur at the clausal level in the raised construction. The result of a two-tailed test is $p=0.4141$ for *kiss*, $p=0.3364$ for *smack*, and $p=0.1274$ for *slap*, which suggests that there is no statistically significant difference.

The formal types for the Body Part referents are as shown in Table 3.60. If we again divide up the formal types into lexical and pronominal forms, we can perform a Fisher's exact-test for *kiss* to assess whether there is a significant difference between the constructions in terms of the morphology of the Body Part NPs. The relevant table is shown in 3.61.¹⁷

It is trivial to perform a test for *slap* and *smack* as there are only lexical Body Part NPs attested for each lexical unit in both constructions. The result of a two-tailed test for *kiss* is $p=0.3167$. This suggests that there is no significant difference between the constructions in terms of lexical versus pronominal encoding of Body Part referents.

In sum, we find that at the fairly crude level of distinguishing between pronominal and lexical expression, there is no discernible correlation between the morphology of the referents and the construction used.

3.2.5 Discourse properties of referents

Prior and subsequent evocation One measure which may reflect the presumed local topic status of the Possessor is the likelihood with which its referent has a prior mention in the discourse. Table 3.62 shows the results for this measure for *kiss* and *slap*. The data here are somewhat surprising. Possessors in the raised construction do not seem to have a prior mention more often than Possessors in the the conflated construction. With *kiss*, 95.1% of Possessors in the conflated construction have a

¹⁷The example classified as having the Body Part encoded as a Definite NP in the conflated construction is the following: *Perhaps she sought some prolongation of that feeling of joy and unity and excitement, the festival holiday feeling, the warm inclusion, yes, even being embraced by those formidable satiny bosoms and kissing the soft powdery cheeks of the old women.*

	kiss		slap		smack	
	Con-flated	Raised	Con-flated	Raised	Con-flated	Raised
Bare/Plural NP	1	0	0	0	0	1
Quantified NP	0	0	1	0	1	0
Indefinite NP	0	0	0	0	0	1
Definite NP	2	0	2	2	0	3
Proper Noun	2	1	0	8	1	1
Possessive NP	1	0	0	3	2	1
1st/2nd Person Pronoun	0	1	0	5	0	12
3rd Person Pronoun	1	17	0	18	1	16
Zero	0	0	0	0	0	1
Non-restrictive relative pronoun	0	0	0	0	0	1
Possessive Pronoun	34	0	17	0	35	0
Total	41	19	20	36	40	37

Table 3.58: Morphology of Possessors with *kiss*

		Conflated	Raised	Total	p-value of Fisher's Exact
kiss	pronominal	35	18	53	.4141
	lexical	6	1	7	
slap	pronominal	17	23	40	.1274
	lexical	3	13	16	
smack	pronominal	36	30	66	.3364
	lexical	4	7	11	

Table 3.59: Fisher's exact test for the morphology of Possessors

		Conflated	Raised
kiss	3rd Person Pronoun	0	1
	Definite NP	1	15
	Quantified NP	0	1
	Possessive NP	40	2
	Total	41	19
slap	Definite NP	0	32
	Possessive NP	20	4
	Total	20	36
smack	Definite NP	1	36
	Possessive NP	39	0
	Quantified NP	0	1
	Total	40	37

Table 3.60: Morphology of Body Parts

		Conflated	Raised	Total	p-value of Fisher's Exact
kiss	pronominal	0	1	1	.3167
	lexical	41	18	59	
slap	pronominal	0	0	0	1.000
	lexical	20	36	56	
smack	pronominal	0	0	0	1.000
	lexical	40	37	77	

Table 3.61: Fisher's exact test for the morphology of Body Parts

prior mention versus 94.7% in the raised construction. With *slap*, the percentages are 95.0% and 97.2%. For *smack*, the numbers are 95.0% and 78.4% respectively.

For *slap* and *kiss*, the differences are not statistically significant according to a Fisher's exact test for the data shown in Table 3.63; $p = 1.0$ in both cases. For *smack*, a two-tailed Fisher's exact test yields $p=0.04230$, which is statistically significant by conventional standards. However, this result is still not linguistically important for the following reason. In the conflated construction the 2 tokens that were not textually evoked were inferable. But of the 8 tokens in the raised construction where the Possessor referent was not textually evoked, only 4 are inferable referents, the other 4 are speech act participants; people co-present in the speech act situation; or omni-relevant referents. Including these four tokens in a broadened category of evoked referents causes the statistical difference to disappear at $p=0.4195$.

Incidentally, the Possessors in both the conflated and the raised constructions seem to be at least as likely to have a prior mention as Possessors that occur by themselves without an overt expression of the Body Part; compare the rows for 'Possessor' in Table 3.62 to those for 'Conflated' and 'Raised'.

Having explored the connection of the Possessor referent to the prior context we may now look at its connection to subsequent context. The most basic measure to look at is the likelihood for the Possessor referent to be mentioned again within the window of context. Table 3.64 shows the relevant data.

Here, too, the data does not lend support to the view that Possessors in the raised construction are more topical. While in the case of *kiss* the Possessors in the raised construction have a subsequent mention more often than those in the conflated construction (94.7% versus 92.7%), the opposite is the case with *slap*. Here 83.3% of Possessors in the raised construction have a subsequent mention compared to 80.0% of Possessors in the conflated construction. A two-tailed Fisher's exact test based on the data in Table 3.65 suggests that the difference is not statistically significant: for *kiss*, $p=1.0$; for *slap*, $p = 0.7327$. For *smack*, the p -value gets closest to the threshold for statistical significance of .05 that I am using, with $p=0.0762$. Notice, though, that the asymmetry in the case of *smack* goes in the wrong direction: although we would expect the Possessor referents that are coded as clause-level arguments in the

	Construction	Textually Evoked		Not Textually Evoked		Total
		n	%	n	%	n
kiss	Conflated	39	95.1	2	4.9	41
	Own Body Part	1	100	0	0	1
	Body Part	0	0	13	100	13
	Possessor	86	91.5	8	8.5	94
	Raised	18	94.7	1	5.3	19
	Total	144	85.7	24	14.3	168
slap	Conflated	19	95.0	1	5.0	20
	Own Body Part	13	100	0	0	13
	Body Part	0	0	7	100	7
	Possessor	22	88.0	3	12.0	25
	Raised	35	97.2	1	2.8	36
	Total	89	88.1	12	11.9	101
smack	Conflated	38	95.0	2	5.0	40
	Own Body Part	7	100	0	0	7
	Possessor	98	78.4	27	21.6	125
	Raised	29	78.4	8	21.6	37
	Total	172	82.3	37	17.7	209

Table 3.62: Prior mention of the Possessor

	kiss			slap			smack		
	Evo-Ked	Not Evo-Ked	Total	Evo-Ked	Not Evo-Ked	Total	Evo-Ked	Not Evo-Ked	Total
Conflated	39	2	41	19	1	20	38	2	40
Raised	18	1	19	35	1	36	29	8	37
	57	3	60	54	2	56	67	10	77

Table 3.63: Distribution of prior evocation status across the conflated and the raised constructions

	Construction	Evoked		Not Evoked		Total
		n	%	n	%	
kiss	Conflated	38	92.7	3	7.3	41
	Own Body Part	1	100	0	0	1
	Possessor	83	88.3	11	11.7	94
	Raised	18	94.7	1	5.3	19
	Total	140	90.3	15	9.7	155
slap	Conflated	16	80.0	4	20.0	20
	Own Body Part	12	92.3	1	7.7	13
	Body Part	0	0.0	7	100.0	7
	Possessor	23	92.0	2	8.0	25
	Raised	30	83.3	6	16.7	36
Total	82	81.2	19	18.8	101	
smack	Conflated	36	90.0	4	10.0	40
	Own Body Part	7	100.0	0	0	7
	Possessor	87	69.6	38	30.4	125
	Raised	27	73.0	10	27.0	37
Total	157	75.1	52	24.9	209	

Table 3.64: Subsequent mention of the Possessor

	kiss			slap			smack		
	Evo- ked	not evo- ked	Total	Evo- ked	not evo- ked	Total	Evo- ked	not evo- ked	Total
Conflated	38	3	41	16	4	20	36	4	40
Raised	18	1	19	30	6	36	27	10	37
	56	4	60	54	2	56	63	14	77

Table 3.65: Distribution of subsequent evocation status across the conflated and the raised constructions

raised construction to be more topical, they seem less likely to have persistence in subsequent discourse.

We now turn to the evocation status of Body Part referents. For both verbs and all constructions it is clearly the case that Body Parts are rather transient referents: they can be inferred on the basis of knowledge about the Possessor but they have neither much of a history nor much of a future in the discourse. As Table 3.66 shows, one Body Part referent in a conflated construction (=0.0%) has an antecedent in the case of *kiss*, while two Body Part referents in the raised construction (=3.0%) do in the case of *slap*. By contrast, of all Possessor referents 85.7% have a prior mention with *kiss*; 88.1% of all in the case of *slap*; and 82.3% in the case of *smack*, as is shown by Table 3.62 above.

Two-tailed Fisher's exact tests show that the number of Body Part referents with prior mentions does not differ significantly for the conflated and the raised constructions; $p = 1.0$ for *kiss*; $p = 0.5402$ for *slap*; and $p=0.1070$ for *smack*.

As is shown in Table 3.67, the situation seems to be more or less the same with subsequent mentions: they are quite infrequent overall. With *slap* only 7.9% of all Body Parts have any subsequent mention compared with 81.2% of Possessors having subsequent mentions. With *kiss* the relevant numbers are 3.2% and 90.3% and with *smack* 3.0% and 75.1%.

Fisher's exacts test show that the number of Body Part referents with subsequent mentions does not differ significantly between the conflated and the raised constructions; for a two tailed test, $p = 0.1684$ with *kiss*; $p = 0.4016$ with *slap*; and $p=0.2022$

with *smack*.

While the likelihood of prior and subsequent mentions does not seem to differ for a given semantic participant across the two constructions, it is clearly the case that within each constructions Possessors are much more likely to have prior and subsequent mentions (cf. Tables 3.62 and 3.64).

		Evoked	Inferrable	Not Evoked	Total
kiss	Conflated	1	40	0	41
	Own Body Part	0	1	0	1
	Possessor	0	0	94	94
	Raised	0	19	0	19
	Total	1	60	94	155
slap	Conflated	0	19	1	20
	Body Part	1	5	1	7
	Own Body Part	0	12	1	13
	Possessor	0	0	25	25
	Raised	2	34	0	36
	Total	3	70	28	101
smack	Conflated	6	31	3	40
	Own Body Part	0	6	1	7
	Possessor	0	0	125	125
	Raised	1	36	0	37
	Total				

Table 3.66: Construction by prior evocation status of Body Part referents

		Evoked	Not Evoked	Total
kiss	conflated	5	36	41
	Own Body Part	0	1	1
	Possessor	0	94	94
	Raised	0	19	19
	Total	5	150	155
slap	Conflated	1	19	20
	Own Body Part	0	13	13
	Body Part	1	6	7
	Possessor	0	25	25
	Raised	6	30	36
	Total	8	93	101
smack	Conflated	5	35	40
	Own Body Part	0	7	7
	Possessor	0	125	125
	Raised	1	36	37
	Total	6	203	209

Table 3.67: Construction by subsequent evocation status of Body Part referents

Number of contextual mentions We now turn to the specific number of prior and subsequent mentions of participant referents as a measure of discourse status. Body Part referents rarely have any prior or subsequent mentions and when they do, they have very few of them. This is shown in Table 3.68

		Conflated		Raised	
		Prior	Subsequent	Prior	Subsequent
kiss	N	41	41	19	19
	Mean	.0732	.1463	.0000	.0000
	Median	.0000	.0000	.0000	.0000
	Mode	.00	.00	.00	.00
	Std. Deviation	.46852	.42196	.00000	.00000
	Variance	.21951	.17805	.00000	.00000
slap	N	20	20	36	36
	Mean	.0000	.05	.0556	.17
	Median	.0000	.00	.0000	.00
	Mode	.00	0	.00	0
	Std. Deviation	.00000	.224	.23231	.378
	Variance	.00000	.050	.05397	.143
smack	N	40	40	37	37
	Mean	.1750	.2250	.0270	.0270
	Median	.0000	.0000	.0000	.0000
	Mode	.00	.00	.00	.00
	Std. Deviation	.44650	.83166	.16440	.16440
	Variance	.19936	.69167	.02703	.02703

Table 3.68: Contextual Mentions of Body Part referents

As Table 3.68 shows, the mean number of any of the contextual mentions is never higher than 0.225, which is the value for subsequent mentions of Body Parts in the conflated construction for *smack*. It should come as no surprise that the data is not normally distributed, which is shown by the Kolmogorov-Smirnov and Shapiro-Wilk tests reported in Tables 3.69 to 3.71.¹⁸

We now turn to Possessor referents, which do have a significant number of previous mentions and subsequent mentions. The relevant data for is presented in Table 3.72.

¹⁸The tests marked as 'n/a' could not be performed because there were no prior or subsequent mentions recorded at all, preventing the establishment of a standard deviation.

	Construction	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Previous Mentions	Conflated	.538	41	.00	.144	41	.000
	Raised	n/a	n/a	n/a	n/a	n/a	n/a
Subsequent Mentions	Conflated	.514	41	.00	.394	41	.000
	Raised	n/a	n/a	n/a	n/a	n/a	n/a

Table 3.69: Normal distribution of number of prior and subsequent mentions of Body Part referents across conflated and raised constructions for *kiss*

	Construction	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Previous Mentions	Conflated	n/a	n/a	n/a	n/a	n/a	n/a
	Raised	.539	36	.000	.246	36	.000
Subsequent Mentions	Conflated	.538	20	.000	.236	20	.000
	Raised	.504	36	.000	.451	36	.000

Table 3.70: Normal distribution of number of prior and subsequent mentions of Body Part referents across conflated and raised constructions for *slap*

	Construction	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Previous Mentions	Conflated	.502	40	.000	.443	40	.000
	Raised	.538	37	.000	.155	37	.000
Subsequent Mentions	Conflated	.482	40	.000	.295	40	.000
	Raised	.538	37	.000	.155	37	.000

Table 3.71: Normal distribution of number of prior and subsequent mentions of Body Part referents across conflated and raised constructions for *smack*

		Conflated		Raised	
		Prior	Subsequent	Prior	Subsequent
kiss	N	41	41	19	19
	Mean	9.10	8.29	8.58	7.11
	Median	9.00	7.00	9.00	7.00
	Mode	6	8	7*	2*
	Std. Deviation	5.911	6.619	4.325	4.533
	Variance	34.940	43.812	18.702	20.544
slap	N	20	20	36	36
	Mean	7.5000	5.2500	6.3333	6.5556
	Median	7.0000	3.5000	5.0000	7.0000
	Mode	3.00	.00	1.00	.00
	Std. Deviation	5.04193	5.18982	5.54462	4.69312
	Variance	25.42105	26.93421	30.74286	22.02540
smack	N	40	40	37	37
	Mean	6.4750	5.3000	5.1081	4.4324
	Median	6.0000	5.0000	4.0000	4.0000
	Mode	1.00	5.00	.00	.00000
	Std. Deviation	5.06869	4.10253	5.05956	4.69362
	Variance	25.69167	16.83077	25.59910	22.03003

Table 3.72: Contextual mentions of Possessors

This table shows that with *kiss* the average number of prior mentions (9.10; 8.58) is greater than that of subsequent mentions (8.29; 7.11) and that the numbers are higher in the conflated construction (9.10; 8.29) than in the raised construction (8.58; 7.11). The situation is parallel for *smack*. The situation is somewhat different with *slap* in that the Possessor referents in the raised construction seem to have fewer prior than subsequent mentions (6.3333; 6.5556).

Before we can test the statistical significance of these numbers, we first have to consider whether the data is normally distributed. Tables 3.73 to 3.75 show that in quite a few cases, the distribution is not normal. One such case consists of the subsequent mentions in the conflated construction with *kiss*. In this one case, the distribution is still normal by Kolmogorov-Smirnov test (0.092) but the more accurate Shapiro-Wilk test shows it to not have a normal distribution (.003). As Figure 3.1 shows, the relatively large number of tokens with 25 subsequent mentions skews the

distribution.¹⁹

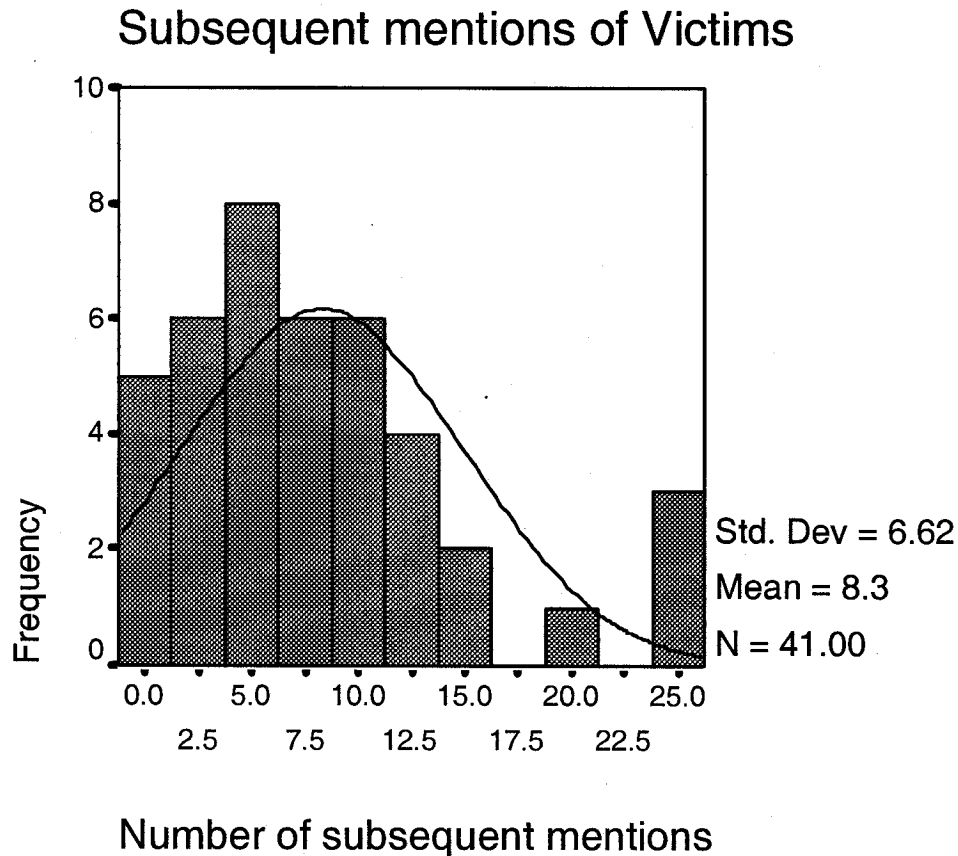


Figure 3.1: Subsequent mentions of Possessors in the conflated construction with *kiss*

Since many of the scores are not normally distributed, I will use the non-parametric Wilcoxon-Mann-Whitney test for making various comparisons between frame elements, constructions, and lexical units. This test does not require normally distributed data. Instead it uses rank orders for assessing the likelihood that two independent groups have been drawn from the same population. The test works by (1) ranking the scores in increasing order while retaining the information on where each

¹⁹Table 3.72 also suggests this: as indicated by the asterisk, there is more than one mode for the number of subsequent mentions of Possessors in the raised construction.

	Construction	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Previous Mentions	Conflated	.115	41	.198	.962	41	.179
	Raised	.118	19	.200	.970	19	.770
Subsequent Mentions	Conflated	.127	41	.092	.907	41	.003
	Raised	.107	19	.200	.961	19	.590

Table 3.73: Normal distribution of number of prior and subsequent mentions of Possessor referents across conflated and raised constructions for *kiss*

	Construction	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Previous Mentions	Conflated	.156	20	.200	.940	20	.244
	Raised	.163	36	.017	.890	36	.002
Subsequent Mentions	Conflated	.184	20	.073	.882	20	.020
	Raised	.107	36	.200	.947	36	.085

Table 3.74: Normal distribution of number of prior and subsequent mentions of Possessors across conflated and raised constructions for *slap*

	Construction	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Previous Mentions	Conflated	.162	40	.010	.902	40	.002
	Raised	.208	37	.000	.862	37	.000
Subsequent Mentions	Conflated	.179	40	.002	.929	40	.015
	Raised	.172	37	.007	.852	37	.000

Table 3.75: Normal distribution of number of prior and subsequent mentions of Possessors across conflated and raised constructions for *smack*

score came from—for tied scores each of the tied observations gets assigned the average of the ranks they would have if no tied had occurred; (2) summing the ranks for each group; (3) determining the average rank for both groups. We would expect the average ranks in each of the sample groups to be about equal if they came from the same population.

COMPARISONS BETWEEN CONSTRUCTIONS We will begin by comparing the number of contextual mentions for a given participant role across the two constructions. We first test whether the number of previous mentions of Possessor referents differs significantly across the two constructions for our three lexical units.

For prior mentions of Possessors with *kiss* across the two constructions, we have groups $A = 19$ and $B = 41$ and we get a Wilcoxon $W = 568.000$ and a p-value of 0.855, which suggests that there is no significant difference. For prior mentions of Possessors with *slap* across the two constructions, we get a similar result. With groups $A = 20$ and $B = 36$, we get Wilcoxon $W = 965.500$ and a $p = .299$. Repeating the test for *smack*, we have groups $A = 37$ and $B = 40$ and we get a Wilcoxon $W = 1316.500$ and a p-value of 0.195. All three tests suggest that there is no significant difference in terms of prior mentions of the Possessor referent between the two constructions.

Next we compare the the number of subsequent mentions of Possessors across constructions. For *kiss*, a Mann-Whitney-Wilcoxon test with groups $A = 19$ and $B = 41$ yields a Wilcoxon $W = 561$ and a z-score of -0.295; $p = 0.768$. Based on these results, we have no reason to reject the null hypothesis that the number of subsequent mentions of Possessor referents with *kiss* is not different between the two constructions. For *slap*, a Mann-Whitney-Wilcoxon test with groups $A = 20$ and $B = 36$ yields a Wilcoxon $W = 507.00$ and a z-score of -1.083; $p = 0.279$. For *smack*, a Mann-Whitney-Wilcoxon test with groups $A = 37$ and $B = 40$ produces a Wilcoxon $W = 1315.500$ and a z-score of -1.307; $p = 0.191$. In this case, too, we have no reason to reject the null hypothesis that the number of subsequent mentions of Possessor referents does not differ across the constructions.

We now go on to repeat the between-construction comparisons for Body Part referents. When we apply the Mann-Whitney-Wilcoxon test to the number of prior mentions of Body Parts with *kiss* with groups $A = 19$ and $B = 41$, we get a Wilcoxon

$W = 570.00$ and a z-score of -0.681 ; $p=0.496$. For *slap*, we have groups $A = 20$ and $B = 36$ and get a Wilcoxon $W = 550.00$, a z-score of -1.064 , and a $p=0.287$. For *smack*, we have groups $A = 37$ and $B = 40$, and get a Wilcoxon W of 1351.500 for a z-score of -1.872 and a p-value of 0.061 . Though the p-value for *smack* is close to significance, we have no reason to reject the null hypothesis that there is no significant difference in the number of prior mentions of Body Part referents between the two constructions with either *kiss* or *slap*.

Applying the Mann-Whitney-Wilcoxon test to the number of subsequent mentions of Body Parts for *kiss*, with groups $A = 19$ and $B = 41$, we get a Wilcoxon $W = 532.00$ and a z-score of -1.576 ; $p=0.115$. For *slap*, with groups $A = 20$ and $B = 36$, we get Wilcoxon $W = 528.00$, $z = -1.254$ and $p=.210$. In the case of *smack*, the group sizes are $A = 37$ and $B = 40$; Wilcoxon $W = 1370.000$, $z=-1.602$, and $p=0.109$. As for prior mentions, we have no reason to reject the null hypothesis that there is no significant difference in the number of subsequent mentions of Body Part referents between the two constructions for either of the lexical units.

COMPARISONS BETWEEN FRAME ELEMENTS We continue our exploration of differences in the number of contextual mentions by holding the construction and the lexical unit constant and comparing the values for the participant referents.

Table 3.76 shows for *kiss* that the frame elements differ from each other significantly in each construction with respect to the number of prior and subsequent mentions, with $p = 0.000$ in all cases.

	Conflated		Raised	
	Prior Mentions	Subsequent Mentions	Prior Mentions	Subsequent Mentions
Group size	41; 41	41; 41	19; 19	19; 19
Mann-Whitney U	48.000	76.500	9.500	9.500
Wilcoxon W	909.000	937.500	199.500	199.500
Z	-7.902	-7.504	-5.405	-5.402
p (2-tailed)	.000	.000	.000	.000

Table 3.76: Within-construction comparison of participants for *kiss*

Table 3.77 shows the results for the lexical unit *slap*. They are entirely parallel and here too $p = .000$ in all cases.

	Conflated		Raised	
	Prior Mentions	Subsequent Mentions	Prior Mentions	Subsequent Mentions
Group size	20; 20	20; 20	36; 36	36; 36
Mann-Whitney U	10.000	43.000	46.000	126.000
Wilcoxon W	220.000	253.000	712.000	792.000
Z	-5.560	-4.721	-7.261	-6.290
p (2-tailed)	.000	.000	.000	.000

Table 3.77: Within-construction comparison of participants for *slap*

In Table 3.78 we find the pattern repeated for *smack*, with $p=.000$ in all cases.

	Conflated		Raised	
	Prior Mentions	Subsequent Mentions	Prior Mentions	Subsequent Mentions
Group size	40; 40	40; 40	37; 37	37; 37
Mann-Whitney U	69.500	116.500	153.000	192.000
Wilcoxon W	889.500	936.500	856.000	895.000
Z	-7.387	-7.004	-6.469	-6.111
p (2-tailed)	.000	.000	.000	.000

Table 3.78: Within-construction comparison of participants for *smack*

For all three lexical units and all measures, it is the case that the sum of ranks for the Possessor referent is much greater than that for the Body Part referents.

COMPARISON BETWEEN LEXICAL UNITS Since we have data on more than one lexical unit, we can check whether there are significant lexical differences between them in terms of the number of contextual mentions that the participant referents have. For the purposes of such a comparison, we will use the pair of lexical units *slap* and *kiss*.

Table 3.79 shows the results of Wilcoxon-Mann-Whitney tests for the raised construction. None of the measures is statistically significant by our conventional stan-

dard of $p < 0.05$ but two of them come close.

Kiss seems to have a slightly greater number of subsequent mentions of Body parts than *slap*, while *slap* seems to have a somewhat greater number of prior mentions of Possessor referents than *kiss*.

	Body Part		Possessor	
	Prior Men- tions	Subsequent Mentions	Prior Men- tions	Subsequent Mentions
Group size	19; 36	19; 36	19; 36	19; 36
Mann-Whitney U	323.000	285.000	236.500	314.500
Wilcoxon W	513.000	475.000	902.500	980.500
Z	-1.037	-1.868	-1.875	-.488
p (2-tailed)	.300	.062	.061	.625

Table 3.79: Comparison between *slap* and *kiss* in the raised construction

Table 3.80 shows the results of Wilcoxon-Mann-Whitney tests for the conflated construction. Here, too, none of the measures is statistically significant by our conventional standard of $p < 0.05$. But the result for subsequent mentions of Possessors comes close to statistical significance: *slap* seems to have more subsequent mentions of Possessors than *kiss*.

	Body Part		Possessor	
	Prior Men- tions	Subsequent Mentions	Prior Men- tions	Subsequent Mentions
Group size	41; 20	41; 20	41; 20	41; 20
Mann-Whitney U	400.000	380.000	350.500	289.000
Wilcoxon W	610.000	590.000	560.500	499.000
Z	-.698	-.893	-.916	-1.864
p (2-tailed)	.485	.372	.360	.062

Table 3.80: Comparison between *slap* and *kiss* in the conflated construction

Overall, the only clear and consistent difference in terms of contextual mentions that we have been able to identify exists between the participants in a given construction: Possessor referents always have significantly more prior and subsequent

mentions than Body Part referents. Clear differences for a given participant role across constructions, and differences between lexical units were not detectable.

Distance to contextual mentions For the number of contextual mentions, we have found significant differences only between participants in a given construction. We will now explore data on the distance between a participant referent in a target sentence and earlier and later mentions of that same referent.

The basic data is summarized in Tables 3.81 to 3.86. Note that I assigned a value of 11 as the closest prior or subsequent mention value when there was no prior or subsequent mention within the window of context. However, I assigned a value of 0 as the last mention when there was no appropriate mention within the window of context. The reason is that the purpose of the last-mention measure is to evaluate a referent's persistence in the subsequent discourse: I decided to err on the side of making tokens that do have a subsequent mention outside the window appear less persistent rather than making tokens that have no such mention appear more persistent by giving it a default score greater than 10.

The basic finding is that in each construction, Possessors score better than Body Parts for each of the three distance measures. For instance, as Table 3.81 shows, with *slap*, Possessors in the conflated construction have their closest preceding mention only 1.15 sentences away, whereas Body Part referents have theirs 11 sentences away: in other words, Body Parts don't have a mention inside the window of 10 sentences of context. Similarly, the next mention of Possessors in the conflated construction with *slap* is 3.1 sentences away while it is 10.95 sentences away for Body Parts. Again, Body Parts do not have a subsequent mention within the window of context. Finally, in terms of the last subsequent mention, Possessors have theirs 5.95 sentences away from the current mention, whereas Body Parts have theirs only 1.60 sentences away. This pattern is repeated in other combinations of lexical unit and construction, as can be seen from Tables 3.82-3.85.

The data that have gone into the summary Tables 3.81-3.86 do not have normal distributions. This is shown by Kolmogorov-Smirnov and Shapiro-Wilk tests performed. I have omitted a display of the results here for brevity's sake. In every case

	Possessor			Body Part		
	Closest Prior	Closest Subsequent	Last Subsequent	Closest Prior	Closest Subsequent	Last Subsequent
N	20	20	20	20	20	20
Mean	1.1500	3.1000	5.9500	11.0000	10.95	1.60
Median	.0000	1.0000	8.5000	11.0000	11.00	.00
Mode	.00	1.00	10.00	11.00	11	0
Std. Dev.	2.56032	4.08978	4.35860	.00000	.224	3.912
Variance	6.55526	16.72632	18.99737	.00000	.050	15.305

Table 3.81: Distance values for contextual mentions of participants in the conflated construction for *slap*

	Possessor			Body Part		
	Closest Prior	Closest Subsequent	Last Subsequent	Closest Prior	Closest Subsequent	Last Subsequent
N	41	41	41	41	41	41
Mean	1.3659	2.39	8.00	10.8537	10.22	.59
Median	1.0000	1.00	9.00	11.0000	11.00	.00
Mode	1.00	1	10	11.00	11	0
Std. Dev.	2.38491	3.145	3.647	.93704	2.525	2.049
Variance	5.68780	9.894	13.300	.87805	6.376	4.199

Table 3.82: Distance values for contextual mentions of participants in the conflated construction for *kiss*

where a score could be computed, $p < 0.05$.

We can now begin to evaluate whether significant differences in terms of distance to contextual mentions exist either between the participant roles in a given construction; a given participant role across constructions; or between the lexical units.

COMPARISONS BETWEEN PARTICIPANTS We begin by comparing distance values between the participant roles for a given construction and a particular lexical unit. The results of the relevant Wilcoxon-Mann-Whitney test are summarized in Tables 3.87-3.92.

As we saw earlier for the number of contextual mentions, there exists a very sig-

	Possessor			Body Part		
	Closest Prior	Closest Subsequent	Last Subsequent	Closest Prior	Closest Subsequent	Last Subsequent
N	40	40	40	40	40	40
Mean	1.7000	2.4500	6.0500	9.9500	10.0250	.5250
Median	1.0000	1.0000	6.5000	11.0000	11.0000	.0000
Mode	.00	.00	10.00	11.00	11.00	.00
Std. Deviation	2.93694	3.24985	3.63706	2.85505	2.75017	1.92137
Variance	8.62564	10.56154	13.22821	8.15128	7.56346	3.69167

Table 3.83: Distance values for contextual mentions of participants in the conflated construction for *smack*

	Possessor			Body Part		
	Closest Prior	Closest Subsequent	Last Subsequent	Closest Prior	Closest Subsequent	Last Subsequent
N	36	36	36	36	36	36
Mean	1.9167	2.5556	6.3611	10.5833	9.89	2.86
Median	1.0000	1.0000	8.5000	11.0000	11.00	.00
Mode	.00	1.00	10.00	11.00	11	0
Std. Deviation	2.87228	3.92388	3.89984	1.74642	2.876	4.574
Variance	8.25000	15.39683	15.20873	3.05000	8.273	20.923

Table 3.84: Distance values for contextual mentions of participants in the raised construction for *slap*

	Possessor			Body Part		
	Closest Prior	Closest Subsequent	Last Subsequent	Closest Prior	Closest Subsequent	Last Subsequent
N	19	19	19	19	19	19
Mean	1.2105	1.37	7.11	11.0000	11.00	.00
Median	1.0000	1.00	9.00	11.0000	11.00	.00
Mode	.00	1	10	11.00	11	0
Std. Deviation	2.48504	2.499	3.635	.00000	.000	.000
Variance	6.17544	6.246	13.211	.00000	.000	.000

Table 3.85: Distance values for contextual mentions of participants in the raised construction for *kiss*

	Possessor			Body Part		
	Closest Prior	Closest Subsequent	Last Subsequent	Closest Prior	Closest Subsequent	Last Subsequent
N	37	37	37	37	37	37
Mean	2.9459	3.9189	4.4054	10.9459	10.7568	.0270
Median	1.0000	1.0000	5.0000	11.0000	11.0000	.0000
Mode	.00	1.00	.00	11.00	11.00	.00
Std. Deviation	4.41554	4.53001	3.99624	.32880	1.47959	.16440
Variance	19.49700	20.52102	15.96997	.10811	2.18919	.02703

Table 3.86: Distance values for contextual mentions of participants in the raised construction for *smack*

	Closest Prior	Closest Subsequent	Last Subsequent
Mann-Whitney U	42.000	130.000	111.500
Wilcoxon W	903.000	991.000	972.500
Z	-8.030	-7.019	-7.211
p (2-tailed)	.000	.000	.000

Table 3.87: Comparison between participants in the conflated construction for *kiss*

	Closest Prior	Closest Subsequent	Last Subsequent
Mann-Whitney U	10.000	42.000	90.000
Wilcoxon W	220.000	252.000	300.000
Z	-5.672	-4.798	-3.239
p (2-tailed)	.000	.000	.001

Table 3.88: Comparison between participants in the conflated construction for *slap*

nificant difference in the distance to contextual mentions for the two participant roles in a given construction. Possessor referents have their closest prior and subsequent mentions closer and their last subsequent mentions farther away than do Body Part referents.

	Closest Prior	Closest Subsequent	Last Subsequent
Mann-Whitney U	86.000	139.000	162.500
Wilcoxon W	906.000	959.000	982.500
Z	-7.278	-6.795	-6.580
p (2-tailed)	.000	.000	.000

Table 3.89: Comparison between participants in the conflated construction for *smack*

	Closest Prior	Closest Subsequent	Last Subsequent
Mann-Whitney U	9.500	9.500	19.000
Wilcoxon W	199.500	199.500	209.000
Z	-5.457	-5.463	-5.201
p (2-tailed)	.000	.000	.000

Table 3.90: Comparison between participants in the raised construction for *kiss*

	Closest Prior	Closest Subsequent	Last Subsequent
Mann-Whitney U	43.000	151.000	402.000
Wilcoxon W	709.000	817.000	1068.000
Z	-7.322	-6.077	-2.875
p (2-tailed)	.000	.000	.004

Table 3.91: Comparison between participants in the raised construction for *slap*

	Closest Prior	Closest Subse- quent	Last Sub- sequent
Mann-Whitney U	152.000	195.000	248.000
Wilcoxon W	855.000	898.000	951.000
Z	-6.533	-6.102	-5.607
p (2-tailed)	.000	.000	.000

Table 3.92: Comparison between participants in the raised construction for *smack*

COMPARISONS BETWEEN CONSTRUCTIONS Next we compare distance values for a given participant role and a particular lexical unit across the two constructions. The results of the relevant Wilcoxon-Mann-Whitney test are summarized in Tables 3.93-3.98.

	Closest Prior	Closest Subsequent	Last Subsequent
Mann-Whitney U	340.000	315.500	289.000
Wilcoxon W	1006.000	981.500	499.000
Z	-1.064	-1.324	-1.528
p (2-tailed)	.287	.185	.127

Table 3.93: Comparison between constructions for Body Part referents with *slap*

	Closest Prior	Closest Subsequent	Last Subsequent
Mann-Whitney U	251.000	280.000	353.500
Wilcoxon W	461.000	946.000	1019.500
Z	-1.966	-1.486	-.114
p (2-tailed)	.049	.137	.909

Table 3.94: Comparison between constructions for Possessor referents with *slap*

Almost all the tests have a p-value greater than 0.05. Overall, there do not seem to be significant differences for a given frame element between constructions in terms of any of the distance measures. The one score that drops below the 0.05 threshold is the distance measure for the closest prior mention of the Possessor referent with *slap*, with $p=0.049$. However, the score is not far from the threshold and given the number of tests we have performed, namely $3*6=18$, we almost have to expect a spurious result that appears to be statistically significant even if it is not.

	Closest Prior	Closest Subsequent	Last Subsequent
Mann-Whitney U	380.000	342.000	342.000
Wilcoxon W	1241.000	1203.000	532.000
Z	-.681	-1.575	-1.575
p (2-tailed)	.496	.115	.115

Table 3.95: Comparison between constructions for Body Part referents with *kiss*

	Closest Prior	Closest Subsequent	Last Subsequent
Mann-Whitney U	352.000	277.500	347.000
Wilcoxon W	542.000	467.500	537.000
Z	-.636	-1.866	-.707
p (2-tailed)	.525	.062	.480

Table 3.96: Comparison between constructions for Possessor referents with *kiss*

	Closest Prior	Closest Subsequent	Last Subsequent
Mann-Whitney U	647.000	667.500	666.000
Wilcoxon W	1467.000	1487.500	1369.000
Z	-1.901	-1.590	-1.623
p (2-tailed)	.057	.112	.105

Table 3.97: Comparison between constructions for Body Part referents with *smack*

	Closest Prior	Closest Subse- quent	Last Sub- sequent
Mann-Whitney U	707.000	632.500	557.000
Wilcoxon W	1527.000	1452.500	1260.000
Z	-.356	-1.128	-1.898
p (2-tailed)	.722	.259	.058

Table 3.98: Comparison between constructions for Possessor referents with *smack*

COMPARISONS BETWEEN LEXICAL UNITS Finally we can compare distance values across lexical units, holding the participant role and the construction constant. As before we'll use the pair *slap* and *kiss* as our example. The results of the relevant Wilcoxon-Mann-Whitney test are summarized in Tables 3.99-3.102.

	Closest Prior	Closest Subsequent	Last Subsequent
Mann-Whitney U	400.000	378.500	391.500
Wilcoxon W	1261.000	1239.500	1252.500
Z	-.698	-.937	-.485
p (2-tailed)	.485	.349	.628

Table 3.99: Comparison between lexical units for Body Part referents in the conflated construction

	Closest Prior	Closest Subsequent	Last Subsequent
Mann-Whitney U	326.500	382.000	327.000
Wilcoxon W	536.500	1243.000	537.000
Z	-1.375	-.451	-1.332
p (2-tailed)	.169	.652	.183

Table 3.100: Comparison between lexical units for Possessor referents in the conflated construction

	Closest Prior	Closest Subsequent	Last Subsequent
Mann-Whitney U	263.000	289.000	297.000
Wilcoxon W	453.000	479.000	963.000
Z	-1.464	-1.023	-.815
p (2-tailed)	.143	.307	.415

Table 3.101: Comparison between lexical units for Possessor referents in the raised construction

	Closest Prior	Closest Subsequent	Last Subsequent
Mann-Whitney U	323.000	285.000	218.500
Wilcoxon W	989.000	951.000	408.500
Z	-1.037	-1.864	-2.940
p (2-tailed)	.300	.062	.003

Table 3.102: Comparison between lexical units for Body Part referents in the raised construction

In all cases but one we find no statistically significant difference. The one exception concerns Body parts in the raised construction. Here it seems that the Body part referents that occur with *slap* have greater persistence, that is, their final mention in context occurs later than with *kiss*.²⁰ But this is so because there are no subsequent mentions of Body parts attested at all with *kiss* in the raised construction. Another reflection of this fact is that the difference in the number of subsequent mentions is close to statistical significance, with $p=0.62$ (cf. Table 3.79).

On the whole, the situation with respect to the distance of contextual mentions is much the same as with respect to the number of contextual mentions: reliable statistical differences seem to exist only between the participant roles within a given construction and lexical unit. The distance values for a given role do not differ significantly across constructions, and, in the case of our sample pair of *kiss* and *slap* there was no difference between the lexical units for particular combinations of construction and participant role.

²⁰A more precise characterization would be that with *slap* the mentions of Body Part referents occur as part of referential chains that extend farther into subsequent discourse than with *kiss*. It is not the referent itself that persists but the activation level of the mental representation that is used in a particular discourse to evoke the referent.

Presence of other possessor and body part referents We will now turn to properties of the context that may contribute indirectly to the inferability of referents. One such property is the presence of other Possessors, that is other Body Part Possessors. Since I assigned cases of self-directed action to a separate category, all tokens for all three verbs in both constructions do in fact have such other Body Part Possessors in their 10 sentence window of context. At minimum, the person doing the slapping, smacking, and kissing is another Body Part Possessor. Consequently, there is no statistically significant difference conceivable across constructions.

The situation is different with Body Parts: not all tokens have contextual mentions of another Body Part, as is shown by Table 3.103.

		Conflated	Raised	Total
kiss	yes	37	16	53
	no	4	3	7
	Total	41	19	60
slap	yes	16	29	45
	no	4	7	11
	Total	20	36	56
smack	yes	21	27	48
	no	19	10	29
	Total	40	37	77

Table 3.103: Presence of other Body Part referents in the 10 sentence window of context

According to a Fisher's exact test, there is no significant difference in the number of tokens that have contextual mentions of other Body Parts: $p=0.6682$ for *kiss*; $p=1.0$ for *slap*; $p=0.09883$ for *smack*.

We may look more specifically for mentions of the same type of Body Part within the 10 sentence window of context. It might be the case that one of the constructions is associated with a notion of contrast between Possessors but identity with respect to affected Body part. Table 3.104 shows how many tokens of both constructions had the same type of Body Part referent mentioned within 10 sentences preceding the target sentence. According to a two-tailed Fisher's exact test, this distribution

is not significantly skewed for *kiss* with a two-tailed P-value of 0.5447. The same holds for *slap* with $p=0.3364$. For *smack*, the p-value is 0.0297, which is statistically significant by conventional standards.

		Conflated	Raised	
kiss	yes	3	0	3
	no	38	19	57
	Total	41	19	60
slap	yes	3	2	5
	no	17	34	51
	Total	20	36	56
smack	yes	8	1	9
	no	32	36	68
	Total	40	37	77

Table 3.104: Mention of same kind of Body Part referent within 10 preceding sentences

Table 3.105 shows the results for the context following the target sentence. Incidentally, note that with *kiss* in all cases where there was a mention of the same kind of Body Part referent, it occurred in the immediately following sentence. According to a two-tailed Fisher's exact test, this distribution is not significantly skewed, $p=0.163$ for *kiss*; $p=1.0$ for *slap*; for *smack*. the p-value is 0.4837.

		conflated	vicbp	Total
kiss	yes	6	0	6
	no	35	19	54
	Total	41	19	50
slap	yes	5	9	14
	no	15	27	42
	Total	20	36	56
smack	yes	6	3	9
	no	34	34	68
	Total	40	37	77

Table 3.105: Mention of same kind of Body Part referent within 10 following sentences

We might look at a narrower context to see if Body Part referents are more

common with one construction than with the other. We begin by looking at the 2 sentences preceding our target sentences in order to see if a Body Part referent occurs in them. Table 3.106 shows how many tokens of both constructions had a Body Part referent mentioned within this narrower window of context. According to a two-tailed Fisher's exact test, there is no significant difference between the two constructions: $p=0.410$ for *kiss*; $p=0.3539$ for *slap*; $p=1$ for *smack*.

We can also look at the 2 sentences following our target sentences in order to see if a Body Part referent occurs in them. Table 3.107 shows how many tokens of both constructions had a Body Part referent mentioned within the two sentences following the target sentence. Again a Fisher's exact test shows no significant difference between the two constructions: $p=1.0$ for *kiss*; $p=0.2555$ for *slap*; $p=0.2631$ for *smack*.

We may wonder if there is a difference between the construction in the grammatical function that the contextual mentions of Body Parts in the 2 sentence window have. Table 3.108 shows the distribution for preceding mentions. If we cluster the grammatical functions into two groups—a more prominent group consisting of subject and direct object versus a less prominent group consisting of oblique and embedded—we can perform a Fisher's exact as shown in Table 3.109 to see whether prior Body Part mentions in one of the constructions tend to have greater grammatical prominence. The two-tailed P value of a Fisher's exact test is $p=0.6653$ for *kiss* and $p=1.0$ for *slap*; for *smack*, $p=0.6199$. There is no reason to reject the null hypothesis that there is no difference in the grammatical prominence of prior Body Part referent mentions.

We can carry out the same test for the grammatical function of Body Part mentions in the 2 sentences following the target mention. The distribution of grammatical functions for the subsequent context is shown in Table 3.110. Again clustering the grammatical functions into two groups we can perform a Fisher's exact on the Table shown in 3.111.

The two-tailed P value for *kiss* is $p=0.6237$; for *slap*, $p=1.0$; for *smack*, $p=0.25$. By conventional standards, we do not have reason to reject the null hypothesis that there is no difference in the grammatical prominence of subsequent Body Part referent mentions.

In this section we have performed statistical tests to check whether the presence

		conflated	vicbp	
kiss	yes	18	6	24
	no	23	13	36
	Total	41	19	60
slap	yes	7	8	15
	no	13	28	41
	Total	20	36	56
smack	yes	9	9	18
	no	31	28	59
	Total	40	37	77

Table 3.106: Mention of a Body Part referent within 2 preceding sentences

		conflated	vicbp	Total
kiss	yes	17	8	25
	no	24	11	35
	Total	41	19	60
slap	yes	5	15	20
	no	15	21	36
	Total	20	36	56
smack	yes	6	10	16
	no	34	27	61
	Total	40	37	77

Table 3.107: Mention of a Body Part referent within 2 following sentences for *kiss*

		conflated	vicbp	Total
kiss	Subject	3	1	4
	Direct Object	8	2	10
	Oblique	5	3	8
	Embedded in NP	2	0	2
	Total	18	6	24
slap	Subject	1	2	3
	Direct Object	2	1	3
	Oblique	4	5	9
	Embedded in NP	0	0	0
	Total	7	8	15
smack	Subject	1	1	2
	Direct Object	5	4	9
	Oblique	1	2	3
	Embedded in NP	1	1	2
	Possessive	0	1	1
	Total	8	9	17

Table 3.108: Grammatical function of Body Part mentions in preceding 2 sentences

		Conflated	Raised	Total
kiss	Higher prominence	11	3	14
	Lower prominence	7	3	10
	Total	18	6	24
slap	Higher prominence	3	3	6
	Lower prominence	4	5	9
	Total	7	8	15
smack	Higher prominence	6	5	11
	Lower prominence	2	4	6
	Total	8	9	17

Table 3.109: Grouped grammatical functions for Fisher's exact test of Body Part mentions in preceding 2 sentences

		conflated	vicbp	Total
kiss	Subject	3	1	4
	Direct Object	9	6	15
	Oblique	4	1	5
	Modifier	1	0	1
	Total	17	8	25
slap	Subject	0	4	4
	Direct Object	3	6	9
	Oblique	1	3	4
	Embedded	1	2	3
	Total	5	15	20
smack	Subject	0	0	0
	Direct Object	6	7	13
	Oblique	0	2	2
	Embedded in NP	0	1	1
	Total	6	10	16

Table 3.110: Grammatical function of Body Part mentions in following 2 sentences

		Conflated	Raised	Total
kiss	Higher prominence	12	7	19
	Lower prominence	5	1	6
	Total	17	8	25
slap	Higher prominence	3	10	13
	Lower prominence	2	5	7
	Total	5	15	20
smack	Higher prominence	6	7	13
	Lower prominence	0	3	3
	Total	6	10	16

Table 3.111: Grouped grammatical functions for Fisher's exact test of Body Part mentions in following 2 sentences

of other Possessor and Body Part referents might correlate with the choice of construction. We found a single significant result: with *smack* the conflated construction seemed to have significantly more mentions of the same kind of Body Part in the preceding 10 sentences than the raised construction. Given that we performed 21 tests at a 95% confidence interval, the one significant result might be the one false positive result that we are likely to get with this number of performed tests.

Overall, we have considered a great many formal and pragmatic factors that might influence the choice between the constructions: the morphology of the NPs referring to the participants; the presence or absence of prior and subsequent mentions of Possessor and Body Part referents, considered as categorical variables; the number of prior and subsequent mentions as a discrete quantitative variable; the presence of other Body Part referents in the 10-sentence context; the presence of the same kind of Body Part referent in the 10-sentence context; the presence of a Body Part referent in narrower 2 sentence contexts preceding and following the target sentence; the grammatical function of this last kind of referent. None of these variables has shown any significant difference between the constructions with either of the verbs. And in most cases, the probability values were very far off from a significant level. The only reliable differences exist between the frame elements Body Part and Possessor: all 14 tests that compared them showed a statistically significant difference. By contrast, of the 95 other tests performed (excluding tests of normality) only 4 showed a significant result, and these may well be false positives given the number of tests performed. Even if they were truly significant results, it would still be the case that there is not enough consistency in the results for the different lexical units and constructions.

All of this suggests that the body part possessor raising alternation is not sensitive to discourse pragmatic factors. This clearly runs counter to Langacker's prediction that raised Possessors ought to be local topics. Moreover, given that the tokens for the raised and unraised sentences in the body part raising alternation were much easier to code, the topicality effect should have been easier to verify for this alternation than in the case of the epistemic raising alternation.

3.3 Change property factoring alternation

We now turn to the change property factoring alternation. As the name suggests, the predicates that occur in this construction mostly denote an inchoative change in the Value of an Attribute of a Possessor (*rise, increase, drop*). However, the factoring out of a parameter or aspect of evaluation or measurement is also possible—though it seems less frequent—with certain other classes of predicates. One of them consists of predicates, mostly adjectives, that denote a static position of an Attribute on a scale. Another consists of predicates denoting a caused change in the Value of an Attribute. An additional semantic class not pointed out before includes verbs of comparison such as *compare*, which allow the dimension or parameter along which items are compared to be conflated with the item or not. The examples in (3.40) illustrate the variety of predicates that is found in the BNC with the Attribute *price* factored out. In the table I include nouns that occur with the same Attribute: at least for the Attribute *price*, the construction seems most frequent with nouns and nouns also exhibit greater type variety.

(3.40) 'X' in price

- a. intransitive: boom, change, come down, compare, drop, ease, fall, go down, go up, grow, increase, jump, move, range, rise, rocket up, soar, start, treble, tumble, vary
- b. transitive: force down, increase, pull up, reduce
- c. adjectives: accessible, comparable, competitive, different, high, identical, lower, reasonable, up, upper mid-range
- d. nouns: adjustment, advance, alteration, bounce, change, climb, collapse, crash, cut, decline, decrease, depreciation, difference, differential, disequilibrium, drop, escalation, explosion, fall, fluctuation, gain, improvement, increase, jump, mark-up, movement, parallelism, plunge, range, recovery, reduction, rise, slow down, slump, stickiness, surge, trend, upsurge variation

The list shows that there is very little if any vocabulary that is specialized to the domain of scalar-property change. Verbs like *double*, *triple*, *treble* etc. may be the ones most at home in this domain. The majority of other predicates have conventionalized metaphorical meanings in the change property domain but come from the motion domain.

The change property alternation is different from the body part possessor raising alternation in certain ways. Overall, it is more complex. For one, it logically involves two relations and three terms rather than one relation and two terms. The notion of scalar property requires not only a Possessor and some Attribute possessed by the Possessor but also a Value of the Attribute.

- (3.41) [The price [of Microsoft shares POSSESSOR] ATTRIBUTE] **rose** today [from \$28.89 VALUE1] [to \$28.91 VALUE2].
- (3.42) [Oil POSSESSOR] is expected to **rise** again [in price ATTRIBUTE] in tomorrow's trading.
- (3.43) [That figure VALUE/ATTRIBUTE] is expected to **increase** in the third quarter.
- (3.44) If Freeport's mine were in Nevada, [the stock POSSESSOR] would be [\$100 VALUE].

Secondly, as noted before, the verbal change property predicates do not allow joint expression with the Possessor as head of the NP.

- (3.45) * (The) Oil of that price is expected to **increase**.

Thirdly, in the case of scalar property change, the distinction between Possessor and Attribute is not always as clear as the distinction between Body Part and Body Part Possessor. For instance, *population*, *production* and *output* are sometimes used like a Possessor having a scalar magnitude Attribute, as is shown by (3.46), (3.47) and (3.48), and other times they are treated like scalar Attributes, as shown in (3.49), (3.50) and (3.51).

- (3.46) [The population] **increased** in number from 122,000 (1859) to 639,000 (1930) to 1,452,000 (1966)
- (3.47) [Apparel production] **increased** in value by about 26 per cent in 2001 compared to 2000.
- (3.48) For all industries and services between 1938 and 1948, [gross output] **increased** in value by 136 per cent.
- (3.49) Kansas **increased** [in population] from 2,688,418 residents in 2000 to 2,694,641 residents in 2001
- (3.50) Cranberries, turkey's fruity complement, have **increased** [in production] by 7 percent this year, Uhl said.
- (3.51) Carton board and other paper and board for packaging **increased** [in output] by 2.1%, to a record 2.95 million tonnes

Finally, the construal of Possessor and Attribute is not the same across all uses. In some cases, there is a simplex Possessor entity having a particular Attribute that changes over time as in (3.52).²¹

- (3.52) The tumor became smaller, she **gained** [in weight] and became stronger.

In other cases there is a multiplex Possessor with a magnitude Attribute related to group size that changes.

- (3.53) However, as the trade routes **grew** [in number], these settlements became centers for merchants, tradespeople, and traders.

In this latter situation it is possible with at least some verbs to omit the Attribute and just use the Possessor as subject, as shown in (3.54).

- (3.54) Road work fatalities **increased** last year.

²¹Notice that *gain* seems to occur only in the Possessor subject construction and does not allow the Attribute subject construction: **His weight gained dramatically*.

(3.55) Profit of saw industry **increased** significantly.

In the case of simplex Possessors and of multiplex Possessors whose Attribute does not relate to group size, omission of the Attribute is usually not possible, which is shown in (3.56).

(3.56) Buston found that if he removed the top-ranking female, the breeding male changed sex and **increased** *(in size) to become the female breeder

(3.57) These lesions **increased** *(in severity).

The above cases of Possessor subjects in the factoring construction are not to be confused with cases where a Value-denoting NP serves as subject of the non-factoring construction. In (3.58), what increases is of course not the incidence of math scores but the average score.

(3.58) Math scores **increased** significantly last year.

3.3.1 Factoring as a non-basic sentence type

Van Oosten (1986) discusses the change property factoring construction as a kind of non-basic sentence type with a non-prototypical subject. Whereas usually subjects have proto-agent and proto-topic properties, the subjects of change property factoring sentences are said to have subjects that make good sentence topics but that are not *primaries*, that is, they do not represent the case role that is highest in the case role hierarchy of the roles that are present.

For instance, in all three of the following examples (3.59a-c), the primary of the verb *go up* is *the price of rice*: Attributes, or more specifically, their Values, can increase but not Possessors, which do not have a direct semantic relation to the predicate at all.

- (3.59) a. The price of rice has **gone up**
 b. Rice has **gone up** in price
 c. Rice has **gone up**

So far Van Oosten's analysis may seem very similar to Langacker's active zone-type of account. However, Van Oosten embeds her treatment in a more elaborate account of topics. Van Oosten divides topics into sentence and discourse topics. Discourse topics are schemata or scenes evoked in a discourse which are relevant at the point of the discourse in question. There may not be any particular word or expression in the sentence that refers explicitly to such a schema or scene, but the schema or scene is nevertheless necessary to fully understand the sentence. A sentence topic by definition actually occurs as a word or a phrase in a sentence. It is the link in the sentence to one of the discourse topics: it is the *representor* of that discourse topic in the sentence.

The application of this distinction to the change property factoring construction is not particularly well described. As I understand what Van Oosten says, in the case of property factoring sentences, the primary, that is the Attribute, is an aspect of the representor, the Possessor. The sentence topic, the Possessor, is not the primary, but instead it is said to be the representor of the primary. In a commodities market context, the primary *the price of rice* is an aspect of the representor *rice*, which presumably represents the primary as a discourse topic rather than a different discourse topic such as the commodities market, which is still needed as background (p. 72).

The account in terms of discourse topic and representors is appealing in that it emphasizes coherence in the discourse even when sentence topics change. Nonetheless, it is difficult to imagine how one could establish sentence and discourse topics independently from each other in a non-circular fashion.

Still, Van Oosten's analysis is relevant in that it points in the same direction as Langacker's Active Zone account: using the change property factoring construction means selecting a more suitable sentence topic over one that is semantically more appropriate but less topical.

3.3.2 English corpus data

For the purposes of investigating the change property factoring alternation, I extracted two sets of sentences from the British National Corpus

- all sentences in which an NP headed by *price* occurred as subject of *increase*
- all sentences in which the PP *in price* occurred as an oblique complement of a change property verb

This method of data gathering was different from what I did for other constructions where I just looked for sentences containing the target verb and then sorted out sentences exemplifying unwanted senses or constructions manually. The reasons for focusing specifically on the combination of *increase* and *price* are as follows. First, *increase* is not as polysemous as some of the other verbs used in the construction such as *rise*, *drop*, which have motion-related senses. Second, *price* is a clear case of an Attribute, unlike other nouns such as *population* that have uses both as property-possessors and as properties, as was discussed above on page 333. For the property factoring sentences I had to leave the verb unspecified because the intransitive verb *increase* occurs in the factoring construction only once with the Attribute *price* as subject—though the factoring construction is quite common with the noun *increase*. One type of use that this way of collecting data does not catch are uses with only a Possessor subject and an implicit Attribute as in (3.60):

(3.60) Gas has **gone up**, again, this weekend

Overall, I collected 55 uses of the change property verb *increase* with the noun *price* heading the subject NP, and 58 sentences in which an intransitive change property verb occurred with the prepositional phrase *in price*. Both samples exhausted the population of such sentences in the BNC. The verbs occurring with *in price* in my sample were the following, repeated from (3.40) above:

(3.61) boom, change, come down, compare, drop, ease, fall, go down, go up, grow, increase, jump, move, range, rise, rocket up, soar, start, treble, tumble, vary

3.3.3 Formal properties

First, let us consider tense in the two constructions. The distribution of tense-aspect combinations is shown in Table 3.112

	Non-factoring		Factoring	
	n	%	n	%
Past	12	21.8	11	19.0
Past Perfect	6	10.9	1	1.7
Past Progressive	0	0.0	1	1.7
Present	7	12.3	19	32.8
Present Progressive	2	3.6	4	6.9
Present Perfect	15	27.3	8	13.8
Untensed	7	12.7	12	20.7
Future	6	10.9	2	3.4
Total	55	100.0	58	100.0

Table 3.112: Tense of *increase*

Most of the texts in the BNC in which the tokens occur are reports on the development of certain products, markets, or companies. Consequently, we mostly find many simple present and past tense as well as present and past perfect sentences. The factoring construction seems to have a higher proportion of simple present tense sentences and more untensed uses but fewer present perfect uses than the non-factoring construction. One possible explanation for this difference may be that the factoring construction occurs more often in generic statements.

We will next consider the morphology of the constituents encoding the Possessor and the Attribute. In the factoring construction, the morphology of the Attribute is always the same, a bare noun, e.g. *in price/value/temperature* etc. In the non-factoring construction, what we find is somewhat surprising. For it is not simple definite NPs that predominate but bare or plural NPs, as is shown in Table 3.113.²²

A closer look at the category bare/plural shows that in fact all the NPs included in the category are plural NPs. Typical examples such are:

- (3.62) a. Only three regions expect prices to **increase** by a meaningful amount.
 b. Prices for petrol will **increase** by 5 per cent.

The plural NPs do not denote the Attribute but rather Values determined/measured

²²Notice that in the four possessive NPs shown in Table 3.112, the possessives do not denote the Attribute-Possessor but rather the Seller of the Possessor, e.g. *The store's prices are very low.*

Morphology	n	%
Definite NP	14	25.5
Restrictive relative pronoun	1	1.8
Non-restrictive relative pronoun	1	1.8
Quantified NP	1	1.8
Bare/Plural NP	34	61.8
Possessive NP	4	7.3
Total	55	100.0

Table 3.113: Morphology of Attribute subject of *increase* in the non-factoring construction

for the Attribute. In (3.62b), there is no distinction involved, say, between retail and wholesale prices. Rather what the sentence means is that the average of price values (of the appropriate kind of price) will increase. Thus, it seems that in the non-factoring construction, Values are more common as subjects than Attributes. Of course, informationally, the difference between an Attribute and a Value-subject sentence is small in these cases since the two kinds of readings of the NPs are related by zero derivation.

The morphology of the Possessors, too, is rather different between the two constructions as Table 3.114 shows. In the factoring construction, around half of all Possessors have definite morphology, in the non-factoring construction only about a quarter do.

Note that indefinite zero realizations of the Possessor in the factoring construction are not possible at all rather than simply unattested. A sentence such as (3.63) is only acceptable if the Possessor is omitted under definite null instantiation and the Possessor is given in the linguistic context.

(3.63) **Increased** in price again.

Similarly, Attributes cannot be realized as zero elements in the non-factoring construction when there is an overt Possessor.

²³Since Possessors are subjects in the factoring construction they are not normally omissible. There is thus a particular grammatical rather than a discourse reason why zero possessors do not appear among the factoring data.

	Factoring		Non-factoring	
	n	%	n	%
Definite Zero	6	10.3	13	23.6
Third person pronoun	8	13.8	0	0.0
This	1	1.7	0	0.0
Definite NP	10	17.2	0	0.0
Non-restrictive relative pronoun	2	3.4	0	0.0
Zero pronoun of supplementary relative clause	2	3.4	0	0.0
Possessive NP	1	1.7	1	1.8
Definite NPs	30	51.7	14	25.5
Indefinite NP	1	1.7	1	1.8
Quantified NP	5	8.6	0	0.0
Bare/plural NP	19	32.8	25	45.5
Indefinite Zero	0*	0* ²³	13	23.6
Gerund	0	0.0	1	1.8
Restrictive relative clause	2	3.4	1	1.8
Reduced relative clause	1	1.7	0	0.0
Total	58	100.0	55	100.0

Table 3.114: Morphology of Possessor of *increase* in the factoring and non-factoring constructions

(3.64) *Of Microsoft **went up** yesterday.

By contrast, Possessors are quite likely to be zero-encoded in the non-factoring construction. In (3.65), the Possessor of the supply price-Attribute is not expressed but is to be co-identified with the referent of *commodity* in the preceding clause.

(3.65) As the flow, or (annual) amount of the commodity increases, the supply price may either **increase** or diminish; or it may even alternately increase and diminish.

Not all zero Possessors have been evoked in previous discourse. Often the hearer has to infer them from context. In (3.66), the phrase *its prices* omits overt mention of a Possessor: the possessive pronoun refers to the food manufacturing industry as the seller of products that have prices. The preceding sentence refers to *the commodities*, but the referent of that expression is not the relevant Possessor. Instead the products

of the food manufacturing industry have to be understood as the Possessor of the price Attribute whose Values are under discussion.

- (3.66) Since 1975, the prices farmers receive for commodities expressed in real terms has fallen by 16 per cent. Input costs have also fallen in real terms, but by only 3 per cent. Output, again in real terms, has increased by 25 per cent. The food manufacturing industry has benefited from the lower costs of the commodities it processes. Its prices have **increased**, however, and the consumer has not benefited from the efforts of farmers.

For another set of zero indefinites the Possessor is even vaguer: it is the class of all consumer goods and services.

- (3.67) Rising prices are perceived as a threat to living standards, so that the faster prices are **increasing** the more likely are workers to wish to join a union in order to maintain their standards of living.

As noted above, the Possessor is quite frequently expressed as a bare or plural nominal, either as a non-head in a compound as in (3.68) or as a complement of a head noun as in (3.69).

- (3.68) In addition, the world oil prices haven't **increased** as predicted, even as recently as two years ago, and that has also had a negative impact on the decisions taken by the world's oil producers.
- (3.69) For example, when the market index rises by 10%, kryptonite futures prices may increase by 20%(a ratio of 2), while the price of phlogiston futures may **increase** by only 5% (a ratio of a half).

As the preceding examples suggest, the Possessor is usually not a specific referent but rather a kind of commodity. It is not clear why this is so. On the one hand it might just be an artifact of the kinds of texts that have been included in the BNC; discussions of company and economic performance do not tend to involve individual items for which price changes over time are tracked. On the other hand it might

be the case that even when such a situation obtains, speakers or writers avoid using Attribute-subject constructions in favor of Possessor-subject predicates such as *get more expensive* or *get cheaper*.

3.3.4 Discourse properties of referents

Presence of prior mentions Given the morphology of the Attribute subjects involved, it is not surprising to find that in the non-factoring construction the subject referents usually have not been evoked in prior discourse but are just inferable from the preceding discourse, as can be seen from Table 3.115. In the factoring construction, the Attribute is even less likely to have been explicitly evoked, as is shown by Table 3.116.

A Fisher's exact test, shown in 3.117, performed on the evocation status of the Attributes across the two constructions yields a two-tailed P value of 0.0265. By conventional criteria, this result is to be considered statistically significant.

Turning now to Possessors, we find that in the non-factoring construction most Possessor referents are either evoked or inferable. However, compared to the Attributes in that construction, the relative proportions between evoked and inferable status are shifted. Possessors are slightly more likely to have been explicitly evoked than to be inferable. 50.9% of the Possessors have a prior mention, and 47.3% are inferable. Of the Attributes, 16.4% were evoked and 78.2% inferable (cf. Table 3.115).

In the factoring construction, Possessors seem to be more likely to have been explicitly evoked than in the non-factoring construction. Compare Tables 3.118 and 3.119. However, a Fisher's exact test, shown in Table 3.120, performed on the number of evoked and non-evoked Possessors in the factoring and the non-factoring constructions yields a two-tailed P value of 0.1866. By conventional criteria, this result is not statistically significant: we have no reason to believe that Possessors are more often explicitly evoked in the factoring construction than in the non-factoring construction.

Number of prior and subsequent mentions We have learned so far from Tables 3.115, 3.116, 3.118 and 3.119 that Possessors are more likely to have prior mentions

	n	%
Evoked	9	16.4
Inferrable	43	78.2
Variable	1	1.8
Non-evoked	2	3.6
Total	55	100

Table 3.115: Evocation status of Attribute of *increase* in the non-factoring construction

	n	%
Evoked	2	3.4
Inferrable	56	96.6
Total	58	100.0

Table 3.116: Evocation status of Attribute in property factoring construction with *increase*

	Factoring	Non-factoring	Total	p-value of Fisher's Exact test
Evoked	2	9	11	.0265
Non-evoked	56	46	102	
Total	58	55	113	

Table 3.117: Fisher's exact test for evocations status of Attributes

	n	%
Evoked	28	50.9
Inferrable	26	47.3
Variable	1	1.8
Total	55	100.0

Table 3.118: Evocation status of Possessor of *increase* in the non-factoring construction

Evoked	37	63.8
Inferrable	18	31.0
Variable	3	5.2
Total	58	100.0

Table 3.119: Evocation status of Possessor in property factoring construction with *increase*

	Factoring	Non-factoring	Total	p-value of Fisher's Exact test
Evoked	37	28	65	.1866
Non-evoked	21	27	48	
Total	58	55	113	

Table 3.120: Fisher's exact test for evocation status of Possessors

than Attributes. For instance, 37 out of 58 Possessors in the factoring construction had a prior mention compared to only 2 Attributes in the same construction. We will now explore how many prior and subsequent mentions Attributes and Possessors have within the ten sentence window of context in the two constructions. The findings are presented in Tables 3.121 to 3.124.

Tables 3.121 and 3.122 suggest that the average number of prior and subsequent mentions for both Possessors and Attributes are higher in the non-factoring construction than in the factoring construction.

If this is a statistically significant result, it represents a very interesting situation in two respects. First, when comparing the same participant across the two constructions, we find that for Attributes the situation is as expected: as subjects in the non-factoring construction they are more connected to prior discourse than as obliques in the factoring construction. The results are, however, surprising with respect to the Possessors in the factoring construction: even though that is the construction where the Possessors are subjects rather than just embedded in another NP, they actually have fewer prior mentions and seem less well connected to prior discourse than in the non-factoring construction.

Second, when we compare the two participants to each other in each construction, then the non-factoring construction appears to use a non-optimal subject. As is shown in Table 3.121, the Attribute is subject in that construction even though on average it less rarely has any prior mentions and, when it does, actually has fewer prior mentions (0.27) than the Possessor (1.44), as well as fewer subsequent mentions (0.55) than the Possessor (1.15).

In evaluating whether the differences between the above scores are statistically significant we cannot use parametric tests because the data points are not normally distributed for any of the eight measures. This is shown by Tables 3.123 and 3.124, which report the results of Kolmogorov-Smirnov tests for a normal distribution. In all cases, the significance level is $0.00 < 0.05$. The skewing of the distributions is more readily seen in the histograms in Figures 3.2 and 3.3, which represent the previous mentions of Attributes in the non-factoring construction and the subsequent mentions of Possessors in the factoring construction, respectively.

In order to test whether Possessors have significantly different numbers of prior mentions in the two constructions, we again use the non-parametric Wilcoxon-Mann-Whitney test for two independent samples. In our case, the null hypothesis is that there is no difference between Possessors across the two constructions. The Wilcoxon Two Sample Test for the Possessor data yields, for groups A = 55 and B = 58, $W = 3035.5$, and a two-tailed p-value ≤ 0.5696 , which strongly suggests that there is no difference in the number of prior mentions.

When we perform a Wilcoxon-Mann-Whitney test for two independent samples for the subsequent mentions, we get $W = 3087.00$ and a two-tailed value of $p=0.131$. Again, by conventional standards, there is no significant difference between the constructions in terms of subsequent mentions of Possessor referents.

We can proceed to perform the same test now for the differences between Attributes in the two constructions. When we perform the test for prior mentions, we get a Wilcoxon W of 3068.000 for the factoring construction. The p-value for a two-tailed test is 0.11, which means that we have no reason to believe that the constructions differ significantly in the number of prior mentions that Attributes have.

The next test to perform is for subsequent mentions of Attributes across the two

	Attribute		Possessor	
	Prior	Subsequent	Prior	Subsequent
N	55	55	55	55
Mean	.2727	.5455	1.4364	1.1455
Median	.0000	.0000	1.0000	.0000
Mode	.00	.00	.00	.00
Std. Deviation	.70592	1.41183	2.37864	2.04955
Variance	.49832	1.99327	5.65791	4.20067

Table 3.121: Prior and subsequent mentions for referents in the non-factoring construction

	Attribute		Possessor	
	Prior	Subsequent	Prior	Subsequent
N	58	58	58	58
Mean	.0345	.1034	1.1034	.4828
Median	.0000	.0000	1.0000	.0000
Mode	.00	.00	1.00	.00
Std. Deviation	.18406	.35981	1.37253	.90304
Variance	.03388	.12946	1.88385	.81549

Table 3.122: Prior and subsequent mentions for referents in the factoring construction

	K-S statistic	df	Sig.
prior mentions Attribute	.540	58	.000
subsequent mentions Attribute	.527	58	.000
prior mentions Possessor	.306	58	.000
subsequent mentions Possessor	.428	58	.000

Table 3.123: Kolmogorov-Smirnov test for normal distribution of number prior and subsequent mentions in the factoring construction

	K-S statistic	df	Sig.
prior mentions Attribute	.469	55	.000
subsequent mentions Attribute	.378	55	.000
prior mentions Possessor	.318	55	.000
subsequent mentions Possessor	.330	55	.000

Table 3.124: Kolmogorov-Smirnov test for normal distribution of number prior and subsequent mentions in the non-factoring construction

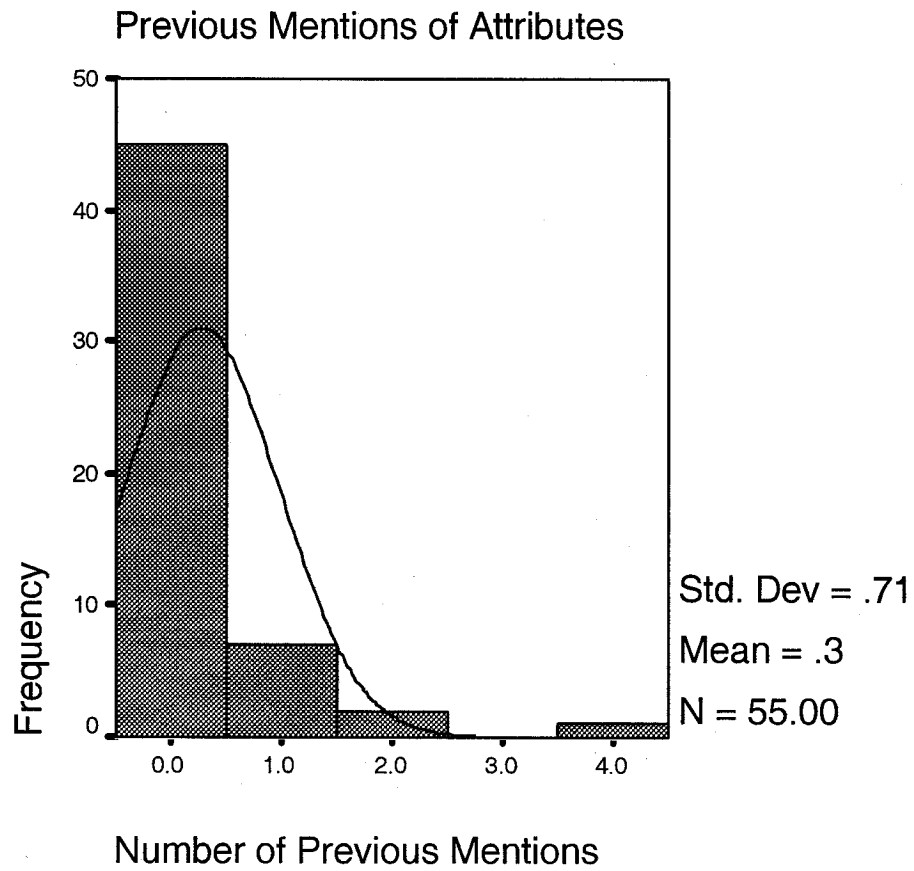


Figure 3.2: Previous mentions of Attributes in the non-factoring construction with *increase*

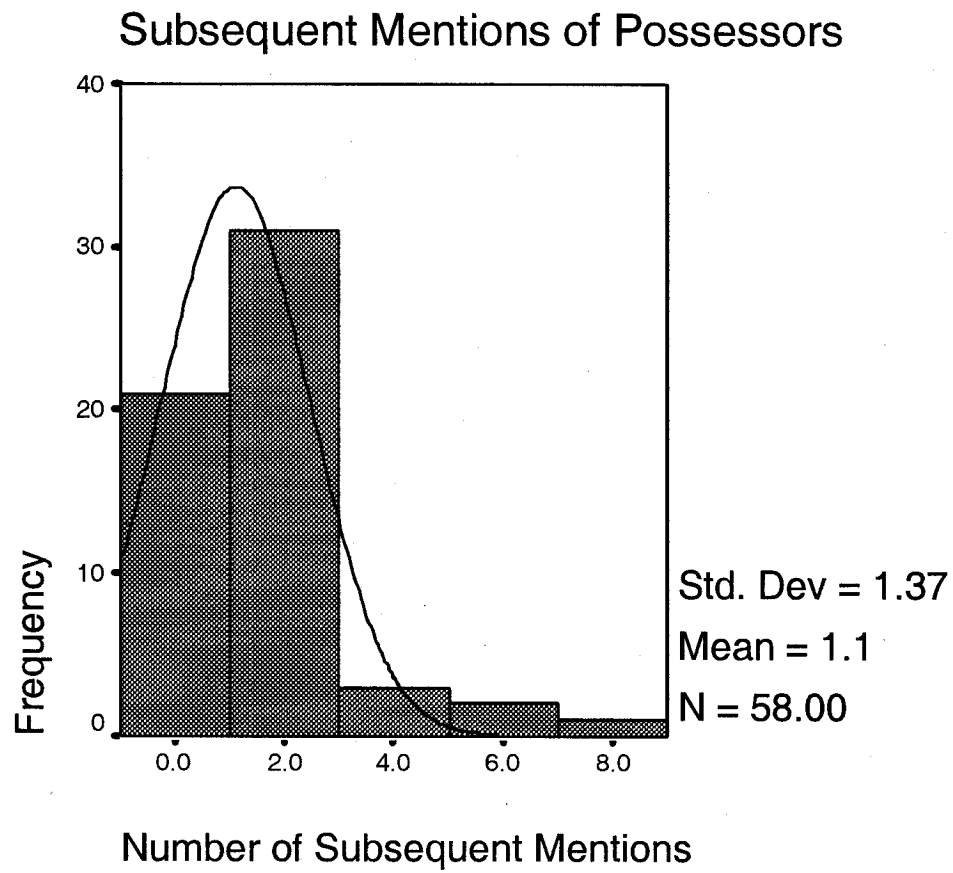


Figure 3.3: Subsequent mentions of Possessors in the factoring construction

constructions. We get a Wilcoxon $W = 3002.000$ for the factoring construction. The two-tailed p -value is 0.009. We thus have reason to believe that in the factoring construction Attributes have significantly fewer subsequent mentions than in the non-factoring construction.

We can also use the Wilcoxon-Mann-Whitney test to assess whether there are significant differences in the number of prior or subsequent mentions that the participant roles have in a given construction.

For prior mentions, we find that there is a significant difference between Attributes and Possessors in the non-factoring construction. The Wilcoxon Two Sample Test for this data, for groups $A = 55$ and $B = 55$, yields $W = 2520$ for the ranks of the Attributes, $p \leq 0.0015$, which strongly suggests that there is a significant difference.

The same test performed for prior mentions in the non-factoring construction, we get for groups $A = 58$ and $B = 58$ a Wilcoxon $W = 2365.00$ for the Attributes in the factoring construction and $p = 0.000$. We may assume then that Attributes and Possessors also differ significantly in terms of prior mentions in the factoring construction.

When we repeat the tests for a difference between Attributes and Possessors in terms of subsequent mentions for the two constructions we get the following results. In the factoring construction, Attributes have significantly fewer subsequent mentions (Wilcoxon $W = 3060.00$ for the rank sums of the Attributes and $p = 0.006$). In the non-factoring construction, there does not seem to be a significant difference (Wilcoxon W for Attributes = 2837.00 and $p = 0.121$).

Presence of other attributes and possessors in the context We next consider whether Attributes and Possessors co-occur with other Attributes and Possessors in the text. The purpose is to determine whether more frequent talk about Attributes and Possessors within the 10 sentence window of context correlates with one of the two constructions.

We find that in slightly more than half of the sentences the same Attribute is mentioned with another Possessor (cf. Table 3.125). Of the remaining cases, most involve a different Attribute with a different Possessor. In (3.70), for instance, in

an urban planning context, several indicators of the state of a neighbourhood are discussed. One of them is house prices, another is the size of the elderly population.

- (3.70) Between 1970 and 1980 owner occupancy was only declining in Fineview, house prices were **increasing** throughout the area and the elderly population was increasing above the average rate in only half the neighbourhoods.

	non-factoring		factoring	
	n	%	n	%
Same Attribute, Different Possessor	28	50.9	27	46.6
Different Attribute, Different Possessor	20	36.4	14	24.1
Different Attribute, Same Possessor	4	7.3	7	12.1
No other Attribute, No other Possessor	3	5.5	10	17.3
Total	55	100.0	58	100.0

Table 3.125: Presence of other Attributes and Possessors with *increase*

What's least likely is that a different Attribute of the same Possessor is mentioned, as in (3.71), where building costs and tender prices are Attributes of houses.

- (3.71) A spokesman for the Housing Executive said: "While building costs have increased the tender price has not **increased** at the same rate.

As Table 3.125 shows, there is no particularly striking difference between the constructions with respect to what other Attributes and Possessors the surrounding discourse mentions. The relative order of the four different context types is the same if we ignore the cases where no other Attribute and Possessor are mentioned. This impression is confirmed by a χ^2 test. With 3 degrees of freedom, $\chi^2 = 5.5887$. For significance at the .05 level, χ^2 should be greater than or equal to 5.99. Thus, by convention we would not consider the distribution significantly out of proportion.

The next factor I considered was whether Attributes or Possessors occurred in the immediate context of the current sentence. I recorded the type of the closest Attribute, Value, or Possessor-referring referent within 2 sentences before and after the target referent and its grammatical function. For the identity of the closest

Attribute or Possessor referent in preceding context, the results were as shown in Table 3.126.

	Non-factoring		Factoring	
	n	%	n	%
N/A	11	20.0	6	10.3
Different Attribute	9	16.4	2	3.4
Same Attribute	11	20.0	1	1.7
Different Value	8	14.5	4	6.9
Same Value	3	5.5	2	3.4
Total A/V	31	56.4	9	15.5
Different Possessor	7	12.7	18	31.0
Same Possessor	6	10.9	25	43.1
Total Possessor	13	23.6	43	74.1
Total	55	100	58	100.0

Table 3.126: Type of closest clause-level argument in preceding two sentences with *increase*

Table 3.126 shows that in terms of preceding local context, the subjects of the two constructions are often part of a chain of topics of the same type. For instance, in non-factoring constructions, 56.4% of all Attribute subjects had as their closest, highest-ranking clause-level antecedent referent one that either denoted an Attribute or a Value. This is illustrated in (3.72).

(3.72) Transport is considerably cheaper than in the UK but *prices* are on the increase. In fact all *prices* are **increasing**. On the whole though, it is a very cheap country to visit if you don't stay in the big hotels.

Only 23.6% of the Attributes had a Possessor as closest highest-ranking preceding referent. In 20% of the cases, there was neither kind of antecedent within the 2-sentence window. Thus, it appears that although based on the larger context Attributes in the non-factoring construction make poorer topics than the Possessors that co-occur with them, in their immediate context Attributes are in fact good topics.

We saw earlier that Possessor referents have about the same number of prior mentions in both constructions and that the slight difference between them suggested by

Tables 3.121 and 3.122 is not significant. Table 3.126 shows that in the more local context of 2 sentences before the target use, Possessors in the factoring construction do make better topics than in the non-factoring construction. In the factoring construction, 74.1% of Possessors are most closely preceded by another Possessor-type referent, with 15.5% having an Attribute/Value-type referent as closest clause-level antecedent, and 10.3% not having either an Attribute/Value or Possessor antecedent within 2 sentences of prior context. In the non-factoring construction, by comparison, only 23.6% of Possessors have a Possessor-type referent as their closest antecedent in the 2-sentence context.

A χ^2 test reducing the number of categories to three—Attribute, Possessor, Neither—confirms that the two constructions differ in regard to which kind of referent is likely to be the closest antecedent: with 2 degrees of freedom and $\chi^2 = 29.5832$, p is less than or equal to 0.001. By convention, we should consider this distribution to be significantly out of proportion.

More evidence that both constructions actually have locally appropriate sentence topics comes from considering the data for the 2-sentence window of context following the target sentence, which is presented in Table 3.127.

As Table 3.127 shows, in following context, too, the constructions differ with respect to which kind of referent is likely to be the nearest subsequent mention. A χ^2 test with the reduced categories Attribute, Possessor, Neither confirms that the constructions differ significantly with respect to which kind of participant referent is likely to be the closest antecedent: with $df. = 2$ and $\chi^2 = 41.0731$, p is less than or equal to 0.001. The distribution is highly significant. It appears that in subsequent context, too, the type of referent that heads the current sentence's subject is most likely to match the closest subsequent mention of an element of the Attribute, Value, Possessor set.

So far we have only considered what the type of the closest highest-ranking preceding or subsequent mention is but not whether such mentions do in fact occupy grammatically prominent positions in the clauses they occur in. The grammatical functions of these mentions are presented in Tables 3.128 and 3.129. Table 3.128 shows the data for the non-factoring construction. In both preceding and subsequent contexts, the mention is most often a subject or a direct object. The share of these

two grammatical functions combined appears even greater when we disregard those tokens that had no prior mention of an Attribute or Possessor within the 2 sentence window of context.

Table 3.129 shows that the same observation holds for the grammatical function of the closest preceding and subsequent Attribute or Possessor mentions in the factoring construction.

Distance of contextual mentions The values for the distance of closest prior, closest subsequent, and last subsequent mentions of both Attributes and Possessors are as shown in Tables 3.130 and 3.131. Note that I assigned a distance value of 11 to those tokens that had no prior or subsequent mention within the standard 10 sentence window of prior and subsequent context that I have been assuming throughout. By contrast, for the last subsequent mention measures I assigned a value of 0 when there was no subsequent mention at all.

Tables 3.130 and 3.131 show that on average the closest contextual mentions of Attributes are more distant from the current mention than the closest contextual mentions of Possessors. The tables also show that for Attributes in both constructions, the closest mention measures all have a Median value of 11. With Possessors, this is the case only for the closest subsequent mentions within the non-factoring construction. In other words, more than half of Attributes lack a contextual mention of the specified kind within the window of context. Further it is the case that all closest mentions measures for both Attributes and Possessors have 11 as the sole mode of their distributions. The tables also show that with both constructions the last subsequent mentions of Possessors are farther away from the target mention than those of Attributes. For all four measures, median and mode have the value 0.

The distributions of the values for the 12 measures presented in Tables 3.130 and 3.131 above are not normal, as is shown by the results of a Kolmogorov-Smirnov test in Table 3.132. An illustration of this is given by Figure 3.4, which shows the histogram for the non-normal distribution of the distance values of closest subsequent mentions of Possessors in the non-factoring construction.

To compare differences between constructions and between the participant roles

	Non-factoring		Factoring	
	n	%	n	%
N/A	18	32.7	17	29.3
Different Attribute	13	23.6	3	5.2
Same Attribute	15	27.3	1	1.7
Different Value	5	9.1	2	3.4
Same Value	1	1.8	2	3.4
Total A/V	34	61.8	8	13.8
Different Possessor	3	5.5	23	39.7
Same Possessor	0	0.0	10	17.2
Total Possessor	3	5.5	33	56.9
Total	55	100.0	58	100.0

Table 3.127: Type of closest clause-level argument in following two sentences with *increase*

	Preceding		Subsequent	
	n	%	n	%
Subject	29	52.7	26	47.3
Direct Object	7	12.7	4	7.3
Oblique	3	5.5	3	5.5
Embedded in NP	4	7.3	4	7.3
Left dislocated element	1	1.8	0	0.0
N/A	11	20.0	18	32.7
Total	55	100.0	55	100.0

Table 3.128: Grammatical function of closest clause-level argument in subsequent and preceding two sentences with *increase* in the non-factoring construction

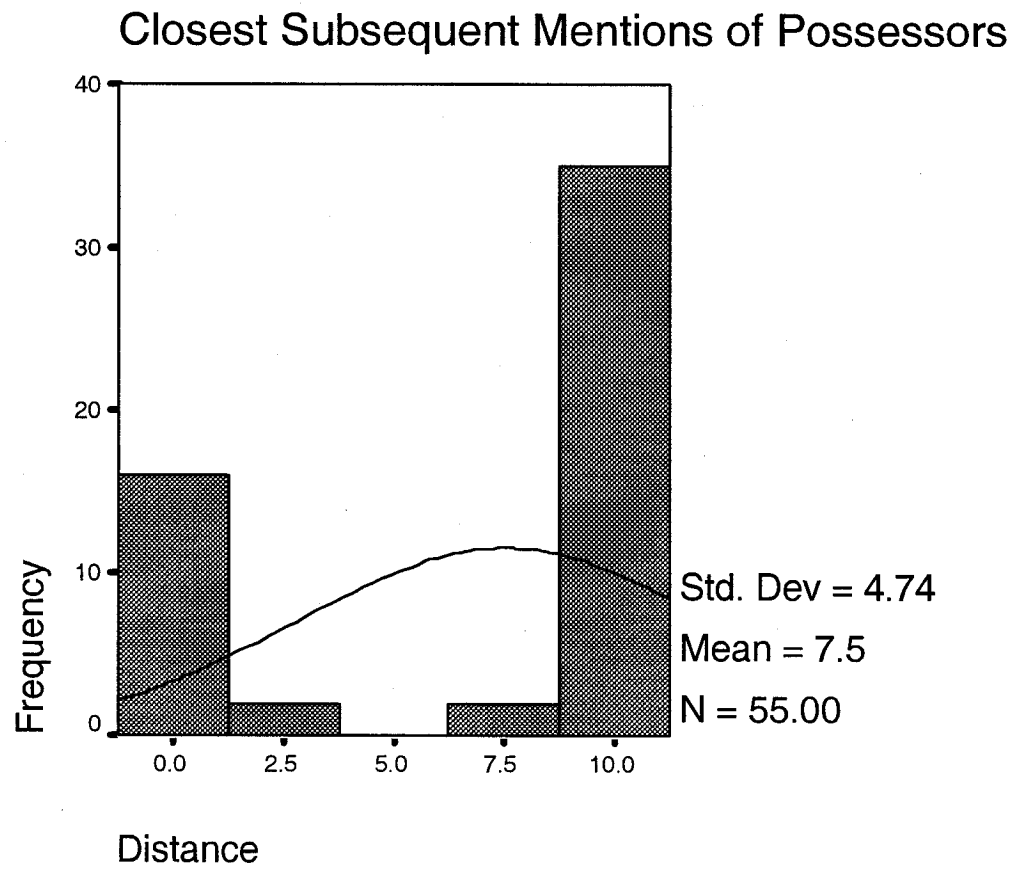


Figure 3.4: Distance of closest subsequent mentions of Possessors in the non-factoring construction

	Preceding		Subsequent	
	n	%	n	%
Subject	31	53.4	31	53.4
Direct object	11	19.0	8	13.8
Oblique	4	6.9	0	0.0
<i>there</i> -construction	2	3.4	2	3.4
Embedded in NP	3	5.2	0	0.0
Other	1	1.7	0	0.0
N/A	6	10.3	17	29.3
Total	58	100.0	58	100.0

Table 3.129: Grammatical function of closest clause-level argument in two preceding and subsequent sentences with *increase* in the property factoring construction

within a construction, we have to use the Wilcoxon-Mann-Whitney test. We will begin by comparing the measures for a given participant across the two constructions.

When we compare Possessors across constructions in terms of distance to prior mention, distance to closest following mention, and distance to last mention, we find that there are no significant differences in all three cases. This is shown by Tables 3.133 and 3.134.

When we use a Wilcoxon-Mann-Whitney test to compare Attributes across constructions in terms of distance to prior mention, distance to closest following mention, and distance to last mention, we find that there are significant differences in all three cases. The results of these tests are shown in Tables 3.135 and 3.136. Attributes in the non-factoring construction have prior mentions that are closer than Attributes in the factoring construction. They also have their next subsequent mentions at closer distances than Attributes in the factoring construction. And, finally, their last subsequent mentions are farther away from the target mentions than is the case with Attributes in the factoring construction.

We now turn to comparing the participants within a given construction with respect to the distance measures. The results of Wilcoxon-Mann-Whitney tests are shown in Tables 3.137-3.140.

In the factoring construction, the two participants differ significantly along all three parameters, as can be seen from Tables 3.137 and 3.138. Possessors have closer prior

	Attributes			Possessors		
	closest prior	closest subsequent	last subsequent	closest prior	closest subsequent	last subsequent
N	55	55	55	55	55	55
Mean	9.5273	8.7636	1.0545	6.2182	7.5273	1.4909
Median	11.0000	11.0000	.0000	7.0000	11.0000	.0000
Mode	11.00	11.00	.00	11.00	11.00	.00
Std. Deviation	3.37120	3.90139	2.30473	4.95420	4.74480	2.91784
Variance	11.36498	15.22088	5.31178	24.54411	22.51313	8.51380

Table 3.130: Distance values for contextual mentions of participants in the non-factoring construction

	Attributes			Possessors		
	closest prior	closest subsequent	last subsequent	closest prior	closest subsequent	last subsequent
N	58	58	58	58	58	58
Mean	10.6552	10.1379	0.18977	4.6724	8.2586	0.6724
Median	11.0000	11.0000	.0000	2.0000	0.000	0.0000
Mode	11.00	11.00	.00	11.00	11.00	0.00
Std. Deviation	1.8406	2.8374	1.06716	4.9147	4.5289	1.88613
Variance	3.388	8.051	1.13884	24.154	20.511	3.55747

Table 3.131: Distance values for contextual mentions of participants in the factoring construction

Attribute		non-factoring			factoring		
		Statistic	df	Sig.	Statistic	df	Sig.
Possessor	prior	.487	55	.000	.540	58	.000
	subsequent	.444	55	.000	.533	58	.000
	last	.422	55	.000	.502	58	.000
Attribute	prior	.324	55	.000	.310	58	.000
	subsequent	.386	55	.000	.452	58	.000
	last	.368	55	.000	.432	58	.000

Table 3.132: Kolmogorov-Smirnov tests for distributions of closest contextual mentions of participants in the non-factoring construction

	Construction	N	Mean Rank	Sum of Ranks
closest prior	Factoring	58	52.71	3057.00
	non-factoring	55	61.53	3384.00
	Total	113		
closest subsequent	Factoring	58	59.36	3443.00
	non-factoring	55	54.51	2998.00
	Total	113		
last subsequent	Factoring	58	53.29	3091.00
	non-factoring	55	60.91	3350.00
	Total	113		

Table 3.133: Difference between distance values for Possessors across constructions

and subsequent mentions and their last mentions within the window are more distant from the current mention than is the case with Attributes. In the non-factoring construction, only one comparison comes up with a significant difference, as shown in Tables 3.139 and 3.140. For Possessors, the closest prior mention is closer than with Attributes ($p=0.000$). In terms of closest subsequent and last subsequent mentions, there is by conventional standards no statistically significant difference discernible.

Summary Of the three raising-type alternations that we have considered, the change property factoring alternation is the only one that gives some clear evidence of topicality playing a role. Attribute referents are significantly different across the factoring and the non-factoring constructions. Specifically, the Attribute referents in

	closest prior	closest subsequent	last subsequent
Mann-Whitney U	1346.000	1458.000	1380.000
Wilcoxon W	3057.000	2998.000	3091.000
Z	-1.503	-.948	-1.591
p (2-tailed)	.133	.343	.112

Table 3.134: Statistics for difference between distance values for Possessors across constructions

	Construction	N	Mean Rank	Sum of Ranks
closest prior	factoring	58	60.97	3536.00
	non-factoring	55	52.82	2905.00
	Total	113		
closest subsequent	factoring	58	61.87	3588.50
	non-factoring	55	51.86	2852.50
	Total	113		
last subsequent	factoring	58	51.66	2996.50
	non-factoring	55	62.63	3444.50
	Total	113		

Table 3.135: Difference between distance values for Attributes across constructions

	closest prior	closest subsequent	last subsequent
Mann-Whitney U	1365.000	1312.500	1285.500
Wilcoxon W	2905.000	2852.500	2996.500
Z	-2.471	-2.440	-2.791
p (2-tailed)	.013	.015	.005

Table 3.136: Statistics for difference between distance values for Attributes across constructions

the non-factoring construction appear to be more topical than those in the factoring construction:

- they are more likely to have been evoked before²⁴
- they have more subsequent mentions

²⁴The number of prior mentions, however, did not seem to be significantly different (cf. p. 345).

	Construction	N	Mean Rank	Sum of Ranks
closest prior	Attribute	58	76.05	4411.00
	Poss	58	40.95	2375.00
	Total	116		
closest subsequent	Attribute	58	64.09	3717.00
	Poss	58	52.91	3069.00
	Total	116		
last subsequent	Attribute	58	54.41	3155.50
	Poss	58	62.59	3630.50
	Total	116		

Table 3.137: Difference between distance values for Attributes and Possessors in the factoring construction

	closest prior	closest subsequent	last subsequent
Mann-Whitney U	664.000	1358.000	1444.500
Wilcoxon W	2375.000	3069.000	3155.500
Z	-6.706	-2.668	-2.189
p (2-tailed)	.000	.008	.029

Table 3.138: Statistics for difference between distance values of Attributes and Possessors in the factoring construction

- they have closer prior and subsequent mentions
- they have last subsequent mentions that are farther away

The above facts are complemented by the observation that Attributes and Possessors are significantly different in terms of almost all measures in the factoring construction but they are more alike in the non-factoring construction. Given that the properties of Possessors in the two constructions are not significantly different, the greater similarity of Attributes and Possessors in the non-factoring construction must be due to the Attributes being more topical, which agrees with the findings above for Attributes. Another piece of evidence for a difference between the two constructions is the data on the presence of other Attributes and Possessors in the narrower 2 sentence context. Both Attributes and Possessors are frequently part of topic chains such as the one in (3.73) where referents of the same type follow each

	Construction	N	Mean Rank	Sum of Ranks
closest prior	Attribute	55	65.35	3594.50
	Poss	55	45.65	2510.50
	Total	110		
closest subsequent	Attribute	55	59.24	3258.00
	Poss	55	51.76	2847.00
	Total	110		
last subsequent	Attribute	55	53.46	2940.50
	Poss	55	57.54	3164.50
	Total	110		

Table 3.139: Difference between distance values for Attribute and Possessors in the non-factoring construction

	closest prior	closest subsequent	last subsequent
Mann-Whitney U	970.500	1307.000	1400.500
Wilcoxon W	2510.500	2847.000	2940.500
Z	-3.825	-1.477	-.835
p (2-tailed)	.000	.140	.404

Table 3.140: Statistics for difference between distance values of Attributes and Possessors in the non-factoring construction

other as sentence topics.

(3.73) Store cattle: Medium weight (300399 kg) *Charolais cross store steer prices increased* by 3.8p to average 160.8p per kg lw in the week ended May 29. *Hereford cross steer prices fell* by 4p to 144.4p per kg lw while *Limousin cross prices fell* by 2.7p to average 155.7p per kg lw. *The GB cull cow price fell* by 1.16p to 109.86p per kg lw during the Thursday/Wednesday week ended June 2.

Due to the way the sample was constructed—for one of the constructions we required a particular verb, while we allowed all comers for the other—it is not possible to perform a regression analysis in order to see how much of the data can be accounted for by the pragmatic factors that we have found rather than, say, the mere relative frequencies of the constructions. Nevertheless, it is clear that with this type of raising

we have at least been able to observe some reliable statistical differences between the discourse properties of the referents in the two constructions.

3.4 Summary

In this chapter, we have examined three raising constructions in terms of the discourse properties of their referents. Counter to the predictions of Langacker's account, there is no consistent correlation between an active zone/reference point semantics and a local topic status of the active zone referent. While the epistemic raising construction and the body part possessor raising construction show no reliable correlation with any discourse pragmatic properties of the argument that alternates between the status of clause-level argument and being embedded in a clause-level argument, we do find a good correlation between the choice of construction and pragmatic properties in the case of the change property factoring alternation. However, in that case the pragmatic difference does not attach to the Possessor referent that is encoded either as a clausal argument or embedded in a clause-level argument. Instead, it is the Attribute, which is always realized as a clause-level dependent of the change property predicate whose properties differ markedly across the constructions.

It seems plausible that the choice between the alternates of the change property factoring alternation depends on the pragmatics of the Possessor and Attribute referents than on a detectable semantic difference since, in coding and analyzing that data, I did not notice the kind of semantic differences that seem to play a role with the epistemic raising alteration and with the Body Part possessor alternation. In the former case, Evidence and Cognizer frame elements appeared significantly more often with the unraised construction than with the raised one. In the latter case, the raised construction provides more detail about the interaction with the Possessor's Body Part than the conflated construction: *slap somebody across the face* specifies more information than does *slap somebody's face*.

The relevant notion of specificity for the difference between body part possessor raising sentences is, of course, strictly concerned with the imagined scene. Pragmatically, it seems that no argument can be made for a discourse tendency to have more

topical Possessors in the raised construction. In fact, pairs like (3.74a-b), brought to my attention by Charles Fillmore, suggest that in some sense the Possessor in the conflated construction has a higher minimum activation level than the Possessor in the raised construction.

- (3.74) a. I feel like **punching** someone's nose.
b. I feel like **punching** someone in the nose.

Sentence (3.74a) naturally fits into a context where the speaker has a particular victim in mind; (3.74b), by contrast, is perfectly compatible with a situation where the speaker is happy with anybody serving as a victim. This difference makes sense given that in (3.74a) the Possessor as syntactic possessor needs to be able to anchor the Body Part referent and thus at least be specific, or identifiable to the speaker. In (3.74b) the victim is realized as a direct object, a grammatical function that is well known to be associated with new referents.

In sum, not all raising constructions support the idea that the choice between valence alternates might be conditioned by the discourse properties of the referents and that generalizations across similar alternations are possible.

Chapter 4

Null instantiation

This chapter is concerned with the conditions under which arguments of predicates can be omitted; how they are then interpreted; and what, if anything, determines whether a predicate has an argument that can be left implicit. If one wanted to think of implicit arguments in terms of alternations, the alternation would involve overt realization versus non-realization.

Before we consider an overview of the types of argument omissions that occur, let me briefly discuss in which sense ‘argument’ is to be taken for our purposes. Linguists (and maybe ordinary speakers, too) do not treat all the kinds of understood information that has no surface realization as unexpressed or omitted in the same way. Linguists usually make a distinction between arguments and adjuncts based on the intuition that some elements have a closer connection to the predicate than others. For instance, expressions of time and place, though they are ontologically necessary for events and compatible with stage-level readings of state predicates, are not felt to be semantically specific to most kinds of events and are called adjuncts. As far as I know, nobody has discussed the absence of time or place setting expressions in the majority of sentences in terms of argument omission.

Though I share the intuition that time and place expressions are not arguments for the majority of verbs and though there exists a vast literature on the topic, it is difficult to formulate a precise account of the argument-adjunct distinction. In the work of FrameNet, complements are classified as arguments and modifiers by

semantic and morpho-syntactic criteria. Semantically, we might distinguish between distinctive elements, arguments, and more peripheral notions such as place, time, manner, etc, which may be ontologically necessary for certain types of events. What FrameNet calls core is very similar to what Koenig et al. 2002 call lexically encoded participant information. That information is said to reduce to two semantic criteria. First, whether participant information is semantically obligatory and entailed for the whole class of situations denoted by a word. Second, whether participant information is specific to a verb or to a restricted verb class to which a verb belongs. Again, the basic intuition here is very much like FrameNet's. The reader is referred to Koenig et al. for discussion of psycholinguistic experiments intended to substantiate the theoretical analysis.

Based on syntax, we would want to include among arguments all those participants that can be realized in the two core-grammatical functions of subject and direct object as well as all participants for which a definite interpretation is required. Also considered core should be all frame elements that always have to be realized or understood as omitted by definite null instantiation (see below). Specification of an idiosyncratic prepositional marker for a frame element realized as an oblique is also a sign of core status: it is intuitively clear that a lexical unit should not say anything about the shape of adjuncts. The verbs *rely (on)* and *depend (on)* in the Contingency frame clearly fulfill this criterion. Unlike with real spatial/locative verbs such as *put*, only one preposition can be used to mark the Determinant frame element of these verbs. And no convincing argument can be made that *on* just happens to be the etymologically and/or metaphorically most appropriate choice given that, for instance, French and German use the Source-marking prepositions *de* and *von*, respectively.

(4.1) This also **depends** {on/*from/*with/*to/*off/*for} the operating system being used.

A further criterion that one might consider is incorporation. One could say that if a frame element can be incorporated then it must be a core argument for all verbs in its class. This would work, for instance, for verbs like *box*, *bag* and *pot* which incorporate a Goal frame element. The Goal frame element is also clearly core for

simple verbs of Placing such as *put*, *place* or *lay*. However, as has been pointed out to me, using incorporation as a coreness criterion may lead to incorrect predictions in the case of Means and Manner frame elements, as illustrated by (4.2) and (4.3).¹

(4.2) He **handed** the envelope to Dunbar, whose face was blank with astonishment.

(4.3) She **hurried** him into the cab.

It is not clear whether adding a criterion of class-specificity would allow one to keep incorporation as a criterion. Manner and Means may seem to be much more general notions than, say, Goal and Source but it is unclear just how generality should be assessed. So, one should probably consider incorporation only as suggestive but not sufficient evidence for core status.

Besides arguments and peripheral elements, one might recognize a further class, extra-thematic elements, which are typically introduced into clauses by special constructions and which are not readily considered ontologically necessary for the event types denoted by the verbs they combine with in the way that, say, time and place are. Rather result phrases, depictives, and comitative expressions introduce additional larger frames. Of these three different types of semantic elements, we will only be interested in analyzing semantically distinctive arguments in terms of their omissibility.²

Finally, I would like to point out right at the beginning that, although the focus of the discussion will be heavily on verbs, much of the discussion in principle also applies to predicates of other grammatical categories. An obvious exception is, as will be seen below, the set of omission types that are licensed by verb-headed constructions.

¹This was pointed out to me by Paul Kay in personal communication.

²The distinctions drawn here appear similar to those made for PPs within Role and Reference Grammar. Van Valin and LaPolla (1997) categorize one kind of PP as argument-marking prepositions. These correspond to core frame elements. A second group consists of adjunct prepositions, which predicate over the core of the clause. This group includes Time and Place setting expressions and corresponds to my peripheral elements. The third group consists of so-called argument-adjunct prepositions. These are predicates in their own right, introduce an argument into the clause and share it with the logical structure of the core, rather than taking the logical structure of the core as an argument. For-benefactive are cited as an example of this type of PP. This type corresponds to my extra-thematic elements.

4.1 Typology of non-overt arguments

There are a considerable number of different types of argument omission. We will be interested in detail in only a few of them, which I will refer to under the heading of null instantiation. In order to clarify what those types are it will be useful to consider the fuller typology given in Figure 4.1.

4.1.1 Ellipsis versus omission

Not all cases in which a part of a sentence is missing are omissions of subcategorized frame elements. I would like to argue in particular that the focal ellipsis construction exemplified in (4.4) is in fact an ellipsis construction rather than an argument omission construction.³

- (4.4) As Schellenberg and Devlin got out, a sergeant emerged from the hut with the radio mast, and hurried towards them. He took in Schellenberg's uniform and got his heels together. "General."
 "And your **name** is \emptyset_{Value} ?"
 "Leber, General. Flight Sergeant."

Unlike the omission constructions we will encounter, what is missing in the case of focal ellipsis is a focus rather than a referent that has active discourse-status and/or bears the pragmatic relation topic to the proposition expressed by the clause.

In context, the illocutionary force of the construction is that of a question. Since the subject omission with imperatives is also associated with illocutionary force, the specification of an illocutionary force by itself does not argue strongly that focal ellipsis is not an omission construction. Gretsche actually (2003) expresses doubt that the focal ellipsis construction itself is of the illocutionary type question. Instances of the German focal ellipsis construction, for instance, do not combine with the particle *denn*, in which respect they behave like declarative clauses rather than questions.

³Unlike some of the constructional omission types, this ellipsis construction does not seem to be widely known and discussed. The only reference I am aware of is an analysis by Gretsche (2003), who discusses the construction in German. The properties of the German and the English focal ellipsis constructions are identical as far as I am able to determine.

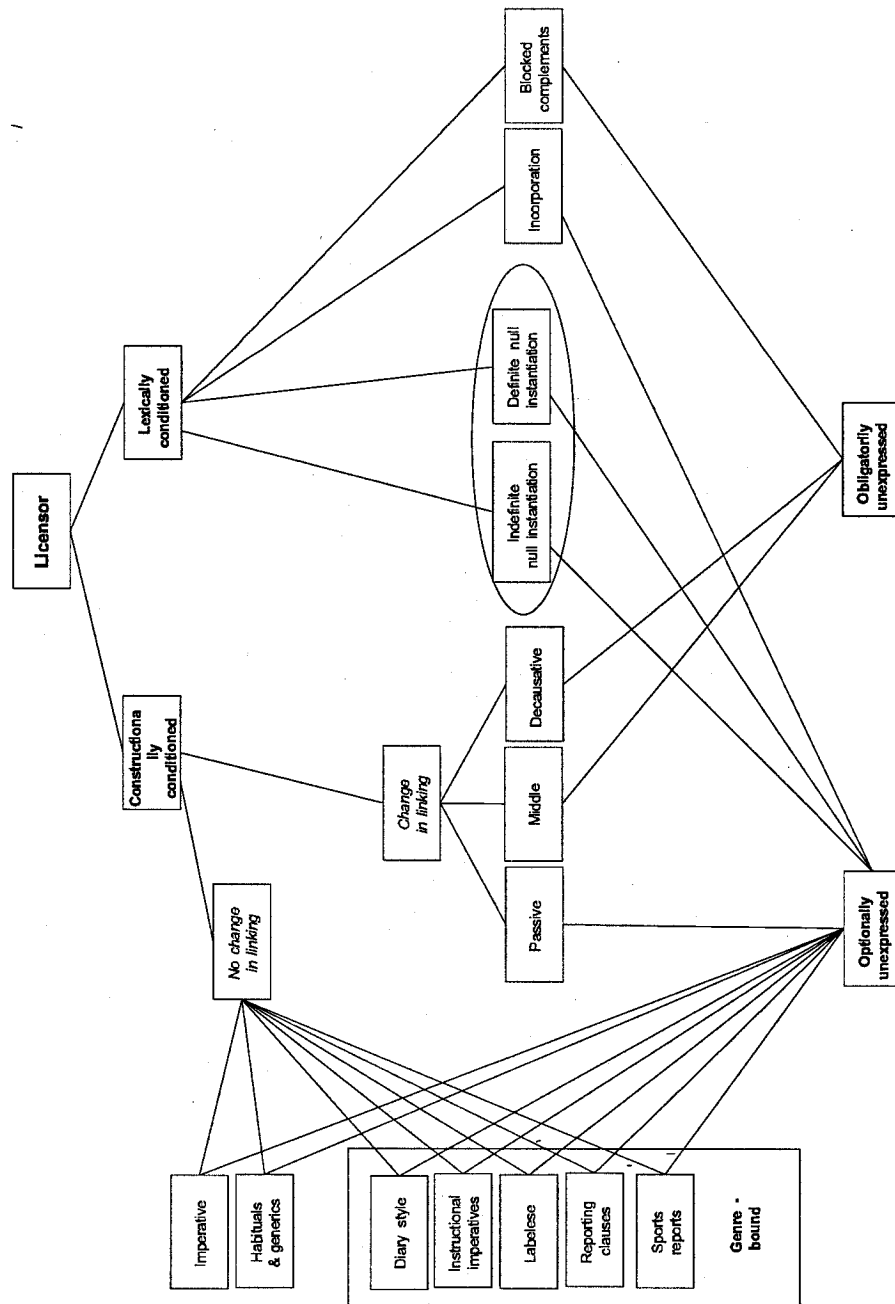


Figure 4.1: Typology of unexpressed arguments

- (4.5) #Das Dessert macht denn Johann.
 ‘So dessert will be made by John.’ (= Gretsche 2003:346, (13a))
- (4.6) Wer macht denn das Dessert?
 ‘So who will make the dessert?’ (= Gretsche 2003:346, (13b))
- (4.7) #Das Dessert macht denn \emptyset
 So dessert will be made by \emptyset ? (= Gretsche 2003:347, (14))

Gretsche argues that the question interpretation derives via Gricean inference from the *completive* focus structure of the sentence.⁴

The ellipsis site is usually at the end of an utterance but it seems it does not have to be at the end of the sentence as a whole, however. As examples (4.8-4.9) show, a vocative or a discourse-level conditional, for instance, may follow the clause. If these elements are not altogether in separate utterances, they are at least not part of the clause propositions missing their focus.

- (4.8) “And your **name** is \emptyset_{Value} , sir?”
- (4.9) “And You **are** \emptyset_{Value} , if I may ask ...”—“Regina Cartwright, pleased to meet you professor.”

The other constructional omission types have no requirement that the site of the zero element be at a clause boundary. Even imperatives do not have to have their zero clause-initially (*Now $\emptyset_{Self-mover}$ go home and practice*).

Focal ellipsis is mostly a main-clause phenomenon but not always. Example (4.10) shows that focal ellipsis can occur in at least some complement clauses.⁵

- (4.10) You were telling me that you don’t like to eat \emptyset_{Value} ?

⁴For further details on this analysis see Gretsche’s paper and the papers on focus structure by Drubig (1997, 2000) cited therein.

⁵Paul Kay pointed out the fact that focal ellipsis can occur in non-main clauses and provided the example.

Still, focal ellipsis cannot occur in complement clauses that are presupposed, as is shown in (4.11).⁶

(4.11) #Only then did I realize that you don't like to eat \emptyset ?

Although they may seem to be, focal ellipses are not merely echo questions with a silent *wh*-word. First, echo-*wh* elements can appear in various positions within the sentence, whereas the gap in focal ellipsis has to be at the very right edge of the clause (and usually actually occurs at the end of the utterance).

(4.12) You **ate** WHAT for dinner?

(4.13) #You **ate** \emptyset_{Food} for dinner?

Second, not all instances of the construction can be replaced/paraphrased with an overt question. In (4.14) the head of a predicate nominal phrase is missing and in (4.15), a particle is questioned. Neither of them are clause-level arguments.

(4.14) A: The glass is half \emptyset ?

B: {Full/Empty}

(4.15) A: You **kicked** the ball \emptyset ?

B: {In/off}

It is even possible to get word-internal focal ellipsis, as shown in (4.16).

(4.16) So the name of the guy is Gold-?

These cases argue most convincingly against treating focal ellipsis as a phenomenon involving the non-realization of valence elements. Another observation that similarly suggests that focal ellipsis is governed by different principles than lexical and constructional argument omissions is the fact that focal ellipsis allows prepositions to omit their objects which can never omit them lexically or in other omission constructions. Compare examples (4.17) and (4.18).

⁶Maybe a more accurate generalization would be that it can occur only in non-matrix clauses that are not pragmatically subordinated.

(4.17) The box **contains** $\emptyset_{\text{Contents}}$?

(4.18) Can I use this box?—I am afraid not. #It's not empty, it **contains** \emptyset

For the purposes of this study, I will thus assume that focal ellipsis exemplifies a phenomenon other than that of omission of a subcategorized valence element.

4.1.2 Obligatory versus optional omission

One of the major dichotomies given in Figure 4.1 may at first appear a bit surprising to the reader: obligatorily versus optionally omitted arguments. What sense could it make to speak of obligatorily omitted arguments?

Consider the phenomenon of what Talmy (1996:241) calls *blocked complements*. Blocked complements are elements that have to be understood in the conceptual background but cannot ever be realized in the syntax. For instance, the conceptually necessary seller of goods cannot be expressed as a complement of *spend*.

(4.19) Government agencies have **spent** more than 50 million dollars (*to/at/with/on/from high-tech companies) during the past five years developing camera surveillance technology.

This phenomenon is not as unusual and idiosyncratic as it may at first appear. German for instance, has a set of verbs prefixed with *zu-* that denote acts of getting hold of an object (*zubeissen* 'bite', *zufassen* 'grasp', *zupacken* 'grasp, grip'). However, distinct from the case of the unprefixated verbs, the object can never be realized overtly.

(4.20) Der Hund hat mich **gebissen**.

The dog bit me.

(4.21) Der Hund hat (*mich/in/auf/gegen/mit/zu/für mich) plötzlich **zugebissen**.

The dog suddenly bit me.

(4.22) Was ist heute in der Schule passiert? *Ein Hund hat **zugebissen**.

What happened at school today?—*A dog bit.

Note that a specific object has to be given in prior context or inferred from it in the case of (4.20-4.22), while the object is only existentially bound in the case of (4.19).

The middle construction, exemplified in (4.23) and (4.24), is another case that involves the obligatory blocking of a conceptually necessary element.

(4.23) The note **reads** like a disturbed poem from the mind of an anarchist.

(4.24) The Hyundai Sonata GLS: **Drives** like a dream!

Clearly, texts don't read themselves and cars don't drive themselves. The Manner-expressions that are so typically found with middle verbs all presuppose an Agent. It is the generic Agent's experience in interacting with the Theme that middle sentences report on.

Decausatives such as (4.26), too, are sometimes considered to have obligatorily unexpressed arguments, in particular Agent arguments.

(4.25) Chief Wiggum **sank** the ship.

(4.26) The ship **sank**.

This is, however, not so clear as these sentences are acceptable when event-causation is involved.

(4.27) The ship **sank** {in the storm/because it was hit by a huge wave}.

We will discuss the possibility of analyzing some of these cases as having the relevant element absent at the conceptual level rather than just at the grammatical level in section (4.2).

4.1.3 Constructionally or lexically licensed argument omission

The other major distinction we need to make concerns the nature of the licensor of the implicit argument. An argument may be left implicit due to the predicate's occurring

in a particular construction or it may be implicit just due to the lexical properties of the predicate itself. This is a very important distinction to keep in mind as the focus of this chapter will later turn on omissions that are licensed specifically by lexical items.

If an omission is constructionally licensed, then all the predicates that can combine with the construction allow the omission. For instance, any verb that has a passive allows omission of the Agent (or whatever frame or semantic role the relevant argument has). By contrast, when a predicate lexically licenses an omission, then that omission is independent of constructional contexts. Of course, in both cases the interaction with other constructions or pragmatic factors may prevent the omission from being felicitous.

A large majority of the omission types arranged in Figure 4.1 involve constructions licensing the non-realization of an argument. Examples for all omission types not yet exemplified are given below.⁷

(4.28) Constructionally licensed omission types

- a. Cisse, off balance, **headed** \emptyset_{Theme} over after a Warnock cross spun to him a couple of yards out. (Sports report)
- b. $\emptyset_{Container}$ **Contains** alcohol (Labelese)

⁷There may be another constructional type having to do with the experiential perfect, as in (i) and (ii).

- (i) Has this lion ever **killed** \emptyset_{Victim} ?
- (ii) Field Goal Kicking Competition. Jamie had never **kicked** \emptyset before, and Jana had about 3 days of "rehearsal" behind her.

It is not clear to me how productive this omission type is. In many cases it sounds preferable to use an overt nominal from the *any-* series for the argument one is considering omitting. The experiential perfect with omission of an argument seems to me to be more readily possible when the activity itself is one carried out as part of a role; (iii) is most readily used, for instance, in a context of construction work where the hearer is asked whether they have been the ones to do the concrete pouring.

- (iii) Have you ever **poured** \emptyset_{Theme} ?

But if this is true, then (iii) may be explained as just due to genericity.

- c. \emptyset_{Agent} **Take** your car to a competent repairer and tell your insurance company immediately. (imperative)
- d. **Boil** \emptyset_{Food} rapidly for 10min (Instructional imperative)
- e. $\emptyset_{Sleeper}$ **Got up** at six and made coffee (Diary style)
- f. "You all look sick," she **scorned** $\emptyset_{Addressee}$ (Reporting clause)
- g. Sue **smokes** $\emptyset_{Substance}$ (Habitual)
- h. Smoking **kills** \emptyset_{Living_being} (Generic)
- i. My bike was **stolen** \emptyset_{Agent} today (Passive)

(4.29) Lexically licensed omission types

- a. In a couple of minutes the robots have collected and **boxed** the goods \emptyset_{Goal} and they are ready for collection or delivery. (Incorporation)
- b. Medtronic has **donated** \$1000 $\emptyset_{Recipient}$ in a match to an employee donation (Definite null instantiation)
- c. A farmer and his wife were lying in bed one evening; she was knitting, and he was **reading** \emptyset_{Text} . (Indefinite null instantiation)

Before moving on to the next factor, I would like to point out that there exist what I will call phantom definite omissions. These arise when a contextual utterance specifies the referent of a frame element that is either not core or core but INI-omissible. Consider the following examples:

(4.30) A: How did you get here?—B: I **walked**.

(4.31) A: How did they get his money?—B: They simply **robbed** him.

(4.32) A: Wow, you fixed the TV!—B: I **whacked** it with a hammer.

The verbs in the B utterances in (4.30-4.32) superficially appear to allow definite null instantiation of Goal, Theme, and Purpose arguments respectively. However, it is clearly not the case that whenever no Goal is expressed overtly for *walk* in a

particular (i.e. non-generic) sentence, it has to be retrievable from context. An easy counterexample is given in (4.33), where the contextual sentence specifies a Landmark along the path rather than the final Goal of movement.

(4.33) A: How did you get past the guard?—B: I **walked** on tip-toes.

Further, for the phantom DNI-reading to be possible, the referent cannot be the explicit discourse topic as in (4.34) but simply has to be part of what is presupposed as in (4.35).

(4.34) A: Where's my soup?—#B: Bill **ate**.

(4.35) A: Why did it take him so long to eat the soup?—B: He **ate** only with a fork.

In example (4.35), of course, B's utterance is an indirect answer describing how the meal was eaten rather than how specifically the soup was eaten. Since the soup was part of the meal and A's question specifically mentions the soup, it may appear that B's answer actually was "He ate the soup with a fork". But I would maintain that that is just an inference.

The fact that the hearer may fill in the referents of the Goal, Theme, Purpose, Landmark, Food/Meal frame elements in (4.30) to (4.35) from their contextual specification is not a lexical fact and there is no sense in treating the verbs as licensing lexical null instantiations of the relative frame elements.

4.1.4 Change in linking

The factor linking creates another significant subdivision between the various types of (constructionally licensed) argument omission. While with most omission types, the other arguments have the same grammatical function as when the omitted argument is overt, in some cases the linking of the overtly realized arguments changes. The passive and middle voice constructions are the clearest examples of this. If one considers decausatives to be a constructional phenomenon, then they, too, need to be included in this group.

4.1.5 Null instantiation

Given the above typology we can now characterize null instantiation as the lexically licensed and optional omission of an argument that is not accompanied by a change in linking to grammatical functions for the overt arguments. Null instantiation is further different from incorporation in that the predicate is not morphologically related to a noun that denotes a prototypical filler of the incorporated argument. Thus, *knit* and *sew* are predicates that null instantiate whereas *box* and *crate* are verbs that incorporate.

4.2 The interpretation omitted arguments

Even a casual inspection of the verbs and constructions that allow argument omission shows that not all cases of omission work the same way. There exist differences in interpretation between the various types of omissions and they are not neatly aligned with parts of the type network.

4.2.1 Constructionally licensed omissions with no change in linking

We will begin by considering constructionally licensed omissions, and specifically the types that involve no change in linking.

Generic and habitual sentences express non-incidental facts. They are concerned not with actual events but with structural events, that is, they impart information about how the world is structured (Langacker 1991). According to Michaelis (2004), citing Krifka et al. (1995), the differences between habitual sentences and gnomic sentences can be traced to characteristic properties of nominal reference. Nominals in gnomic sentences such as (4.36) have attributive reference, leading to contingency readings. Habitual sentences like (4.37) do not have contingency readings, since such sentences attribute properties to specific entities.⁸

⁸Cross-linguistically generic sentences never seem to be marked by dedicated morpho-syntax but they appear to be recognized as such only on the basis of a mismatch between perfective verbal

(4.36) Usually bears **attack** \emptyset when they are surprised or are protecting their young.

(4.37) I have never **smoked** \emptyset in my entire life.

Generic and habitual sentences such as (4.37) are the only kind of constructionally licensed argument omissions involving no change in linking where the referent is not only not specific but possibly non-referential.

Like the other constructionally licensed omission types that involve no linking change, generic or habitual construal allows for the omission of many arguments that cannot be omitted just based on the identity of the lexical unit. An example is the unrealized Theme of *take* in (4.38).

(4.38) People who always **take** \emptyset and never give must be ejected from your life.

Still, even generic construal may have its limits as a licenser of argument omission. It seems unacceptable to not express Contents in (4.39) when using the verb *contain* with the intended interpretation *This box contains something*.

(4.39) Can I use this box?—I am afraid not. #It's not empty, it **contains** \emptyset

With the exception of generics/habituals and the subjects of imperatives, all the constructions licensing argument omission with no change in linking have affiliations with particular genres. The referents of their implicit arguments seem to be definite.

- the agent of an imperative is an addressee in the speech context (cf. (4.28c))
- the omitted subject in narrative/diary style is often the author/speaker, but may be another topical referent that has been textually evoked (Haegeman 1990) (cf. (4.28e))
- the Evaluatee of a judgment verb can only be omitted when she is the addressee of the speaker/judge (cf. (4.28f))

Aktionsart and the syntactic context in which that verb appears (Michaelis 2004, citing Dahl 1995).

- the arguments omitted in labelese refer to the container or object that the label is affixed to (cf. (4.28b))
- the object of an instructional imperative is an entity that the reader is manipulating as he reads the instructions (cf. (4.28d); Culy 1996, Bender 1999)
- in reports of sports events, objects that are used in play, especially balls, are frequently omitted (cf. (4.28a))

With all of these constructions it is in principle possible to use an overt definite NP or a pronoun, although, in many cases using an overt form may be unusual, as, for instance, in the case of the subjects of imperatives.

The omissions occurring in the case of diary style, labelese, instructional imperatives, and sports reports differ from the other kinds in that they do not seem to be based on strictly local considerations of topicality. For zero subjects in diary style, Haegeman (1990) suggests that the notion discourse topic is relevant.

An example in support of that analysis is found in (4.40), where a series of clauses with zero subject referring to the author Virginia Woolf is interrupted by reference to *L.*, the author's husband Leonard.

- (4.40) ... \emptyset **read** Michelet; \emptyset **wrote** to Desmond about his poetess; *L.* **out** at Fabians; \emptyset **played** gramophone; ... (adapted from Haegeman 1990:158, (1a))

The notion of discourse topic may also be relevant to labelese and instructional imperatives. Similar to example (4.40) from diary style, we find that in recipes, overt non-co-referents intervening between the last mention of a zero referent and a co-referential zero mention do not prevent the correct interpretation of the zero referent.

- (4.41) Sweet Lassi₁ - Ingredients: 1 Serving Plain yogurt- 1 cup Sugar- 2 table-
spoons Ice Cubes- 4 Method: Blend all the ingredients in an electric blender.
Serve \emptyset_1 cold.

In (4.41), the final product is maybe co-referential with the name of the recipe. If we don't want to make that assumption, then we would have to say that the zero refers to what was produced by the blending process, even though the product was never mentioned, which would be a case of inference rather than co-reference.

The interpretation of omitted arguments in labelese contexts may also change between clauses without an intervening overt mention of the new zero referent. In (4.42), it is clear that the referent changes: the argument omitted from *comb* is the user's hair but the next omission with *apply* is the dab of balm.

- (4.42) After **shampooing** \emptyset , squeeze a dab of balm into the palm, **comb** \emptyset and **apply** \emptyset evenly throughout wet hair starting at the roots down to the ends

A use of simple personal pronouns in all the places where an argument is omitted in (4.42) would not be acceptable as the expected co-reference relations do not hold.

- (4.43) After **shampooing** it, squeeze a dab of balm into the palm, **comb** it and **apply** it evenly throughout wet hair starting at the roots down to the ends.

Sports reports seem still different from both diary style and recipes. Consider example (4.44), which occurs at a point in the text where the omitted referent, the ball, has not been mentioned overtly even a single time despite the fact that a lot of actions such as shots on goal or throw-ins involving the ball have been reported.

- (4.44) Hughes, however, twice almost broke through as Bradford flirted dangerously with the offside trap, although the £5million man's blushes were spared by the linesman's flag on a third occasion when he **hammered** \emptyset wide of Gary Walsh's exposed net.

It is a notable characteristic of the argument omissions in sports reports that they are heavily focused on particular entities. I have not actually found a case where in a reports of a soccer or tennis match a referent other than the ball is omitted.⁹

⁹In soccer or basketball reports, the goal/net referents are also sometimes not expressed overtly, as in (4.45).

For instance, it seems that regardless of any context one might supply, the first clause of (4.47) cannot be used in a case where a player smashes their racquet into the net out of frustration, which should be plausible given that each player uses just one racquet and given that racquets are of central importance in the game.

- (4.47) But she **smashed** \emptyset **into the net** with the court gaping to set up another break of serve for Williams in the next game, who took advantage ruthlessly to hold and break again for victory.

We will see later that with lexically licensed omissions with definite interpretation the referent is not really restricted: the omitted referent can be anything that meets the predicate's selectional restrictions and is accessible in the discourse.

The focus on a single entity does not constitute a difference between sports reports and diary style, labelise, or instructional imperatives. Still, the omission of overt reference to balls in sports reports seems to be motivated differently from that found in diaries, instructional imperatives, or recipes: the actual ball-referents are unimportant in terms of the overall narrative, the report is not *about* the ball(s) per se.¹⁰ What allows the omissions in sports reports is simply the fact that the ball-in-play at any given time is a unique and salient entity in the scenario. In soccer and even more so in tennis, it is likely that different balls get used during a game, but they are not kept track of individually. Reference is made only to the ball-in-play. It is also worth noting that in the context of sports reports and instructional imperatives the omission of the

-
- (4.45) The power play put Brown on the board early in the second. Paul Esdale, an offensive-minded defenseman with great stick skills, fired a shot from the point which Les Haggett **tipped in** \emptyset_{Goal} to make it 2-1.

However, these omissions are licensed by the general lexical mechanisms for omitting prepositional objects. The match report contexts do not, for instance, override argument omission with prepositions such as *into* or *onto* that do not lexically specify it.

- (4.46) In a similar play on the opposite side of the field, Chin sent a low cross into the box that Dorman **fired into** *(the goal) at 30:38.

¹⁰This is also suggested by the combined facts that the ball referents are omitted only in object rather than subject position and that as inanimate entities, they would also be unlikely topics in a context containing human beings. Of course, the same is true of recipe contexts. But there we would want to say that the text is about the dish.

central referent(s) is more likely with some predicates than with others. For instance, in a sports context, while the ball referent may be unexpressed as an argument of a verb that expresses causing it to move, it is not readily omitted as an argument of a predicate expressing a non-canonical event taking place during the game. In (4.49), for instance, the NP *the ball* cannot be omitted since handling the ball is not a legal move for field players.¹¹

- (4.49) On 23 minutes Morecambe were awarded a penalty when Gavin Peers was adjudged to have **handled** *(the ball) in the area.

This restriction is similar to what is found with recipes (Culy 1996) but different, for instance, from diary style where there is no restriction on the predicates with which a zero subject may occur.

Thus, the genre-oriented constructions all seem to be subtly different. While all allow omissions that heavily fall on particular entities that are of global importance in the discourse, the status of the entity is not the same in all cases: the ball-in-play does not seem to be a plausible discourse-topic in sports reports. Moreover, in labelese, recipes, and sports reports, omissions seem to require that the referent be mentioned as part some kind of canonical action. There is no similar requirement for diary style. Recipes and sports reports allow successive zero reference even though the actual object may be not be quite identical to the one that was omitted before. In sports reports, balls are not kept track of individually; in recipes, the omitted argument tracks the main object being manipulated regardless of the current ontological state, say, as a mixture of ingredients, a batter, a baked cake, or slices of cake.

¹¹In this connection notice that, as pointed out to me by Charles Fillmore, a requirement for canonical actions also seems to hold with respect to the possibility of using a simple present tense in a sports broadcast.

- (4.48) Now that the middle innings are over with, the managers are bringing in their heavy artillery from the bullpen. Right-hander Mike Timlin relieves Arroyo and sets the Cardinals down in order. Millar has had a rough night in the field but he makes a nice pick on an errant Timlin throw for the third out. Kelly Clarkson comes in and finishes up her God Bless America rendition quicker than you can say, "Irish Tenor." #A fan **jumps** over the outfield wall, **strips**, and **runs** across the field naked.

The various constructions also differ with regard to their compatibility with other semantic and syntactic phenomena. For instance, omitted objects of instructional imperatives and labelese can be predicated over by depictives and resultatives and they can also bind reflexives, as is shown by example (4.50).

- (4.50) Chill dough, then **roll** \emptyset_{Theme} to 1/4"-thick and **spread** \emptyset_{Goal} with date filling and **turn** \emptyset_{Theme} over on itself, making a jelly-roll.

As pointed out in the case of the generics and habituais, these examples clearly violate Fillmore's (1986) observation for lexically licensed omission that there usually is no definite omission of Patient/Theme arguments. And in fact, they also violate the, in my judgment, even more robust, generalization that the Theme of a resultative construction is not omissible (e.g. **And now you have to **cook** \emptyset tender*).

The zero objects of instructional imperatives also are able to bind not only purpose clauses oriented towards the Agent but also ones that are oriented towards them, the Patient-object. Compare (4.51) to (4.52).

- (4.51) Powdered Sponge. Just **sprinkle** \emptyset on to absorb even the worst kinds of messes. Then sweep it up and throw it away.

- (4.52) I **bought** it to give me peace after stressful days.

Sentences with omitted subjects in diary style seem to behave exactly like sentences with an overt subject. The omitted subjects can occur with Instruments and Means expressions as well as Agent-oriented Manner expression. They also can control purpose clauses, bind reflexives, and have depictives predicate over them, as illustrated in (4.53).

- (4.53) $\emptyset_{Self-mover}$ **Went** to the store without a hat on to buy myself groceries and caught cold.

Reporting clauses, match reports, and habituais and generics, which allow omission of objects, share the following properties: they only allow for purpose clauses

oriented towards the Agent but not benefit clauses oriented towards the Object. Depictives cannot predicate over the zero objects, only over the overt subjects. The zero objects also cannot bind reflexives, which the overt subjects, of course, are able to do.

(4.54) #Sue likes to **eat** \emptyset raw

(4.55) # He **throws** \emptyset to bounce once and then strike the wicket.

(4.56) # “Oh, you’ve done this sort of thing before,” she **reproached** \emptyset about myself.

Finally, the omitted arguments differ with respect to referent accessibility. While imperatives, diary style, recipe imperatives, labelise, and reporting clauses involve discourse-active given referents, match reports involve inactive but unique referents. Habituals and generics involve quantification and the omitted arguments should be treated as variables that may or may not be bound in the relevant component contexts that go into the generalization; on the ‘top’ level, however, there are no specific referents involved. Decausatives, to the degree that one wants to include Agents in their conceptualization at all, involve inaccessible Agent referents.

4.2.2 Valence changing constructions and null instantiation

We now turn to argument omission with valence changing constructions such as middles, causatives, and decausatives. Intuitively, one might like to say that whereas the omission types we have looked at so far have the same conceptual background as fully transitive uses, valence changing constructions are accompanied by changes in the conceptual semantics and specifically involve different sets of conceptually necessary participants as well as different sets of sub-events in their event structure.¹²

However, this suggested distinction is not as clear as it may seem. In the case of causative-inchoative pairs, one might also take the view that inchoatives just take

¹²Causativization typically correlates with the adding of an argument but not always. Exceptions are reported by Ackerman & Moore 1994 and by Dixon 2000.

a particular perspective on the the event such that they background any agentive or causing forces in the scenario but do not actually exclude them. Consider the sentence pairs below in (4.57) and (4.58) involving the verb *break*. According to Talmy (2000:69-70) the (a) sentence in (4.57) exemplifies Causing-Event Causation and that in (4.58) Agent (voluntary) or Author (involuntary) causation, whereas both of the (b) sentences exemplify Resulting-Event Causation.

- (4.57) a. The impact of the ball **broke** the vase.
 b. The vase **broke** from the impact of the ball.
- (4.58) a. I **broke** the vase
 b. The vase **broke** because of me

Given that the same thematic roles seem to be realized in the subjects of the (a) and the oblique phrases of the (b) sentences one might choose to treat these pairs of sentences as having the same conceptual background.

Passives clearly do not involve the elimination of Agents from the conceptual background. Even in so-called short passives that have no oblique expression of an Agent in a *by*-phrase, the Agent can at least semantically control the subjects of purpose clauses. In addition, short passive sentences welcome the expression of peripheral semantic roles such as Means and Instrument, as shown in (4.60-4.61), which inherently require that an Agent be part of the scenario, as well as of Agent-oriented Manner expressions such as *deliberately*, as shown in (4.62).

- (4.59) New Harbour was **built** \emptyset_{Agent} in order to cope with the trading activities.
- (4.60) The base was **assembled** \emptyset_{Agent} by attaching the two side arms (the arms that hold the scope) to the circular base.
- (4.61) Previously, every door within each of the four sites was **opened** \emptyset_{Agent} with its own individual key.
- (4.62) However, they did not understand what he meant because the information was deliberately **hidden** \emptyset_{Agent} from them.

Different from passives and inchoatives, the middle construction, is a valence changing construction that prevents the overt realization of an element that is conceptually necessary for the non-middle uses.

- (4.63) Like other hard cheeses, it **cooks** easily (*for/to you) and will transform simple dishes.

Despite the impossibility of expressing an Agent, I would maintain that the notion of an Agent is still conceptually necessary: middles report on the generic Agent's experience in interacting with the Theme. The notion of ease or difficulty that is relevant in (4.63) refers to the experience of an Agent. Moreover, as shown in (4.64) and (4.65), respectively, middle sentences allow the expression of Instrument and Means frame elements, two frame elements which presuppose a volitional Agent.

- (4.64) Equipment boxes **mount** easily with external flanges

- (4.65) It **mounts** by sliding in the mounting bars into an empty P4 socket mount, placing the unit down on the chip and tightening the screws.

Middles do, however, not seem to allow the use of Agent-oriented Manner adverbs such as *deliberately*.

- (4.66) #The booklet deliberately **sells** for very little money.

We should note, too, that the behavior of the English middle and passive constructions is not necessarily representative of all similar constructions in other languages. Koenig et al. (2002) mention, for instance, that the Spanish passive can be followed by purpose clauses even though it never allows the expression of an Agent.

- (4.67) El florero se **vendió** rápidamente (*por la organización) para recaudar fondos para los niños.

The vase was sold quickly (*by the organization) to gather money for the children.

An issue that is very relevant to how implicit arguments are to be represented is the discourse status of implicit referents. Koenig & Mauner (1999) argue that implicit arguments are *a-definites*, which they say represent a kind of NP-type that is distinct both from definites and indefinites. In an analysis that draws largely on Discourse Representation Theory, they argue that a-definites fill an argument position of a predicate but do not introduce a discourse marker for that position.¹³ As a result of that, the referent of an implicit argument cannot serve as the antecedent of a pronoun in a subsequent clause. As an example, Koenig & Mauner use the short passive:

- (4.68) #The first female president was **murdered** e_j . He $_j$ was from the Berry, it seems.

Importantly, implicit arguments are not the only kind of a-definites. Koenig & Mauner (1999) also analyze French *on* and German *man* as a-definites. Consider here a pair of French examples contrasting the a-definite *on* with the indefinite *quelqu'un*, both of which are translated as *someone* in English.

- (4.69) Somebody assassinated the president. He was from the Berry, it seems.
- a. On a **assassiné** la présidente. #Il était du Berry, paraît-il
- b. Quelqu'un a **assassiné** la présidente. Il était du Berry, paraît-il (= (8-9) Koenig & Mauner 1999: 213)

The basis of Koenig & Mauner's (1999) suggestion that implicit arguments introduce no discourse marker and are therefore pragmatically inert is data on cross-clausal anaphora such as (4.69). Intra-clausal phenomena such as the control of purpose-clauses for passives and the ability of the a-definite *on* to antecede a reflexive, as shown in (4.70), are handled as lexical phenomena.

- (4.70) On s'est encore **lavé** dans ma salle de bain.
'Somebody washed himself/herself in my bathroom again.' (= (18) Koenig & Mauner 1999: 217)

¹³Along with introducing a restriction on the referent of the discourse marker they introduce, these are the main functions of non-quantificational NPs or pronominals within DRT.

Thus, implicit arguments are a-definites that can enter into a grammatically induced relation of coreference, provided this relation is not mediated by discourse marker equality clauses.

Subsequent pronominal reference with indefinite *they* (Condoravdi 1989), which seems to contradict the claim of pragmatic inertness of a-definites, is treated as inference rather than co-reference. Consider the sequence (4.71a-b).

- (4.71) a. The rebel priest was **tortured** for three days
 b. They wanted him to reveal where the insurgents were hiding out.

The inference involved in this example is specifically said to be of a type that Koenig & Mauner (1999) call factoring, inspired by Hobbs et al. 1993. Factoring inferences are ones that are not grammatically induced but based simply on the desire for a coherent discourse involving a minimal number of distinct entities and events. The factoring inferences occurring with indefinite *they* are said to be different from the ones that occur when an implicit argument is targeted by a following definite NP. The latter cases are treated just as instances of bridging inferences (Clark & Haviland 1977) such as the one shown in (4.72).

- (4.72) I bought a new car last week. The engine is already giving me trouble.

According to Koenig & Mauner (1999), the referents of definite pronouns can sometimes also be accommodated in discourse through bridging inferences. They give examples (4.73) and (4.74).

- (4.73) John became a guitarist because he thought it was a beautiful instrument

- (4.74) I saw Mr Smith the other day; you know SHE died last year.

Koenig & Mauner (1999) suggest that the accommodation of definite pronouns is a rarer event than that of definite NPs because of the pronouns' low descriptive content.

It is not clear that (4.73) and (4.74) require the same kind of inference, and how well motivated the posited difference between factoring and bridging inference really

is. With both (4.73) and (4.71b), the hearer first has to process the remainder of the clause to have any idea what the referent might be. The understanding that is gained is simply that the referent participates in a certain kind of event and must have appropriate ontological properties. Then, given that the pronominal form suggests high accessibility of this referent, the hearer needs to re-examine prior context for a plausible event or state in which the referent could feature. The only difference between (4.71b) and (4.73) is that the former sentence is non-generic and the referent is existentially bound while the latter sentence is generic and has a non-specific (type) referent.¹⁴

(4.74) is different from the previous two cases only in that the prior context contains no clear frame trigger: *Mr* is used with men of any marital status. In the absence of a contextually available frame, contrastive stress is used to activate a him-her pair that gets resolved with cultural knowledge to a marriage relation. If the first clause explicitly mentioned that Mr Smith was a widower then pronominal reference to his wife in the second clause would be possible without the need for stress. In (4.78), it would actually be odd to contrastively stress the pronoun *she*.

(4.78) If you run into my neighbor, Mr. Smith, be very nice to him. He was recently **widowed**; she died just a month ago.

Koenig & Mauner (1999) also predict that subsequent inference to an implicit

¹⁴One might actually be justified in treating the so-called indefinite *they* as an a-definite in Koenig & Mauner's (1999) terms, which the *it* in (4.73) is not. *They* does, for instance, not allow cross-sentential anaphora when that anaphora involves places the 'referent' in another frame (4.75a-b). It also seems that indefinite *they* cannot receive stress unlike *someone* (4.76a-b). Finally, it seems odd to claim that one knows the exact identity of *they*, which is not the case for *someone*.

- (4.75) a. Someone's building a new subdivision by the lake. God, how I hate him for that.
 b. ??They're building a new subdivision by the lake. God, how I hate them for that.
- (4.76) a. (Looking at room-mate) SOMEONE ate my ice-cream again
 b. (Looking at room-mate) *THEY ate my ice-cream again.
- (4.77) a. Someone mugged my brother when he was in town and I know exactly who {she/it was/they were}.
 b. ??They mugged my brother when he was in town and I know exactly who {that/it was/they were/}.

Agent should be more difficult with middles such as (4.79) than with short passives like (4.80) because the former's lexical representation is said not to encode an Agent.¹⁵

(4.79) The antique vase had **sold** to a wealthy woman. The owner/He charged a lot of money for it.

(4.80) The antique vase was **sold** to a wealthy woman. The owner/He charged a lot of money for it.

According to Koenig & Mauner (1999), both the pronominal and the continuation with a definite NP should be easier to process for the passive than either kind of continuation for the middle. I think that this claim might not stand up to empirical verification: it seems that at least the continuation of the middle with a definite NP should be easier to process than the continuation of the passive with a pronoun. Whether and how large a difference there might be between the sentences when both are continued with the same NP type is, however, an interesting question.

Another point deserves comment about the pair of examples (4.79) and (4.80). The example of a middle that Koenig & Mauner (1999) give in (4.79) is actually a rather marginal example of a middle, if it is not entirely lexicalized. It is not generic, it has a specific token as subject rather than a type, and there is no manner adverb present. Most verbs cannot be used in this way in a middle construction:

(4.81) #This loaf of bread is **cutting** in half easily.

(4.82) # The new Chrysler **drives** on bumpy roads.

With typical examples of middles, inference to implicit Agents either with indefinite *they* or a definite NP should be much more difficult than in (4.79). It seems to me that the best continuation for example (4.83) is (4.84e), which has an indefinite plural NP as subject.

¹⁵Note that while Koenig & Mauner (1999) and Mauner & Koenig (2000) discuss passive and middle verbs as lexical items, their analysis could probably be recast so that passive and middle are constructions that specify a status as syntactically active or inactive for an Agent argument, respectively. As such they might be linking constructions or morphological constructions that derive lexical entries.

- (4.83) It **drives** comfortably like a sedan, tows like a tough pickup, off-roads like a rugged sport/ute, and moves bulky cargo or as many as seven people like a minivan.
- (4.84) a. #They will have lots of fun with it.
 b. #The driver will have lots of fun with it.
 c. #The drivers will have lots of fun with it.
 d. ?#A driver will have lots of fun with it.
 e. Drivers will have lots of fun with it.

So far, we have seen that there are indeed differences between middles, passives, and decausatives with respect to the prominence of the Agent and its ability to be accessed again in subsequent discourse. However, it is not clear that the observed differences need to be represented by either having the Agent be present or absent in the semantic (as opposed to a conceptual) representation. Even canonical middles such as *The text reads well*, which don't allow purpose clauses and are not readily followed by clauses with pronominal arguments whose referents need to be bound to the implicit Agent of the middle, may well have an Agent as part of their semantic representation, given that the co-occurring Manner adverbs all require that an Agent be part of the scenario. The Agents of middles may be just as discourse inactive as those of short passives and what makes them more likely to be discourse inactive may be the genericity of the construction. Maurer & Koenig (2000) have indeed adopted a representation that explicitly divorces the presence of the Agent argument at the semantic level from its syntactic activeness. They include a list of syntactically active arguments in the lexical representation of verbs: the presence or absence of the Agent on that list captures the difference between passives and middles.¹⁶ Maurer & Koenig (2000) also reports on additional psycholinguistic experimentation that takes

¹⁶This may or may not represent a change in the analysis from the 1999 paper. There the difference between passives and middles was attributed to a difference in argument structure. Argument structure is distinguished from a conceptual level but it is not made clear whether a specifically linguistic-semantic level is assumed. It seems to be, as the authors refer the reader to the Maurer & Koenig 2000 paper for discussion of experimental evidence.

into account the time course of processing of middles and passives rather than just the overall difficulty measured by reaction times as in the research reported by Koenig & Mauner (1999). The results of Mauner & Koenig 2000 confirm the earlier finding that passives followed by purpose clauses are processed faster than middles. They also suggest that speakers access verb argument structure immediately when they encounter the relevant verb and that the encoding of implicit arguments in passives and middles is not reducible to generic conceptual knowledge representations that may be accessed only once the sentence or clause as a whole has been taken in.

There is thus distributional and processing evidence to believe that implicit arguments may be present in the lexical semantic representations of passives and middles and are not inferred directly from generic conceptual knowledge when they need to be activated by later discourse. Differences between the constructions in terms of the ease with which implicit arguments can be identified with intra- and cross-clausal arguments are attributed specifically to a list of syntactically active arguments that is part of the lexical representation.

While Koenig & Mauner 1999 and Mauner & Koenig 2000 suggest that at the conceptual and linguistic-semantic level the representation of the events denoted by middles and passives are the same, Härtl (2003) argues that at least in the case of German particle verbs which obligatorily omit mention of their reference object and which alternate with base verbs occurring with full prepositional phrases, the reference object (that is, the prepositional landmark) is absent already in their conceptual representation and that, when that object needs to be activated in subsequent discourse, conceptual rather than linguistic knowledge is involved.

Härtl notes that full PPs and particles cannot be conjoined; (4.85) has no acceptable reading where the landmark of *auf* 'on_{horizontal}' is the same as that of *an* 'on_{vertical}'.

- (4.85) *Er **klebte** Etiketten **auf** und an den Projektor
 'He stuck labels on \emptyset and on the projector' (= (5b), Härtl 2003, p.7)

Härtl also points out that particle verbs cannot introduce a referent, as shown in

(4.86).¹⁷ This is true at least for compositionally formed particle verbs. Verbs with idiosyncratic meanings such as *auflegen* in (4.87) may allow later reference to a frame-inferable object.

(4.86) Maria wollte ein Preisschild **aufkleben**. #Aber die Flasche war zu glatt.
‘Maria wanted to stick on a price-tag. But the bottle was too smooth’ (= (6a+a’), Härtl 2003, p.7)

(4.87) Maria wollte eine Schallplatte **auflegen**. Aber das Gerät war defekt.
Maria wanted to put on a record. But the device was broken. (= (6b+b’), Härtl 2003, p.7)

Further, Härtl says that no co-referential relationship can be established between a pronominal expression and a potential implicit reference object with particle verbs (cf. (4.88)). That is said to be possible only in the case of predicates that are equipped with restricted sortal requirements on the entity left implicit in the grammar (cf. (4.89)).

(4.88) Hans **wirft** die Briefe **ein**. *Der wird schon gelb sein. [the mail box]
Hans puts the letters in. It will surely be yellow. (= (7a), Härtl 2003, p.8)

(4.89) Sheila **heiratet**. Der wird schon ein Idiot sein.
Sheila is getting married. He will surely be an idiot. (= (7b), Härtl 2003, p.8)

Härtl also argues that in German weak bridging pronouns, which are homophonous with the definite article and always receive their interpretation via implicatures drawn from domain-specific world knowledge and which are not subject to the conditions of grammatical feature agreement— they occur in neuter only—are preferred with constructions where no nominal expression is explicitly introduced in the discourse.

(4.90) Sheila **heiratet**. Das wird schon ein Idiot sein.
Sheila is getting married. That’ll be some idiot.

¹⁷The sentence would be good if the bottle had been mentioned before.

- (4.91) Peter **liest**. Das wird schon ein Unsinn sein.
Peter is reading. That'll be some nonsense.

He notes that in the case of particle verb combinations, the interpretation of a weak pronoun referring to a potential reference object is blocked.

- (4.92) Hans **wirft** die Briefe **ein**. *Das wird schon gelb sein. [the mail box]
Hans puts the letters in. That will surely be yellow. (= (9a), Härtl 2003, p.8)

- (4.93) Maria will eine Schallplatte **auflegen**. *Das wird schon nicht defekt sein.
'Maria wants to put on a record. That surely won't be broken.'

Despite the inability to introduce referents, according to Härtl, any kind of particle verb, whether compositional or idiomatic, seems to have "the information structural potential to denote an adequate background set containing the explicit reference object of a full prepositional phrase". Consider examples (4.94) and (4.95).

- (4.94) Da ist ein Briefkasten. **Wirf** doch den Brief **ein**.
There's a mailbox. Why don't you put the letter in?

- (4.95) A: Er klebte ein Etikett auf den Computer. B: Nein er **klebte** ein Namensschild **auf**.
A: He put a tag onto the computer. B: No, he put a name label on.

Härtl also gives sentence (4.96) as evidence that sometimes one can get at the semantic/conceptual entity presupposed by particle verbs.

- (4.96) Michael **lehnt** das Fahrrad **an**, aber ich sage Dir nicht woran.

Härtl concludes that the zero reference objects (landmarks) of particle verbs behave like a-definites in the sense of Koenig & Mauner 1999: only under specific contextual circumstances can the landmark object be activated in subsequent discourse. Härtl further notes that the reference objects of particle verbs are always implicit, in

distinction to, for instance, the agents of passives. This last consideration suggests to him that it is unlikely that “the language system provides an argument slot which can never be realized” (p. 10).

Based on these observations, Härtl sets out to test whether the cases where conceptual activation of reference object entities occurs in later discourse are part of canonical processing or whether it is the result of some special procedure/event, which he calls re-analysis, “which is achieved by means of meaning postulates anchored in the conceptual knowledge base and recruited to manage special contextual requirements” (p. 10).

For Härtl, if the reference object is not activated by default, one has to conclude that a potential reference object has to be inferred non-compositionally on the basis of extra-linguistic domain knowledge as soon as the context requires it. And importantly, “the truth conditions underlying particle verbs themselves would not require the existence of a corresponding entity, which *per se* would not be part of the conventional interpretation of particle expressions.”

Härtl conducts two experiments involving object categorization tasks with line drawings. Two sets of sentences were created, one with particle verbs, another with verbs taking full preposition phrase complements. A sample pair is given in (4.97).

- (4.97) a. Henry druckt die Adresse auf nachdem (*) der Computer repariert worden war.
Henry prints the address on after (*) the computer had been fixed.
- b. Henry druckt die Adresse auf das(*) Papier und versorgt die Kollegen.
Henry prints the address on the (*) paper and caters the colleagues
(= Härtl 2001: 16, (15a-b))

As shown in (4.97), the pairs of sentences started out the same up until and through the word that is ambiguous between being a particle or a preposition. The next word was then chosen to disambiguate clearly between the two, or at least to provide a strong bias. (The asterisks in (4.97a-b) follow the disambiguating word.) Also, after the disambiguating word, a line drawing was presented on a computer screen.

The picture either matched or did not match the/a typical object relevant to the scene in terms of sortal requirements. In order to ensure conceptual access to the extra-linguistic representation of the objects denoted by the drawing, participants were asked to categorize the corresponding object with regard to its size, i.e. whether the object fits on a sheet of A4 paper in the real world. In addition to the primary categorization task, subjects also had to perform a secondary task regarding the content of the stimulus sentences but no details of this secondary task are given.

The same basic experiment was carried out twice because of a potential problem posed by the agreement morphology on the article that serves to cue the PP construction. Accessing the conceptual representation of the drawing might be facilitated simply by the agreement between the grammatical category of the article and the ontological category of the drawing linguistically denoted by a noun as well. Härtl says “the full prepositional phrase is accompanied with a linguistic form—an article indicating a noun phrase—which parallels the linguistic features of the picture expressions realized as a non as well”. This seems to mean that the article cues “thinghood” explicitly.

To weaken the facilitating effect of the article, the experimental design is carried out in such a way that the prepositional object is not the canonical object. “The idea is that now with regard to the grammatical integration of the picture representation into the running sentence, processing of is nearly equated for both the particle and preposition condition” (p. 13).

However the non-canonical object and article may now give rise to the concern that the non-canonicity of the object—cued by the gender mismatch—interferes and leads to slower responses. To control for this, the experiment is simply re-run (with different subjects) such that the gender of the expression in the sentence and the gender of the canonical object agree. Note that the canonical object itself is not mentioned; only its gender is matched.

In the first experiment with the non-canonical article, only full prepositions are found to prime a potential reference object. The difference in congruency between picture and sortal requirements of the verb played no role in the case of particles. Härtl concludes that with prepositions but not with particles participants seem to have been

aware of the procedural necessity to conceptually activate an entity that matches the sortal and ontological requirements of a potential reference object. In particular, the effect of the congruent drawing seems to cause inhibition and slower reaction times on the object size categorization task. Härtl considers the longer reaction times a reflex of processing difficulties stemming from competition for conceptual activation between two conceptual entities—one activated by the drawing, the other activated by the prepositional complement.

In the second experiment, an article is used in the prepositional condition which matches the article of the canonical object for the verb. A sample pair is given in (4.98). Note that the word in square brackets refers to the picture shown on the screen.

- (4.98) a. Maria klebt das Etikett auf nachdem [FLASCHE] der Winzer den Wein verkostete.
 Maria sticks the label on after [BOTTLE] the wine-grower the wine tasted.
- b. Maria klebt das Etikett auf die [FLASCHE] Ware und betrachtet das Bild.
 Maria sticks the label on the_{fem.} [BOTTLE_{fem.}] product and watches the picture.
 (= Härtl 2001:21, (16a-b))

Though the results were not as clear as in the first experiment, the data seemed to confirm that incongruous objects were categorized faster than congruous objects in the context of a full preposition only. No reliable difference in categorizing congruous v. incongruous objects was found with particles.

Härtl's interpretation of the experiments is that no entity corresponding to a landmark of a preposition is activated conceptually in the actual course of processing of particle construction. Therefore, for particles, in contrast to prepositions, no effect occurs as a result of the presentation of a pictorial representation. He also concludes that no grammatical information is available while processing conceptual features.

Which fact is taken to explain why a gender mismatch between the article (determining the prepositional reference object) and the canonical picture name does not influence conceptual processing.

While Härtl's study and design is very interesting I think there are some problems with it where particle verbs are concerned. One problem has to do with the assumption that there are compositional, non-idiosyncratic particle verbs. My own native speaker intuition tells me that in fact many of the particle verbs that Härtl discusses are lexicalized. For instance, *einwerfen* with an object referring to a piece of mail functions much like English 'post [a piece of mail]'. Other objects used with the same verb do not behave the same way:

(4.99) Kannst Du mir einen Euro leihen?—Leider nicht. #Ich habe meinen letzten Euro gerade **eingeworfen**.

Can you lend me a euro?—Unfortunately not. I put in my last one.

(4.99) is bad because *einwerfen* when used with money is not lexicalized to mean something like 'put money in some vending machine'. (4.99) is only felicitous as an definite null instantiation, as when both speaker and hearer are standing in front of the vending machine. Thus, the contrast that Härtl intends between compositional particle verbs and idiosyncratic ones does not seem quite convincing.

Another problem affects 15 of the 20 experimental sentences that Härtl used for particle verbs: the understanding of the landmark is not an existential one as in the case of a-definites but a definite one that needs to be resolved in context. Some examples are given below, with verb and particle underlined.

(4.100) Andrea strickt das Teil an nachdem die Tante das Geld holte.

'Andrea {knits on/is knitting on} the piece after the aunt had fetched the money'

(4.101) Gerda koppelt den Waggon an nachdem die Lieferung verfrachtet worden war.

'Gerda {attaches/is attaching} (lit. coupled on) the waggon after the order had been loaded'

- (4.102) Chris **leimt** den Henkel **an** nachdem der Nachbar das Geschirr besichtigte.
 ‘Chris {glues/is gluing} on the handle after the neighbor inspected the dishes.’
- (4.103) Henry **druckt** die Adresse **auf** nachdem der Computer repariert worden war.
 ‘Henry {prints on/is printing on} the address after the computer had been repaired’.

Sentences (4.100-4.102) do not have landmarks that need to be merely existentially bound. They need to be given in prior discourse or be physically present in the speech context. They are unlike the much discussed *Brief einwerfen* ‘post letter’, which has an idiosyncratic meaning not requiring an actual container-goal. Thus, it seems to me that Härtl’s experiment is not really evaluating what he meant to evaluate.

It is not clear to me what led Härtl used to identify these particle-verbs as having implicit landmarks with a-definite semantics. Possibly, his test of whether a particle verb is able to introduce a referent at first use misled him: verbs that license definite null instantiate also pass it. In any event, the subjects in the experiments may well have processed the particle sentences faster on average simply because they assumed that they were given a sentence taken from a larger coherent discourse which would resolve the goal. Notice that the object following the main verb and preceding the particle or preposition was always definite, also prompting the reader to imagine additional prior context that would make the NP referents familiar.

Apart from questions about the linguistic analysis, there also seem to be some problems with the experimental design and analysis.¹⁸ There are several assumptions that Härtl makes which are problematic. One is that processing difficulty is the only important factor for processing. Processing difficulty is a very broad measure, and so negative evidence does not mean anything. If one task takes a lot longer to process than another, one might be able to say the processes involved are different. But if there is no difference in reaction time one cannot claim that there is no processing

¹⁸The following points are the result of consultation with my ICSI colleague Shweta Narayan. Any mischaracterizations of Härtl’s work are, of course, my own fault.

effect involved. Härtl also assumes that neutral context sentences are interpreted in a natural way, a belief that is not warranted according to Clark 1996, 1997.

The foregoing observations do not account for why an effect was found with the prepositions. But it's possible that in the particle case, subjects did not have to think of the actual object (at the basic-level), so none of the possible target objects were primed, whereas in the preposition case subjects knew a noun would follow so they started activating the actual possible members of the (superordinate) category indicated by the rest of the sentence. Basically, in the preposition case subjects knew there would be an object so they tried to find one. In the particle case subjects knew the object would not be linguistically specified, so they were not expecting to be given one. Which does not mean that it was not conceptually active; it could have been active at a level that does not have rich imagery (a superordinate level); alternatively, it could have been active but with the awareness that it was not going to be specified by the interlocutor.

Given these concerns about Härtl's experiments, we should be hesitant to embrace his conclusions. Nevertheless, the theoretical possibility that certain constructions may involve the removal of conceptual participants at least from a linguistic level of representation should be kept in mind. We can have greater confidence in Koenig & Mauner (1999) and Mauner & Koenig's (2000) results for English passives and middles, showing that while for both constructions agent arguments are present in the semantic representation, in middle constructions the implicit arguments are less syntactically active in certain regards than in the passive construction.

4.2.3 Lexically licensed omissions

Blocked complements

Blocked complements are semantically necessary elements that can never be overtly realized. This was earlier illustrated with the English verb *spend*, which doesn't allow expression of a Seller—except indirectly in a location phrase:

(4.104) You can **spend** \$100 at a skateshop and get a real board, or you can throw

\$20 away at Wal-Mart for a useless piece of garbage.

A similar example of a blocked complement is the German verb *schwindeln* 'cheat' in (4.105).

- (4.105) Frank hat **geschwindelt**.
'Frank cheated'

Cheating is often done to disadvantage a Victim but in German the Victim cannot be expressed with the basic verb *schwindeln* (cf. (4.106)). A prefixed verb such as *beschwindeln* must be used to express a Victim, as in (4.107).

- (4.106) *Frank hat {mir/mir gegenüber/an mich/auf mich/zu mir} **geschwindelt**
'Frank cheated me.'
- (4.107) Frank hat mich **beschwindelt**.

Similarly, the verb *zuschlagen* 'hit' cannot express the Victim but a definite Victim has to be retrieved from context.

- (4.108) Peter hat zuerst **zugeschlagen!**
Peter was the first to hit $\emptyset_{definite}$

One might argue that verbs with blocked complements simply have lexical representations separate from their conceptual semantic background and that the lexical representation does not include the unrealizable argument. However, there are two problems with this view. First, given that it is usually assumed that predicates' lexical entries have nothing to say about the presence or form of adjuncts, one wonders why these semantic participants should not even be expressible by adjunct PPs. Second, in the case of the German *zu*-verbs (*zubeissen* 'bite', *zupacken* 'grasp'), hearers are instructed to retrieve a specific discourse referent, which would be a rather unusual requirement for an adjunct-like element. Thus, we must entertain the possibility that a lexical entry contains semantically distinctive arguments that cannot be projected into the syntax. Note that in the German case, the predicates with blocked

complements could not possibly be derived from other entries (e.g. *schwindeln*₂ or *zubeissen*₂) with the identical set of word-forms.¹⁹

Incorporation

For my purposes here, I will speak of incorporation when a verb is morphologically related to a noun whose referent's type constrains the type of one of the verb's core frame element.^{20,21} Simple examples are verbs of placing such as *bag*, *box*, *crate*, for which the default goal locations are understood to be bags, boxes, and crates respectively. The morphological relation for English is typically one of conversion such that the verb is zero-derived from the noun. In other languages, the relation may be more complicated. In German, for instance, the equivalents of the verbs *bag*, *box*, *crate* carry the prefix *ein-* 'in', indicating the container-hood of the goal location: *eintüten*, *einschachteln*, *einkisten*.

Incorporated elements are typically not very prominent in discourse. The arguments do not receive overt expression except when a specific subtype of the incorporated type is to be indicated, as in (4.109) and (4.110).

(4.109) Once I had them all rolled I flash froze them and **bagged** them into a ziplock.

¹⁹It is interesting to note in this context that Talmy generally likes to account for different verb uses in terms of lexicalization differences rather than by either assuming deletion of surface material that would produce the correct meaning or by relying only on speaker's facility for semantic interpretation, even when drawing on context and world knowledge (2000: 23-24). Talmy (2000: 35) also seems to view the lexicalization account as more straightforward than accounts that attribute the meaning of one or more of a verb's uses to grammatical constructions (Goldberg 1995, Michaelis & Ruppenhofer 2001) because of the problem of capturing which verbs and constructions can and actually do combine.

²⁰There are other verbs in English that seem to incorporate a frame element but where this is marked morphologically. For instance, the verb *misbehave* might be said to incorporate a Manner frame element which needs to be expressed overtly in adverbial form with related verbs such as *act*, *behave*, *conduct oneself*.

²¹The term unfortunately invites some confusion because of the use of the term incorporation in the Boas-Sapir sense for the grammatical rather than lexical incorporation phenomena in languages like Mohawk. In this work, however, I will continue using the term in my sense as it is also well established in the frame semantic community. Eventually, the frame semantic terminology should be changed as it involves a much shorter tradition of use.

- (4.110) There, the students took huge boxes of chocolate chips and **bagged** them into smaller bags so that they could be distributed to families.

Incorporated elements are only existentially bound; in negative clauses, their existence is not entailed.

- (4.111) I can't **bag** the groceries as we're out of bags!

As we will see, incorporation is basically very similar to indefinite null instantiation. The latter differs only in that there is no morphological relation between the verb and a noun whose referent's type would serve as a selectional constraint on one of the verb's arguments.

Null instantiation

Fillmore (1986) distinguishes two kinds of lexically licensed argument omissions, the indefinite (INI) and the definite type (DNI). The two types of lexically governed null instantiation are exemplified in (4.112) below.

- (4.112) a. John **contributed** \$50 $\emptyset_{\text{Recipient}}$ (DNI)
 b. We **ate** $\emptyset_{\text{Food/Meal}}$ at 2 (INI)

Notice first that, as Fillmore 1986 points out, what is relevant for lexical null instantiation are word senses rather than dictionary headwords. For instance, *give* can omit the Recipient under definite interpretation when it is used in a charitable-donation sense but not when it expresses transfer of physical possession:

- (4.113) Somebody from the Red Cross came by and asked for a donation. I **gave** \$30.
 (4.114) *John needed the scissors so I **gave** them.

Further it is important to keep in mind that lexical null instantiation is tied to what Krifka et al. (1995) call particular as opposed to characterizing sentences. The

former are about particular events, properties of particular referents, etc. while the latter express generalizations or, in the broadest sense, talk about what the world is like. We saw earlier that in generic/habitual sentences such as (4.115) a vast number of arguments can be omitted that must be realized overtly in particular sentences like (4.116).

(4.115) A bully is someone who **injures** and maims and wreaks terror for the sheer pleasure of causing others suffering.

(4.116) The three bomb blasts **injured** more than 100 people.

The two kinds of null instantiation involve different processes of recovering the unrealized argument. In sentences with INI there need not be any overt mention of an indefinite NP in the linguistic context nor does there have to be a referent of the kind denoted by the omitted argument in the physical discourse setting. However, it seems to be a requirement for cases of DNI that either a specific entity has been mentioned or it is inferable from prior discourse context or that one is present in the physical discourse setting. These properties correlate with the differential felicitousness of the two types of omission in various discourse contexts.

- With indefinite null instantiation, both speaker and hearer can claim ignorance of the exact identity of the referent; not so in the case of definite null instantiation.

(4.117) Speaker

- Kim was **reading** \emptyset but I don't know what he was reading
- #Kim donated** \$20 \emptyset but I wonder who she gave it to.

(4.118) Hearer

- a. A: She's **knitting**. B: I'm curious to know exactly what she's knitting. Do you know?
 - b. A: She's **knitting**. B: #I must have missed this but could you remind me what she's knitting?
 - c. A: I **contributed** \$20. B: #I'm curious to know exactly which organization you contributed to? Do you know?
 - d. A: I **contributed** \$20. B: I must have missed this but could you remind me which organization you contributed to?
- Neither definite nor indefinite null instantiation is possible when the argument to be omitted is explicitly topicalized and would be expected to be the primary sentence topic of the clause containing the null instantiating verb.

(4.119) What about the bananas? #John ate \emptyset .

(4.120) What about the Red Cross? - #I gave \$20 \emptyset .

In this regard, the lexical null instantiations behave the same as some of the constructions with indefinite or more specific a-definite interpretation for the omitted argument.

(4.121) What about the criminals?—#The store owner was **locked up** in his store room \emptyset_{Agent} .

Note that topicality or discourse-active status, as opposed to bearing the pragmatic relation topic to the proposition, is, however, no problem for DNI and INI omissions.

(4.122) Did the guy from the Red Cross come by?—Yes, and I **gave** \$25 $\emptyset_{Recipient}$.

(4.123) John really loves Sue very much but he wants to **marry** $\emptyset_{Partner_2}$ only after graduation.

- A speaker can opt to identify a previously non-specific entity in the context following indefinite null instantiation. Such an identification makes no sense in the case of definite omission.

(4.124) I've **eaten** \emptyset already. It was delicious

(4.125) I gave \$20 \emptyset . #It was some kind of charity.

- In contexts of definite omission, a definite pronoun can usually substitute for the zero but not in the case of indefinite omission.

(4.126) Did the Red Cross call you about a donation? -Yes, and I **contributed** \$20 to them.

(4.127) Are you hungry? -#No, I've **eaten** it

- When an INI-verb heads a negated clause, existence of a referent for the null instantiated argument is not entailed, which we saw is also the case for incorporated arguments (cf. (4.111)). By contrast, in the case of DNI verbs, the null-instantiated argument is entailed to have a referent. Consider the felicity of the the continuations in (4.128) and (4.129).

(4.128) I'm not **cooking**. Do you see any food around here?

(4.129) I didn't **contribute**. #How could I? There are no charities in this country.

- Sentences (4.130) and (4.131) show that a DNI element doesn't have to be a discourse topic in the way that this is claimed to be necessary for recipe zeroes (see Haegeman 1990): another referent can be topicalized but the need for anaphoric identifiability of the argument omitted under DNI is preserved.

(4.130) The house I **donated** $\emptyset_{\text{Recipient}}$

(4.131) The troops they **withdrew** $\emptyset_{\text{Source}}$ but the the munitions they left behind.

Not surprisingly, the same is also true for arguments omitted under INI. In (4.132), no specific Theme needs to be given even though the frame element Theme is a core conceptual participant of loading (and more generally verbs in the Filling frame).

(4.132) The truck I **loaded** Ø first.

- It seems that depictives cannot predicate either over arguments null instantiated under DNI or INI.

(4.133) *I like to **eat** Ø raw/hot.

(4.134) *I like to **read** Ø hot off the press.

(4.135) *You **promised** me Ø as a reward!

(4.136) *She **taught** me Ø as the most important skill in this business.

(4.137) *We **attacked** Ø in their homes.

Incorporated arguments also are not available for secondary predication. For sentence (4.139) it is not possible to get the interpretation that the saddle is wet, regardless of where the depictive phrase is placed.

(4.138) He **put** the saddle, still wet, back onto the animal.

(4.139) #He **re-saddled**, still wet, the animal.

- Goldberg & Jackendoff (2004) cite verbs of emission, especially bodily emission, and food ingestion as cases where resultatives are predicated of implicit arguments. They give the examples in (4.140).

(4.140) a. Bill **spit/urinated/coughed/sneezed** out the window.

b. Bill **sweated/bled** on the floor.

c. The toilet **leaked** through the floor into the kitchen below.

d. Bill **ate** off the floor.

e. Bill **drank** from the hose. (= Goldberg & Jackendoff 2004:537, (12a-e))

Goldberg & Jackendoff note that of these cases some but not all—e.g. *cough into*— might be treated as involving cognate objects. Further we may note that some of the cases seem to be idiomatic. Example (4.141), for instance, shows that verbs that like *eat* in (4.140d) describe acts of having a meal, do not comfortably combine with the phrase *off the floor*.

(4.141) ?*His house is so clean you could **lunch/dine/snack** off the floor.

Also, while in (4.140e) the resultative and the predicate could be substituted by various containers and verbs of fluid-ingestion, it seems that the sentence does not primarily assert a motion of the fluid caused by the means of a drinking action. In fact, all the attested cases the lexical verb entails that motion of a Theme happens but the action itself is not primarily seen as a means to achieve translational motion.

Apart from the above verb classes, it seems that resultatives are not generally available with arguments omitted under either a definite or an indefinite interpretation.

(4.142) #Bill **cleaned** \emptyset spotless.

(4.143) #This morning Smithers decided to **shave** \emptyset clean.

(4.144) #I think today I **ironed** \emptyset to perfection.

- It seems clear that an argument omitted under INI is not able to control a benefit clause, as shown by (4.145).²² There are relatively few cases of DNI where the argument is human. In the case of charitable giving, it may in some cases seem

²²I am using the term benefit clause here as a complementary term to Agent-oriented purpose clauses such as (i).

(i) One day he **called** to tell me to come quickly, I had to sign papers.

Note that benefit-clauses are not merely reduced relative clauses that are part of object NPs since they can also attach to pronouns and demonstratives as is shown by (ii) and (iii).

(ii) I am **knitting** this to keep my neck warm in the winter.

(iii) We **brought** it to cool down our dorm room at night.

that a purpose clause can be controlled by an argument omitted under DNI, as in (4.146). However, with verbs of this type, there seems to exist a general metonymy between Recipient and Charitable cause such that the Recipient does not actually have to be identifiable, as shown by example (4.147). In example (4.148), *send* does not have a charitable donation background and it seems that a null instantiated Recipient/Goal cannot be construed as the subject of a purpose clause.

(4.145) *She's **knitting** Ø to keep my neck warm in the winter.

(4.146) The Worthington Public Library got its first home of its own in 1927, when Elizabeth Jones Deshler (pictured in the portrait in the Worthington Room and at left) donated money for a building on the north-east corner of the village green in memory of her grandfather, Colonel Kilbourne. Four years later, she **donated** money Ø to build the library's north and south wings.

(4.147) A: I heard that Bill Gates **gave** \$100m for some AIDS vaccine trials.
B: Oh, that's good news. I hope they're making progress.

(4.148) #When they got married, she **sent** a book Ø to enjoy.

Overall, it seems we can follow Allerton 1975 who characterizes indefinite argument omissions as cases where 'we are faced with an object that is not recoverable, because it has not been thought of and it is not expected that the listener should concern himself with it' (p. 214). Of course, as the discussion of the implicit agents of constructions such as the passive or the middle suggested, it may be that the interlocutors do not track a specific token filling a role but that they are still perfectly well aware that the role exists and that it has some filler.

Indefinite lexical null instantiation is clearly not anaphoric whereas definite null instantiation seems to be. We will return to the topic of how zero arguments are interpreted under definite null instantiation after a discussion of the kinds of verbs that exhibit the two kinds of omissions.

In the case of verbs with indefinite null instantiation, the listener is invited to assume that an entity of an appropriate type participates in the event but is not given any further information about that entity. The appropriate type of object is typically strongly/uniquely associated with the eventuality denoted by the verb in one of the following ways.

- the noun is a nominalization of the verb or a derivationally related event noun (sing a song); these are central examples of the cognate object construction
- the noun denotes a subtype of the eventuality denoted by the verb (dance the waltz) and there may also exist a denominal verb (to waltz)
- the noun denotes an artifact whose canonical purpose is to be used in the kind of intentional action denoted by the verb and which is by default inferred to be used in each such event (cut the bread with a knife; read a book)
- The noun denotes a canonical product of the process denoted by the verb (paint a picture, knit a sweater, sew a shirt, write a paper/book)

With regard to the fourth class, note that there may be cross-linguistic variation regarding what, if anything, counts as a canonical product. For instance, while there is no canonical object for *build* in English, its German equivalent *bauen* has 'house, family home' as a canonical object.²³

(4.149) Sie haben Ø **gebaut** und wollen nun einen Garten anlegen?

Sie haben gebaut und wollen nun einen Garten anlegen ?
 You have built and want now a garden start ?

You built a house and now want to start a garden?

(4.150) *You **built** Ø and now want to start a garden?

²³Johanna Nichols pointed out to me that one can use *build* in English without an object when building is contrasted with, for instance, buying a house: *We couldn't find a house we liked so we're going to build.*

The opposite situation obtains for *use* and *nehmen* 'take', the verbs associated with taking drugs. While English allows argument omission in this sense of *use*, the same is not possible for the German counterpart.

(4.151) At 26, Paul has been addicted to crank for years and sent to jail at least twice. By now, he's **used** \emptyset so long he's not quite sure of time anymore. (Sacramento Bee, October 8, 2000, "A Madness Called Meth")

(4.152) Er **nahm** *(Drogen).
He took *(drugs)

Another interesting class of INI-verbs has to do with actions that result in a status-change. Consider example (4.153).

(4.153) If I were to guess ... I would say that you will **marry** \emptyset_{Other} within three years.

Predicates like *marry*, *get engaged* and *divorce* differ from related predicates such as *be related* in that they have to do with marital status and a change in marital status is interesting enough to be reported without providing further details. Similarly, there is a distinction between *go out* and *see* in their dating related senses and the verb *date*: only the latter can be used (generically) to indicate that somebody is looking for a partner and willing to go out on dates (*She isn't dating right now and she doesn't give much thought to finding Mr. Right*). The same kind of phenomenon can be observed with the verb *retire*: unlike *resign*, *step down*, *abdicate*, *quit*, *retire* is used for indicating that somebody ends their working career altogether. Again, this counts as a status change and reporting it is of interest even if the last job held is not specified, as in (4.154).

(4.154) Great drama. I grew up in a drag racing family. My father **retired** last year and this year he is going to travel to all 23 NHRA races in his motor home.

Among the verbs that lexically license definite null instantiation there is a significant subset in which the referent to be recovered must not simply be part of prior discourse but is specifically restricted to being anchored to the subject referent.

- With certain verbs the object referent is typically the same as the agent referent. For instance, grooming is mostly done to oneself and in English verbs of grooming mostly occur without a reflexive pronoun.
- With certain verbs that denote necessarily symmetrical events such as *meet*, a plural subject referent has to be understood as identifying both sides when there is no oblique expression present that could be understood to express Side-2 to the Side-1 expressed by the subject. The events involved are construed as reciprocal, with the actors acting on each other. Levin (1993) refers to such verbs as “Verbs of Social Interaction” and “Verbs of Contiguous Location”. A further class is illustrated by the lexical unit *coincide*, which at first glance may seem to be a temporal counterpart of Levin’s Verbs of Contiguous Location. But *coincide* is actually not based on proximity/juxtaposition. Rather it is part of a group of predicates that indicate that an Entity₁ shares structure or properties with an Entity₂. (e.g. *coterminous*, *coextensive*, *coessential*, etc.). These predicates also allow DNI of Entity₂.
- Certain verbs denote the relationship between a whole and a uniquely or near-uniquely identifiable part. Verbs of body-movement like *shrug* [*shoulders*] and *blink* [*eyes*] are good examples.

For other verbs, the missing argument is a location that is given in the discourse. For instance, with most verbs of departing, fleeing/escaping, removing, or arriving the source or goal location, respectively, can usually be omitted:

- (4.155) The troops began to **evacuate** Ø on 7th December and the last troops left Ø on 22nd December.
- (4.156) He was carrying a false passport when he **arrived** Ø from London

- (4.157) The government **withdrew** the troops \emptyset .
- (4.158) But although the crowd has danced its way through the set , once the band are off stage , everyone **scarpers** \emptyset .

Interestingly, when verbs of arriving, departing, or passing are used in a fictive motion sense, null instantiation is a lot less common than when they are used with sentient self-movers. For instance, among 200 uses of the sequence “the highway passes” which were retrieved from the Internet as part of a Web-search yielding 2220 total matches, none exhibited null instantiation of the Landmark at all.

- (4.159) After crossing the Murray River, the highway **passes** through the only place not bypassed on the route - Albury.

This is statistically very different from the situation for *pass* in the relevant sense overall. Table 4.4 in section 4.3.3 will show that 25% of Landmarks are omitted under DNI.

Similarly, I inspected 126 matches of the sequence “the road arrives”, which were also retrieved with the help of the Google search engine. Among these, there were no cases of fictive motion either; among all uses of *arrive*, the omission rate is 62.2% (cf. Table 4.4). The 7 uses that did lack a Goal all described situations in which the road itself was under construction rather than being the path that an unexpressed Cognizer follows.

- (4.160) However, Claudio’s worst fears are realized: disease transmitted by the proffered gifts of knives and axes in 1968 begins to decimate them long before the road **arrives**. The ironic reality that first contact itself mimics the deluge to follow is not lost on the student
- (4.161) Small farmers are limited in their access to land since they must stay near roads; most property is claimed as soon as the road **arrives**.

The difference between the concrete and the fictive motion uses seems to lie in the fact that with the latter the construal is generic but also from the point of view of

the path as opposed to the Landmark or Goal: there is no viewer in these locations in whose consciousness the locations themselves could already be activated.

Another group of DNI verbs denote events that are necessarily understood as part of larger event chains.

(4.162) I'm glad I **got even** with that jerk.

(4.163) Demeke was **arraigned** at a New York court last night but German prosecutors said they may ask for him to be extradited.

Some of these verbs typically require that a specific previous event be understood. This previous event is lexically presupposed and is often expressed as a complement of the verb as is shown in (4.164) and (4.165), respectively.

(4.164) a. I never **got even** with him → I have reason to seek revenge
 b. Demeke was not **arraigned** this morning. → Demeke is currently suspected of having committed a crime.

(4.165) a. I **got even** with that jerk [for always stealing my newspaper]
 b. Demeke was **arraigned** at a New York court last night [on an indictment charging him with murder] but German prosecutors said they may ask for him to be extradited

Other clear examples of verbs allowing definite null instantiation are ones having to do with rewarding and punishing, of which we may consider taking revenge verbs a sub-type, and response/reaction verbs (*respond*, *reply*, *answer*, *react*).

One large special class of verbs with presupposed background events, is the one noted by Fillmore 1986, namely that of aspectual verbs.²⁴ These verbs talk about the stage to which an event has progressed or how process flow is managed. As such they necessarily presuppose an event whose process control is at issue. Almost all of these verbs allow omission of the non-finite complement which names the larger background event.

²⁴This and other verb classes as well as certain constructions are discussed by Levinson 1983 as presupposition-triggers (181-184).

(4.166) They just **started** Ø.

(4.167) **Go on** Ø.

(4.168) We'll **finish** Ø by six.

Note that one needs to include here predicates that talk about preparation (*ready, prepare*) as well as predicates that talk about being finished (*done, through (with)*).

(4.169) ...so when you are **ready** Ø, you call her and then make your descent into the mikva .

(4.170) By the time farm ministers are **through** Ø, the cutting edge of the Mac-Sharry reform will be blunted.

With verbs of Grooming such as *shave, wash, shower* and *bathe*, a non-overt patient is interpreted as identical to the subject. However, this is not just typically the case but necessarily. That is, exchanges like the following are impossible:

(4.171) Did you bathe the patient?—No, not today. #But I **showered** Ø

Another set of verbs that superficially seems to allow null instantiation of a reflexive object are verbs like *relax, refresh* and even *invigorate*.

(4.172) He was **relaxing** with a pint and a score of the Resurrection symphony.

(4.173) Come **relax** yourself with a hot oil rubdown.

(4.174) Relax in complete privacy in your suite, or **invigorate** with a brisk workout in our indoor pool completely with its own whirlpool and sauna.

(4.175) Refresh and **invigorate** yourself with a combination of robust Black Tea leaves steeped with a blend of Bergamot.

(4.176) If you are doing this on one of Pennsic's notoriously cold nights, you can **refresh** with more hot water.

Another small cluster of verbs that allow reflexive DNI consists of words that have to do with changing one's appearance so as not to be recognizable *mask*, *masquerade*, *disguise*, and *dress up*. The two conceptually related verbs *dress* and *hide* also allow reflexive DNI. No other members of their respective frames (e.g., *put on*, *pull on*, *don*; *conceal*, *obscure*) seem to allow DNI, reflexive or non-reflexive.

(4.177) We all went off to **dress** \emptyset for the banquet.

(4.178) Luther hid everywhere. He **hid** \emptyset behind trees. He **hid** \emptyset under straw bales.

The restriction to identity the referent of the null object with the subject distinguishes verbs of grooming, relaxing, masking from the kinds of verbs licensing definite null instantiation discussed above. Though the restriction might be captured either diacritically in a lexical entry or by a special lexical rule, there is cross-linguistic evidence that intransitive uses of grooming verbs can be basic.

Talmy (2000) argues that in some cases, an action complex such as grooming can be taken to manifest either locally in the body and movements of a single actor, a situation he calls monadic *Personation* type, or distributively, with the actor's body acting on that of a further participant, a situation he calls dyadic *Personation* type. The two distinct *Personation* types may be derived from each other. In French, the dyadic *Personation* type is basic and self-directed grooming is reported with reflexive object pronouns. In Atsugewi, however, the monadic self-directed construal is basic and special benefactive morphology (*iray*) is needed to augment the valence of a grooming verb to allow specifying a person other than the groomer as the one being groomed.

(4.179) French

a. Jeanne se **lavait**

'Jeanne washed'

b. Jeanne la **lavait**

'Jeanne_i washed her_j'

(4.180) Atsugewi (Talmy 2000:91)

a. s-'-w-cu-spał^a

'I combed my hair'

b. m-'-w-cu-spał-iray-isahk

'I combed your hair'

In English there is no morphology to suggest a direction of derivation. Absent psychological or psycholinguistic evidence, it is reasonable to treat both uses as separate lexicalizations as Talmy does.²⁵ This treatment could, of course, be applied to verbs like *relax*, which also require that a use without overt object be interpreted as being self-directed.

Finally, we should take note of the body-related verbs *ache*, *hurt*, *itch*. These can occur with or without expressing the Experiencer as a clause-level argument.

(4.181) the x-ray of my chest which has been **aching** me for weeks, showed nothing, thank goodness.

(4.182) You may want to place a small pillow between your knees if your back is **aching**.

Since the Experiencer is necessarily the Body-part possessor, their identity is typically recoverable within the clause. The subjects of these verbs are conditions or body parts and the objects, when overt, are consequently not co-referential with the subjects and therefore are not reflexives.

Let us return now to the topic of how arguments omitted under definite null instantiation are interpreted. While indefinite null instantiation is not readily anaphoric—it does not tolerate antecedent co-reference as shown in (4.119) above—definite null instantiation is usually assumed to be anaphoric in some sense. However, there is quite

²⁵Clearly, an account in terms of personation does not seem promising for the majority of null instantiations. While it is plausible that verbs meaning 'behave' can cross-linguistically or intra-linguistically vary in personation type (*behave* v. *behave oneself*), it is difficult to conceive, for instance, that the verb meaning 'drive (a vehicle)' could be basically intransitive in some language and would have to undergo derivation to yield a transitive verb *drive* that can occur with an object.

a range of variation among the individual verb classes in the characteristics of the null argument. Consider again verbs like *meet* and *shave*. With these verbs the antecedent itself or the element that licenses the inference to the referent is always the subject of the clause, which makes these cases of ‘anaphora’ quite distinct from pronouns, which are usually assumed to be the overt forms that DNI omissions alternate with. According to Brown (1983) the antecedent as well as the subsequent mention of the referent of an unstressed pronoun in written English are on average 1.7 clauses away from the current mention. In the case of verbs like *meet* and *shave*, however, the presumptive antecedent would be a clause-mate, that is, 0 clauses away. In the terminology of generative linguistics, the omitted arguments of *meet* and *shave* are null *anaphors* rather than *pronominals*.

In other cases, the missing argument does not have an overt antecedent in the linguistic context or is not physically present in the extra-linguistic context. Especially in the case of locations, there need not be an overt mention as long as the referent location has been bound to some prior event. Discourses such as the following are possible:

- (4.183) Recently, an Albanian youth was stabbed by an Arab, and was left bleeding to death on the ground while the ambulance waited for the police to **arrive** \emptyset .

In example (4.183), from a piece about the immigrant situation in the Swedish city of Malmö, there is no overt mention of the exact place where the stabbing takes place. While the context tells the reader that the incident takes place in Malmö, that cannot be the zero Goal of *arrive* as the ambulance presumably is a local one that is not going to Malmö from somewhere else. Moreover, the speaker may very well not know where exactly the scene of the stabbing was. All that is needed is a notion of identifiability in the sense of Lambrecht (1994)—that a discourse representation of a particular referent already be stored in a hearer’s mind or be constructed—rather than the real-world identifiability that allows the conversationalist to subsequently recognize the referent in an encounter, which Lyons (1999: 7) seems to take as the criterion for identifiability. Nonetheless, cases of DNI require that the referent be

given either in the physical context, in prior discourse, or via associative inference. DNI cannot be used to introduce a specific referent that is only identifiable to the speaker but has not been mentioned before or is inferable. Compare:

(4.184) [In the library] I am looking for a book. It's a first edition of the Canterbury Tales and I left it on this desk last night.

(4.185) [Police officer to headquarters] We are **following** Ø. #It's a a blue Toyota Camry going west on MLK. The driver ran a red light.

That cases like (4.183) have to count as DNI even though no specific token of a location can be identified is also suggested by applying the various tests suggested earlier for distinguishing between DNI and INI omission. It would be odd for the Speaker to simply add *I actually wonder where the police arrived/were going* or *I actually don't know where the police were going/arrived*: for the purposes of the discourse, the answer *wherever the stabbing occurred* seems to be presupposed and as the level of detail that is relevant. Also, if the arrival was negated as in (4.186), modified from (4.183) above, the particular location of the accident scene would still be understood as the place where the police didn't arrive:

(4.186) Recently, an Albanian youth was stabbed by an Arab, and was left bleeding to death on the ground while the ambulance waited for the police, who never **arrived** Ø.

It seems that DNI-omissible location elements not only might lack an overt antecedent in principle, but do in fact frequently lack one. For instance, in a random sample from the BNC the verb *cross* occurred without an explicit object 22 times in various senses/uses having to do with crossing boundaries or spaces. In thirteen of these cases, there was no explicit reference within the 10 clause windows before and after the relevant mention of the boundary or space that had been crossed. In two cases there was only a subsequent mention, in three there was only an antecedent mention, and a mere four cases had an antecedent as well as a subsequent mention within a 10 clause window. Though in some of the cases without an antecedent mention in the

10 clause window, the antecedent just happened to fall outside the window—in one case, the antecedent, a train compartment, was mentioned 60 clauses earlier—in the majority of cases the space or boundary was to be inferred via frame/world knowledge. For instance, in (4.187) what is crossed on a trip from London to Paris is the English Channel, while in (4.188) the reader infers that getting from the doorway to the dressing-table requires crossing the room.

(4.187) In the late afternoon of 21 August, he **crossed** to Paris and joined Baldwin at the Ritz shortly before midnight.

(4.188) He stood in the doorway adjusting his tie, looking immaculate in a charcoal-grey suit. “You scared the life out of me!” **Crossing** to her dressing-table, he checked the knot in the mirror.

Compared to the above examples of DNI, the omitted recipients of the verb *contribute* seem to behave in discourse very much like the overt unstressed pronouns of Brown’s (1983) study. Brown reports an average look-back to a prior mention of 1.72 clauses (1983:329). In a sample of 200 sentences of the verb *contribute* in the BNC, there were 22 non-generic/habitual sentences without an overt Recipient. In 21 of the 22 sentences, the null Recipient had an overt antecedent that was on average 1.2 sentences away; only in one sentence, there was no antecedent within a ten-sentence window of prior context.²⁶

While it may seem tempting to restrict the notion of definite null instantiation to those cases that are similar to pronouns and thus to exclude grooming verbs and verbs that allow locations to be unexpressed, I will resist this temptation. The reason is that, as far as I am aware, it has not been demonstrated that pronouns themselves always have the same lookback and topic persistence (that is distance to prior and subsequent mentions) across all predicates they co-occur with. For instance, reflexive pronouns normally have a lookback of 0 clauses.

(4.189) I **cut** *myself* on the knife

²⁶Notice that the counting of sentences probably under-counts clauses somewhat so the actual number measured in clauses might be closer to that for Brown’s unstressed pronouns.

Also, pronouns can occur without an antecedent in exophoric reference such as (4.190) or via associative inference such as (4.191):

(4.190) John got really sick and had to **vomit**. *It* got all over the carpet.

(4.191) My sister's getting **married** and *he* happens to be an old friend of mine.

Though one might want to say that (4.190) and (4.191) constitute inference and not co-reference, it is also the case that co-reference can be conceptualized as a relationship between conceptual representations and linguistic expressions (Bosch 1988) rather than as one directly between linguistic expressions. Under the former conceptualization, inferables differ from 'true' cases of anaphora only in that their representation has not been explicitly activated by a linguistic expression before. One might thus say that in cases like (4.183), the zero argument of motion verbs is a non-overt inferable similar to the overt ones in (4.190) and (4.191).

A final issue that deserves consideration is whether, as suggested by Jacobs 1994, among verbs allowing definite null instantiation of an argument, there also is a distinction, as observed among verbs licensing indefinite null instantiation (cf. *knit* v. *drink*), between (a) verbs that have the same selectional restrictions for null arguments as for overt arguments and (b) verbs that have narrower selectional restrictions for implicit arguments. Jacobs discusses only one example, the German verb *akzeptieren* 'accept' (1994:302). He points out that when it occurs without a post-verbal argument (which we may call Content), what is accepted can only be something like a proposal but not, for instance, one's being handicapped or being defeated. From the Frame semantic point of view, however, these are separate senses. In the case of accepting a proposal, *akzeptieren* shares a conceptual background with *zustimmen* 'agree [to]' and *einwilligen* 'consent', whereas in the case of accepting, for instance, a loss or defeat, *akzeptieren* evokes the same background as *sich abfinden mit* 'resign oneself to' or *hinnehmen, tolerieren* 'put up with'.

In sum, there is no compelling reason to exclude any of the cases of argument omission discussed in this section from DNI status, given the parallelism between their behavior and that of overt pronouns and anaphors. There is also no clear evidence

that it is necessary to sub-divide DNI-verbs into ones that have narrower selectional restrictions and others that exhibit the regular range of selectional restrictions.

4.2.4 Representation of omitted arguments

We may begin our consideration how a lexicon and grammar should handle null instantiation by first asking the more fundamental question whether there is anything in particular to be represented. The null-instantiation related phenomena that were discussed in the previous sub-sections seem to suggest that the relationship between syntax and conceptual background knowledge is mediated by one or more linguistic levels, at one of which null instantiation is 'effected'. However, there are analyses that make no such assumption. For instance, Rice's (1988) approach to null instantiation denies the possibility of making a sharp lexical distinction between NI verbs and verbs that do not allow NI. For her, there seems to be a direct mapping from a conceptual background structure onto a syntactic surface form. There is no lexical argument structure or syntactic structure from which anything could be omitted or subtracted along the way.

... object omission is neither a process nor does it represent two separate versions of a verb, a transitive and an intransitive one. Rather, certain construals of transitive events are such that they focus on the active participant and leave the acted-upon participant unspecified and, most importantly, to be filled in with a default value. Omitted objects are still objects, which is to say that they are still present at some level of organization, perhaps not at a lexical or syntactic level, but certainly at a conceptual one. (Rice 1988:203)

On this account, the null instantiating uses are just that: uses, not senses. All that is required is that verbs have a rich lexical-conceptual knowledge associated with them that makes a null instantiating use more or less likely given principles of event/scene construal.

A similar point is also made by Ono & Thompson 2000 and Thompson & Hopper 2001, with emphasis however on the role of frequency of usage. In their study of zero subject anaphora in Japanese, Ono and Thompson argue that what linguists think

of in terms of argument structure just reflects the frequency with which certain kinds of elements tend to be associated with uses of a predicate.

What this means is that, WITH FREQUENT REPEATED USE, for certain verbs, certain associations DO in fact get reified by grammars-the grammars of languages often do pick out certain associations-these are what we call "core" roles , or part of the "case frame". As shown by Bybee 1985, Clancy forthcoming, Du Bois 1987, Hopper 1987, 1988, and others, frequent repetition often leads to grammaticization. This is an important fact about languages. But we want to stress that it's not merely a structural fact, it's primarily a pragmatic fact because it is a direct result of frequency in USE. (p. 484, emphasis in original)

The consequences of this thinking for the analysis of Japanese zero arguments are taken to be as follows

- argument structure in Japanese is primarily semantic and pragmatic, rather than structural
- the core/non-core distinction is only loosely grammaticized and there are no obligatory argument slots that must be filled by an overt argument or a zero
- the sense that speakers have of an "intended referent" can be accounted for by well-known inferential processes that are said to be pervasive in language use and comprehension

Ono & Thompson present as further evidence for the ill-foundedness of the concept of argument structure that for most of the predicates in their conversational Japanese data it is unclear what their arguments might be. In particular, they argue that while one might want to resolve such cases of underspecification by positing several argument structures, it is not a good solution in so far as there do not seem to be distinct meanings involved.

Assuming that language use and language processing are cross-culturally very much the same, the arguments about Japanese subject zeros should also be applicable to null instantiated non-subject arguments in English. However, the relevance of the pragmatic principles would presumably have to be tempered by language specific

Verb	Total	null instantiation (%)
arrive	196	122 (62.2)
ask (Addressee of begging)	88	29 (33.0)
ask (Addressee of querying)	176	114 (64.8)
bathe	104	59 (56.7)
contribute (Recipient)	122	23 (18.9)
contribute (Theme)	122	86 (70.5)
cross (Landmark)	141	25 (17.7)
dress	78	62 (79.5)
eat	125	96 (76.8)
enter	80	16 (20.0)
govern	60	20 (33.3)
grab	137	4 (2.9)
know	202	62 (30.7)
obey	188	43 (22.9)
pass (Exam)	22	4 (18.2)
pass (Landmark)	56	14 (25.0)
reap	51	15 (29.4)
squeeze	76	11 (14.5)
steal (Theme)	106	11 (10.4)

Table 4.1: Percentage of null instantiating uses in given sense

grammatical facts: the differences between English and Japanese are due to the presence or absence of grammaticalized argument structures. Nevertheless, to sharpen our understanding of the English data it is useful to see to what extent the reasoning about Japanese might explain English facts.

Based on the kinds of evidence presented in Bybee and Hopper 2001, one can certainly agree with Ono and Thompson that frequency plays a great role in language in general. In the case of null instantiation, the sheer frequency of null instantiating uses of some verbs, as shown in Table 4.1 might convince us that they have separate representations. However, one wonders where such frequency knowledge is stored if there are no argument structures or if not all predicates have them.

Further, the relationship between frequency and obligatoriness is probably more complex than Ono and Thompson suggest. For instance, according to Ono and Thompson, the agents of *eat* are understood as obligatory arguments because people tend to mention these referents frequently—conversationalists are interested in people and what they do, as evidenced by the frequency with which speakers refer to people IN DISCOURSE IN GENERAL. While this reasoning might go through for human agents, it might not be able to capture obligatory elements that are infrequent in discourse overall. For instance, *formulate*, *phrase*, and *word* require a manner phrase but it is unclear how prominent manner really is in discourse. Similarly, some locative phrases are optional setting adjuncts (*I ran into John this morning in the city*) whereas others are obligatory location or goal arguments (*John lives in Miami*). The frequency of three-dimensional spaces in discourse cannot serve as a clue to the status of such an element with respect to any particular predicate.

One might be tempted to say instead that the sense of obligatoriness then just depends on the frequency with which an element shows up with a particular predicate. That would, however, make the account circular. On the one hand, a certain element of a conceptual background is important to speakers so they will mention it frequently when using a predicate evoking that background. On the other hand, the importance of that element is taken to be demonstrated by its frequent occurrence with the predicate at issue.

For a language like English one also needs to record the particular interpretational conventions associated with null instantiations: in non-generic sentences a zero argument of a given verb can have either only a definite or an indefinite interpretation. However, this could not be done unless there was some kind of argument structure or valence. Definite or indefinite interpretation could not be associated generally with all instances filling the given semantic role as overt realizations do not necessarily have to match the interpretation of the zero argument in terms of definiteness.

Finally, the pervasive inferential processes that Ono and Thompson invoke to explain how it is that Japanese speakers sometimes have the sense that there is a specific referent do not seem to be relevant in the same way to English. At best, they are called upon only with certain lexical items, which would again amount to a

lexical fact. If truly general cognitive or communicative principles were operating by themselves in English, they should not show such lexical affinities. Their operation should lead to e.g. general subject pro-drop when topic continuity is given, as for instance under the CONTINUE condition described by Walker & Prince 1996 in terms of Centering Theory. One should be able to say or write things such as the following:

- (4.192) Rob Van Dam nailed Goldust in the chest with a huge kick. Then Ø **hit** him with the rolling thunder and then Ø **pinned** him but only got a two count.

By similar reasoning all languages should, like Hungarian (Németh 2000) or spoken Brazilian Portuguese (Schwenter & Silva 2003), have lexically much less restricted object omission when the intended referent is predictable and no ambiguity is possible. Consider the following Hungarian example:

- (4.193) (Péter, András and Jakab are distributing bread to children.)
 Péter: Mindeki Ø **kapott**?
 András: Nem. Pálnak nincsen.
 Jakab: Péter Ø **adott** Pálnak. Biztosan már Ø **megette**.
 Péter: Has everybody got [bread]?
 András: No. Pál does not have any.
 Jakab: Péter gave [bread] to Pál. He must have eaten [it] already. (Németh 2000: 1675)

In example (4.193), we find both an INI and a DNI omission in Jakab's utterance. The object of *give* is interpreted partitively as 'some bread' and the object of *eat* is the quantity of bread introduced into the discourse by the clause headed by *give*. The omissions here are bound to the specific prior context. The verb *eat* does not have a lexical entry with the patient specified to be of type BREAD. And the verb *eat* inflected for the definite conjugation is not special in being able to license the definite omission

of an argument.^{27,28} Clearly, English is very different: indefinite omissions have no partitive readings and outside of particular constructional contexts verbs do not freely license definite and indefinite omissions. The cases of constructional/contextual null instantiation e.g. in recipes or in soccer match reports may come closest to the Hungarian state of affairs for definite omissions. However, these contexts are clearly restricted. And despite Rice's suggestion that there is fluidity, the situation in English simply cannot be properly described without reference to specific lexical items.

Given the foregoing discussion, how are we to analyze lexically specific and constructional argument omissions linguistically? Let us begin by considering the minimum requirements that a representation should meet.

- (A) First, lexically licensed null instantiation should differ in its representation from the broad constructional types such as subject drop in diary style, object omission in instructional imperatives and *labelese*, given that the latter are not lexically specific.
- (B) Indefinite interpretation needs to be distinguished from definite interpretation.
- (C) The accessibility status of the omitted referents should be reflected. In particular, referents of arguments omitted under DNI should be tracked as discourse active rather than, say, merely identifiable. In other words, their analysis should be more along the lines of a pronoun than along the lines of a definite NP.
- (D) Verbs that impose narrower selectional restrictions on objects omitted under INI should be differentiated from verbs that maintain the same range of

²⁷In Hungarian the difference between the definite and the indefinite conjugations allows verbs to license both kinds of null instantiation very freely and with dependence on the context. It seems conceivable for languages like English that one of the two types is available freely while the other type is lexically restricted.

²⁸Mithun 2000 reports that in Central Alaskan Yup'ik, an ergative language, patients of many transitive verbs can be omitted when intransitive inflectional morphology is used. Interestingly, the interpretation of the omitted argument then has to be definite. An anti-passive has to be used to omit a patient with an indefinite interpretation. Clearly, it would be interesting to study how alignment types, licensing type, and interpretation of omitted arguments pattern typologically.

selectional restrictions for both overt and implicit arguments.

What is less clear is whether and how the additional syntactic facts that we noticed for the various omission types should be taken into consideration. Table 4.2 provides a summary of the relevant properties. The Table shows, for instance, that lexically null instantiated object arguments cannot bind reflexives, be predicated over by depictives, and cannot co-occur with Patient-related purpose phrases (P) but only with Agent-related ones (A).

Based on Table 4.2 we may choose not to treat decausatives at all as involving argument omission given that Agent-related expressions and constructions are not possible and that inference of an Agent is very difficult. It is more difficult to decide how middles and passives ought to be distinguished. One possible treatment is to say that middles have the same conceptual background as active and passive sentences but take a different, restricted perspective on it, which different from passives, excludes the Agent from the focus of attention. Agent-related manner adverbs such as *deliberate* and *carefully* as well as Purpose-phrases require access to the psychology of the Agent: they are therefore incompatible with middles. Instruments and Means actions are different. They do presuppose Agents, to be sure, but they also constitute links between Agent and Patient: it is through the Instrument and/or Means action that the Agent acts on the Patient. Instrument and Means thus are not as exclusively Agent-related as Purpose and adverbs of mental attitude. Koenig & Mauner's (1999) findings suggesting that Agents of middles are less readily accessed than those of passives are more of a descriptive nature: the Agents of middles are attributed to the difference in middle and passive *semantics* rather than to particular pragmatic specifications for such referents by these constructions.

The omitted subjects of diary style as well as the omitted arguments of labelise and the objects of recipe imperatives are omissible as discourse topics. In that way, they differ from ordinary imperatives, with which they share most of their behaviors and co-occurrence possibilities. These constructions do not seem to foreground or background any participants or sub-events in a way different from active sentences with overt subject arguments. The four constructions do of course differ with respect to the argument

Omission Type	Omitted Argument	Instrument & Means	Manner of Agent	Purpose/Benefit	Depictive	Reflexive controller	Discourse status
Decausative	A	n	n	O	O	O	inaccessible
Middle	A	y	n	O	O	O	inaccessible
Passive	A	y	y	A, O	O	O	inaccessible
Imperative (Subject)	A	y	y	A, O	A, O	A, O	situationally given
Labellese	A/S, O	y	y	A, O	A, O	A, O	discourse topic
Diary style	A/S	y	y	A/S, O	A/S, O	A, O	discourse topic
Recipe imperatives	O	y	y	A, O	A, O	A, O	discourse topic
Reporting clauses	O	y	y	A	A	A	situationally given
Match reports	O	y	y	A	A	A	uniquely identifiable
Habituals & Generics	O	y	y	A	A	A	inaccessible
INI	O	y	y	A	A	A	inaccessible
DNI	O	y	y	A	A	A	given (textually or situationally)

Table 4.2: Syntactic properties of various argument omission types

whose omission they license. The constructions allowing omission of an argument qua discourse topic are not all equal: in diary style and recipe imperatives the argument has to be subject and object, respectively, whereas in the case of *labelese*, discourse topics may be omitted both as subjects and objects.

The remaining omission types—reporting clauses, match reports, habituals/generics, lexically licensed definite and indefinite null instantiation—share a lot of behaviors and co-occurrence possibilities. They do differ, however, in discourse status. Referents of unexpressed addressees in reporting clauses and DNI-omitted referents must be given, whereas the omitted referents in match reports just need to be uniquely identifiable. The referents of arguments omitted under indefinite null instantiation or as part of generic/habitual sentences are inaccessible even though their type is known given the predicate's selectional constraints.

Let us now return to the question how we represent argument omission in the lexicon and grammar. Let us first turn to *marry* as an example where the null instantiated uses have the same selectional restrictions as uses with an overt object. One possible lexical treatment is along the lines of Jacobs 1994, who assumes only a single lexical entry with multiple listed syntactic and semantic valences. The semantic representation would not itself be the input to projection into syntax. Instead a separate construct, an argument structure list called SYN-VAL, keeps track of those elements that are syntactically expressed. Null instantiating lexical entries then exhibit a mismatch between the elements appearing in the semantic representation and those appearing on the argument structure list (SYN-VAL). Consider the representations for the uses of German *heiraten* 'marry' with or without an overt Partner₂ frame element in (4.194).

- (4.194) a. SYN-VAL₁: /nom/akk
 SYN-VAL₂: /nom
 b. SEM(SYN-VAL₁): $\lambda x^2 \lambda y^1$ [HEIRAT(x)(y)]
 SEM(SYN-VAL₂): λy^1 [HEIRAT(u)(y)] (= Jacobs 1994:298, (20a-b))

SYN-VAL₂ and its associated semantics SEM(SYN-VAL₂) belong to the null-instantiating use of *heiraten*. The null instantiated element is present in the semantic represen-

tation as a free variable u that is neutral with respect to definiteness (*definitheit-sneutral*). As Jacobs points out (1994: 298, fn. 39), assuming subtypes of variables presupposes that the logical representation can be sensitive to the discourse status of referents. Jacobs envisions these representations being used together with Discourse Representation Theory (Kamp 1981).

(4.195) We also got a chance to see the Vermeer on view at Sotheby's. The painting **went** for 16 million pounds at auction last week.

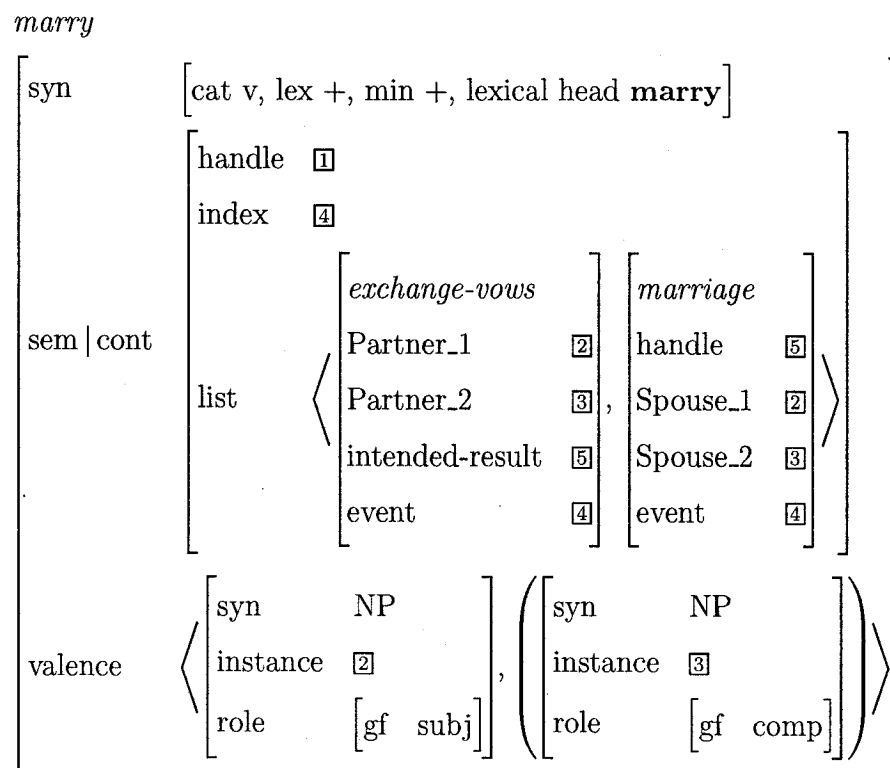
A simple treatment of the INI-verb *marry* in a Construction Grammar-style lexical entry is shown in Figure 4.2. The optionality of the post-verbal complement is indicated simply by placing that argument in parentheses on the valence list.²⁹

The fact that the Partner₂ argument cannot be focal is not specifically encoded here. I will assume that it is a general requirement of the grammar that focus be associated with overt phonological content. The existential binding of the Partner₂ argument is also not specified in the lexical entry. I will assume that the fillers of any ontologically necessary roles of the semantic frame associated with a particular lexical unit are existentially quantified over. (This does, of course, presuppose that negated and modalized sentences involve distinct mental spaces or possible worlds within which it makes sense to talk about existence.) Notice that there is no stipulation of any particular discourse status for an INI-omitted referent. This is as it should be given that INI-verbs fit into contexts where the referent has been identified and is discourse active as well as in discourses where no referent has been identified (and consequently none is discourse-active). We can also assume as a general principle that zero-forms cannot introduce referents into the discourse. Any cases of reference such as (4.196) and (4.197) that seem to be anaphoric to a zero element are really only inference-based.

(4.196) John got really sick and had to **vomit**. *It* got all over the carpet.

(4.197) My sister's getting **married** and *he* happens to be an old friend of mine.

²⁹I am treating this sense of *marry* as involving only a complement rather than an object since it does not occur in the passive unlike *marry*₂ 'perform a marriage ceremony'.

Figure 4.2: Minimal lexical entry for *marry*

We now turn to INI uses with narrower selectional restrictions than the uses with overt NPs. It should be a fairly uncontroversial move to assume separate lexical entries given the earlier discussion and desiderata A-D. An example of such a lexical units is the INI sense for *drink*, ‘drink alcohol’, which is exemplified in (4.198).³⁰

- (4.198) a. I **drink** only socially.
 b. Thanks, I don’t **drink**.
 c. Do you **drink** because you are shy with other people?
 d. John’s father **drinks**.

Note that the basic representation needed for such cases of INI should be very similar to what is need to account for predicates that have blocked complements. For instance, *go [for]* evokes as its background the commercial transaction frame but it cannot express a Seller participant as one of its syntactic dependents. The Seller is just existentially bound.

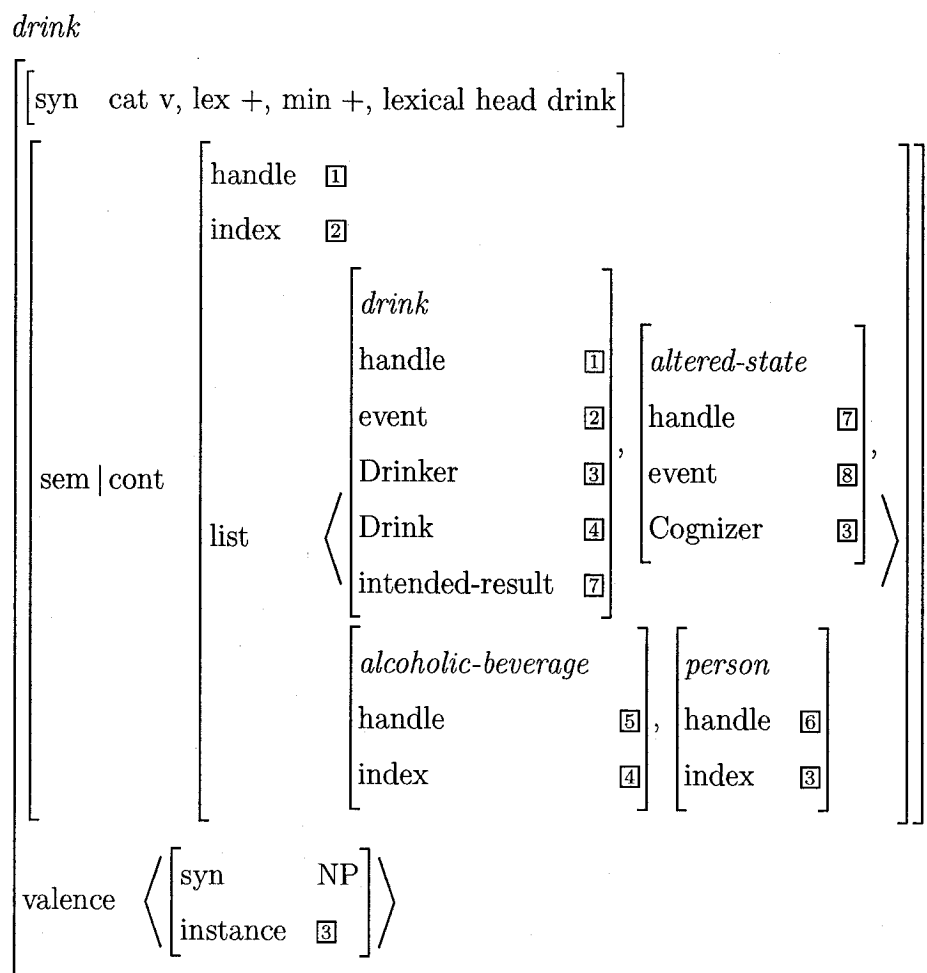
Figure 4.3 provides a minimal lexical entry for the alcohol-specific sense of *drink*. The selectional restrictions on the zero argument as well those on the subject are pre-specified on the list of predications. Note that the drinking event specified is the same as the general one, that is, I assume that the drinking of alcohol is not a distinct kind of bodily event from drinking non-alcoholic beverages.

Let us next turn to verbs that have a blocked DNI-complement such as the German verb *zubeissen* ‘bite’. Such lexical units are similar to the alcohol sense of *drink* in not occurring with overt objects. They are different, however, in that one does needs to make a specification about the discourse pragmatic status of the unexpressed referent. In Figure 4.4, this is done by making use of the background frames feature which is

³⁰Besides these uses of *drink* in (4.198) English also has verbs such as *tipple*, *imbibe* and *booze* which seem to be dedicated to alcohol consumption. These verbs can have overt objects or null instantiate them. If we assumed that *drink* in the alcohol-sense also can take overt objects in examples like (i), then we could place the alcohol sense of *drink* in the same frame with *booze*, *imbibe*, *tipple*, and possibly certain multi-word expressions such as *do [rum/tequila/etc] shots*.

- (i) Do you **drink** too much?

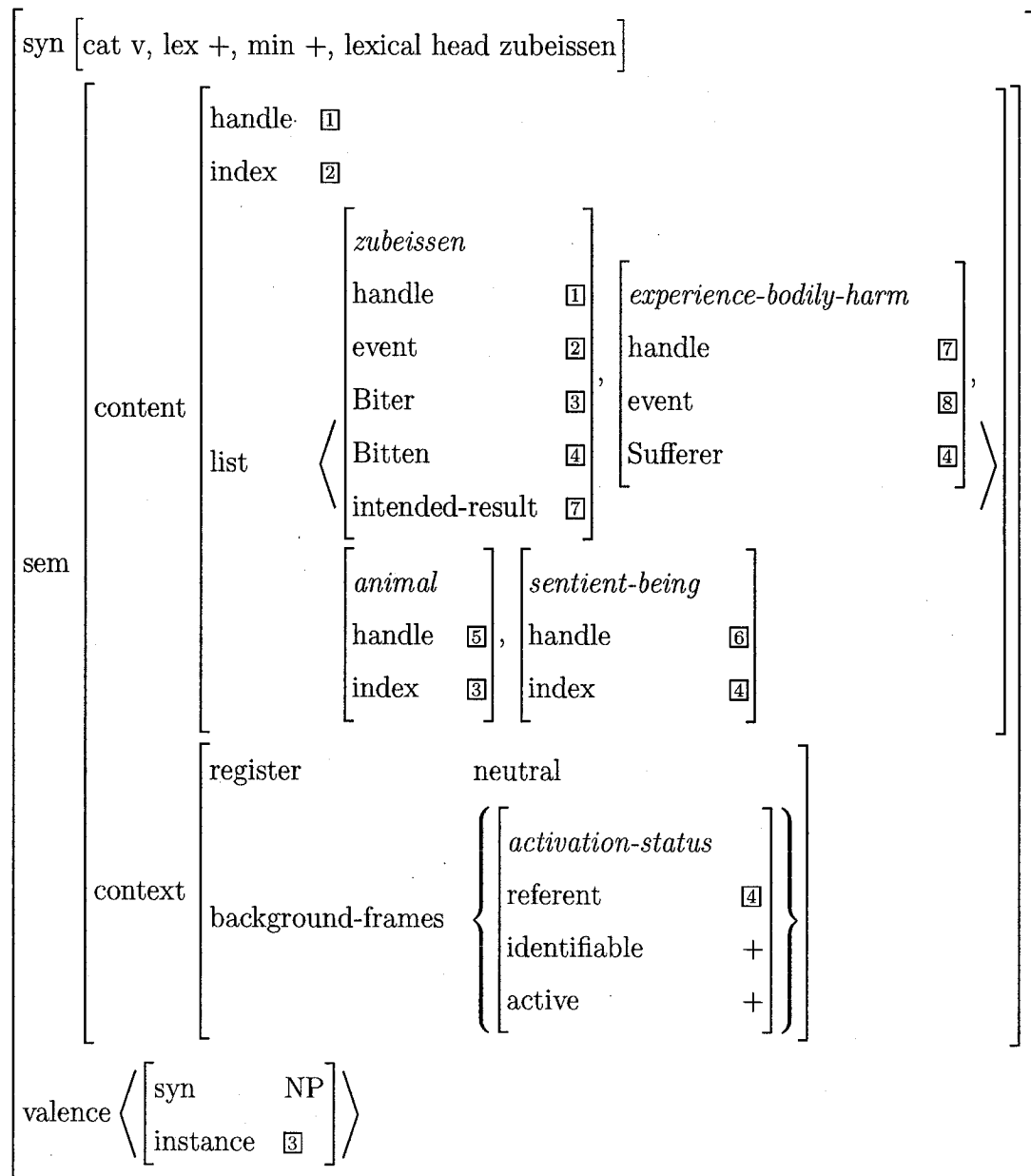
The alcohol-related uses of *drink* would then after all not be a case where a separate sense is ‘intransitive’, that is, consists only of null instantiating uses.

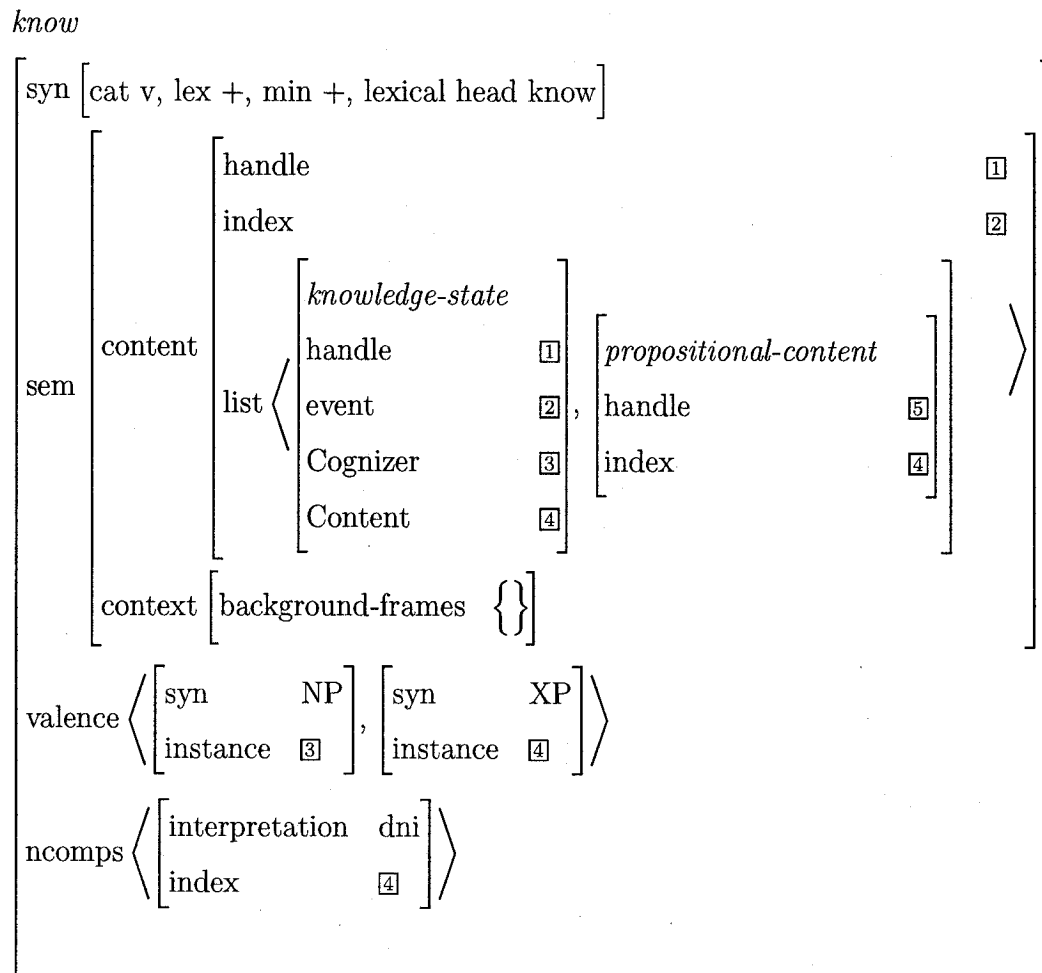
Figure 4.3: Minimal lexical entry for *drink*₂ 'drink alcohol'

one of the attributes of the context feature. The context feature is the place to specify the contextual conditions in which an utterance containing the linguistic sign at issue is felicitous. The context feature envisioned here is inspired by the eponymous HPSG feature, which has been used for purposes such as encoding information structure (e.g., Engdahl & Vallduví 1996; Melnik 2002) and usage conditions for honorifics (Pollard & Sag 1994: 92-97).³¹ Another conceivable use, indicated in Figure 4.4 is to record register information under the context feature. For current purposes all that is needed is a background-frames feature that can include among the set which is its value, assertions about discourse pragmatic properties of referents. In the case of a blocked complement with anaphoric interpretation we can simply put an activation-status predication about the referent into the background frames set.

Now that we have seen how blocked complements can be dealt with, the of DNI verbs that do occur with overt objects requires just one more mechanism. We could either derive special DNI-lexical entries by an HPSG-type lexical rule or use a two storied template construction of the kind employed by Kay (MS). Either way, we need a mechanism that record which argument is omissible and that it's interpretation is of the anaphoric type. Figure 4.5 shows the basic entry for the verb *know* (in the *wissen/savoir* sense). The entry contains a new book-keeping feature, *n(ull)comp(lement)s*, adapted from unpublished work by Paul Kay. The value of this feature is a set of feature structures with two attributes, interpretation and index. The interpretation type is specified as *dni* in 4.5 and the index, which may be understood as a pointer to a referent in the discourse model, is that of the Content argument. Now a lexical rule could derive from the entry in Figure 4.5 the new entry shown in Figure 4.6. The effect of the rule would be threefold: it would remove the feature structure specifying the interpretation for a given argument from the *ncomps* list, remove the argument bearing the same index from the valence list, and insert into the background frames set a predication specifying that the referent pointed to by the relevant index is identifiable and active.

³¹Since, for instance, the social reality of honor is not just reflected in the honorifics but in a sense constituted by them, an alternative and equally valid characterization of the context feature is that the information specified there is asserted about the context.

zubeissenFigure 4.4: Minimal lexical entry for *zubeissen* 'bite'

Figure 4.5: Basic lexical entry for *know*

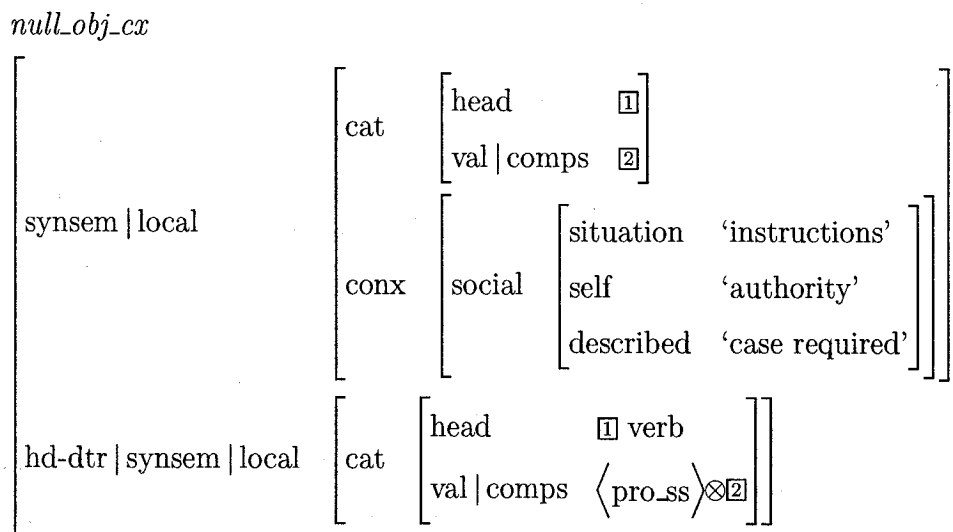


Figure 4.7: Bender 1999's HPSG analysis of instructional imperatives

The representations provided here for lexical omissions meet desiderata B-D. We will briefly consider one of the constructional omission types to see how desideratum A, that constructional and lexical omission should have a different representation, could be met. Figure 4.7 displays Bender 1999's HPSG analysis of instructional imperatives. Overriding HPSG's default valence principle, the construction builds a phrasal sign whose valence is missing the pro-synsem that represents the omissible argument on the valence list of the lexical verb.

Insofar as Bender's proposed construction is phrasal, it is representationally distinct from the treatment suggested above for the lexical omission types. Another benefit of the constructional solution is that the verb's ARG-ST list in the phrases daughter has the full set of elements that are present when the verb is used with an overt object. Now, given that the ordering of elements on the ARG-ST list is what governs HPSG's treatment of reflexive binding, the given treatment of instructional imperatives predicts that omitted arguments should still be able to bind reflexives. This is indeed the case, as was shown earlier in example (4.50). By contrast, the lexical analysis given for INI, DNI, and lexical blocking cases would, cast within the HPSG-FrameNet treatment, correctly predict that in these cases binding of reflexives is not possible.

Of course, an alternative analysis of this particular fact is also possible: one could simply say that reflexive binding works differently in the context of instructional imperatives. However, if phenomena other than reflexive-binding could be found to be aligned in this way, then there truly would be an advantage in distinguishing lexical and constructional omissions. As Table 4.2 shows, there are some candidate phenomena, namely the (in)ability to be predicated over by depictives (and resultatives) and the (in)ability to control the subject of an object-oriented purpose clause. Pursuing this issue is beyond the scope of this work and so I will leave it open for future research.

Summary As we have seen in this section, the argument omission types of English differ from each other in subtle ways. For current purposes, I will treat as null instantiation all those cases where a use of a verb has the same linking properties as another except that a semantically distinctive argument is not realized. Depending on the verb (and the constructional context) the unrealized argument is understood only with regard to its type or it has to be identified with a particular referent in prior discourse or in the physical setting of the communicative exchange. Based on the observed distinctions, I proposed some desiderata for the kinds of representations they should be given. For the lexical null instantiation types, I also offered sample lexical representations that differentiate all the types. Underived lexical entries were assumed only for verbs with blocked complements such as German *zubeissen* and null-instantiating INI verbs such as *drink* that have narrower selectional restrictions than their counterparts with overt objects.

Though I will continue to use the processual-sounding term null instantiation, I wish to remain agnostic on the issue how exactly null instantiation is processed and how lexical restrictions are stored. Some of these issues might be able to be resolved by psycholinguistic experiments. In the remainder of this chapter, I will focus only on linguistic factors that correlate with the availability of a null instantiating use of a predicate. The conditions under which speakers actually use null instantiating predicates as opposed to realizing arguments overtly will not concern us for the moment. It is clear, though, that all null instantiating uses are context-dependent. As noted

earlier, an argument can never be omitted if it is in focus, for instance, as when it is the target of a *who* or *what* question (Who did you give money to?—*I contributed \$23). Similarly, in defective contexts such as psycholinguistic experimentation involving sentence completion tasks, speakers are highly unlikely to use null instantiation given that there is no context to recover referents from (Roland and Jurafsky 2002).³²

4.3 Possible explanations

The discussion in section 4.2 of the interpretation of omissions with different kinds of predicates may suggest that there is some regularity to be found regarding which verbs do or do not allow omission and what the interpretation of any omissible argument will be. At the same time it has variously been pointed out that in English argument omission seems to be basically unpredictable. For instance, Fillmore 1986 says about definite null instantiation

From the examination of near-synonyms which differ with respect to our feature, we must conclude that DNC [definite null complementation, JR] are not explainable by *semantic* facts. (emphasis in original)

Nonetheless, a sizable amount of research has been devoted to looking for generalizations concerning which verbs allow what kind of null instantiations. We will now review some generalizations that have been proposed in the literature as well as some new ones.

4.3.1 Lemma frequency

Several different interactions between frequency and null instantiation are conceivable. First, a certain base line frequency might be required to allow omission of a core argument. In order for clauses containing verbs with lower frequencies to be processed quickly, the arguments need to be overt even if the referents of the arguments involved

³²Though this is not addressed by Roland and Jurafsky's study, it seems that speakers should avoid null instantiation in sentence completion or production tasks only, or more strongly, when definite (or anaphoric) interpretation is required. Cases of indefinite null instantiation such as *eat* should occur in experimental contexts.

are highly constrained as to their type, that is, selectional restrictions. Second, the two types of null instantiation might be associated with verbs of different frequency ranges, though it is not immediately clear why that should be so.

It seems, however, that there is no interaction at all between the frequency of a lemma³³ and the ability to license omission nor between lemma frequency and the type of omission. As Table 4.3 shows, lemmas with very high token frequency allow null instantiation as well as lemmas with very low token frequency. It is also not the case that definite or indefinite interpretation is associated with a particular part of the frequency spectrum. It is possible, though unlikely, that if one instead measured frequency of the relevant verb sense, one might find some association.

4.3.2 Degree of polysemy

Another factor that one might expect to figure in the distribution of null instantiation is the number of different senses a verb has. For instance, it would make sense if verbs that were highly polysemous always required overt objects to facilitate the identification of the intended verb sense with the help of the object NP.

However, as Table 4.3 suggests, based on the American Heritage Dictionary's and WordNet³⁴ entries for the transitive uses of the verbs listed, very polysemous verbs still allow null instantiation and there also doesn't seem to be an association between NI type and how polysemous a verb is. Both DNI and INI verbs can be polysemous to a great or a small degree.

This outcome maybe is not surprising after all because the tendency of common words to have many senses is well documented. For instance, Zipf (1949) states that the number of meanings of a word is proportional to the square root of the

³³I use the term lemma to mean the set of inflected forms of a lexical unit consisting of one or more lexemes. For instance, *mix up* is a lemma consisting of the two lexemes 'mix' and 'up' with the inflected forms *mix up*, *mixes up*, *mixed up*, *mixing up*. A lemma is paired with a frame in a lexical unit, which corresponds to a word sense in traditional terminology. For instance, the one-lexeme lemma *charge* participates in one lexical unit that belongs to a commercial frame and in another lexical unit that belongs to the Judgment_communication frame.

³⁴The WordNet version used is 2.0. Since WordNet does not group senses into transitive and intransitive senses, I inspected the senses and example sentences manually to identify transitive senses.

Lemma	NI-type	Frequency	AHD senses	WordNet senses
get	—	220,773	16	29
know	DNI (content)	184,638	9	8
take	—	178,398	32	37
use	INI (drugs)	116,374	5	5
ask	DNI (addressee)	60,574	8	6
follow	DNI	45,636	13	22
write	INI	33,298	11	6
read	INI	26,503	15	8
pass	DNI (exam;landmark)	19,459	19	12
arrive	DNI	14,017	3	2
enter	DNI	13,988	11	6
eat	INI (meal)	13,857	6	5
cross	DNI (landmark)	6,615	9	6
contribute	INI (theme)	6,050	3	1
contribute	DNI (recipient)	6,050	3	1
drink	INI	6,026	6	1
sing	INI	5,862	6	1
dress	DNI	4,817	10	13
steal*	INI (goods)	4,416	5	2
pack	INI (luggage/goods)	2,916	10	11
shrug	DNI	2,666	1	1
grab	DNI (theme)	2,655	5	5
govern	DNI (dominion)	2,431	6	4
bet	INI (asset)	2,133	4	2
squeeze	DNI (theme)	1,869	9	9
knit	INI	1,820	4	3
obey	DNI (authority/command)	1,272	2	1
donate	DNI (recipient)	918	1	1
blink	DNI	810	4	1
bake	INI	692	2	3
reap	INI (harvest)	460	4	2
iron	INI	361	3	1
bathe	DNI	128	4	2
baby-sit	INI	61	1	1

Table 4.3: Lemma frequency in the British National Corpus, polysemy, and interpretation type for select null instantiating verbs

frequency. Given the association between lemma frequency and polysemy we might have expected to pick up any association between null instantiation and polysemy indirectly when considering lemma frequency.

4.3.3 Definiteness of the argument

Another possible frequency effect of null instantiation would be for the type of null instantiation and the likelihood of null instantiation to depend on the degree to which an argument is preferentially definite or indefinite. In other words, if the referents of an overt argument of a verb are predominantly (in)definite we would expect that if the verb can omit that argument at all, the interpretation of the omitted argument will be (in)definite. Further, we might speculate that the likelihood of a verb to allow omission is greater, the greater the disparity between definite and indefinite uses. Thus, if an overt argument of a predicate is definite 70% of the time and indefinite 30% of the time, then (1) that argument should be more likely to be omitted than if the proportion was 60% to 40% and (2) the interpretation under omission ought to be definite rather than indefinite.

Table 4.4 shows that, for instance, in the case of *arrive*, the interpretation for a null instantiated goal location is definite and of all objects occurring with *arrive* 97.5% are definite. The Figure also shows that with the exception of *reap* and *carve*, the interpretation of the null instantiating uses agrees with the majority of all the uses, null instantiating and overt ones combined. Further, if one compares the interpretation only for overt arguments one finds that there the same type of interpretation that the null instantiating uses have predominates. For instance, among all uses of *pass* in the exam sense, there are 59.1% overt uses with a definite interpretation and 22.7% overt uses with an indefinite interpretation.

Note that a clear frequency bias in one direction does not constitute a sufficient condition for allowing omission of that argument. For instance, none of the verbs in Table 4.5 allow null instantiation even though some of them have a very clear bias towards either definite or indefinite objects. A comparison with the verbs in Table 4.4 shows that some of the verbs that do not allow null instantiation are more strongly

Verb	NI interpretation	total tokens	NI out of total (%)	uses with same definiteness value out of total (%)	overt uses with same definiteness as NI (% of overt)
arrive	DNI	196	122 (62.2)	191 (97.5)	70/74 (94.6)
bathe	DNI	104	59 (56.7)	102 (98.1)	40/45 (88.9)
blink	DNI	99	84 (84.8)	98 (89.0)	14/15 (93.3)
carve (figure, decoration)	INI	47	4 (8.5)	22 (46.8)	17/43 (39.5)
contribute (Recipient)	DNI	122	23 (18.9)	83 (68.2)	60/99 (60.6)
contribute (Theme)	INI	122	86 (70.5)	117 (95.9)	32/36 (88.9)
cross	DNI	141	25 (17.7)	122 (86.5)	97/116 (83.6)
donate (Theme)	INI	201	9 (4.5)	141 (70.2)	132/192 (68.7)
donate (Recipient)	DNI	159	77 (48.4)	143 (89.9)	66/82 (80.5)
dress	DNI	78	62 (79.5)	77 (98.7)	15/16 (93.4)
eat	INI	125	96 (76.8)	111 (88.8)	15/29 (51.7)
enter	DNI	80	16 (20.0)	75 (93.7)	59/64 (92.1)
govern	DNI	60	20 (33.3)	53 (88.5)	33/40 (82.5)
grab	DNI	137	4 (2.9)	118 (86.1)	124/133 (93.2)
obey	DNI	188	43 (22.9)	143 (76.1)	100/145 (69.0)
pass (Exam)	DNI	22	4 (18.2)	17 (77.3)	13/18 (72.2)
pass (Landmark)	DNI	56	14 (25.0)	43 (76.7)	29/42 (69.0)
reap	INI	51	15 (29.4)	25 (49.0)	16/36 (44.4)
shrug	DNI	112	100 (89.3)	112 (100.0)	12/12 (100.0)
spew (vomit)	INI	17	15 (88.2)	16 (94.1)	1/2 (50.0)
squeeze	DNI	76	11 (14.5)	74 (97.4)	63/65 (96.9)

Table 4.4: Null instantiation type and overall definiteness for select lexical units

biased than some of the verbs that do allow it. Thus, it seems that the strength of bias per se has no influence on the likelihood that a verb will allow null instantiation.

Verb	% of definite	n
emit (light, sound, particles, odour)	15.7	108
lower (prices, temperature)	73.7	57
lower (physical object)	94.0	33
disappoint	56.0	59
impress	54.0	109
spew (emit, not vomit)	33.3	24

Table 4.5: Definiteness bias for some verb senses without the possibility of null instantiation

4.3.4 Thematic role

The thematic role identity of an argument may seem like a natural place to look for generalizations among null instantiating predicates. Curiously, it has not been discussed much in the literature. Fillmore (1986) only notes that *Patients* (or *Themes*) do not seem to be readily omissible. From the examples he discusses it is clear, however, that thematic roles do not allow generalizations about null instantiation.

Consider, for instance, the thematic role *Agent*. While agents are frequently subjects and subjects are frequently omissible, these omissions always depend on particular constructional contexts. There is no lexically licensed *Agent* omission of the kind suggested in (4.200).

- (4.199) a. \emptyset Go home!
 b. Smithers was beaten \emptyset
 c. \emptyset To go would be foolish.

- (4.200) a. What did you do this morning?—# \emptyset **went** to the store and bought bread.

In the case of the thematic role Goal, some verbs allow its omission under indefinite null instantiation (4.201), others allow its omission under definite null instantiation (4.202), and still others cannot omit it at all (4.203).

- (4.201) a. Adam **left** Paris (for New York).
 b. Sue **drove** past the park (to her aunt's house)
- (4.202) a. Smithers **arrived** (in Paris).
 b. Brooks **landed** (in Doha) at eight.
- (4.203) a. John **headed** *(home).
 b. The expedition **reached** *(the oasis) before nightfall.

Similar demonstrations could be given for the other classical thematic role types such as Experiencer, Source, Recipient etc. Trying to predict or characterize null instantiation in terms of thematic role type is not fruitful.

4.3.5 Verb neutrality and basic level status of the object

Rice (1988) summarizes her ideas on object omission as follows:

There seems to be, then, a happy medium for verbs that license object omission. Verbs that are very neutral but that furthermore sustain a wide variety of complements, tend always to require complements (e.g. John loves *(lima beans/Country and Western music).) Verbs that are neutral but whose objects are restricted to one or two possible semantic domains may generally omit them (John bet (five dollars/his entire pension fund).) Finally, verbs that are quite specific with regard to their complement or verbs that reveal something about the manner in which the specified activity is carried out almost always require overt objects (John manicured *(his nails).) Thus neither extremely schematic nor extremely specific verb-complement pairs encourage object omission. (p. 297)

One general problem with Rice's account is that it is not clear how basic-level status or neutrality and variety of complements are to be measured. Judging from her examples, a second problem is that Rice's account does not distinguish indefinite, definite, and constructional omissions even though the types are rather different from each other and what characterizes one type need not apply to the other. For instance, the notion of being (proto-) typical does not apply in an obvious way to definite omissions (e.g. *I know*). And similarly, with verbs of body (part) motion, there does not seem to be a constraint against omission despite the fact that the possible body part objects are highly restricted: contrary to Rice's prediction, *blink* and *shrug* overwhelmingly tend to omit their objects, *one's eyes* and *one's shoulders*.

4.3.6 Differential object marking

Schwenter and Silva (2002) have argued that in spoken Brazilian Portuguese, where null instantiation is lexically unrestricted, animate and specific objects are least likely to be omitted. They suggest that this can be viewed as a kind of differential object marking (Bossong 1985, Aissen 2000): In the domain of lexical NPs, less typical objects need to get more morpho-syntactic marking than typical objects. (For instance, in Spanish, lexical, specific, animate objects need to be marked with *a*.) In the pronominal domain, less typical objects cannot be omitted whereas typical objects can.³⁵

Ruppenhofer & Baker 2003 explored the applicability of the factors animacy and specificity to English object omissions. They examined 361 verbs from 27 semantic frames. They found a tendency in the direction predicted by the differential object marking hypothesis. Among the verb types that allowed omission of objects that represent animate referents, there were more verbs whose omitted object had an indefinite interpretation (e.g. *impress*, *disappoint*) than verbs whose omitted object had a definite interpretation (e.g. *obey*, *acquit*). This could be seen as a differential object

³⁵Aissen 2000 also predicts the existence of differential subject marking effects based on her Optimality Theory analysis. She mentions Guugu Yimidhrr, Punjabi, and Dyirbal as languages exhibiting differential subject marking effects and suggests, more generally, that split ergative systems could and should be thought of in terms of differential subject marking. For the purposes at hand, we will not concern ourselves further with differential subject marking.

	Zero	Pronoun	Total
Inanimate	30	0	66
Animate	1	1	6
Total	31	2	72

Table 4.6: Realization of highly accessible referents for ordering sense of *follow*

marking effect in so far as less prototypical objects with animate and specific referents proved more resistant to zero coding than more typical objects with non-animate and/or non-specific referents. However, the number of verbs with animate objects in Ruppenhofer & Baker's sample were small and the results were not statistically significant.

Ruppenhofer & Baker speculated that the differential object marking hypothesis might hold at the level of sentence tokens rather than lexical units, that is verb senses. They showed the plausibility of this assumption by studying two senses of *follow*. In their sample of 1000 instances of *follow* there were 72 cases of the ordering sense of the verb occurring in an active and potentially fully transitive form. Out of these 72 cases, 31 (43.1%) had a null instantiated object:

- (4.204) An appraisal of the project **follows** \emptyset whereby cost estimates, ...
- (4.205) The Werner Report of October 1970 ... assumed that economic union would automatically **follow** \emptyset .

There was only one pronominal object with this sense of *follow*. In other words, when the object is highly given, it is typically null instantiated, as is shown in Table 4.6.

By comparison, *follow* occurs 131 times in an active and potentially fully transitive form in the Co-theme sense, where a typically animate Self-mover follows the motion of a Cotheme, which is typically an animate participant, though it need not be. In this sense, only 19 (or 14.5%) of the tokens of *follow* have a null object, as shown by Table 4.7.

	Zero	Pronoun	Total
Inanimate	5	2	25
Animate	14	61	106
Total	19	63	131

Table 4.7: Realization of highly accessible referents for co-theme sense of *follow*

(4.206) The bottle of green snug in my jacket, I return to my family. Winchell **follows** \emptyset

(4.207) If they do, should I **follow** \emptyset ?

Besides the 19 null instantiations there are 63 pronominal objects, which suggests that with the Co-theme sense of *follow* highly given objects are typically realized pronominally.

Beyond this initial exploration no systematic study has been done and it remains unclear if there is a systematic negative effect of the conjunction of animacy and specificity on the use of null instantiation. In any event, an effect of animacy and specificity on omission rates could only be a small part of an overall explanation of null instantiation that applies to all types of predicates and arguments.

4.3.7 Taxonomic verb classes

Fellbaum and Kegl (1989) suggest that a verb's behavior with respect to indefinite object omission can largely be predicted from its place in a taxonomic hierarchy such as WordNet, and, by implication, that groups of semantically similar verbs should behave the same way with respect to object omission. They leave aside the issue of whether a taxonomic analysis also applies to verbs allowing definite object omission.

Their central example is the verb *eat*. Fellbaum and Kegl argue that *eat* when it is used without an object has the special sense 'eat a meal', which is distinct from an 'ingest food' sense. In the former meaning, *eat* is said to pattern with the verbs *dine*, *lunch*, *snack*, and *breakfast*, whereas in the latter it is said to pattern with the

verbs *gobble*, *gulp*, and *nibble*. The taxonomic difference between the verbs in these two different groups is then reduced to a syntactic difference: the manner verbs are said to have some kind of obligatory PP adjuncts in their underlying structure that require the presence of an object.³⁶ The meal(time)-related verbs, by contrast, lack such a manner component and are therefore free to omit their objects.

While Fellbaum and Kegl's (1989) taxonomic approach seems to pick up on a real difference between verbs that do allow null instantiation and verbs that do not, it is limited in several ways. It addresses only cases of INI. It can only be as good as its taxonomic base. Consider the fact that WordNet organizes words by part of speech. For example, *object* and *objection* are part of different sub-hierarchies but when talking about the null instantiation properties of these two words—both allow DNI—it would make sense to do so at the same time. Another challenge consists in working out hyponym and hypernym relationships between verbs. And an even more formidable issue is the question of whether the inventory of relationships between verbs is the right one.

4.3.8 Selectional strength

Resnik's (1993, 1996) theory of object omission is strongly centered on the identity of the predicate. The basic intuition is that certain verbs carry enough information about their objects that they do not need to overtly express them. Coming from a computational background, Resnik has formalized the notion of selectional strength as an information-theoretic metric and has tested his theory in quantitative work.

Though selectional strength may seem to characterize only cases of indefinite null instantiation, Resnik (1993, 1996) also applies the notion of selectional strength to definite null instantiation. The basic distinction between the two verb types is said to be that the verbs allowing indefinite object omission select even more strongly and therefore do not require overt antecedents in the discourse for their objects.

As Resnik acknowledges, the selectional strength account by itself does not seem

³⁶No motivation for calling the implicit argument a PP rather than any other phrasal category is given in the paper.

sufficient. He finds that some verbs allowing definite omission have greater selectional strength than some of the verbs allowing indefinite omission. Furthermore, while he found great selectional strength to be a necessary condition for object omission in experiments, it was not by itself a sufficient condition for omission. Certain verbs that select their objects extremely strongly still do not allow omission of those objects. The verb *devein*, for instance, very strongly selects an object of type *shrimp* but cannot omit it.

In fact it may be desirable to find an account of null instantiation that does not directly rely on selectional strength at all but appeals to the states, relations, properties and events evoked by predicates. Resnik's notion of selectional strength is not really concerned with the specific semantics of predicates but rather with the processing of one aspect of a predicate's semantics, its selectional restrictions. But if null-instantiation was really a by-product of processing then we would not expect to find that certain very strong selectors cannot license omission. Consider again the example of *devein*. This verb basically occurs only with the object *shrimp* but there are no non-habitual/generic uses where the object can be omitted, contrary to what a selectional strength account should predict. On the other hand, *devein* behaves like other words relating to emptying containers and clearing areas of some substance or items:

bone.v, clear.v, core.v, debug.v, deforest.v, defrost.v, degrease.v, delouse.v, denude.v, descale.v, disembowel.v, divest.v, drain.v, empty.v, emptying.n, eviscerate.v, expurgate.v, gut.v, peel.v, purge.v, rid.v, scalp.v, skin.v, strip.v, unload.v, void.v

These words all belong to the Emptying frame and, in non-generic sentences, they have to take an overt Source frame element as their direct object as in (4.208).

(4.208) Pat **cleared** the table of dishes.

Many of these verbs seem to be rather strong selectors. In the BNC, *bone* occurs with the names of fish or meat in preparation for eating; *core*'s overt objects are noun phrases headed by *apple*, *lamb*, or *turkey*. This kind of range does not seem to

be significantly larger than that of, say, *knit* or *sew*. Yet while *knit* and *sew* can omit their object, *bone* and *core* cannot. If my intuition about the strength of selection for *core* and *bone* is right, then it is unexpected on a processing-style account that considers only the absolute strength of selection that these verbs should not allow null instantiation. Resnik has not demonstrated that strong selectors as a class tend to allow null instantiation and that *devein* really is just an exception.

Resnik's account also does not account for the regularity in the interpretation of null instantiation across frames. The processing that Resnik has in mind depends on individual lexical units rather than on classes and if lexical units in the same frame have different selectional strengths then they should be able to have different null instantiation interpretations for the same frame element. But this does not seem to occur.

An additional argument against a processing account is that one does not seem to find an association between argument omission and a lemma's degree of polysemy or its frequency: there does not seem to be cutoff point on one side of which verbs do not allow null instantiation. There also is no clear boundary in terms of selectional strength between the two types of lexical null instantiation, or between null instantiating verbs and verbs that do not allow null instantiation, as Resnik points out himself (1993:86).

On the other hand, Resnik 1993 reports that NI verbs for which the object is more readily inferable will omit that argument correspondingly more frequently than NI verbs for which the object is less easily inferred. This seems similar to the kind of effect that Bybee 2001b reports for French liaison: the probability that a word-final consonant will appear before a vowel-initial word depends on the frequency of the sequence.

In sum, although there are some aspects of the selectional strength account that point to an automatic, probabilistic process, it seems clear that selectional strength, if it plays a role at all, is subordinate to semantics. No necessary or sufficient conditions on null instantiation in terms of selectional strength can be stated.

4.3.9 Aspect

Aspect and transitivity

Building on Resnik (1993) and Resnik (1996), Olsen and Resnik (1997) pursue an approach that is situated against the theoretical background of Hopper & Thomp-sons (1980) transitivity hypothesis. Thus, while Resnik's earlier work was focused on properties of the verb alone, the later work with Olsen is oriented towards the properties of whole clauses.

The gist of Olsen and Resnik's argument is that definite omissions are associated with higher transitivity than indefinite omissions and both are located between transitive sentences on the high end of the scale and intransitives on the the low end. More specifically, Olsen and Resnik argue that there is a connection on the one hand between telicity, high object individuation and definite interpretation of omitted ob-jects and between atelicity, low object individuation, and indefinite interpretation of omitted objects. Olsen and Resnik use the term telicity to refer to the Aktionsart of the verb and individuation of an object is said to depend on its individuation from the agent (subject), its individuation from other objects of its class or kind, and in-dividuation from the verb, that is, the extent to which the direct object replicates information provided by the verb.

Olsen & Resnik's findings replicate mostly those of Resnik's earlier work. They find that while objects with any kind of omission select more strongly than the average verb, implicit object verbs permitting the indefinite reading select more strongly than do verbs that require the definite reading. Because of this lesser selection strength, definite-omission verbs are said to require a discourse antecedent for their omitted objects.

Olsen & Resnik also conclude, as did Resnik 1993, that although some strongly selecting verbs fail to appear with implicit objects a verb allowing object omission is necessarily a strong selector. However, Olsen & Resnik suggest in passing that the exceptional cases, that is the verbs that are strong selectors but do not omit their objects, do not do so because of their aspectual properties. They do not discuss any relevant examples.

Aspectual templates and argument realization

Rappaport Hovav and Levin (1998) (henceforth RH&L) present a theory of argument realization based on the combination of verb meanings with a set of universal aspectual templates. This theory also makes predictions regarding null instantiations.

According to RH&L's theory, the idiosyncratic aspects of meaning (what distinguishes e.g. *jog* from *run* from *trot*) are recorded in the lexicon. Next to the lexicon there exists a fixed set of lexical semantic templates provided by Universal Grammar. These templates consist of various combinations of semantically primitive predicates. The templates correspond to a large degree to well-known event types such as accomplishment, etc. Using a verb then means combining the idiosyncratic lexical information with an event structure template.

The event types that RH&L recognize are:

Activity [x ACT_{<MANNER>}] (run, walk)

State [x<STATE>]

Achievement [BECOME_[x<STATE>]]]

Accomplishment (complex-causative) [[x ACT_{<MANNER>}] CAUSE [BECOME_[y<STATE>]]]]

Accomplishment [x CAUSE [BECOME [y <STATE>]]]

Combining verbs with templates means replacing the constant in the template (italicized in the list above) with the meaning of the verb. Constants can be modifiers of another predicate in the template (in this function the constants are written as subscripts) or they can function as predicates within the template.

Constants (i.e. lexically specified verb meanings) contribute the following information:

(4.209) based on its ontological type, each constant is associated with some basic template in accordance with a so-called canonical realization rule

(4.210) the number of participants

The number of participants that a constant specifies may exceed the number of slots in the template (e.g. activity verbs with two participants such as *sweep*).

RH&L call participants that are licensed by the verb meaning as well as by the template **STRUCTURE PARTICIPANTS**; participants licensed only by the constant are called **CONSTANT PARTICIPANTS**. According to RH&L, the two kinds differ in how their realization is licensed.

When a constant is inserted into a template that is not the one lexically specified for it, RH&L speak of **Template Augmentation**: ‘Event structure template may be freely augmented up to other possible templates in the basic inventory of event structure templates’ (p. 111). In other words, no new templates can be created, which is said to account for the failure of accomplishment verbs to receive an activity interpretation: the only available activity template is too small for the template an accomplishment verb has lexically and in fitting verbs into other templates, no information contained in their basic template can be thrown away.

RH&L posit two well-formedness conditions on the syntactic realization of (basic or augmented) templates.

(4.211) **Subevent Identification Condition** Each subevent in the event structure must be identified by a lexical head (e.g., a V, an A, or a P) in the syntax (p. 112)

(4.212) **Argument Realization Condition**

- a. There must be an argument XP in the syntax for each structure participant in the event structure.^{37 38}
- b. Each argument XP in the syntax must be associated with an identified subevent in the event structure.

As discussed by Rappaport Hovav and Levin 1998 and 1999, two interesting predictions regarding null instantiation arise from these constraints. First, all constant participants should in principle be omissible. This would apply to the objects of transitive activity verbs as well as the non-subject complements of bivalent stative verbs.

³⁷Constant participants are said to just be subject to a recoverability condition: they must be prototypical participants such as a floor in the case of *sweep*.

³⁸Linking rules determine the particular realization on the basis of an argument’s position in the lexical semantic representation

The reason is that the activity and the state templates and the achievement template which contains a state template have only one slot, namely the one for their active-form subject. The objects of these verbs are thus only CONSTANT participants, that is, they are part of the conceptual semantics associated with the constants but do not have to be mapped into the syntax by the subevent identification or the argument realization conditions. However, such constant participants need to be recoverable either by being prototypical or by being prominent in the context. RH&L do not specify whether what's recoverable from context can be both a type or a specific entity or whether it has to be one or the other.

Second, objects of resultative constructions need to be realized because they are structure participants of the resultant state predicate. This applies not only to resultative constructions involving an overt secondary predicate naming the resultant state but also to lexical accomplishment verbs like *break*, *dry*: while the verb can identify both sub-events, the subject cannot discharge the realization requirement for the object. The predictions on null instantiation thus fall in line with RH&L's observation that 'result verbs show a much narrower range of variation in meaning and syntactic context than manner verbs' (p. 101). However, it is not clear if the account given is meant to extend to accomplishment verbs like *sew*, *paint*, *write* that do not involve a change of state of an existing entity but rather talk about coming into existence. One could extend the account by choosing EXIST as the state predicate inside the accomplishment template. However, then the template account would then make an incorrect prediction because creation verbs can in fact omit their objects, typically in the progressive: *Mother was sewing in the bedroom*.

In an earlier analysis of null objects that also involved the notion of templates, Brisson 1994 treats creation verbs like *write* as just having two lexical entries. No alternation—or, alternatively, construction—is posited. The intransitive uses are really separate and monadic in Brisson's terms. One potential criticism of this approach is deflected by Brisson: while one might want to know what was written whenever an act of writing has been asserted, according to Brisson such assumptions reflect world knowledge. People will assume that something specific was written but there is also random writing in the sand on the beach. Similarly, she argues, while people will assume that

when walking is reported some self-mover must have gone from and to some place, the same assumption wouldn't be made if swimming had been reported.

Apart from any problems one might detect with analyses of specific predicates, the account given by RH&L has a more general problem: it is circular in that it uses the representational facts—an argument is either a constant or a structure participant—to predict omissibility even though omissibility seems to be the only factor that motivates the distinction in the first place. Neither RH&L 1998, 1999, or Levin 2000 mention any other phenomenon that can be explained in terms of the two two types of participants.

4.4 A possible synthesis: semantic frames

While all of the analyses reviewed in the previous section have noticed relevant facts and made suggestive proposals for generalizations, it remains to be seen how they can be integrated into a more comprehensive account that covers all verbs and also extends to other parts of speech.

Apart from the more processing-related features of frequency and polysemy, the set of explanations explored in the previous section crucially involve two notions: the denotational properties of the missing object and the event structure of the predicate. While Resnik and his collaborators have made an attempt to integrate the two in terms of clausal transitivity, I would like to suggest here that a more fruitful approach is to combine the two via verb classes organized by semantic frames.

4.4.1 Why aspectual classes are not enough

While RH&L's template analysis for verbal predicates allows one to capture several interesting generalizations about null instantiation, it cannot offer a full account of the phenomenon. First, it is necessarily incomplete in the sense that it cannot predict which verbs actually allow NI. I believe that full predictability is not attainable under any account—there really is lexical idiosyncrasy at work. Second, the account implies that the same type of interpretation applies to all omissible objects of verbs in a particular aspectual class but this is not the case.

Consider the case of simple state predicates. At least some relational state predicates should be able to omit their non-subject argument. We do indeed find this to be the case, with the omissible Ground elements of simple stative predicates mostly being interpreted as definite as in the case of (4.213) and (4.214).

(4.213) Everybody **knows** \emptyset .

(4.214) **Ready** \emptyset ?

A DNI reading also seems to be the only possibility in the case of atemporal predicates that denote an interrelationship: relational nouns and adjectives—the latter notably including all comparative forms—allow DNI for their Ground elements.

(4.215) The pattern was **similar** \emptyset .

(4.216) Jim was **older** \emptyset .

(4.217) You should see the **mother-in-law** \emptyset !

Nonetheless, there are some predicates denoting stative relationships that can omit their Ground arguments under an indefinite interpretation. The adjectives *married* and *divorced* as well as the verb *marry* can be used without mention of the spouse.

(4.218) She was **married** \emptyset .

(4.219) He fell in love and **married** \emptyset .

Intuitively, it is clear, though, that these uses of the adjectives *married* and *divorced* belong in the same set with *single*, that is, they talk about an individual person's marital status rather than about the relationship itself. Similarly, the verb *marry*, when used without an object simply signals an individual's change in marital status. The same kind of situation also obtains with *rent*, *buy* and *own*. When used to indicate whether one falls into the home renter, buyer, or owner category, these verbs can be used without an object:

- (4.220) First of all, I not only live in Niagara County, but I have my own place. I don't live with my parents and I don't **rent** Ø, I **own** Ø. I can truly claim to be a bonafide taxpayer in every sense of the word.
- (4.221) For many, the decision whether to **buy** Ø or **rent** Ø is one of basic financial expediency.

Though in (4.220), one assumes on the basis of world knowledge that the speaker owns only one apartment and lives in it, that particular apartment is not really at issue. In context what is important is that the speaker, who is making the case for the office of county manager, presents himself as a local property tax payer who would share the financial burden of decisions he would make as an office holder.

While some of the above uses may be felt to be somewhat special, there are some stative predicates that regularly allow indefinite null instantiation. Predicates like *full*, *chock-full*, *empty* and *overcrowded* are informative even when their oblique ground element is not present. Interestingly, however, these verbs take the Ground as subject rather than the Figure. Further, there exists a large class of predicates that systematically illustrate the difference between a specific relational interpretation and a non-specific generic one, namely evaluative adjectives. Examples (4.222a-b) talk about particular events and agents. The smartness that is attributed to the agent is not a general attribute. By comparison, (4.222d) and (4.222e) are most readily understood as generalizations. In (4.222d), *driving this way* is presented as the smart thing to do anytime and for anybody; in (4.222e) *Pam* is smart in everything she does. (4.222c) is ambiguous. Either there was a period of time when it was smart for everybody to be riding a bike; or there was a period of time when it was smart for a particular person, omitted under DNI, to be riding their bike; or there was a single occasion when it was smart of an individual or a group of people to ride a bike.

- (4.222) a. It was **smart** of you to sell your shares.
 b. You were **smart** to sell your shares.
 c. Riding a bike was **smart**.

- d. Driving this way is **smart**.
- e. Pam is **smart**.

Clearly, an analysis at the level of aspectual class is too coarse to adequately capture the differences between state predicates.

The omission pattern for activity verbs similarly seems to require more semantics than aspectual classes by themselves offer. Compare examples (4.223-4.228), which require a definite interpretation of the object, to those in (4.229) and (4.230), which require an indefinite interpretation.

- (4.223) John is **playing** Ø. (the piano)
- (4.224) I dropped my rope at the base of one of the trees and started to **rake** Ø, struggling to keep from slipping on the steep slope with its loose pine-needle surface.
- (4.225) We re-arranged the room, **swept** Ø and mopped, cleaned the windows, dusted the lockers and changed into combat kit for the day.
- (4.226) He found Indy to be spiking a temp of 101, and when he **palpated** Ø, he found what he thought was a fairly large, hard impaction at the pelvic flexure.
- (4.227) While he was **scratching** Ø, the whole tree broke with a snap!
- (4.228) Here, let me **mop** Ø.
- (4.229) Kim is **sewing** Ø.
- (4.230) I've been **writing** Ø all morning.

We now turn to the internally more complex aspectual classes. Based on the analysis of simple state predicates, we should expect that of the aspectually complex predicates which embed relational state predicates, some will allow for null instantiation of their Ground element. Further, the interpretation should be definite in most

cases. This seems indeed to be the case. Consider the achievement predicates in (4.231) and (4.232).

(4.231) Smithers **won** Ø.

(4.232) Sue **left** Ø.

As seen in the case of simple state predicates, the aspectual treatment alone does not capture the cases where being in a certain type of relationship is by itself informative (e.g., *single*, *divorced*, *married*). Among achievement verbs, there are some that denote entering into such a state and they, too, can omit mention of a Ground element under an indefinite interpretation.

(4.233) Peter's dad **married** Ø late in life.

However, the class of accomplishment predicates as traditionally conceived presents a problem. Levin 2000 argues that among accomplishment predicates one needs to distinguish at least between verbs of creation and consumption (4.234-4.235); verbs of causation (4.236); and verbs of motion that occur with PPs that bound the path (4.237).

(4.234) Tom **carved** a statue.

(4.235) Smithers **ate** a sandwich.

(4.236) Sue **broke** the vase.

(4.237) Kim **jogged** to the store.

Levin 2000 argues that the group of accomplishments exemplified by *jog to the store* is not causative because they are not transitive and the subevents of moving along the path and leaving the source or arriving at the goal location are not independent from each other, as they are expected to be for causatives.

It is not clear what representation should be chosen for these verbs. However, we may note that the null instantiation possibilities of such verb phrases behave in a

familiar way. If the preposition is 'stative' (that is, it does not belong to the set *to, at, into, onto, toward, etc*), the Ground element can be missing under definite interpretation.³⁹

(4.238) They drove **through/in** (tunnel)
They're through/in (the tunnel) now

(4.239) They rode **into/to/toward** *(the national park).
They're into/to/toward *(the national park) now.

Semantically similar simple verbal predicates such as *cross* and *enter* also allow omission of an object under a definite interpretation, as in (4.240).

(4.240) The survey found that on an average weekday 80,457 people crossed the Tappan Zee Bridge heading eastbound. Of these, 98 percent (79,107 people) **crossed** *O* in automobiles and 2 (1,350 people) percent **crossed** *O* in express buses.

Verbs of creation and consumption also do not involve causation for Levin (2000) because there is again a necessary temporal dependence between the subevents: for instance, eating and the vanishing of food unfold together. Simple object-less uses of creation verbs behave aspectually just like activities; missing objects are interested as indefinite.

(4.241) Erica is **painting** \emptyset .

(4.242) Max was **reading** \emptyset .

However, it seems that at least the verb *eat* has a use in which it occurs without an object but is still an accomplishment of the consumption type, as is shown in (4.243).

³⁹The set of prepositions that allows null instantiations may at first appear to be the same that Aske (1989) identifies for Spanish as heading telic path phrases which predicate an end-point location of the moving Figure in distinction to other path prepositions which predicate over the whole motion event. However, prepositions such as *through* and *over* in English have uses in which they function just like telic prepositions: *We drove through in less than 10 minutes.*

- (4.243) I got the letter right before lunch, so I ate \emptyset in 5 minutes, still completely numb with shock...jotted off a quick e-mail to Ben, who is on a Watson right now, and Norm, who is on a Fulbright.

The use of the meal-sense of *eat* in (4.243) behaves like an accomplishment on the sentence level. Regardless of whether one treats *eat* as including the resultant state of food-attrition in its lexical representation or as being basically an activity and combining with a larger template to yield a complex event with a resultant state, the INI-interpretation of the object is not predicted.⁴⁰

The third group, verbs of causation, have representations like the following, repeated from section 4.3.9 above:

- (4.247) [[x ACT_{<MANNER>}] CAUSE [BECOME[y <STATE>]]](e.g., *break, dry*)

- (4.248) [x CAUSE [BECOME [y <STATE>]]] (e.g., *bag, box, butter*)

According to Levin 2000, the important characteristic of these verbs is that the action of the Agent/Cause can be temporally independent from the change of state. For verbs of this type we expect not to find any null instantiation if Levin's Argument Realization condition is correct. The reason is that the object is a structure participant of the embedded state predicate. That prediction seems in fact to be borne out, as indicated by the examples in (4.249).

⁴⁰The label "verbs of creation or consumption" seems too broad to single out only verbs like *sew, knit, read* etc. There are other verbs denoting creation or consumption events that never seem to allow null instantiation.

- (4.244) *He **destroyed** yesterday \emptyset .

- (4.245) *John **created** \emptyset with Bill.

Unlike, *sew, knit* etc. *create* and *destroy* do not include a particular notion of action over time by a particular process and they may indeed cover instantaneous acts of creation or destruction.

- (4.246) The bomb **destroyed** three cars waiting in a queue to enter the coalition headquarters.

- (4.249) a. *We heated up the plastic to 300 degrees and then we **cooled** \emptyset again.
 b. What happened to my vase?—*Tom **broke** \emptyset .
 c. What are you doing at the moment?—*I'm in the kitchen, **cooling** \emptyset .
 d. What happened?—*I **broke** \emptyset but I've already cleared away the mess.

More generally, Levin's treatment of causatives correctly predicts that resultatives as a class cannot omit their objects at all, except in the special contexts of recipes and match reports. Even lexicalized such *drive crazy* and *stab to death* never seem to allow omission of the object.

Overall, the data suggest that while some pretty reliable generalizations about null instantiation can be formulated in terms of Vendlerian aspectual classes, that might not be the most appropriate level of description overall. In particular, it seems that semantically richer classes are what is needed.

Neither lexical aspect nor sentence aspect by itself are predictive of whether null instantiation is possible and what the interpretation will be. Lexical aspect does, for instance, not predict that *marry* allows omission of its object under an indefinite interpretation. In the case of, e.g., accomplishment sentence aspect, we find indefinite zero objects (e.g., *eat in 5 minutes*) as well as definite definite ones (*cross in less than a day*). These findings regarding null instantiation are in line with Levin's assessment, based on consideration of a variety of verb classes and constructional contexts, that aspectual notions by themselves cannot explain all regularities found in argument expression (2000: 420, et passim).

4.4.2 Frame semantics

Clearly, verb classes are useful in making generalizations and predictions about null instantiation. However, they need to have a more specific semantics than just one defined in terms of a causal or an event structure. We will now turn to semantic frames for verb classes that are based on semantic criteria beyond event structure.

A frame is an intuitive construct that allows one to formalize the links between semantics and syntax in the process of lexical analysis. Semantic frames are schematic

representations of situations involving various participants, props, and other conceptual roles, each of which is a frame element. The semantic arguments of predicates—which can belong to any part of speech—correspond to the frame elements of the frame (or frames) associated with those words.

As an example, consider the Revenge frame. An Agent performs a Response_action on a Victim as a punishment for an earlier action, the Injury, that was inflicted on an Injured_party. The Agent need not be identical to the Injured_party but needs to consider the prior action of the Victim a wrong. Importantly, the punishment of the Victim by the Agent is seen as justified by individual or group opinion rather than by law. Lexical units in this frame include *avenge.v*, *avenger.n*, *get even.v*, *retaliate.v*, *retaliation.n*, *retribution.n*, *retributive.a*, *retributory.a*, *revenge.v*, *revenge.n*, *revengeful.a*, *revenger.n*, *vengeance.n*, *vengeful.a*, and *vindictive.a*. Some example sentences with the lexical unit *avenge* are given below.

- (4.250) [His brothers AGENT] **avenged** [him INJURED_PARTY].
- (4.251) With this, [El Cid AGENT] at once **avenged** [the death of his son INJURY].
- (4.252) [Hook AGENT] tries to **avenge** [himself INJURED_PARTY] [on Peter Pan VICTIM] [by becoming a second and better father RESPONSE_ACTION].

Frame semantics is the basis of the FrameNet project, an NSF-funded lexicon-building effort. The project studies words; describe the frames or conceptual structures which underlie them; examines sentences, using very large corpus of contemporary English that contain these words (BNC and newswire texts); and records how information from the associated frames are expressed in these sentences. Of special value to the concern here is the fact that FrameNet not only records the identity, grammatical function and phrase type of syntactically realized material but it also records how frame elements that are not realized have to be understood, that is, whether they are omitted under a specific or non-specific interpretation.

When looking at the annotated data provided for the lexical units in FrameNet frames, several things are noticeable.

First, in the vast majority frames it seems to be the case that if the predicates can omit a given argument/frame element at all, then the interpretation will be the same across all predicates. For instance, in the Revenge frame with all the predicates that allow omission of the Reason frame element, the interpretation of the frame element is definite. Likewise, with verbs having to do with departing from a location (e.g. *leave*, *withdraw*, *depart*, the interpretation of the omitted argument is always definite. (For more evidence along these lines consider the group of ten randomly selected frames discussed in Appendix B.) As far as I know the hypothesis about the constant interpretation of a given omitted frame element for all predicates in a frame is contradicted clearly only by status change predicates such as *marry* or *retire*: with the other predicates in their respective frames, the omitted arguments get a definite interpretation.

Second, generalizations about null instantiation with verbs typically carry over to frame members of different parts of speech. For instance, both *depart* and *departure* allow omission of a Source location under definite interpretation when they are used to report particular instances of this event. In quantified uses of nouns, of course, the DNI understanding is lost. If verbs allowing indefinite null instantiation occur with related nominalizations in the same frame, then the nominalizations will typically also allow INI:

- (4.253) Out of the 60 000 visits I get each months, only a very small number of people made a **donation** to support this project.
- (4.254) If we don't see a report that she **donated** to the earthquake in Iran, or to the children orphaned by British bombs in Iraq, are we to assume that therefore she does not deem them worthy of aid?

It is more difficult to approach the relationship from the side of the nouns. A noun like *destruction* can be used without overt mention of an Undergoer, as is shown in (4.255) which refers to a specific act, namely the destruction of two giant Buddha statues by the Taliban government of Afghanistan in 2001. The verb *destroy* does not allow the same.

(4.255) Upon his arrival in Pakistan on Saturday, Annan said he would convey the world's outrage at the **destruction** Ø.

(4.256) #The Taliban finally **destroyed** Ø.

As (4.257) shows, gerunds similarly often allow null instantiation of frame elements that the base verb cannot omit in non-generic contexts, even when the event itself has been mentioned before or is inferable.

(4.257) EU foreign policy chief Javier Solana condemned the **killing** as “very, very bad news” for the Middle East peace process.

(4.258) #EU foreign policy chief Javier Solana condemned as “very, very bad news” for the Middle East peace process that they **killed** Ø.

It seems to be a general property at least of event nouns and nominalizations that they require a lot fewer participants to be expressed than corresponding verbs.

Third, the majority of predicates in a frame will actually exhibit or lack the possibility of omission for a given frame element. For instance, all 21 verbs that are listed in the Departing frame allow omission of the Source frame element under a definite interpretation.

Fourth, frame semantics allows us to capture similarities across frames that share a background scenario but highlight different parts of that background. As observed by Fillmore 1986, many aspectual verbs allow null instantiation. On a frame semantic analysis, they share an event schema as a background. Against this background, the lexical units in one frame denote being done with an event (*done, through, finished (with)*), others denote preparing for it (*prepare, get ready*), and so forth. All of the lexical units in these frames will require a definite interpretation for their event argument if they can omit it at all.⁴¹

⁴¹In fact, this analysis could maybe be extended to auxiliaries such as perfect *have* and progressive *be*. Instead of talking about ellipsis, one could then interpret examples like the following as null instantiation:

(4.259) I haven't done my homework. Have you?

(4.260) Sue is going to Canada for spring break.—She is?

A slightly different relationship is that between the frames of Revenge and Rewards and Punishments. The former frame is an elaboration of the latter. While words like *punish* are neutral with respect to the basis of the punitive action, Revenge words specifically rule out that the punishment is inflicted as a result of a legal decision. Rather, the punishment is inflicted based on a personal sense of injury to the individual or his or her group. Despite these semantic differences, the words in both frames behave the same way with regard to the action that the patient of the punishment or revenge committed earlier: it has to be recoverable from the context or the predicates are used infelicitously.

- (4.262) He was severely **punished**. → Nobody had ever found him doing anything wrong.
- (4.263) Kim **took revenge** on Pat. → Pat had never given offense to Kim.

This fact is predicted by FrameNet's treatment of the Revenge frame as inheriting the Rewards and Punishments frame: frames lower in the hierarchy of frames inherit all the properties of their parent frames and add properties of their own.

Now the concepts of revenge or punishment are not the only ones that crucially rely on the notion of trigger and response. Another frame that involves the same notion is Communication_response including such words as *reply*, *answer*, *respond*. There is also a construction involving the particle *back* which is used for indicating that an action is taken in response to a previous action. In all these cases the trigger action has to be recoverable from the context.

- (4.264) Bradford TEC **responded** immediately

This kind of analysis should not be surprising in grammatical frameworks that assign a valence to auxiliaries. Also, the problem of sloppy co-reference observed in (4.259), where the hearer is most likely asked about her homework rather than about the speaker's, also occurs with more straightforward cases of null instantiation.

- (4.261) A: My dad is coming to visit.—B: I know Ø.

In (4.261), interlocutor B knows that speaker A's father is coming to visit rather than B's own father. Thus, cases of ellipsis and null instantiation are not in principle different with regard to whether there is an element in prior discourse that could be inserted in place of the zero.

(4.265) She smiled **back**.

Like response scenarios, some reversal scenarios also allow null instantiations that the simple scenarios contained in them do not necessarily license.

(4.266) I'm done here. I'll **replace** the books \emptyset and then we can leave.

(4.267) No—each lady had personally **placed** her cake {on the long table/* \emptyset }—without assistance

(4.268) Later they returned, **unchained** Terry \emptyset and took him upstairs.

(4.269) At the Raja's insistence Tara was **chained** {to a 500 year old tree/* \emptyset }.

Note that in the case of response actions, reversals, and shared background frames with aspectual verbs we are essentially dealing with semantic notions. These generalizations could not be made by aspectual analysis. The aspectual predicates are of various aspectual types—*done* is stative, *finish* is an accomplishment—and so are response actions in the *back* construction—*smiling back* is an activity and *hitting back* is punctual. Reversal predicates all seem to be accomplishments but there is certainly no aspectual characteristic that distinguishes them from other accomplishments that do not allow NI.

Fifth, members of a frame typically belong to the same aspectual class. While FrameNet does not mark lexical units for Vendlerian aspectual class, such classes could be derived, for instance, from the verbal Frame Description or from information about sub-frame structure. For instance, lexical units like *die*, *pass away* in the Death frame involve the transition to a final state of being dead. The lexical units *prepare* and *get ready* mentioned above in the Activity_prepare frame are activities, while the lexical units in the Activity_done_state frame (*through*, *finished*, *done*) are states.

Sixth, the effects of selectional strength are captured indirectly by differences in frame membership. The prototypical example in most discussions of selectional strength is the contrast between the verb *eat* and verbs such as *devour* or *guzzle*. While the former allows indefinite null instantiation of its object, the latter two do not

allow any null instantiation. Under the selectional strength account, this difference reflects the fact that meal-type objects of *eat* are particularly strongly selected for. By comparison, *guzzle* and *devour* do not call any particular kind of object to mind very strongly. In a frame semantic analysis, these facts would be analyzed in a similar way to what Kegl and Fellbaum 1989 suggested: there is a sense of *eat* where it means 'eat a meal' and belongs in a frame together with *lunch*, *breakfast*, *dine*, *feast*, *snack*, *sup*. None of the verbs in that frame require a food to be mentioned specifically. In another sense, *eat* belongs to a Food ingestion frame that contains words like *nibble*, *devour*, *gobble*, *swig*, *swill*, which do not allow null instantiation of their object in particular contexts.

Seventh, frame identity also is responsible for the findings concerning argument definiteness: the uses with overt arguments and the null instantiating ones belong to the same frame. We expect lexical units in a frame to have similar definiteness profiles given that they talk about very similar kinds of events and participants and are likely to be used in similar ways in discourse.

Finally, exploring null instantiation on the basis of frames is better practice than trying to find a common account for semantically dissimilar words that may not share anything except aspectual class, if that, and the interpretation under omission of their object. Regularities can be more easily detected and semantic differences between predicates become clearer when explored against the background of frames. Frames can and should also serve as a background for cross-linguistic study. If the regularities and sub-regularities that can be found for English also exist in other languages we could show that lexically unrestricted argument omission is not as idiosyncratic and haphazard as earlier studies sometimes made it appear. Some preliminary comparisons between English and German look promising. Consider the semantic frame of Kidnapping, where a Perpetrator seizes and takes control of a Victim and holds him or her against his will. In English, the lexical units in the frame include *abduct.v*, *abducted.a*, *abduction.n*, *abductor.n*, *kidnap.v*, *kidnapped.a*, *kidnapper.n*, *kidnapping.n*, *nab.v*, *shanghai.v*, *snatch.v*, and *snatcher.n*. In German, the relevant lexical units are *entführen*, *entführt.a*, *gekidnappt.a*, *kidnappen*, *sich greifen*, *Entführung.n*, *Kidnapping.n*. Like their English counterparts, the German lexical units cannot omit

reference to the *Victim*; null instantiation is not possible.⁴² The findings are the same for the other nine frames discussed in Appendix B.

4.4.3 Lexical pragmatics

In this chapter we have noted on several occasions that omission and pragmatics interact, both for constructional and lexical omission types. We found, for instance, that:

- Omitted arguments can never be focal.
- For definite omissions, it is the case that with some of the genre-related constructions that do not involve a change of linking, the referents seem to have to be discourse topics; with other constructional omission types and lexical DNI, that is not necessary.

But in addition to these discourse-pragmatic connections, there is another kind of pragmatic influence. If we assume that the basic job of semantics is to tell us who is doing what to whom and against which background, then we still lack an explanation how it is that speakers are sometimes happy with not having all the details filled in. This particular problem of course only affects lexically licensed indefinite null instantiation since with definite omissions the hearer does have to fill in the missing details from context. With indefinite omissions such as (4.270), however, speaker and hearer simply pass over the fact that some details are not spelled out.

(4.270) What's Kim doing?—**Cooking** ∅.

This fact has of course been noted before, as the earlier quote by Allerton 1975 shows (cf. 408). However, I think there is a subtle distinction to be made between interpreting null instantiation as lack of interest in the identity of a specific participant and thinking about it as a situation in which the parts of the full scenario that are specified are by themselves informative enough to be communicated. I believe that

⁴²As we saw earlier, nominalizations will behave differently: *My neighbor was arrested for kidnapping.*

a pragmatic notion of informativeness has a role to play in argument (and modifier) realization in addition to frame semantic and to syntactic considerations.

Goldberg & Ackerman have argued for the relevance of such a notion in the case of so-called obligatory adjuncts. Consider the contrast between sentences (4.271a) and (4.271b).

- (4.271) a. #This house was **built**.
 b. The man was **deceived**.

Absent any context, (4.271a) seems to be infelicitous whereas the structurally identical clause in (4.271b) is perfectly acceptable. However, the clause in (4.271a) can be 'saved' by expanding it with any of the phrases given in (4.272).

- (4.272) This house was **built** {in 1815/with federal money/according to a plan drawn up by Charles E. Pfeiffer}.

The reason for the contrast between (4.271a) and (4.272) cannot lie in argument omission: the prepositional phrases are adjuncts. As Goldberg & Ackerman point out, (4.271a) is grammatical but outside of very specialized contexts that do not readily come to mind it is not informative. An example of a context where (4.271a) by itself would be appropriate is given in (4.273).

- (4.273) A: Look at these two houses: one of them was wished onto the property by a sorcerer, the other one was built the old fashioned way. Can you tell me which is which?
 B: (pointing at one house) **THIS** house was **built**.

Goldberg & Ackerman have characterized the notion of informativeness that is at issue here in terms of Grice's maxim of quantity or Horn's R-principle ('make your contribution necessary; say no more than you must'). Utterances that assert no more than what their lexico-grammatical parts presuppose are not felicitous—except in rare circumstances where, for instance, one of the presuppositions has been explicitly questioned or contradicted (cf. (4.273)). In the case of examples like (4.271a),

the existential presupposition of the definite article and the fact that a house is an artefact and therefore has to have been created, in particular built, add up to all the same information that the assertion would impart. Cases of obligatory adjuncts thus demonstrate that licensing a valid syntactic construct which encodes all the participants that are canonically part of the evoked frame is not necessarily enough to produce a felicitous utterance. The sentence also needs to be informative within its discourse context. Achieving discourse felicity may require expressing adjuncts.

How does Goldberg & Ackerman's analysis relate to indefinite null instantiation? We might think that it explains why certain verbs allow indefinite null instantiation only when these uses by convention have narrower selectional restrictions than the uses with overt NPs, as is the case with *drink*.

(4.274) Chief Wiggum has been **drinking** \emptyset for years.

Sentence (4.274) cannot just mean that Chief Wiggum has been drinking liquids for years because that much can be inferred just from the ontological properties of the referents: humans need to drink liquids to survive. But using (4.274) to report that somebody has been having an alcohol problems for years is informative as this fact cannot be inferred from lexico-grammatically cued properties of the subject referent. However, consider now examples (4.275) and (4.276).

(4.275) A: Can't John help us?—B: He just started **making** \emptyset

(4.276) A: Can't John help us?—B: He just started **making** \emptyset with his bare hands.

The informativeness account as stated might not explain why neither (4.275) nor (4.276) are acceptable. Unlike, say, breathing, making things is not something that people do all the time and reporting the fact that somebody is creating something ought to be informative. But even if B's utterances in (4.275) and (4.276) are only intended in context to explain why John has no time to help, they are still not felicitous.

It seems then that not only is external informativeness relative to the discourse required—i.e. that the relation between the frame and the overtly referred-to participants not be one that can be easily inferred or has been stated before—but also

internal informativeness. The framal scenario itself needs to be fleshed out to a sufficient degree to allow the hearer to form a fairly specific idea of what is going on in the scene. Clearly, *make* is a much more general verb than say *cook* or *knit*, both of which can be used without an object. In general, it seems safe to predict that very general predicates such as *create*, *do*, *make*, *say* cannot allow indefinite null instantiation. Of course, if a more specific interpretation is given to the argument when it is null instantiated, as is the case for German *bauen*, which can be used to mean ‘build a building’, then null instantiating uses will occur with verbs whose uses with overt NPs cover a large number of specific scenarios. This explanation could also be taken to apply to the earlier example of *drink*.

Clearly, what I called internal and external informativeness cannot be wholly separated. The inferences that are needed to reason about how an utterance asserting a state or event fits into the ongoing discourse—whether it is really saying something—crucially depend on what specifically is going on. Interestingly, there seem to be verbs about which one cannot readily reason only in terms of internal or external informativeness. Consider the sentences (4.277a-f), which illustrate the German verb *wohnen* ‘live, dwell, be housed’.⁴³

- (4.277) a. Sie **wohnt** in Owingen.
‘She lives in Owingen’
- b. Sie **wohnt** bei Katharina Dören.
‘She lives with Katarina Dören’
- c. Die Familie Schilling **wohnt** in einem Einfamilienhaus.
The Schilling family lives in a one-family home.
- d. Er **wohnte** luxuriös und hatte einen Hausmeister und einen Chauffeur.
‘He lived in luxury and had a caretaker and a chauffeur’
- e. Und ich muß sagen, er **wohnt** schön :-).
‘And I have to say: he lives nicely’

⁴³The basic argument also applies to English *live* and generally to verbs like *reside* and *dwell*. I use German *wohnen* here because it lacks the polysemy with a ‘be alive’ sense that English *live* exhibits.

f. Man isst teuer, kauft teuer ein und **wohnt** teuer. Wer hierher zieht, muss schon mehr als ordentlich verdienen.

You eat expensively, shop expensively, and live expensively. Those who move here have to be making more than decent money.

It is hard to imagine a use of *wohnen* with only the subject denoting the Resident and no other complement or modifier. *Wohnen* is not the kind of activity that is easily pictured. It is also by itself not particularly newsworthy: nobody ever reports that they spent the week-end 'living'. A simple clause with only the Resident cannot even be used to deny that somebody is homeless.

(4.278) A: Ist der Mann obdachlos?—B: #Nein, er **wohnt**.

'A: Is that man homeless.—B: No, he lives.'

This fact makes *wohnen* appear different from obligatory adjunct cases, involving external informativeness, where the basic clause can be used felicitously if just the right context can be constructed.⁴⁴ On the other hand, it seems to be the case *wohnen* does not require a particular kind of frame element to be realized. Examples (4.277a-f) show that specifications of municipalities, co-residents, types of building, characterizations of the interior design of the residence, characterizations of the setting of the residence, and characterizations of the cost of housing all are able to make the proposition of a clause headed by *wohnen* a worthwhile thing to assert. Unless one posits a macro-role, say Housing Circumstances, for these different notions, it seems that *wohnen* behaves like an obligatory adjunct verb in not being picky about how the basic Residence scenario that is evoked is fleshed out.

The notion of informativeness discussed here is, of course, meant to capture the same kinds of observations that underlie the selectional strength account and

⁴⁴There is one playful such use attested by the satirist Gerhard Polt.

- (i) Wenn die Mieten weiter so steigen, dann **wohne** ich nicht mehr.
'If rents keep rising like this, I won't dwell anymore'.

The understanding here is slightly different from ordinary stative *wohnen*: (i) suggests that *wohnen* is a habitual activity like smoking that can be discontinued.

Rice's observation that "neither extremely schematic nor extremely specific verb-complement pairs encourage object omission". The notion of informativeness is, however, subtly different from those two. First, whereas these other accounts focus on the missing object, the informativeness account focuses on what can be understood from the what is overtly specified. Second, the notion of informativeness is clearly not taken to be a predictive factor; it may be a necessary factor for allowing INI but it is not a sufficient factor. Allowing or not allowing INI is clearly a lexical fact.

4.4.4 Possible complications

There are at least two phenomena that are possibly problematic for the above analysis. The first one is incorporation. Compare sentences (4.279) and (4.280).

(4.279) John **put** the books *(on the living room table).

(4.280) We **bagged** the groceries \emptyset and left.

Whereas the verb *put* in (4.279) needs to express all three of its arguments, the verb *bag* can omit specification of a goal location.⁴⁵ All that is known about the goal of *bag* is that it is of a certain type, namely of type bag. Thus, incorporation can change the realization requirements for a particular frame element/argument even though the we would prefer for the basic frame membership not to change and even though the aspectual properties are no different from similar verbs that have not incorporated that same argument. Still we can note that cases of incorporation never lead to a situation where for some lexical units of the frame the argument is incorporated and interpreted as an indefinite omission whereas for other lexical units that same argument is omitted under definite interpretation.

The second complication for the analysis presented above is argument or frame element conflation, that is, the expression of information about more than one frame element in a single constituent.

⁴⁵ *Put* does in fact have one specific use where it can occur without a Goal; when talking about a dish being prepared, one can for instance say, *You should put more salt*. However, this fact clearly needs to be recorded separately from the entry for the general Placing sense of *put*.

(4.281) The court **acquitted** a man who used cannabis strictly on medical grounds.

(4.282) The treatment **cured** my colitis.

In example (4.281), for instance, the relative clause gives the charges on which the defendant was tried. In example (4.282), the possessive determiner indicates who the patient was that was treated. If the object NPs had just been *John Doe* and *the colitis*, we would have to assume that the CHARGES and the PATIENT are given in the prior linguistic contexts of the two sentences.

It is possible to treat cases of conflation as cases of null instantiation where the linguistic material that provides the missing referent just happens to be part of the same clause as the null instantiated element. This is the analysis adopted, for instance, by Michaelis & Ruppenhofer 2001 for Theme-incorporating German *be*-prefixed verbs evoking Filling scenarios (e.g., *bereifen* ‘put tires on a vehicle’). Under that treatment, the analysis of null instantiation given above remains unchanged. This treatment is plausible in so far as the inference that information about a second frame element is expressed in the relevant constituent can, in principle, be defeated. In (4.283), which is a variation on (4.281), the Charge that the cannabis user is acquitted of is trespassing rather than cannabis use.

(4.283) Seven of the protesters were convicted for illegal trespassing. The court **acquitted** a medical marijuana user, holding that he had a right to protest on FDA property.

If, despite the defeasibility facts, one nevertheless chooses to consider conflation a phenomenon *sui generis*, then one might claim that a single constituent fills both frame element roles at the same time and that there is in fact no null instantiation at all. Either way, it seems that the generalizations we made about null instantiation in terms of frames do hold up.

4.4.5 Summary

So far accounts of null instantiation have been of two types. The first type of analysis concerned itself either with the analysis of disparate sets of predicates that shared

nothing but the type of interpretation of certain omitted arguments. The second type inspired by a trend to study the interaction of aspect and argument linking looked at aspectual classes as the locus of generalizations. I have suggested here that a more fruitful account consists of investigating frames and frame elements. This strategy is not only methodologically beneficial but also leads to an account for the lexical units of all major lexical categories at the same times and it yields very reliable generalizations. Not all of these generalizations are parochial to individual frames, as the discussion of trigger/response frames such as Revenge or Communication_response showed. However, the degree to which generalizations can become larger will depend on whether the data and frames are descriptively adequate even if larger generalizations over the hierarchy of frames cannot be made. Finally, frames tend to capture the observed regularities with respect to single factors such as lexical aspect, definiteness of a certain argument, and selectional strength without relying on any of these factors as the sole criterion for whether a certain argument will be omissible and under which interpretation. And most importantly, the evidence we have considered here suggests that null instantiation is not primarily dependent on aspects of language processing but rather on the semantics and micro-pragmatics of frames and lexical units.

Chapter 5

Conclusion

In this thesis I suggested that it was methodologically sound and analytically fruitful to apply discourse-pragmatic methods developed in the study of sentence-type constructions and the distribution of NP form-types to the study of argument structure alternations.

The empirical work reported here shows that some of our intuitive pragmatic theories are simply not borne out, while others are. Specifically, the analysis of three members of the family of raising alternations that I investigated suggested that these do not all pattern in the same way, with only the change property factoring alternation providing clear consistent evidence for a role of discourse-pragmatics in the constructional choice.

The research on the dative alternation further showed that the kinds of pragmatic factors that can be relevant to argument structure constructions at least includes pragmatic relation and discourse status. The data on the change property factoring alternation also suggest a role relative salience as measured by the number and closeness of contextual mentions. A factor that was not found to exert an influence in the data examined here was activation source, that is, whether a referent had been textually evoked or was situationally given. It remains to be seen whether there are argument structure constructions that are sensitive to this factor.

Another interesting finding observed, for instance, in the case of the change property factoring alternation and in the dative alternation is that in some cases the

pragmatic properties of the overall less prominent argument vary more than those of the more prominent one. The differences between the properties of Attributes in the factoring and non-factoring constructions are far more significant than the differences between the properties of Possessors. Similarly, Themes differ more across constructions in the dative alternation than Recipients. The role of the less prominent argument seems to me to have been somewhat underestimated in the literature. Thompson (1990) compares only Recipients across the double object and caused-motion constructions but omits such a comparison for Themes.

The use of discourse-pragmatic methodology not only helps us study the usage conditions of alternates but it can also make contributions to the semantic-syntactic understanding of the constructions themselves. This was argued in the case of the dative alternation where it was suggested that Basilico's (1998) account of certain asymmetries between the alternates in terms of differences between lexical representations that were said to be associated with distinct embedded focus structures does not provide a complete account. I argued that an analysis in terms of the topic-pragmatic relation for Recipients in the double object construction might be more successful.

The research carried out here also suggests that argument-structure to discourse-status mappings are constructionally relative and that an adequate account needs constructions that capture constraints on particular thematic role to grammatical-function mappings. In particular, the Recipients in the double object construction seem to require topic status and are strongly adverse to being focal, neither of which is property of objects in general. Lexical null instantiation phenomena, both lexical and constructional, also showed the need to incorporate specifications of discourse status into constructions and lexical entries.

As I hope this work has shown, careful corpus study of argument structure constructions also has methodological benefits. Rare patterns that are not readily accessible to introspection are brought to the analyst's attention. Combined with statistical modelling, corpus-based work also keeps the analyst honest as they are confronted with the type and number of tokens that their explanations can account for.

Finally, based on the results found in this work, it seems clear that other argument structure constructions also deserve to be investigated relative to discourse-pragmatic properties. Prominent candidates are the members of *spray/load*-alternation. Another alternation that is less well researched is the mental property alteration exemplified in (5.1).

- (5.1) a. *You* were **smart** to sell at that price.
 b. To sell at that price was **smart** *of you*.
 c. It was **smart** *of you* to sell at that price.

One of the most interesting cases to be looked into is the family of resultative argument constructions, which includes the *Verb X into NP/VPing* construction with which Basilico's transformation predicates combine. Basilico treats transformation predicates as the categorical members of the creation/transformation. We need not agree with his analysis of the alternate arguments structures in terms of embedded focus structures, but the intuition that underlies it seems to be sound. From a functional point of view we may note that the resultative construction contains what we may call a 'linchpin' argument. On the one hand the object position is one of the places where new referents get introduced (Du Bois 1987). On the other hand, Lambrecht's Principle of the Separation of Reference and Role requires that a referent not be introduced and predicated over in the same clause. Because of this tension, we should expect that the 'host' of the secondary predicate would have to strike a balance between two statuses: the referent should be uniquely identifiable, but not evoked. Future research will test whether this prediction is borne out.

Bibliography

- ACHARD, MICHEL. 2000. The Distribution of French Raising Constructions. In *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, ed. by Lisa J. Conathan, Jeff Good, Darya Kavitskaya, Alyssa B. Wulf, and Alan C.L. Yu, volume 26, 1–16, Berkeley, CA. Berkeley Linguistics Society.
- ACKERMAN, FARRELL, and JOHN MOORE. 1994. Valence and the semantics of causatives. *Proceedings of BLS* 20.1–13.
- AISSEN, JUDITH, 2000. Differential object marking: Iconicity vs. economy. unpublished manuscript.
- ALLERTON, D.J. 1975. Deletion and proform reduction. *Journal of Linguistics* 11.213–237.
- ARIEL, MIRA. 1988. Referring and Accessibility. *Journal of Linguistics* 24.65–87.
- ARIEL, MIRA. 1990. *Accessing Noun Phrase Antecedents*. London, UK: Routledge.
- ARIEL, MIRA. 1996. Referring Expressions and the +/- Coreference Distinction. In (Fretheim and Gundel 1996), 13–36.
- ARNOLD, JENNIFER E., THOMAS WASOW, ANTHONY LOSONGCO, and RYAN GINSTROM. 2000. Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language* 76.28–55.
- ASKE, JON. 1989. Path predicates in english and spanish: A closer look. *Proceedings of BLS* 15.1–14.

- AYGEN-TOSUN, GULSAT. 1999. Specificity and subject-object positions/scope interactions in Turkish. In *Proceedings of the 1 International Conference in Turkish Linguistics*, Manchester. Manchester University.
- BENDER, EMILY. 1999. Constituting Context: Null Objects in English Recipes Revisited. *Penn Working Papers in Linguistics* 6.53–68.
- BIBER, DOUGLAS. 1995. *Dimensions of register variation: a cross-linguistic comparison*. Cambridge, UK: Cambridge University Press.
- BIBER, DOUGLAS, STIG JOHANSSON, GEOFFREY LEECH, SUSAN CONRAD, and EDWARD FINEGAN. 1999. *The Longman grammar of spoken and written English*. London: Longman.
- BIRNER, BETTY. 1997. The linguistic realization of inferrable information. *Language and Communication* 17.133–147.
- BIRNER, BETTY J., and GREGORY WARD. 1998. *Information Status and Non-canonical Word Order in English*. Amsterdam/Philadelphia: John Benjamins.
- BOAS, HANS C., 2000. *Resultative Constructions in English and German*. Chapel Hill: University of North Carolina at Chapel Hill dissertation.
- BORKIN, ANN. 1973. *To be and not to be*. *CLS* 9.44–56.
- BOSCH, PETER. 1988. Representing and accessing focussed referents. *Journal of Semantics* 3.207–231.
- BRISSON, CHRISTINE. 1994. The Licensing of Unexpressed Objects in English Verbs. *CLS* 30.90–102.
- BROWN, CHERYL. 1983. Topic Continuity in Written English Narrative. In (Givón 1983a), chapter 7, 315–341.
- BYBEE, JOAN. 1985. *Morphology: A Study in the Relationship between Meaning and Form*. Philadelphia: Benjamins.

- BYBEE, JOAN. 2001. Frequency effects on french liaison. In *Frequency and the Emergence of Linguistic Structure*, ed. by Joan Bybee and Paul Hopper, Typological Studies in Language, 337–360. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- BYBEE, JOAN, and PAUL HOPPER (eds.) 2001. *Frequency and the Emergence of Linguistic Structure*. Typological Studies in Language. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- BYBEE, JOAN, and SANDRA THOMPSON. 2000. Three Frequency Effects in Syntax. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 23.378–388.
- CARLSON, GREGORY N., and FRANCIS JEFFRY PELLETIER. 1995. The generic book. In *The Generic Book*, ed. by Gregory N. Carlson and Francis Jeffrey Pelletier. Chicago: The University of Chicago Press.
- CHAFE, WALLACE. 1976. Givenness, Contrastiveness, Definiteness, Subjects, Topics, and Point of View. In *Subject and Topic*, ed. by Charles Li, 25–55. New York: Academic Press.
- CHAFE, WALLACE. 1987. Cognitive constraints on information flow. In *Coherence and Grounding in Discourse*, ed. by Russell Tomlin, 21–51. Amsterdam: John Benjamins.
- CLANCY, PATRICIA M. 1980. Referential Choice in English and Japanese. In *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production*, ed. by Wallace L. Chafe, volume III of *Advances in Discourse Processes*, chapter 4, 127–202. Norwood, New Jersey: Ablex Publishing Corporation.
- CLARK, HERBERT H. 1996. *Using language*. Cambridge, UK: Cambridge University Press.
- CLARK, HERBERT H. 1997. Dogmas of understanding. *Discourse Processes* 23.567–598.

- CLARK, HERBERT H., and SUSAN E. HAVILAND. 1977. Comprehension and the given-new contract. In *Discourse-Production and Comprehension*, ed. by Roy O. Freedle, 1–40. Norwood, NJ: Ablex.
- CONDORAVDI, CLEO. 1989. Indefinite and generic pronouns. In *Proceedings of the Eighth West Coast Conference on Formal Linguistics*, ed. by E. Jane Fee and K. Hunt, 71–84, Stanford, CA. CSI.
- COOREMAN, ANN. 1983. Topic continuity and the voicing system of an ergative language: Chamorro. In (Givón 1983a), chapter 10, 425–490.
- COOREMAN, ANN, BARBARA FOX, and TALMY GIVÓN. 1988. The Discourse Definition of Ergativity. *Studies in Language* 8.1–34.
- CULY, CHRISTOPHER. 1996. Null object in English recipes. *Language Variation and Change* 8.91–124.
- DAHL, ÖSTEN. 1995. The marking of the episodic/generic distinction in tense-aspect systems. In (Carlson and Pelletier 1995), 412–425.
- DAHL, ÖSTEN, and KARI FRAURUD. 1996. Animacy in grammar and discourse. In (Fretheim and Gundel 1996), 47–64.
- DELANCEY, SCOTT. 1981. An Interpretation of Split Ergativity and Related Phenomena. *Language* 57.627–657.
- DIXON, R.M.W. 1994. *Ergativity*. Cambridge, UK: Cambridge University Press.
- DIXON, R.M.W. 2000. A typology of causatives: form, syntax and meaning. In (Dixon and Aikhenvald 2000), chapter 2, 30–83.
- DIXON, R.M.W., and ALEXANDRA AIKHENVALD. 2000. Changing valency: case studies in transitivity. In *Changing valency: case studies in transitivity*, ed. by R.M.W. Dixon and Alexandra Aikhenvald. Cambridge, UK: Cambridge University Press.

- DIXON, R.M.W., and ALEXANDRA Y. AIKHENVALD. 1997. A typology of argument-determined constructions. In *Essays on language function and language type*, 71–113. Amsterdam: John Benjamins.
- DOWTY, DAVID. 1979. *Word Meaning and Montague Grammar*. Dordrecht: Reidel.
- DOWTY, DAVID. 1991. Thematic Proto-roles and Argument Selection. *Language* 67.547–619.
- DRUBIG, BERNHARD. 1997. Some cross-categorical generalizations of focus structure. Arbeitspapiere des sonderforschungsbereiches 340, Universität Tübingen, Tübingen.
- DRUBIG, BERNHARD, 2000. Towards a typology of focus and focus constructions. Unpublished manuscript, available at <http://www.sfb441.uni-tuebingen.de/b2/b2pubs.html>.
- DU BOIS, JOHN W. 1980. Beyond Definiteness: The Trace of Identity in Discourse. In *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production*, ed. by Wallace L. Chafe, volume III of *Advances in Discourse Processes*, chapter 5, 203–274. Norwood, New Jersey: Ablex Publishing Corporation.
- DU BOIS, JOHN W. 1987. The discourse basis of ergativity. *Language* 63.805–55.
- ENGDAHL, ELISABET, and ENRIC VALLDUVÍ. 1996. Information packaging in HPSG. Working Papers in Cognitive Science 12, University of Edinburgh.
- ENGELBERG, STEFAN. 2000. *Verben, Ereignisse und das Lexikon*. Number 414 in *Linguistische Arbeiten*. Tübingen: Max Niemayer Verlag.
- ERTESCHIK-SHIR, NOMI. 1979. Discourse constraints on dative movement. In *Discourse and Syntax*, ed. by Talmy Givón, volume 12 of *Syntax and Semantics*, 441–467. New York, NY: Academic Press.
- FAUCONNIER, GILLES. 1985. *Mental Spaces*. Chicago: University of Chicago Press.

- FAUCONNIER, GILLES. 1994. *Mental Spaces: Aspects of Meaning Construction in Natural Language*. Cambridge: Cambridge University Press.
- FAUCONNIER, GILLES. 1997. *Mappings in Thought and Language*. Cambridge, UK: Cambridge University Press.
- FAUCONNIER, GILLES, and EVE SWEETSER. 1996. Spaces, worlds, and grammar. In *Spaces, Worlds, and Grammar*, ed. by Gilles Fauconnier and Eve Sweetser, Cognitive Theory of Language and Culture. Chicago: Chicago University Press.
- FELLBAUM, CHRISTIANE, and JUDY KEGL. 1989. Taxonomic Structures And Cross-Category Linking in the Lexicon. In *Proceedings of the Sixth Eastern States Conference on Linguistics*, ed. by Kenneth de Jong and Yongkyoon No, 93–104, Columbus, Ohio. Ohio State University.
- FIELD, ANDY. 2000. *Discovering Statistics using SPSS for Windows*. London, UK: Sage Publications.
- FILLMORE, CHARLES J. 1968. The Case for Case. In *Universals in Linguistic Theory*, ed. by E. Bach and R. Harms, 1–88. New York: Holt, Rinehart & Winston.
- FILLMORE, CHARLES J. 1970. The grammar of *Hitting* and *Breaking*. In *Readings in transformational grammar*, ed. by R. Jacobs and P. Rosenbaum, 120–133. Waltham, MA: Ginn & Co.
- FILLMORE, CHARLES J. 1986. Pragmatically Controlled Zero Anaphora. *Berkeley Linguistics Society* 12.95–107.
- FOLEY, WILLIAM, and ROBERT VAN VALIN. 1984. *Functional Syntax and Universal Grammar*. Cambridge: Cambridge University Press.
- FRAURUD, KARI. 1990. Definiteness and the processing of NPs in natural discourse. *Journal of Semantics* 7.395–434.
- FRAURUD, KARI. 1996. Cognitive Ontology and NP Form. In (Fretheim and Gundel 1996), 65–87.

- FRETHEIM, THORSTEIN, and JEANNETTE K. GUNDEL. 1996. Reference and referent accessibility. In *Reference and referent accessibility*, ed. by Thorstein Fretheim and Jeannette K. Gundel, Pragmatics and Beyond. Amsterdam/Philadelphia: John Benjamins.
- GEORG BOSSONG. 1985. *Differentielle Objektmarkierung in den neuiranischen Sprachen*. Empirische Universalienforschung. Tübingen: Gunter Narr Verlag.
- GIVÓN, TALMY. 1979. *On Understanding Grammar*. Academic Press.
- GIVÓN, TALMY. 1983a. Topic continuity in discourse: A quantitative cross-language study. In *Topic Continuity in Discourse: A Quantitative Cross-Language Study*, ed. by Talmy Givón. Amsterdam: John Benjamins.
- GIVÓN, TALMY. 1983b. Topic continuity in discourse: An introduction. In (Givón 1983a), chapter 1, 5–41.
- GIVÓN, TALMY. 1983c. Topic continuity in spoken english. In (Givón 1983a), chapter 4, 347–363.
- GIVÓN, TALMY. 1984a. Direct object and dative shifting: semantic and pragmatic case. In *Objects: Towards a Theory of Grammatical Relations*, ed. by Frans Plank. London: Academic Press.
- GIVÓN, TALMY. 1984b. *Syntax: A Functional-Typological Introduction*. Amsterdam: John Benjamins.
- GIVÓN, TALMY. 1995. *Functionalism and Grammar*. Amsterdam/Philadelphia: John Benjamins.
- GOLDBERG, ADELE E. 1995. *A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- GOLDBERG, ADELE E. 1996. Making One's Way Through the Data. In (Masayoshi Shibatani and Sandra A. Thompson 1996), chapter 2, 29–54.

- GOLDBERG, ADELE E. 2001. Patient arguments of causative verbs can be omitted: the role of information structure in argument distribution. *Language Sciences* 23.503–524.
- GOLDBERG, ADELE E. 2002. Surface generalizations: An alternative to alternations. *Cognitive Linguistics* 13.327–356.
- GOLDBERG, ADELE E., and RAY JACKENDOFF. 2004. The English Resultative as a Family of Constructions. *Language* 80.532–568.
- GREGORY, MICHELLE L., and LAURA A. MICHAELIS. 2001. Topicalization and left-dislocation: A functional opposition revisited. *Journal of Pragmatics* 33.1665–1706. earlier on, I cited this with '00' in the name.
- GRETSCH, PETRA. 2003. Omission impossible? topic and focus in focal ellipsis. In *The Interfaces: Deriving and interpreting omitted structures*, ed. by Kerstin Schwabe and Susanne Winkler, *Linguistik Aktuell*, 341–365. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- GUNDEL, JEANNETTE K., NANCY HEDBERG, and RON ZACHARSKI. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69.274–307.
- HAEGEMAN, LILIANE. 1990. Understood subjects in english diaries: On the relevance of theoretical syntax for the study of register variation. *Multilingua* 9.157–199.
- HALE, KENNETH, and SAMUEL JAY KEYSER. 1993. On Argument Structure and the Lexical Expression of Syntactic Relations. In *The View from Building 20*, ed. by Kenneth Hale and Samuel Jay Keyser, *Current Studies in Linguistics* 24, 53–110. Cambridge, MA: MIT Press.
- HALLIDAY, MICHAEL A., and RUQAIYA HASAN. 1976. *Cohesion in English*. London: Longman.
- HÄRTL, HOLDEN. 2003. The conceptual inactiveness of implicit arguments: Evidence from particle verbs and object categorization. *Journal of Semantics* 20.1–33.

- HASPELMATH, MARTIN. 1990. The grammaticization of passive morphology. *Studies in Language* 14.73–92.
- HOBBS, J., D. APPELT, and P. MARTIN. 1993. Interpretation as abduction. *Artificial intelligence* 63.69–142.
- HOPPER, PAUL. 1987. Emergent grammar. *Proceedings of the Berkeley Linguistics Society* 13.139–57.
- HOPPER, PAUL. 1988. Emergent grammar and the *a priori* principle. In *Linguistics in context: Connecting observation and understanding*, ed. by Deborah Tannen, 117–134. Norwood, N.J.: Ablex.
- HOVAV, MALKA RAPPAPORT, and BETH LEVIN, 1999. Two types of compositionally derived events. <http://www-csli.stanford.edu/beth/pubs.html>.
- HUDDLESTON, RODNEY, and GEOFFREY K. PULLUM. 2002. *The Cambridge Grammar of the English Language*. Cambridge, UK: Cambridge University Press.
- ISRAEL, MICHAEL. 1996. The way constructions grow. In *Conceptual Structure, Discourse, and Language*, ed. by Adele E. Goldberg, 217–230. Stanford: CSLI.
- JACOBS, JOACHIM. 1994. Das lexikalische fundament der unterscheidung von obligatorischen und fakultativen ergänzungen. *Zeitschrift für Germanistische Linguistik* 22.284–319.
- KAMP, HANS. 1981. A theory of truth and semantic representation. In *Formal methods in the study of language*, ed. by Jeroen Groenendijk, Theo Janssen, and Martin Stokhof, 277–322. Amsterdam: Mathematisch Centrum.
- KAY, PAUL, Manuscript. Argument Structure Constructions and the Argument-Adjunct Distinction. <http://www.icsi.berkeley.edu/~kay/>.
- KAY, PAUL, and CHARLES J. FILLMORE. 1999. Grammatical constructions and linguistic generalizations: the what's x doing y? construction. *Language* 75.1–33.

- KOENIG, JEAN-PIERRE, and ANTHONY DAVIS. 2001. Linking from the outside in. In *Proceedings of the Third Annual Texas Linguistic Forum: Perspectives on Argument Structure*. Texas Linguistics Society.
- KOENIG, JEAN-PIERRE, and GAIL MAUNER. 1999. A-definites and the discourse status of implicit arguments. *Journal of Semantics* 16.207–236.
- KOENIG, JEAN-PIERRE, GAIL MAUNER, and BRETON BIENVENUE. 2002. Class selectivity and the lexical encoding of participant information. *Brain and Language* 81.234–245.
- KRIFKA, MANFRED, FRANCIS JEFFRY PELLETIER, GREGORY N. CARLSON, ALICE TER MEULEN, GODEHARD LINK, and GENNARO CHIERCHIA. 1995. Genericity: An introduction. In (Carlson and Pelletier 1995), 1–124.
- LAMBRECHT, KNUD. 1994. *Information structure and sentence form*. New York, N.Y.: Cambridge University Press.
- LAMBRECHT, KNUD. 1995. The pragmatics of case. In *Essays in Semantics and Pragmatics*, ed. by Masayoshi Shibatani and Sandra A. Thompson, 145–190. Amsterdam/Philadelphia: John Benjamins.
- LAMBRECHT, KNUD. 2001. A framework for the analysis of cleft constructions. *Linguistics* 39.463–516.
- LAMBRECHT, KNUD, and MARIA POLINSKY. 1997. Typological variation in sentence-focus constructions. In *Proceedings of the Chicago Linguistic Society*, volume 33, 189–206, Chicago. Chicago Linguistic Society.
- LANGACKER, RONALD W. 1987. *Foundations of Cognitive Grammar*, volume I: Theoretical Prerequisites. Stanford: Stanford University Press.
- LANGACKER, RONALD W. 1991. *Foundations of Cognitive Grammar*, volume II: Descriptive Application. Stanford, CA: Stanford University Press.
- LANGACKER, RONALD W. 1995. Raising and transparency. *Language* 71.1–62.

- LANGACKER, RONALD W. 2000. *Grammar and Conceptualization*. Cognitive Linguistics Research. Berlin: Mouton de Gruyter.
- LARSON, RICHARD K. 1988. On the double object construction. *Linguistic Inquiry* 335–391.
- LEVIN, BETH. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.
- LEVIN, BETH. 2000. Aspect, lexical semantic representation, and argument expression. *Proceedings of the Berkeley Linguistics Society* 25.413–429.
- LEVINSON, STEPHEN C. 1983. *Pragmatics*. Cambridge, UK: Cambridge University Press.
- LIPPMANN, LAURA. 1997. *Baltimore Blues*. New York, NY: Avon Books.
- LÖBNER, SEBASTIAN. 1985. Definites. *Journal of Semantics* 4.279–326.
- LYONS, CHRISTOPHER. 1999. *Definiteness*. Cambridge, UK: Cambridge University Press.
- MASAYHOSHI SHIBATANI AND SANDRA A. THOMPSON. 1996. Grammatical constructions. In *Grammatical Constructions*, ed. by Masayhoshi Shibatani and Sandra A. Thompson. Oxford, UK: Oxford University Press.
- MAUNER, GAIL, and JEAN-PIERRE KOENIG. 2000. Linguistic versus conceptual sources of implicit agents in sentence comprehension. *Journal of Memory and Language* 43.110–134.
- MELNIK, NURIT, 2002. *Verb-initial Constructions in Modern Hebrew*. Berkeley: University of California dissertation.
- MICHAELIS, LAURA A. 2004. Type shifting in construction grammar: An integrated approach to aspectual coercion. *Cognitive Linguistics* 15.1–67.

- MICHAELIS, LAURA A., and HARTWELL S. FRANCIS. in press. Lexical subjects and the conflation strategy. In *Topics in the Grammar-Pragmatics Interface: Papers in Honor of Jeanette K. Gundel*, ed. by Nancy Hedberg and Ron Zacharski. Amsterdam: John Benjamins Publishing Company.
- MICHAELIS, LAURA A., and KNUD LAMBRECHT. 1996. Toward a Construction-Based Model of Language Use: The Case of Nominal Extraposition. *Language* 72.215–247.
- MICHAELIS, LAURA A., and JOSEF RUPPENHOFER. 2001. *Beyond Alternations: A Constructional Model of the German Applicative Pattern*. Stanford Monographs in Linguistics. Stanford, CA: CSLI.
- MITHUN, MARIANNE. 2000. Valency changing derivation in central alaskan yup'ik. In (Dixon and Aikhenvald 2000), chapter 3, 84–114.
- MITTWOCH, ANITA. 1982. On the difference between eating and eating something: Activities versus accomplishments. *Linguistic Inquiry* 13.113–22.
- NÉMETH T., ENIKŐ. 2000. Occurrence and identification of implicit arguments in Hungarian. *Journal of Pragmatics* 32.1657–82.
- O'CONNOR, MARY CATHERINE. 1996. The Situated Interpretation of Possessor Raising. In (Masayoshi Shibatani and Sandra A. Thompson 1996), chapter 6, 125–156.
- OLSEN, MARI BROMAN, and PHILIP RESNIK. 1997. Implicit Object Constructions and the (In)transitivity Continuum. *CLS* 33.327–336.
- ONO, TSUYOSHI, and SANDRA A. THOMPSON. 2000. Deconstructing "Zero Anaphora" in Japanese. *BLS* 23.481–491.
- OOSTEN, JEANNE VAN. 1986. *The nature of subjects, Topics, and Agents: A Cognitive Explanation*. Bloomington, IN: Indiana University Linguistics Club.

- POESIO, MASSIMO, and RENATA VIEIRA. 1998. A Corpus-based Investigation of Definite Description Use. *Computational Linguistics* 24.183–216.
- POLINSKY, MARIA. 1996. Situation perspective: On the relations of thematic roles, discourse categories, and grammatical relations to figure and ground. In *Conceptual Structure, Discourse and Language*, ed. by Adele E. Goldberg, 401–419. Stanford, CA: CSLI.
- POLINSKY, MARIA. 1998. A non-syntactic account of some asymmetries in the double object construction. In *Discourse and Cognition: Bridging the gap*, ed. by Jean-Pierre Koenig, 403–422. Stanford: CSLI.
- POLLARD, CARL, and IVAN A. SAG. 1994. *Head-Driven Phrase Structure Grammar*. Studies in Contemporary Linguistics. Chicago: University of Chicago Press.
- POSTAL, PAUL M. 1974. *On Raising: one rule of English grammar and its theoretical implications*. Cambridge, MA: MIT Press.
- PRINCE, ELLEN. 1981. Toward a taxonomy of given-new information. In *Radical Pragmatics*, ed. by Peter Cole, 223–56. Academic Press.
- PRINCE, ELLEN. 1992. The ZPG letter: Subjects, definiteness, and information-status. In *Discourse Description: Diverse Analyses of a Fundraising Text*, ed. by Sandra A. Thompson and William Mann, 295–325. John Benjamins.
- PRINCE, ELLEN. 1997. On the Functions of Left-Dislocation in English Discourse. In *Directions in functional linguistics*, ed. by Akio Kamio, 117–44. Philadelphia/Amsterdam: John Benjamins.
- RAPPAPORT HOVAV, MALKA, and BETH LEVIN. 1998. Building Verb Meanings. In *The Projection of Arguments: Lexical and Compositional Factors*, ed. by Miriam Butt and Wilhelm Geuder, 97–134. Stanford: CSLI.
- RESNIK, PHILIP, 1993. *Selection and information: A class-based approach to lexical relationships*. Philadelphia: University of Pennsylvania dissertation. <ftp://ftp.cis.upenn.edu/pub/ircs/tr/93-42.ps.Z>.

- RESNIK, PHILIP. 1996. Selectional constraints: An information-theoretic model and its computational realization. *Cognition* 61.127–59.
- RICE, SALLY. 1988. Unlikely Lexical Entries. *Proceedings of the Fourteenth Meeting of the Berkeley Linguistics Society* 14.202–212.
- ROLAND, DOUGLAS, and DANIEL JURAFSKY. 2002. Verb sense and verb sub-categorization probabilities. In *Lexical Basis of Sentence Processing: Formal, Computational, and Experimental Issues*, ed. by Suzanne Stevenson and Paola Merlo, 325–46, Amsterdam. John Benjamins.
- RUBBA, JOHANNA E. 1996. Alternate grounds in the interpretation of deictic expressions. In (Fauconnier and Sweetser 1996), 227–261.
- RUPPENHOFER, JOSEF, and COLLIN F. BAKER, 2003. The semantics and pragmatics of implicit arguments. Handout for paper presented at the 77th Annual Meeting of the Linguistic Society of America held in Atlanta, GA.
- SCHWENTER, SCOTT A., and GLÁUCIA SILVA. 2003. Anaphoric Direct Objects in Spoken Brazilian Portuguese: Semantics and Pragmatics. *Revista Internacional de Lingüística Iberoamericana* 2.99–123.
- SHESKIN, DAVID J. 2004. *Handbook of parametric and nonparametric statistical procedures*. Boca Raton: Chapman & Hall/CRC, 3rd edition.
- SIWIERSKA, ANNA. 1993. Syntactic weight versus information structure and word order variation in polish. *Journal of Linguistics* 29.233–65.
- SMITH, CARLOTA S. 1997. *The Parameter of Aspect*, volume 43 of *Studies in Linguistics and Philosophy*. Dordrecht: Kluwer Academic Publishers, 2nd edition.
- TALMY, LEONARD. 1996. The Windowing of Attention in Language. In *Grammatical Constructions*, ed. by Masayoshi Shibatani and Sandra A. Thompson, 235–288. Oxford, UK: Oxford University Press.

- TALMY, LEONARD. 2000. *Toward a Cognitive Semantics*, volume II: Typology and Process in Concept Structuring. Cambridge, MA: MIT Press.
- THOMPSON, SANDRA A. 1990. Information flow and dative shift in english discourse. In *Development and Diversity: Language Variation across Time and Space*, ed. by Jerold A. Edmondson, Crawford Feagin, and Peter Mühlhäusler, 239–253. Arlington, TX: Summer Institute of Linguistics.
- THOMPSON, SANDRA A. 1997. Discourse motivations for the core-oblique distinction as a language universal. In *Direction in functional linguistics*, ed. by Akio Kamio, 60–82. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- THOMPSON, SANDRA A., and PAUL J. HOPPER. 2001. Transitivity, clause structure, and argument structure: Evidence from conversation. In *Frequency and the Emergence of Linguistic Structure*, ed. by Joan Bybee and Paul Hopper, Typological Studies in Language, 27–60. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- TOOLE, JANINE. 1996. The effect of genre on referential choice. In (Fretheim and Gundel 1996), 263–290.
- VALIN, ROBERT D. VAN, and RANDY J. LAPOLLA. 1997. *Syntax: Structure, meaning and function*. Cambridge, UK: Cambridge University Press.
- VAN HOEK, KAREN. 1997. *Anaphora and Conceptual Structure*. Cognitive Theory of Language and Culture. Chicago: University of Chicago Press.
- WALKER, MARILYN A., and ELLEN F. PRINCE. 1996. A Bilateral Approach to Givenness: A Hearer-Status Algorithm and a Centering Algorithm. In (Fretheim and Gundel 1996), 291–306.
- WASOW, THOMAS. 1997. Remarks on grammatical weight. *Language Variation and Change* 9.81–105.
- WASOW, THOMAS. 2002. *Postverbal Behavior*. Stanford, CA: CSLI Publications.

WUNDERLICH, DIETER. 1997. Cause and the Structure of Verbs. *Linguistic Inquiry* 28.27-68.

ZIPF, GEORGE K. 1949. *Human Behavior and the Principle of Least Effort*. Cambridge, MA: Addison-Wesley.

Appendix A

Stativity and definiteness

In my discussion in section 4.4.1, I mentioned that stative verbs can be expected to give rise primarily to definite null instantiation, if they allow it at all. In section 4.3.3, I suggested that the interpretation of a null instantiated argument generally is the same as the majority of overt uses of that argument occurring with the verb in question. Combining these two observations, one might expect to find that the complements of stative verbs tend to be definite.

In order to test whether stative verbs or sentences are indeed clearly associated with definite NP objects, I took a random sample of 200 clauses from the British National Corpus. As Table A shows, 22.5% of all clauses were copular sentences with a predicate nominal, adjective or PP; 70% of the clauses were headed by a clearly verbal form; the remainder were modifying participial clauses.

Among the clauses headed by verbal forms were 59 active-voice clauses with a simple or complex transitive verb that was neither used as a cognate object verb (*tell a tale*), a support verb (*take a pace*) or part of an idiom (*have the sniffles*). The objects occurring in these clauses have an unusual distribution, as can be seen in Table (A).

First, there were more definite than indefinite objects in the sample.² Second, stative sentences had more indefinite objects than definite objects while the situation

¹What I treated as an adverb was the word *home*, used as in *Pat's not home*.

²I considered as definite NPs personal and reflexive pronouns, possessive NPs, NPs with the definite article, all NPs containing or consisting of a demonstrative, proper nouns, and cases of DNI.

Predicate	Frequency	Percentage
Adjective	37	18,5
Noun	4	2,0
Adverb ¹	1	,5
Preposition	3	1,5
Verb	140	70,0
Past Participle	1	,5
Ing form	14	7,0
Total	200	100,0

Table A.1: Distribution of predicate types in a random sample from the BNC

NP-type	Stative	Non-stative	Total
Definite NP	12	20	32
Indefinite NP	19	8	27
Total	31	28	59

Table A.2: Relationship between stative sentence aspect and definiteness of NP objects

was the opposite for non-stative sentences. The result is similar if one considers lexical aspect instead of sentence aspect: there were 9 occurrences of lexically stative verbs with an indefinite object and 6 with a definite object.³ Thus, it seems that while stative verbs that allow definite null instantiation strongly tend to have definite objects, this is not true of stative verbs overall.

³If one further excluded uses of quasi-presentational *have* (*The next jeep had two French Commando wounded aboard*) and *be* in the existential construction, the number of uses of lexically stative verbs with an indefinite object would drop to 4 as compared to 6 uses with a definite object.

Appendix B

Frames and Null Instantiation

B.1 Abundance

A **Location** is filled or covered with the **Theme**. The **Location** is realized as the **External Argument**, and the **Theme** either as PP complement headed by *with*, *in* or *of*.

- Location: Subject, CNI
- Theme: Oblique, INI with some predicates

Lexical Units: adorned.a, asphalted.a, bedecked.a, bejewelled.a, bespattered.a, blanketed.a, brimming.a, brushed.a, buttered.a, chock-a-block.a, chock-full.a, cloaked.a, coated.a, covered.a, crammed.a, crawl.v, crawling.a, crowded.a, dabbed.a, decked.a, decorated.a, dotted.a, draped.a, drizzled.a, dusted.a, embellished.a, festooned.a, filled.a, full.a, gilded.a, glazed.a, heaped.a, hung.a, injected.a, jammed.a, jostling.a, lacquered.a, lined.a, ornamented.a, overcrowded.a, overfilled.a, painted.a, panelled.a, paved.a, plastered.a, replete.a, smeared.a, spattered.a, splattered.a, sprinkled.a, strewn.a, studded.a, stuffed.a, surfaced.a, swarm.v, swarming.a, teem.v, teeming.a, throng.v, thronged.a, thronging.a, tiled.a, varnished.a, wallpapered.a

B.2 Health_response

A **PROTAGONIST** is sensitive to a **TRIGGER**, which has the potential to cause some kind of reaction in the **PROTAGONIST**.

- Protagonist: Subject, CNI
- Trigger: Oblique, DNI

Lexical Units: allergic.a, allergy.n, sensitive.a, sensitivity.n, susceptibility.n, susceptible.a

B.3 Word_relations

Definition: One linguistic unit, **Sign_1**, has a relation to another linguistic unit, **Sign_2**. This relation may be semantic, phonological, orthographic, or distributional in nature. The parameter along which the two signs are set in relation to each other is incorporated in the meaning of the target words in this frame. When the relation is symmetrical, the two linguistic units can be expressed jointly, **Signs**. Otherwise they have to be expressed disjointly. Ordinarily it is assumed that one linguistic unit has exactly one meaning but in cases where this is not so, a **Condition** may be specified that helps pick out the relevant sense.

Lexical Units: antonym.n, collocater.n, collocater.v, contrary.n, holonym.n, homograph.n, homonym.n, homophone.n, homophonous.a, hypernym.n, hyponym.n, meronym.n, synonym.n, synonymous.a

B.4 Kidnapping

Definition: The words in this frame describe situations in which a **Perpetrator** carries off and holds the **Victim** against his or her will by force.

- Perpetrator: Subject, CNI

- Victim: Direct Object, no NI

Lexical Units *abduct.v*, *abducted.a*, *abduction.n*, *abductor.n*, *kidnap.v*, *kidnapped.a*, *kidnapper.n*, *kidnapping.n*, *nab.v*, *shanghai.v*, *snatch.v*, *snatcher.n*

B.5 Filling

Definition: These are words relating to filling containers and covering areas with some thing, things or substance, the **Theme**. The area or container can appear as the direct object with all these verbs, and is designated **Goal** because it is the goal of motion of the **Theme**. Corresponding to its nuclear argument status, it is also affected in some crucial way, unlike goals in other frames.

- Agent: Subject, CNI
- Goal: Direct Object, no NI
- Theme: Oblique, INI

Lexical Units *adorn.v*, *anoint.v*, *asphalt.v*, *brush.v*, *butter.v*, *coat.v*, *cover.v*, *cram.v*, *crowd.v*, *dab.v*, *daub.v*, *douse.v*, *drape.v*, *drizzle.v*, *dust.v*, *embellish.v*, *fill.v*, *flood.v*, *gild.v*, *glaze.v*, *hang.v*, *heap.v*, *inject.v*, *jam.v*, *load.v*, *pack.v*, *paint.v*, *panel.v*, *pave.v*, *pile.v*, *plant.v*, *plaster.v*, *pump.v*, *scatter.v*, *seed.v*, *shower.v*, *smear.v*, *sow.v*, *splatter.v*, *splash.v*, *splatter.v*, *spray.v*, *spread.v*, *sprinkle.v*, *squirt.v*, *strew.v*, *stuff.v*, *suffuse.v*, *surface.v*, *tile.v*, *varnish.v*, *wallpaper.v*, *wrap.v*

B.6 Manufacturing

Definition: A **Manufacturer** produced a **Product** from a **Resource** for commercial purposes.

- Manufacturer: Subject, CNI
- Product: Direct, Object no NI

- Resources: Oblique, INI

Lexical Units: make.v, manufacture.n, manufacture.v, manufacturer.n, manufacturing.n, produce.v

B.7 Warrant

Definition: The existence of a *State_of_Affairs* is sufficient reason for taking an *Action*. The agent who is justified in taking the suggested *Action* is not part of the scenario, however.

- (B.1) The underwater sequences **warrant** the price of the admission alone.
- (B.2) 'Desperate situations **call** for desperate measures ,' said Mr Eames grandly.

- *State_of_Affairs*: Subject, CNI
- *Action*: with verbs: Direct Object, no NI; with nouns: DNI (with grounds)

Lexical Units call.v, deserve.v, grounds.n, justify.v, merit.v, warrant.n, warrant.v, worth.a

B.8 Measure_duration

Definition: This frame contains (transparent) nouns that are *Units* of measurement for time. The frame elements are *Count*, *Unit* and *Process*.

- (B.3) [Seven COUNT] [**hours** UNIT] [of torrential rain PROCESS] reduced the pitch to a lake
- (B.4) I had to listen to [**hours** UNIT] [of complaints PROCESS]

- *Count*: pre-determiner, INI
- *Process*: Oblique, DNI

- Unit: —

Lexical Units: century.n, day.n, decade.n, hour.n, millenium.n, minute.n, month.n, second.n, week.n, year.n

B.9 Quitting

Definition: An Employee voluntarily leaves the service of an Employer.

- Employee: Subject, CNI
- Employer: Direct Object: DNI¹

Lexical Units: give notice.v, quit.v, resign.v, resignation.n, retire.v*, retirement.n, stand down.v

B.10 Damaging

Definition: An Agent affects a Patient in such a way that the Patient (or some Subregion of the Patient) ends up in a non-canonical state. Often this non-canonical state is undesirable, and some lexical units (marked with the Negative semantic type) specifically indicate that the Patient is negatively affected.

- Agent: Subject, CNI
- Patient: Direct Object, no NI
- Subregion: Oblique, INI

Lexical Units chip.v, damage.v, deface.v, dent.v, key.v, mar.v, nick.v, sabotage.v, score.v, scratch.v, vandalise.v, vandalism.n

¹*retire* allows INI, which runs counter to our predictions. However, it does have a special meaning in that one does not just retire from one's job but from working life altogether. *Retire* is thus similar to *marry* in that it indicates an important status change within the culture and might thus be informative enough when predicated of its subject protagonist.

Appendix C

Coding

- **Animacy**

- animate (human)
- organization
- not animate

In most analyses, the animate and organization categories were merged into a single animate class.

- **Constituent length**

The number of space separated words that belong to the constituent. Thus, *car maker* consists of two words, *carmaker* and *car-maker* of one. Constituent length includes appositives in the constituents that they are in apposition to. Prepositions are not included in the constituent length. (But their presence is reconstructible from other data.)

- **Grammatical Function of XP**

This feature was used for arguments as well as their antecedent and subsequent mentions.

- subject

- object
- first/indirect object
- complement (that-clauses, VP_{to}, etc.)
- oblique (PP)
- possessive
- modifier
- dependent of nominal
- post-copular NP
- vocative
- appositive

Vocative, post-copular NP, and dependent of NP are included here only because they are needed for antecedent and subsequent mentions. Possessives may be an 'argument' of a verbal predicate when that predicate is controlled by a noun as in *his attempt to flee*.

• **XP Morphology**

- definite NP (*the N*)
- NP with proximal demonstrative determiner (*this N*)
- NP with distal demonstrative determiner (*that N*)
- proper noun (*John, the UK, the Smoky Mountains*)
- possessively determined NP (*my brother, John's brother*, but not *brother of John*)
- possessive pronoun (*mine, yours*)
- possessive determiner (*his, my*)
- speech act participant (that is, first or second person pronoun)
- non-speech act participant (that is, third person pronoun)

- distal demonstrative pronoun (*that*)
- proximal demonstrative pronoun (*this*)
- reflexive and reciprocal pronouns (*myself, yourself*)
- zero rel. pronoun of supplementary relative clauses (into the Dinkhik region, \emptyset **bordering** Afghanistan)
- non-restrictive relative pronoun (*which, who, that*¹)
- headless wh-clause (*what may be this year's best book*)
- definite zero
- indefinite NP (*a N*)
- quantified NP (*all students, every student, each student, ...*)
- bare plural or mass NP (*bears, milk*)
- indefinite zero
- indefinite pronoun (*one*)
- interrogative pronoun (*who, what, which*)
- restrictive relative pronoun (*who, which*)
- reduced relative (*nothing to **fear**; the policeman **following** the suspect*)
- zero in restrictive *that*-relative clauses
- pronoun of free relative (*what really **scares** me is his hair*)
- expletive (*it, there*)
- negative determiner (*no, NEG+any*)
- negative pronoun (*none, no one, NEG+anybody*)
- VP_{to} (I am **delighted** *to meet you*)
- VPing/Gerund (Smithers **delighted** *in teasing his grandson*)
- sentential gerund (*Ards winning it would **please** me*)

¹Though frowned upon by prescriptive grammar, *that* does have some marginal uses in non-restrictive relative clauses.

- finite clause (including *that*-clauses)
- relative clause (*who*, *which* not *that*)
- embedded *wh*-questions
- *for-to*-clause
- *whether*-clauses (embedded *yes/no*-questions introduced by *if* or *whether*)
- *when*-temporal clause
- parenthetical *as* (as he seems to claim)
- adverbial clause (*because I said so*)
- anaphoric adverb (*so*)
- deictic adverb *here*
- deictic adverb *there*
- anaphoric adverb *here*
- anaphoric adverb *there*
- other

- **Givenness**

This feature was intended especially for dealing with the cases where a referent has no direct co-referential link to preceding context but is instead inferred or anchored in another referent. The basic set of relations that I assumed were:

- evoked, textually (*The doctor just got in. He will see you now*)
- situationally evoked (*Have you seen this ?*)
- unused/omni-relevant (*Traffic's bad today because the president is in town*)
- inferable (*My car is quite old now and the engine is starting to give me trouble.*)
- brand new (*Guess what? I found a fifty-dollar bill on the sidewalk.*)

- **Types of inferential links**

For NPs referring to inferable referents, I also recorded what gave rise to the inference.

- frame/script (*How is that new restaurant next door? The food is great but the service is lousy.*)
- verbally encoded event, state, or proposition (*I don't think my HD-28 weighs four pounds but that's what he said; Sue denied the relationship several times but her denials fell on deaf ears*)
- other nominally coded referents
 - * inferred referent is subset of evoked set (in addition to subsets, this includes subtypes as well as parts of wholes) (*The Pars goals came from a couple of errors in the Hibs defence. The first came from a corner; That involved six people. Five of them were nurses from Jennie Stuart Medical Center*)
 - * co-subset (in addition to co-subsets, this includes co-subtypes as well as co-parts of wholes) (*How's the remodeling job going?—We're finished with the roof, now we're working on the basement*)
 - * superset of evoked set (in addition to supersets, this includes super-types as well as wholes) (*George W. Bush is the 43rd President. Up to now, all presidents have been men*)
 - * situationally given entities (exophora) (*The patient vomited and it got on the nurse's shoes*)
 - * a relation that is spelled out as part of the NP (normally a restrictive modifying clause, e.g. *the guy I work with*)
 - * kinship and possessive of strict ownership, that is possession by human (*My car broke down*)
 - * attribute/property (*'Lantana' hides thorns in its beauty*)
 - * other (this category was a catch-all for unclear or ambiguous cases)