OXFORD
LINGUISTICS

CROSS-LINGUISTIC VARIATION AND EFFICIENCY

John A. Hawkins

Cross-linguistic Variation and Efficiency

This page intentionally left blank

Cross-linguistic Variation and Efficiency

JOHN A. HAWKINS

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries.

First Edition published in 2014

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data

Data available

Library of Congress Control Number: 2013941078

ISBN 978–0–19–966499–3 (Hbk.)
978–0–19–966500–6 (Pbk.)

Printed and bound by
CPI Group (UK) Ltd., Croydon, CR0 4YY

The manufacturer's authorized representative in the EU for product safety is
Oxford University Press España S.A. of Parque Empresarial San Fernando de Henares,
Avenida de Castilla, 2 – 28830 Madrid (www.oup.es/en or product.safety@oup.com).
OUP España S.A. also acts as importer into Spain of products made by the manufacturer.

To the memory of my father, Eric William Hawkins,
qui a éveillé ma passion pour les langues étrangères

This page intentionally left blank

Contents

This page intentionally left blank

Preface

Have the rules of grammar been shaped by ease of processing and communicative efficiency? Can the differences between languages be explained by a usage-based theory of grammar? A growing body of research in several areas of the language sciences over the past fifteen years has answered these questions in the affirmative. My 2004 OUP book *Efficiency and Complexity in Grammars* can be situated within this research tradition. In it I gave detailed evidence for the profound role of performance in shaping grammars and I showed how many typological patterns and universals of grammar can be explained in this way. Not only is grammar *not* autonomous from performance, I argued, but the rules and conventions of grammars appear to be systematically aligned with preferences that can be observed in language use.

The present book is an updating and extension of the research program that was laid out in 2004. Since then my collaborators and I have investigated many new areas of grammar and of performance, other linguists and psycholinguists have tested some of my claims and predictions further, and there has been invaluable theoretical discussion and critical feedback in the literature. Meanwhile the fields of language processing, of grammar, language typology, and historical linguistics, all of which I have been trying to integrate, have each moved forward. It is time to bring this research program up to date, to present new data, and to address certain issues. At the same time I wish to carry forward the basic Performance–Grammar Correspondence Hypothesis (PGCH) of the 2004 book and the three principles that gave substance to it: Minimize Domains (MiD); Minimize Forms (MiF); and Maximize Online Processing (MaOP). These principles will be repeated here with illustrative supporting data. A considerable amount of new material has been added to the present book with the result that this is now a new book rather than a second edition.

The research program that my collaborators and I have been engaged in for over twenty years (see below for detailed acknowledgments) has been broadly based and involves an empirical and interdisciplinary approach to cross-linguistic variation. We have been systematically comparing variation patterns *within* and *across* languages, i.e., in usage and in grammars. At the same time we make extensive use of generative principles of the kind that Chomsky and his followers have given us, and of typologists' generalizations as developed by Joseph Greenberg, Matthew Dryer, Martin Haspelmath, Bernard Comrie, and many others. These grammatical principles and patterns

have been integrated with models of language processing in psycholinguistics and with experimental and empirical usage data collected by linguists and psycholinguists.

There are two reasons why this methodology has proved fruitful. First, a general correlation has emerged: the patterns of preference that one finds in performance in languages possessing several structures of a given type—for example, the preferences for different word orders or for different relative clause types when relativizing on different positions—appear to be the same patterns found in the fixed conventions of grammars, in languages with fewer structures of the same type (e.g., with more fixed word orders and more restrictive relativization options). These preferences, and the quantitative distribution of less preferred structures, show striking correspondences between usage and grammar.

Second, this correlation has far-reaching consequences for language universals, for the theory of grammar, and for psycholinguistic models of language processing. It provides an argument against the autonomy of grammar from performance (see Chomsky 1965). It enables us to make predictions from performance data for grammatical conventions, and the grammatical patterns predicted are often unexpected from grammatical considerations alone. It helps us understand both why there are universal patterns across languages, and why there are often exceptions to these and when they will occur. It adds a much-needed cross-linguistic component to theories of language processing, and provides both usage data and grammatical regularities from languages that are very different from those on which current processing theories have been built. These data can lead to a rethinking of a number of processing assumptions derived from more familiar languages. They can also pinpoint precise areas that should now be tested using experimental paradigms on native speakers of other languages.

Joseph Greenberg was the first to draw attention to correlating patterns between performance and grammars (in morphosyntax, in his 1966 book *Language Universals, with Special Reference to Feature Hierarchies*, Mouton, The Hague). In my 1994 book *A Performance Theory of Order and Constituency* (CUP) I argued that the preferred word orders in languages with choices are those that are productively conventionalized as fixed orders in grammars permitting less freedom. And in the present book and in 2004 I examine many more grammatical areas in a systematic test of the Performance–Grammar Correspondence Hypothesis. Specifically, the data from these books support the following conclusions:

- Grammars have conventionalized syntactic structures in proportion to their degrees of preference in performance, as evidenced by patterns of

selection in corpora and by ease of processing in psycholinguistic experiments.

- These common preferences of performance and grammars are structured by general principles of efficiency and complexity that are clearly visible in both usage data and grammatical conventions; three of these principles are defined and illustrated here: Minimize Domains, Minimize Forms, and Maximize Online Processing.
- Greater descriptive and explanatory adequacy can be achieved when efficiency and complexity principles are incorporated into the theory of grammar; stipulations are avoided, many exceptions can be explained, and improved formalisms incorporating significant generalizations from both performance and grammars can be proposed.
- Psycholinguistic models can benefit from looking at languages that are structurally very different from English; psycholinguistics has traditionally not paid enough attention to cross-linguistic variation, though this is starting to change; more corpus data and experimental data need to be gathered from different languages, whose usage preferences sometimes differ radically from those of English, and psycholinguists can also look to grammatical rules and conventions for ideas about processing, precisely because of the correspondence between them; psycholinguists can in the process help linguists to better understand the rules and conventions that are the cornerstone of the field of linguistics.
- The patterns of cross-linguistic variation presented here point to multiple processing factors interacting with one another and to degrees of preference and relative strength that can usefully guide theorizing about the interaction of principles in a multi-factor model.

Some of these conclusions will be controversial, I realize. But the claimed autonomy of grammar from performance, which many linguists still believe in, does need to be subjected to empirical test, and grammars and grammatical variation do need to be examined from a processing perspective. This is what the research program described here and in previous books is all about. And my finding is that there is a deep correspondence between performance data and grammars. Psycholinguistics, meanwhile, can benefit from becoming more cross-linguistic. There is a long tradition of cross-linguistic work in language acquisition, spearheaded by Dan Slobin and his colleagues. Language processing needs to follow suit. Psycholinguists can also help linguists with their fundamental task of better understanding how linguistic rules and conventions work, and why they have the properties they do, precisely because they have been profoundly shaped by processing.

There are many people I am indebted to for feedback and assistance of various kinds while writing this book. I must first acknowledge the detailed and extremely useful reviews of the first draft provided by Martin Haspelmath and Chris Cummins. Their comments led to significant rewriting and to many improvements. With respect to data collection and theory development I must also mention especially Matthew Dryer, Luna Filipović, Kaoru Horie, Stephen Matthews, Fritz Newmeyer, and Tom Wasow, all of whom have helped me with key aspects of the present book. Many others provided feedback and help on various points, including Gontzal Aldai, Elaine Andersen, Raul Aranovich, Theresa Biberauer, Paula Buttery, Bernard Comrie, Peter Culicover, Patrick Farrell, Fred Field, Giuliana Fiorentino, Elaine Francis, Ted Gibson, Laura Gonnerman, Anders Holmberg, Florian Jaeger, Ed Keenan, Ruth Kempson, Johannes Kizach, Ekkehard König, Lewis Lawyer, Christian Lehmann, Maryellen MacDonald, Brian MacWhinney, Edith Moravcsik, William O'Grady, Masha Polinsky, Beatrice Primus, Ian Roberts, Dan Slobin, Jae Song, Aaron Sonnenschein, Lynne Stallings, Harry Tily, Peter Trudgill, Sten Vikner, Caroline Williams, and Fuyun Wu. I am extremely grateful to all these individuals, none of whom is of course implicated in the theory presented here or in any errors that remain. I must also thank John Davey of Oxford University Press for his support and professionalism in getting this book published. John is the best linguistics editor in publishing today and our field owes him an enormous debt.

At an institutional level I must thank the German Max Planck Society which has supported my work over a long period, first at the psycholinguistics institute in Nijmegen and then at the evolutionary anthropology institute in Leipzig. Most of the 2004 book was written in Leipzig and I am grateful to that institute, and to Bernard Comrie in particular, for the opportunity to complete it there. The University of Southern California in Los Angeles also supported me for many years prior to 2004. Thereafter the University of Cambridge and the University of California Davis have provided generous support for the research and writing that have culminated in this book, in the form of research grants, research assistance, travel support, and teaching buyouts. I thank them both.

JAH
Davis, California

List of Figures and Tables

Figures

Tables

This page intentionally left blank

List of Abbreviations

A	appositive
ABL	ablative
Abs/ABS	absolutive
Acc/ACC	accusative
Adj	adjective
AdjP	adjective phrase
Adv	adverb
Ag	agent
AH	Accessibility Hierarchy
AP	argument precedence
Art	article
Asp	aspect
Aux/AUX	auxiliary verb
C	complementizer
CL	classifier
Classif/CLASSIF	classifier
Comit/COMIT	comitative
Comp	complementizer
Cont/CONT	continuative
Correl/CORREL	correlative
CRD	constituent recognition domain
CV	consonant–vowel (syllable)
CVC	consonant–vowel–consonant (syllable)
Dat/DAT	dative
Def/DEF	definite determiner
Dist.Past	distant past
DO	direct object
DP	determiner phrase
DS	different subject
EIC	Early Immediate Constituents
Erg/ERG	ergative
F	form
Fem/FEM	feminine

FGD	filler gap domain
F:P	form–property pairing
FOFC	Final-Over-Final Constraint
Fut	future
G	grandmother node
Gen/GEN	genitive
Gen-DO	genitive within a DO
Gen-SU	genitive within a SU
H	hierarchy
HM	head before modifier
IC	immediate constituent
ICL	longer IC
ICS	shorter IC
Impf/IMPF	imperfective
Instr	instrument
IO	indirect object
L TO R	left to right
LD	Lexical Domain
Loc/LOC	locative
M	mother node
Masc	masculine
MaOP	Maximize Online Processing
MH	modifier before head
MiD	Minimize Domains
MiF	Minimize Forms
mPP	a PP with head-initial preposition
MPT	manner before place before time
N	noun
Neut	neuter
Nom/NOM	nominative
Nomlz/NOMLZ	nominalizer
NP	noun phrase
NPo	a direct object head-final NP
NRel	noun before relative clause
NSRC	non-subject relative clause
O	(direct) object
O	operator
OBJ	object
OBL	oblique
oNP	a direct-object head-initial NP

OP	online property
OP/UP	online property to ultimate property (ratios)
OS	object before subject
OV	object before verb
OVS	object before verb before subject
P	preposition or postposition
P	property
Part/PART	particle
Pat	patient
PCD	Phrasal Combination Domain
Pd	dependent PP
PGCH	Performance–Grammar Correspondence Hypothesis
Pi	independent PP
Pl/PL	plural
Plur	plural
Poss/POSS	possessive
Possp/POSSP	possessive phrase
PP	prepositional or postpositional phrase
PPL	longer PP
PPm	a PP with head-final postposition
PPS	shorter PP
Pres/PRES	present
Pret/PRET	preterite
PS	phrase structure
Q	quantifier
R	restrictive
R-case	rich case marking
Recip	recipient
Refl/REFL	reflexive
Rel	relative clause
Rel/REL	relativizer
RelN	relative clause before noun
S	sentence
S	subject
S′	S bar
Sg/SG	singular
Sing	singular
SO	subject before object

SOV	subject before object before verb
SPLT	syntactic prediction locality theory
SVO	subject before verb before object
SU/SUBJ	subject
T	tense
UG	universal grammar
UP	ultimate property
V	verb
vf	finite V
VO	verb before object
VOS	verb before object before subject
VP	verb phrase
VSO	verb before subject before object
WALS	*World Atlas of Language Structures*
X	an oblique phrase
XP	an arbitrary phrase

1

Language variation and the Performance–Grammar Correspondence Hypothesis

1.1 The Performance–Grammar Correspondence Hypothesis

This book explores the kinds of universals of language that Greenberg introduced in his seminal 1963 paper on word order. They took the form of implicational statements defining patterns of cross-linguistic variation: if a language has some property (or set of properties) P, then it also has (or generally has) property Q. For example, if a language has subject–object–verb (SOV) order, as in Japanese, it generally has postpositional phrases ([the movies **to**] **went**), rather than prepositional phrases as in English (**went** [**to** the movies]). When implicational universals were incorporated into generative grammar, in the Government–Binding theory of Chomsky (1981), they became known as 'parameters' and the innateness claimed for absolute universals (Chomsky 1965; Hoekstra and Kooij 1988) was extended to the parameters (Lightfoot 1991; J. D. Fodor 2001). It was proposed that the child's linguistic environment 'triggered' one innate parameter rather than another.

The status of these variation-defining universals in generative grammar has been questioned following the publication of Newmeyer's *Possible and Probable Languages: A Generative Perspective on Linguistic Typology* (2005). Newmeyer argued that the major parameters proposed hitherto, the head ordering parameter, the pro-drop parameter, and others, had systematic exceptions across languages, were probabilistic, and were not part of UG, which is concerned with defining possible versus impossible languages. Haspelmath (2008b) has given a similar critique of parameters. In effect, these authors recognize what Greenberg (1963) first recognized: the majority of his implicational statements hold only with more than chance frequency, and most of those he formulated as exceptionless have turned out to have exceptions (Dryer 1992). Clearly, if these parameters are not correct descriptively, they are not innate either, and the kind of environmental trigger theory for

language acquisition built around them fails, if the basic premise fails (the existence of exceptionless parameters of variation).

The question now arises: where do we go from here in order to better understand cross-linguistic grammatical variation? A number of generative theorists are trying to improve the empirical adequacy of earlier predictions. Cinque (2005) is a laudable example which combines Kayne's (1994) antisymmetry principle with painstaking typological work. The work of Biberauer, Holmberg, and Roberts (2007, 2008) developing the 'Final-over-Final Constraint' is another welcome example of research bridging formal grammar, typology, and also historical linguistics in this case (see §§5.6–5.8 in this volume for a detailed discussion and assessment). A different research program, more in line with Newmeyer's (2005) proposals, is the one I presented in my 2004 book *Efficiency and Complexity in Grammars*, which built on my 1994 *A Performance Theory of Order and Constituency*. The present study updates this general program and incorporates many new findings, ideas, and criticisms. Together with several collaborators I have been pursuing a strongly empirical and interdisciplinary approach to language universals, comparing variation patterns *within* and *across* languages. We have been examining variation both in usage (performance) and in grammars. This program makes extensive use of both generative principles and typologists' generalizations (Comrie 1989; Croft 2003), and integrates them with psycholinguistic models and findings.

There are two reasons why this has proved fruitful. First, a general correlation has emerged: the patterns of preference that one finds in performance in languages possessing several structures of a given type (different word orders, relative clauses, etc.) look very much like the patterns found in the fixed conventions of grammars, in languages with fewer structures of the same type. In other words, grammars appear to have conventionalized the structural variants that speakers prefer to use. Numerous examples will be summarized in §1.2 and many more will be discussed in greater detail in subsequent chapters.

Second, this correlation has far-reaching consequences for language universals and for the theory of grammar. It makes predictions from performance data for grammatical conventions, and the grammatical patterns predicted are often unexpected from grammatical considerations alone. It helps us understand why these patterns are found across languages, and also why there are sometimes exceptions and when they occur.

Greenberg (1966) was the first to draw attention to correlating patterns between performance and grammars in his discussion of markedness hierarchies like Singular > Plural > Dual > Trial/Paucal. Morphological inventories across grammars and declining allomorphy provided evidence for the

universal hierarchies, while declining frequencies of use in languages with rich inventories suggested not only a correlation with performance but a possibly causal role for it in the evolution of the grammatical regularities themselves (Greenberg 1995: 163–4). Givón (1979: 26–31) meanwhile observed that performance preferences in one language, e.g., for definite subjects, may correspond to an actual categorical requirement in another. In Hawkins (1994) I argued that the preferred word orders in languages that allow flexibility are those that are productively conventionalized as fixed orders in languages with less flexibility. The 2004 book examined many more grammatical areas and proposed the following general hypothesis:

(1.1) *Performance–Grammar Correspondence Hypothesis (PGCH)*
Grammars have conventionalized syntactic structures in proportion to their degree of preference in performance, as evidenced by patterns of selection in corpora and by ease of processing in psycholinguistic experiments.

In the past fifteen years we have seen mounting evidence from several areas of the language sciences for the role of performance in shaping grammars, and even for a basic correspondence between them along these lines. Haspelmath (1999a) has proposed a theory of diachrony in which usage preferences lead to changing grammatical conventions over time. Bybee and Hopper (2001) document the clear role of frequency in the emergence of grammatical structure. There have been intriguing computer simulations of language evolution, exemplified by Kirby (1999), in which processing preferences of the kind assumed for word order in Hawkins (1990, 1994) are incorporated in the simulation and lead to the emergence of observed grammatical types after numerous iterations (corresponding to successive generations of language users). There have been developments in Optimality Theory, exemplified by Haspelmath (1999a) and Aissen (1999), in which functional motivations ultimately related to processing are provided for many of the basic constraints. Stochastic Optimality Theory (Bresnan et al. 2001; Manning 2003) incorporates the preferences of performance ('soft constraints') as well as grammatical conventions ('hard constraints'). Newmeyer (2005) advocates replacing generative parameters with principles derived from language processing, while Phillips (1996) and Kempson et al. (2001) incorporate the online processing of language into the rules and representations of the grammar itself.

But despite this growing evidence for performance–grammar correspondences, the precise extent to which grammars have been shaped by performance is still a matter of debate. There are different opinions in the publications

cited here. Even the question of whether performance has shaped grammars at all is still debated. See §3.6 for a discussion and critique of recent work within Chomsky's Minimalist Program which explicitly denies such a causative role for performance and which maintains the asymmetry between competence and performance that was first advocated in Chomsky (1965), as detailed in §1.2.

In this book I adopt a data-driven approach and focus on the empirical evidence for the PGCH. My goal is to try to abstract away from current disagreements and unresolved issues in grammatical models and in processing theories and to convince the next generation of researchers that there is a real generalization here that needs to be incorporated into theories of grammatical universals. In the next sections I briefly summarize a range of observed performance–grammar correspondences supporting the PGCH (§1.2) and I define its predictions and consequences (§1.3). Subsequent chapters examine many areas of grammatical variation across languages in more detail from this perspective.

1.2 Examples of proposed performance–grammar correspondences

The Keenan and Comrie (1977) Accessibility Hierarchy (SU>DO>IO/OBL>GEN; see Comrie 1989) has been much discussed in this context. Grammatical cut-off points in relative clause formation possibilities across languages follow the hierarchy, and Keenan and Comrie argued for an explanation in terms of declining ease of processing down the lower positions on the hierarchy. As evidence they pointed to usage data from languages with many relativizable positions, especially English. In such languages corpus frequencies declined down the hierarchy while processing load and working memory demands have been shown to increase under experimental conditions (Keenan 1975; Keenan and S. Hawkins 1987; Hawkins 1999; Diessel and Tomasello 2006); see also §2.2.3.

More generally, filler-gap dependency hierarchies for relativization and Wh-movement across grammars appear to be structured by the increasing complexity of the permitted gap environments in the lower positions of these hierarchies. The grammatical cut-off points in these increasingly complex clause-embedding positions for gaps correspond to declining processing ease in languages with numerous gap-containing environments (including subjacency-violating languages like Akan: Saah and Goodluck 1995); see Hawkins (1999, 2004: ch. 7) and §8.1.

Reverse hierarchies across languages for conventionalized gaps in simpler relativization domains and resumptive pronouns in more complex environments (Hawkins 1999) match the performance distribution of gaps to pronouns

within languages such as Hebrew and Cantonese in which both are grammatical (in some syntactic positions), gaps being preferred in the simpler and pronouns in the more complex relatives (Ariel 1999; Matthews and Yip 2003; Hawkins 2004); see §2.2.3.

Parallel function effects (whereby the head of the relative matches the position relativized on) have been shown to facilitate relative clause processing and acquisition (Sheldon 1974; MacWhinney 1982; Clancy et al. 1986). They also extend relativization possibilities beyond normal constraints holding in languages such as Basque and Hebrew (Aldai 2003; Cole 1976; Hawkins 2004: 190); see §2.2.3.

Declining acceptability of increasingly complex center embeddings, in languages in which these are grammatical, is matched by hierarchies of permitted center embeddings across grammars, with cut-offs down these hierarchies (Hawkins 1994: 315–21); see §5.5.

(Nominative) subject before (accusative) object ordering is massively preferred in the performance of languages in which both SO and OS are grammatical (Japanese, Korean, Finnish, German) and is also massively preferred as a basic order or as the only order across grammars (Hawkins 1994; Gibson 1998; Tomlin 1986; Primus 1999; Miyamoto 2006); see §3.5 and §8.5.

Markedness hierarchies of case (Nom>Acc>Dat>Other) and number (Sing>Plur>Dual>Trial), etc., correspond to performance frequency hierarchies in languages with rich morphological inventories (Greenberg 1966; Croft 2003; Hawkins 2004: 64–8); see §2.2.1.

Performance preferences in favor of a definite rather than an indefinite grammatical subject, e.g., in English, correspond to a categorical requirement for a definite subject in others (e.g., in Krio: Givón 1979).

Performance preferences for subjects that obey the Person Hierarchy (1st, 2nd > 3rd) in English (whereby *The boy hit me* is preferably passivized to *I was hit by the boy*) have been conventionalized into a grammatical/ungrammatical distinction in languages such as Lummi (Bresnan et al. 2001). Sentences corresponding to *The boy hit me* are ungrammatical in Lummi.

The distinction between zero agreement in local NP environments versus explicit agreement non-locally in the grammar of Warlpiri matches the environments in which zero and explicit forms are preferred in performance in languages with choices: for example, in the distribution of zero and explicit relativizers in English (Hawkins 2004: 160); see §2.2.2 and §6.3.1.

These are just the tip of a large iceberg of performance-motivated cross-linguistic patterns. If these correspondences are valid, then the classic picture of the performance–grammar relationship presented in Chomsky (1965) needs to be revised. For Chomsky the competence grammar was an integral

part of a performance model, but it was not shaped by performance in any way:

> Acceptability . . . belongs to the study of performance . . . The unacceptable grammatical sentences often cannot be used, for reasons having to do . . . with memory limitations, intonational and stylistic factors, . . . and so on. . . . it would be quite impossible to characterize unacceptable sentences in grammatical terms . . . we cannot formulate particular rules of the grammar in such a way as to exclude them. (Chomsky 1965: 11–12)

Chomsky claimed (and still claims: see §3.6) that grammar is autonomous from performance and that UG is innate (see Newmeyer 1998 for a full summary and discussion of these points). The PGCH in (1) is built on the opposite assumption that grammatical rules *have* incorporated properties that reflect memory limitations and other forms of complexity and efficiency that we observe in performance. This alternative is supported by the correspondences above, and makes predictions for occurring and non-occurring grammars and for frequent and less frequent ones. It accounts for many cross-linguistic patterns that are not predicted by grammar-only theories, and for exceptions to those that are predicted. Subsequent chapters will illustrate the PGCH and this research method in greater detail.

1.3 Predictions and consequences of the Performance–Grammar Correspondence Hypothesis

The PGCH makes predictions, which we must define. In order to test them we need performance data and grammatical data from a range of languages involving the same grammatical structures. This research program proceeds as follows. First, find a language whose grammar generates or permits a plurality of structural alternatives of a common type. They may involve alternative orderings of the same constituents with the same or similar domination relations in the phrase structure tree, e.g., different orderings of NP and PP constituents in the free-ordering postverbal domain of Hungarian, or [PP NP V_{VP}] vs [NP PP V_{VP}] in a verb-final language like Japanese. Or they may involve alternative relative clauses with and without an explicit relativizer, as in English (*the Danes whom/that he taught* vs *the Danes he taught*), or alternations between relativizations on a direct object using a gap strategy vs a resumptive pronoun strategy, as in Hebrew.

Second, check for the distribution of these same structural patterns in the grammatical conventions across languages. The PGCH predicts that when the

grammar of one language is more restrictive and eliminates one or more structural options that are permitted by the grammar of another, the restriction will be in accordance with performance preferences. The preferred structure will be retained and 'fixed' as a grammatical convention; the dispreferred structures will be removed. Either they will be eliminated altogether from the output of the grammar or they may be retained in some marginal form as lexical exceptions or as limited construction types. So, for example, if there is a general preference in performance for constituent orderings that minimize the number of words on the basis of which phrase structure groupings can be recognized, as I argued in Hawkins (1994), then I expect the fixed word orders of grammars to respect this same preference. They should permit rapid immediate constituent (IC) recognition in the normal case. Numerous adjacency effects are thereby predicted between sister categories in grammars, based on their (average) relative weights and based on the information that they provide about phrase structure online (through, e.g., head projection). Similarly if the absence of the relativizer in English performance is strongly associated with the adjacency of the relative clause to the head noun, while its presence is found frequently when the relative clause is both adjacent and non-adjacent to the head, then I expect that grammars that actually remove the zero option altogether will do so in a way that reflects the patterns of preference in performance: they will either remove the zero option and require an explicit relativizer only when the relative clause is non-adjacent to the head, or they will remove zero under both adjacency and non-adjacency, but they will not remove zero and require explicit relativizers only when they are adjacent to the head. And if the gap relativization strategy in Hebrew performance provides evidence for a structural proximity preference to the head noun, compared with the resumptive pronoun strategy, then it is predicted that the distribution of gaps compared to pronouns across grammars should be in this same direction, with gaps being more or equally proximate to their head nouns.

These are illustrations of the research strategy of this book. The major predictions of the PGCH that were systematically tested in Hawkins (2004) are the following:

(1.2) *Grammatical predictions of the PGCH* (Hawkins 2004)

(a) If a structure A is preferred over an A' of the same structural type in performance, then A will be more productively grammaticalized, in proportion to its degree of preference; if A and A' are more equally favored, then A and A' will both be productive in grammars.

(b) If there is a preference ranking A>B>C>D among structures of a common type in performance, then there will be a corresponding

hierarchy of grammatical conventions (with cut-off points and declining frequencies of languages).

(c) If two preferences P and P' are in (partial) opposition, then there will be variation in performance and grammars, with both P and P' being realized, each in proportion to its degree of motivation in a given language structure.

These predictions will be exemplified further in this book and it will be argued that the PGCH (1.1) is strongly supported descriptively. This approach can also provide answers to explanatory questions that are rarely raised in the generative literature such as: Why should there be a head-ordering principle defining head-initial and head-final language types (Hawkins 1990, 1994, 2004)? Why are there heads at all in phrase structure (Hawkins 1993, 1994)? Why are some categories adjacent and others not (Hawkins 2001, 2003, 2004)? And why is there a subjacency constraint and why is it parameterized the way it is (Hawkins 1999, 2004)?

These questions can now be asked, and informative answers can be given, within this framework. The basic empirical method involves conducting a simple test: Are there, or are there not, parallels between universal patterns across grammars and patterns of preference and processing ease within languages? The data of this book suggest that there are and the descriptive and explanatory benefits for which I argue then follow.

Let me end this chapter with some general remarks on bigger issues that are raised by this research program. We can distinguish between variation-defining universals and absolute universals of the form 'all languages (or no languages) have property P.' These latter have been at the core of Universal Grammar (UG) in generative theories from Chomsky (1965) through Chomsky (1995) and beyond. Increasingly the range of such absolute universals has been scaled back; cf., e.g., Hauser, Chomsky, and Fitch (2002). Whatever absolute universals of syntax are currently recognized can, in principle, be innately grounded in the species. But innate grammatical knowledge is not, I suggest, plausible for variation-defining universals, both because innate parameters are not plausible in principle and because they are largely probabilistic (Newmeyer 2005). But notice that it is still plausible to think in terms of Elman et al.'s (1996) 'architectural innateness' as constraining the data of performance, which can then evolve into variant conventions of grammar. The architectural innateness of the human language faculty enters into grammars indirectly in this way. Absolute universals can also, in principle, be innately grounded as a result of processing constraints on grammars. When complexity and efficiency levels are comparable and tolerable, we get the variation between grammars that we will see. But within and beyond certain

thresholds I would expect universals of the kind 'all languages have X' and 'no languages have X', as a result of processability interacting with the other determinants of grammars. The Performance–Grammar Correspondence Hypothesis is no less relevant to absolute universals, therefore, with the extremes of simplicity/complexity and (in)efficiency being inferable from actually occurring usage data. Some interesting proposals have been made recently by Mobbs (2008) for incorporating the efficiency proposals of this book into Chomsky's (1995) Minimalist Program. The efficiency principles are now recast as general cognitive constraints on the 'internal computations' integrating linguistic and other mental entities, rather than as principles of performance as such, and are seen as having shaped cross-linguistic parameters in a way not unlike that proposed by Hawkins (2004). This proposal, which brings the two research traditions closer together, is discussed and critiqued in §3.6.

There can also be innate grammatical and representational knowledge of quite specific properties, of the kind summarized in Pinker and Jackendoff 's (2009) response to Hauser, Chomsky, and Fitch (2002). Much of phonetics, semantics, and cognition are presumably innately grounded and there are numerous properties unique to human language as a result. See Newmeyer (2005) for the role of conceptual structure in shaping absolute universals, and also Bach and Chao (2009) for a discussion of semantically based universals.

The precise causes underlying the observed preferences in performance require more attention than I can give them here. Much of psycholinguistics is currently trying to develop appropriate models for the kinds of performance data I discuss in this book. To what extent do the preferences result from parsing and comprehension, and to what extent are they production-driven? What is the role of frequency sensitivity and of prior learning in online processing (see, e.g., Reali and Christiansen 2006a,b)? What is the relationship between predictive and 'surprisal' metrics on the one hand (Levy 2008; Jaeger 2006) and more 'integration'-based ones on the other (Gibson 1998, 2000)? These issues are discussed further in Chapter 3.

A performance explanation for universals has consequences for learning and for learnability, since it reduces the role of an innate grammar. UG is no longer available in the relevant areas (head ordering, subjacency, etc.) to make up for the claimed poverty of the stimulus and to solve the negative evidence problem (Bowerman 1988). The result is increased learning from positive data, something that Tomasello (2003), connectionist modelers like MacDonald (1999), and also linguists like Culicover (1999) have been arguing for independently. These converging developments enable us to see the data of experience as less impoverished and more learnable than previously thought. The grammaticality facts of Culicover's book, for example, pose learnability problems that are

just as severe as those for which Hoekstra and Kooij (1988) invoke an innate UG, yet Culicover's data involve language-particular subtleties of English that cannot possibly be innate (*the student is likely to pass the exam* is a grammatical raising construction in English whereas **the student is probable to pass the exam* is not). See Hawkins (2004: 272–6) and §3.6.2 for further discussion of these issues.

The explanation for cross-linguistic patterns that I propose here also requires a theory of diachrony that can translate the preferences of performance into fixed conventions of grammars. Grammars can be seen as complex adaptive systems (Gell-Mann 1992), with ease of processing driving the adaptation in response to prior changes. But we need to better understand the 'adaptive mechanisms' (Kirby 1999) by which grammatical conventions emerge out of performance variants. How do grammatical categories and the rule types of particular models end up encoding performance preferences? And what constraints and filters are there on this translation from performance to grammar? I outlined some major ways in which grammars respond to processing in Hawkins (1994: 19–24) (e.g., by incorporating movement rules applying to some categories rather than others, defining certain orderings rather than others, constraining the applicability of rules in certain environments, and so on), and these are recast here in updated form in §4.3.

Before turning to these and other issues raised by this research program it is necessary to summarize the basic efficiency principles themselves that are proposed in Hawkins (2004) and that are at the heart of the Performance–Grammar Correspondence Hypothesis in (1.1). This is the purpose of Chapter 2, to which I now turn.

2

Three general efficiency principles

In Hawkins (2004) I proposed three general efficiency principles, which I shall summarize briefly here together with illustrative data sets supporting them (§§2.1–2.3). I then discuss how I see the relationship between efficiency on the one hand and simplicity/complexity on the other (§2.4). The last section (§2.5) looks at some competing efficiencies in grammars and in performance and draws attention to issues of cooperation and competition between them.

2.1 Efficiency principle 1: Minimize Domains

One clear principle of efficiency, evident in both performance and grammars, involves the size of the syntactic domain in which a given grammatical relation can be processed. How great is the distance separating interrelated items and how much other material needs to be processed simultaneously as this relation is processed? In those languages and structures in which domain sizes can vary in performance, we see a clear preference for the smallest possible domains. In those languages and structures in which domain sizes have been grammatically fixed, we see the same preference in the conventions. The organizing principle here is defined as follows in Hawkins (2004: 31):

(2.1) *Minimize Domains* (MiD)
The human processor prefers to minimize the connected sequences of linguistic forms and their conventionally associated syntactic and semantic properties in which relations of combination and/or dependency are processed. The degree of this preference is proportional to the number of relations whose domains can be minimized in competing sequences or structures, and to the extent of the minimization difference in each domain.

Combination = Two categories A and B are in a relation of combination iff they occur within the same syntactic mother phrase or maximal projection (phrasal combination), or if they occur within the same lexical co-occurrence frame (lexical combination).

Dependency = A category B is dependent on a category A iff the parsing of B requires access to A for the assignment of syntactic or semantic properties to B with respect to which B is zero-specified or ambiguously or polysemously specified.

Consider the ordering of words. Words and phrases have to be assembled into the kinds of groupings that are represented by tree structure diagrams as they are parsed and produced in the linear string of speech. Assigning phrase structure can typically be accomplished on the basis of less than all the words dominated by each phrase. Some orderings reduce the number of words needed to recognize a mother phrase M and its immediate constituent daughters (ICs), making phrasal combination faster. Compare (2.2a) and (b):

(2.2) a. The man [$_{VP}$ looked [$_{PP_1}$ for his son] [$_{PP_2}$ in the dark and quite derelict building]]

1 2 3 4 5

b. The man [$_{VP}$ looked [$_{PP_2}$ in the dark and quite derelict building] [$_{PP_1}$ for his son]]

1 2 3 4 5 6 7 8 9

The three items, V, PP1, PP2 can be recognized on the basis of five words in (2.2a), compared with nine in (2.2b), assuming that (head) categories such as P immediately project to mother nodes such as PP, enabling the parser to construct and recognize them online. Minimize Domains predicts that Phrasal Combination Domains (PCDs) should be as short as possible, and the degree of this preference should be proportional to the minimization difference between competing orderings. This principle (a particular instance of Minimize Domains) is called Early Immediate Constituents (EIC):

(2.3) *Phrasal Combination Domain* (PCD)
The PCD for a mother node M and its I(mmediate) C(onstituent)s consists of the smallest string of terminal elements (plus all M-dominated non-terminals over the terminals) on the basis of which the processor can construct M and its ICs.

(2.4) *Early Immediate Constituents* (EIC) [Hawkins 1994: 69–83]
The human processor prefers linear orders that minimize PCDs (by maximizing their IC-to-word ratios), in proportion to the minimization difference between competing orders.

In concrete terms EIC amounts to a preference for short before long phrases in head-initial structures like those of English, e.g. for short before long PPs in (2.2). These orders will have higher 'IC-to-word ratios,' i.e., they will permit more ICs to be recognized on the basis of fewer words in the terminal string. The IC-to-word ratio for the VP in (2.2a) is 3/5 or 60 percent (five words required for the recognition of three ICs). The comparable ratio for (2.2b) is 3/9 or 33 percent (nine words required for the same three ICs). (For comparable benefits within a Production Model, see Hawkins 2004: 106.)

Structures like (2.2) were selected from a corpus on the basis of a permutation test (Hawkins 2000, 2001): the two PPs had to be permutable with truth-conditional equivalence (i.e., the speaker had a choice). Only 15 percent (58/394) of these English sequences had long before short. Among those with at least a one-word weight difference (excluding 71 with equal weight), 82 percent had short before long, and there was a gradual reduction in the long before short orders, the bigger the weight difference (PPS = shorter PP, PPL = longer PP):

(2.5)	n = 323	PPL > PPS by 1 word	by 2–4	by 5–6	by 7+
	[V PPS PPL]	60% (58)	86% (108)	94% (31)	99% (68)
	[V PPL PPS]	40% (38)	14% (17)	6% (2)	1% (1)

In §2.5.1 and §3.5 below I discuss other structures of English that reveal a similar weight preference, e.g., Extraposition and Heavy NP Shift.

A possible explanation for the distribution in (2.5) can be given in terms of reduced processing demands in working memory. If, in (2.2a), the same phrase structure information can be derived from a five-word viewing window rather than nine words, then phrase structure processing can be accomplished sooner, there will be fewer additional (phonological, morphological, syntactic, and semantic) decisions that need to be made simultaneously with this one, and there will be less structural complexity to compute and fewer competing structural decisions to resolve (cf. Lewis and Vasishth 2005). Overall fewer demands will be made on working memory and on the computational system. (2.2a) is more efficient, therefore, but not because some claimed capacity limit has been breached. All the attested orderings in (2.5) are clearly within whatever limit there is. The graded nature of these data point instead to a preference for reducing simultaneous processing demands when combining words into phrases. More generally we can hypothesize that the processing of all syntactic and semantic relations prefers minimal domains, which is what MiD predicts.

Grammatical conventions of ordering across languages reveal the same degrees of preference for minimal domains. The relative quantities of attested languages reflect the preferences, as do hierarchies of co-occurring word orders (see, e.g., §5.5). An efficiency approach can also explain exceptions to the majority patterns and to grammatical principles such as consistent head

ordering. These points will be illustrated in greater detail in Chapter 5. Consider in this context a brief example.

Greenberg (1963) examined alternative verb positions across languages and their correlations with prepositions and postpositions in phrases corresponding to (2.6):

(2.6) a. $[_{VP}$ went$[_{PP}$ to the movies$]]$ (1 2)
b. $[[$the movies to $_{PP}]$ went $_{VP}]$ (1 2)
c. $[_{VP}$ went $[$the movies to $_{PP}]]$ (1 2 3 4)
d. $[[_{PP}$ to the movies$]$ went $_{VP}]$ (1 2 3 4)

(2.6a) is the English order, (2.6b) is the Japanese order, and these two sequences, with adjacent lexical heads (V and P), are massively preferred in language samples over the inconsistently ordered heads in (2.6c) and (2.6d). (2.7) summarizes the distribution using the database of Dryer's (1992) paper on the 'Greenbergian correlations' (see Hawkins 2004: 124):

(2.7) a. $[_{VP}$ V $[_{PP}$ P NP$]]$ = 161 (41%)
IC-to-word: 2/2 = 100%
b. $[[$NP P $_{PP}]$ V $_{VP}]$ = 204 (52%)
IC-to-word: 2/2 = 100%
c. $[_{VP}$ V $[$NP P $_{PP}]]$ = 18 (5%)
IC-to-word: 2/4 = 50%
d. $[[_{PP}$ P NP$]$ V $_{VP}]$ = 6 (2%)
IC-to-word: 2/4 = 50%
Assume: V = 1 word; P = 1; NP = 2
MiD-preferred (2.7a)+(b) = 365/389 (94%)

The adjacency of V and P and consistent head ordering to the left or right guarantees the smallest possible string of words (indicated by the word totals in (2.6)) and the highest IC-to-word ratios for the recognition and construction of VP and its two immediate constituents (ICs), namely V and PP. Non-adjacent V and P in (2.7c,d) require longer and less efficient strings for the parsing of phrase structure—i.e., positioning heads next to one another provides a minimal Phrasal Combination Domain for the construction of VP and its daughters. Similarly, in the performance data of (2.5), orderings are preferred in which heads are adjacent or as close to one another as possible.

Consistent head ordering in grammars can be argued to derive from Minimize Domains (2.1), therefore. Conventions of ordering have emerged out of performance preferences, and one and the same principle explains both the preferred conventions of grammar and the preferred structural selections in performance (in languages and structures in which speakers have a choice). MiD can also explain why there are two productive mirror image types here, head-initial and head-final languages exemplified by (2.7a) and (b): they are

equally good strategies for phrase structure comprehension and production (Hawkins 2004: 123–6; see further §5.1 and §7.1).

There are many similar grammatical universals that can be explained as conventionalizations of the gradient preferences defined by MiD. Consider relative clause formation. It involves a dependency between the head of the relative clause and the position relativized on, i.e., the gap, subcategorizer, or resumptive pronoun within the clause that is coindexed with the head (see Hawkins 1999, 2004: ch. 7 for a summary of the different formalizations and theories here and see §2.2.3). I have argued that various hierarchies can be set up on the basis of increasing domain sizes for relative clause processing, measured in terms of the smallest number of nodes and structural relations that must be computed in order to match the relative clause head with the coindexed gap, subcategorizer, or resumptive pronoun. Numerous correlating patterns from performance and grammars in different languages are presented in Hawkins (1999, 2004) that support this general explanation.

2.2 Efficiency principle 2: Minimize Forms

A second general efficiency principle defended at length in Hawkins (2004) involves the minimization not of domains of connected words but of individual linguistic forms. It is defined in (2.8):

(2.8) *Minimize Forms* (MiF)
The human processor prefers to minimize the formal complexity of each linguistic form F (its phoneme, morpheme, word, or phrasal units) and the number of forms with unique conventionalized property assignments, thereby assigning more properties to fewer forms. These minimizations apply in proportion to the ease with which a given property P can be assigned in processing to a given F.

The basic premise of MiF is that processing linguistic forms and their conventionalized property assignments (i.e., their meanings and grammatical properties) requires effort. Minimizing the forms required for property assignments reduces this effort by exploiting information that is already active in processing, through its accessibility and high frequency, and through inferencing strategies of various kinds. These processing enrichments avoid duplication in communication and are efficient.

MiF is visible in two complementary sets of variation data. One involves complexity differences between surface forms (morphology and syntax), with preferences for minimal expression (e.g., zero) in proportion to their frequency of occurrence and hence their degree of expectedness. Nominative

case is more frequent than accusative and singular number is more frequent than plural. Correspondingly, nominative and singular are more often expressed by zero forms than accusative and plural respectively (compare, for example, singular *cat* with plural *cats* in English). The overall number of formal units that speakers need to produce in communication is reduced when the more frequent and expected property values are assigned zero. Another dataset involves the number and nature of lexical and grammatical distinctions that languages conventionalize. The preferences are again in proportion to their efficiency, including frequency of use. There are preferred lexicalization patterns across languages, certain grammatical distinctions are cross-linguistically preferred (certain numbers, tenses, aspects, causativity, some basic speech acts, thematic roles like Agent and Patient), and so on. Let us consider these two sets of variation data in slightly more detail.

Examples abound suggesting that a reduction in form processing is efficient, as long as relevant information can be recovered. Consider the use of pronouns versus full NPs (*he*/*she* versus *the professor*). The high versus low accessibility of entities referred to in discourse correlates with less versus more formal structure respectively (see Ariel 1990). Also relevant here are Zipf (1949) effects (e.g., the shorter *TV* replaces the high-frequency *television*), compounds (*paper plate* is reduced from *plate made of paper; paper factory* from *factory that makes paper*: see Sperber and Wilson's 1995 theory of relevant real-world knowledge activated in the processing of these minimal structures), coordinate deletions (*John cooked 0 and Fred ate the pizza*; see Hawkins 2004: 93–5 and van Oirsouw 1987), and control structures involving understood subjects of verbs within non-finite subordinate clauses (whose controllers are in a structurally accessible matrix clause position; see Hawkins 2004: 97–101). Filler-gap dependencies in relative clauses (*the book*$_i$ *that I read* O_i) are also plausibly motivated by (2.8). Gaps can be identified by reference to the filler with which they are coindexed. The result is a more minimal structure than the resumptive pronoun counterparts (*the book*$_i$ *that I read it*$_i$) in languages such as Hebrew. The advantage of minimization disappears, however, in more complex environments in which processing demands and processing domains become larger (Hawkins 2004: 182–6; Ariel 1999).

Form reduction is supported further by the Economy Principle of Haiman (1983) and by the data he summarizes from numerous languages. It is also reminiscent of Grice's (1975) second Quantity maxim for pragmatic inferencing ('Do not make your contribution more informative than is required'), and more specifically of Levinson's (2000) Minimization principle derived from it ('Say as little as necessary', that is, produce the minimal linguistic information sufficient to achieve your communicational ends).

Principle (2.8) adds a second factor to this efficiency logic, beyond the simplicity or complexity of the surface forms themselves, and is defined in terms of properties that are conventionally associated with forms. It is not efficient to have a distinct form F for every possible property P that one might wish to express in everyday communication. To do so would greatly increase the number of form–property pairs in a language, so choices have to be made over which properties get priority for unique assignment to forms. The remaining properties are then assigned to forms that are ambiguous, vague, or zero-specified with respect to the property in question and/or these properties are derived by combinatorial and compositional processes that apply to the basic meaningful units.

There are numerous semantic and syntactic properties that are frequently occurring across languages and that have priority in grammatical and lexical conventions. The property of causation is often invoked in everyday language use and is regularly conventionalized in the morphology, syntax, or lexicon (Comrie 1989; Shibatani 1976). Agenthood and patienthood are also frequently expressed and given systematic (albeit partially different) formal expression in ergative-absolutive, nominative-accusative, and active languages (Primus 1999). The very frequent speech acts (asserting, commanding, and questioning) are each given distinct formal expression across grammars, whereas less frequent speech acts such as bequeathing or baptizing are assigned separate lexical items, but not a uniquely distinctive construction in the syntax (Sadock and Zwicky 1985). Within the lexicon the property associated with *teacher* is frequently used in performance, that of *teacher who is late for class* much less so. The event of *X hitting Y* is frequently selected, that of *X hitting Y with X's right hand* less so. The more frequently selected properties are conventionalized in single lexemes or unique categories and constructions in all these examples. Less frequently used properties must then be expressed through word and phrase combinations and their meanings must be derived by semantic composition. This makes the expression of more frequently used meanings shorter, that of less frequently used meanings longer, and this pattern matches the first pattern of less versus more complexity in the surface forms themselves correlating with relative frequency. Both patterns make utterances shorter and the communication of meanings more efficient overall, which is why I have chosen to capture them both in one common Minimize Forms principle (2.8).

MiF asserts further that form minimizations apply in proportion to the ease with which properties can be assigned to forms in processing. The mechanisms by which this is done are various and include processes variously described as processing enrichments, inferences, and implicatures (Grice 1975; Levinson 1983, 2000; Sperber and Wilson 1995). These processes often involve

accessing real-world knowledge and information given in a larger text or discourse (see Hawkins 2004: 44–9). They also include grammatically based enrichments to underspecified forms and sentence-internal dependencies of various sorts (e.g., filler-gap dependencies). The limits on these processes, e.g., limits on what can be readily inferred (see again Hawkins 2004: 44–9), provide a check on how far minimization can go (one cannot minimize everything and assign all properties through enrichment) and it enables us to make some testable predictions for grammars and performance:

(2.9) *Form Minimization Predictions*

a. The formal complexity of each F is reduced in proportion to the frequency of that F and/or the processing ease of assigning a given P to a reduced F (e.g., to zero).

b. The number of unique F:P1 pairings in a language is reduced by grammaticalizing or lexicalizing a given F:P1 in proportion to the frequency and preferred expressiveness of that P1 in performance.

In effect, form minimizations require compensating mechanisms in processing. (2.9a) asserts that frequency and processing ease regulate reductions in surface form (their associated properties are readily inferable), while frequency and preferred expressiveness regulate the grammaticalization and lexicalization preferences of (2.9b).

2.2.1 Greenberg's markedness hierarchies

The cross-linguistic effects of both (2.9a) and (b) can be seen in Greenberg's (1966) markedness hierarchies, such as those in (2.10) and their reformulations and revisions by later authors as shown:

(2.10) Sing > Plur > Dual > Trial/Paucal (for number) [Greenberg 1966; Croft 2003]
Nom/Abs > Acc/Erg > Dat > Other (for case marking) [Primus 1999]
Masc/Fem > Neut (for gender) [Hawkins 1998]
Positive > Comparative > Superlative [Greenberg 1966]

Greenberg pointed out that these grammatical hierarchies also define performance frequency rankings for the relevant properties in each domain. The frequencies of number inflections on nouns in a corpus of Sanskrit, for example, were:

(2.11) Singular = 70.3%; Plural = 25.1%; Dual = 4.6%

The other hierarchies had similar frequency correlates. In other words, these hierarchies appear to be **performance frequency rankings** defined on entities within common grammatical and/or semantic domains. The ultimate causes of the frequencies are presumably quite diverse and will involve a number of factors such as real-world frequencies of occurrence, communicative biases in favor of animates rather than inanimates, conceptual and communicative biases in favor of talking about singular entities, and syntactic and semantic complexity (cf. Haspelmath 2008a for discussion of the diverse causes of frequency effects). What is significant for grammars is that these performance rankings, whatever their precise causality in each case, are reflected in cross-linguistic patterns conventionalizing morphosyntax and allomorphy in accordance with (2.9a,b).

(2.12) *Quantitative Formal Marking Prediction*
For each hierarchy H the amount of formal marking (i.e., phonological and morphological complexity) will be greater or equal down each hierarchy position.

(2.12) follows from (2.9a). In Manam the 3rd singular suffix on nouns is zero, the 3rd plural is *-di*, the 3rd dual is *-di-a-ru*, and the 3rd paucal is *-di-a-to* (Lichtenberk 1983). The amount of formal marking increases from singular to plural and from plural to dual, and is equal from dual to paucal, in accordance with the hierarchy in (2.10). Similarly English singular nouns are zero-marked whereas plurals are formally marked, generally with an *-s* allomorph.

(2.13) *Morphological Inventory Prediction*
For each hierarchy H (A > B > C) if a language assigns at least one morpheme uniquely to C, then it assigns at least one uniquely to B; if it assigns at least one uniquely to B, it does so to A.

(2.13) follows from (2.9b). A distinct dual implies a distinct plural and singular in the grammar of Sanskrit, and a distinct dative implies a distinct accusative and nominative in the case grammar of Latin and German (or a distinct ergative and absolutive in Basque; cf. Primus 1999). A unique number or case assignment low in the hierarchy implies unique and differentiated numbers and cases in all higher positions.

(2.14) *Declining Distinctions Prediction*
For each hierarchy H the number of morphological distinctions expressed through feature combinations at a given level of H will be greater than or equal to the number expressed at each lower position.

(2.14) also follows from (2.9b). For example, when gender features combine with and partition number we find that unique gender-distinctive pronouns

can exist for the singular and not for the plural in English (*he*/*she*/*it* vs *they*), whereas the converse uniqueness is not found (with a gender-distinctive plural, but gender-neutral singular). See Hawkins (2004: 68–79) for detailed testing of (2.13) and (2.14) on most of the dialects of a single language family, Germanic, in its historical evolution.

More generally, (2.13) and (2.14) lead to a general principle of cross-linguistic morphology:

(2.15) *Morphologization*
A morphological distinction will be grammaticalized in proportion to the performance frequency with which it can uniquely identify a given subset of entities {E} in a grammatical and/or semantic domain D.

This principle enables us to make sense of 'markedness reversals.' For example, in certain nouns in Welsh whose referents are much more frequently plural than singular, like 'leaves' and 'beans,' it is the singular form that is morphologically more complex than the plural, i.e., *deilen* ('leaf') vs *dail* ('leaves'), *ffäen* ('bean') vs *ffa* ('beans'); cf. Haspelmath (2002: 244).

2.2.2 Wasow et al.'s relativizer omission data

A fascinating recent study by Wasow, Jaeger, and Orr (2011) on relativizer omission in English in non-subject relative clauses (NSRCs) provides syntactic data of relevance to Minimize Forms (2.8), i.e., in structures like *the man I know* versus *the man who(m)*/*that I know*. Their paper shows that there is a clear correlation between the frequency with which a given determiner, prenominal adjective, or head noun is actually followed by NSRCs in various corpora and the frequency of relativizer omission. For example, relativizers in NSRCs are omitted in 66 percent of NPs with the definite article, but in only 25 percent after the indefinite *a(n)*; in 76 percent after universal *all* and 85 percent after universal *every*, but in just 36 percent after existential *some*; in 90 percent of NSRCs following a prenominal superlative adjective *last* and 82 percent following *first*, but in just 26 percent following an adjective like *little*; in 87 percent of NSRCs following the head noun *way* and 86 percent after *time*, but in just 43 percent following *people*; and so on. Matching these relativizer omission frequencies are frequencies for the actual occurrence vs non-occurrence of an NSRC after the relevant determiner, adjective, and head noun. For example, NPs with definite articles have over five times more NSRCs than NPs with *a(n)*, specifically 5.93 percent for the definite (1813/30,587) versus 1.18 percent (302/45,698) for the indefinite. This greater

frequency of occurrence for accompanying NSRCs matches the greater frequency of relativizer omission (66% for *the*, 25% for *a(n)*).

In order to make sense of this correlation, Wasow et al. argue, first, that the frequencies with which NPs of different types contain NSRCs can be understood in terms of their semantic and pragmatic properties. NSRCs provide a clause that restricts the reference of the nominal head, and this restriction is often necessary in discourse in order to provide the required referential uniqueness for *the* or for superlative adjectives, or in order to delimit a universal claim so that it can be true. These (independently explainable) frequencies make the occurrence of a following relative clause more or less predictable in performance, and the more predictable a relative clause is, the less necessary the relativizer becomes in order to construct it in comprehension and production. Building on an insight in Fox and Thompson (2007), Wasow et al. propose the following Predictability Hypothesis: 'In environments where an NSRC is more predictable, relativizers are less frequent.'

These syntactic data and the Predictability Hypothesis are further instances, as I see it, of Minimize Forms and of the prediction in (2.9a). By omitting a relativizer, the formal complexity of the relative clause is reduced. In the process an explicit signal of relative clause status is removed and an additional dependency is set up between the reduced clause and the head noun whereby the head needs to be accessed in parsing in order to recognize that this clause is a relative clause and in order to attach it to the head (see the parsing definition of dependency given in (2.1) and the reasoning behind this definition in Hawkins 2004: 18–25). (2.9a) asserts that this reduction can happen in proportion to the frequency of a syntactic form F (here a relative clause) and in proportion to its ease of processing, either from greater frequency or from some other factor which renders the syntactic and semantic properties (P) of this F recognizable in its reduced form. This provides a syntactic parallel to the morphological data of the last section (§2.2.1). The phonological complexity of morphemes could be reduced in proportion to the frequencies of occurrence for the properties assigned to them. The morphosyntactic complexity of relative clauses can similarly be reduced in proportion to their frequencies of occurrence following the relevant determiners or adjectives or nouns in question. These frequencies make their associated properties more predictable in performance.

Predictability can sometimes lead to default interpretations such as 'zero on nouns means (the more frequently used) singular interpretation rather than a plural' or 'zero is assigned absolutive rather than ergative case,' and these defaults can be conventionalized in grammars (see §4.4). A zero singular has been conventionalized for English nouns (*cat* versus *cats*). A zero absolutive has been conventionalized in many ergative languages such as Avar

(see §8.1). English has not conventionalized a default interpretation for clauses, however, to the effect that 'a zero-marked clause is assigned relative clause status,' because other types of (main and subordinate) clauses also have zero initially (see Hawkins 2004: ch. 6) and because the predictability of a following relative clause is highly gradient and sensitive to numerous and diverse co-occurring items within the preceding noun phrase, all of which rules out a straightforward formal rule corresponding to 'zero noun receives singular interpretation' (see further §4.3). The co-occurrence frequencies and predictabilities to which Wasow et al. have drawn attention can be seen clearly in performance, however, in different rates for relativizer omission that are statistically significant and these rates support the form minimization prediction (2.9a).

2.2.3 Gaps and resumptive pronouns in relative clauses

Another pattern that is relevant for Minimize Forms (2.8) and (2.9a) was first pointed out by Keenan and Comrie (1977) and involves data from relative clauses that have a gap in the position relativized on ([–Case] in their terminology) versus resumptive pronouns (as a type of [+Case] relativization strategy). The difference between the two can be illustrated with the following pair from Hebrew (Ariel 1990). (2.16a) contains the gap, (2.16b) the resumptive pronoun.

(2.16) a. Shoshana hi ha-isha*i* [she-nili ohevet 0*i*] (Hebrew)
Shoshana is the-woman that-Nili loves

b. Shoshana hi ha-isha*i* [she-nili ohevet ota*i*]
that-Nili loves her

The choice between zero and an explicit element now involves a position internal to the relative clause relevant for processing of the head's role within the relative clause, as opposed to the initial relativizer considered in the last section which serves to construct the relative clause and make it recognizable as such. The distribution of gaps to resumptive pronouns across languages follows the Keenan and Comrie Accessibility Hierarchy (AH), but in a reverse manner that is surprising from a purely grammatical perspective. Their hierarchy was initially proposed on the basis of 'cutoffs' across languages; that is whether relativization on a given position was possible at all in a language, using any strategy for relative clause formation. I summarize the AH in (2.17), with English examples for each position.

(2.17) *Accessibility Hierarchy* (AH): SU>DO >IO/OBL>GEN

a. the professor*i* [that 0*i* wrote the letter] *SU*
b. the professor*i* [that the student knows 0*i*] *DO*
c. the professor*i* [that the student showed the book to 0*i*] *IO/OBL*
d. the professor*i* [that the student knows his*i* son] *GEN*

The cross-linguistic pattern in their data is: if a gap strategy is grammatical on a low position of AH in a language, it is grammatical on all higher positions; if a resumptive pronoun is grammatical on a high position, it is grammatical on all lower positions (that can be relativized at all); see Hawkins (1999, 2004: 186–90) for details. This can be seen in Table 2.1 in which I quantify the distribution of gaps to pronouns for twenty-four languages in the Keenan–Comrie

TABLE 2.1 Languages combining [–Case] gaps with [+Case] pronouns (Keenan and Comrie 1977)

	SU	DO	IO/OBL	GEN
Aoban	gap	pro	pro	pro
Arabic	gap	pro	pro	pro
Gilbertese	gap	pro	pro	pro
Kera	gap	pro	pro	pro
Chinese (Peking)	gap	gap/pro	pro	pro
Genoese	gap	gap/pro	pro	pro
Hebrew	gap	gap/pro	pro	pro
Persian	gap	gap/pro	pro	pro
Tongan	gap	gap/pro	pro	pro
Fulani	gap	gap	pro	pro
Greek	gap	gap	pro	pro
Welsh	gap	gap	pro	pro
Zurich German	gap	gap	pro	pro
Toba Batak	gap	*	pro	pro
Hausa	gap	gap	gap/pro	pro
Shona	gap	gap	gap/pro	pro
Minang-Kabau	gap	*	*/pro	pro
Korean	gap	gap	gap	pro
Roviana	gap	gap	gap	pro
Turkish	gap	gap	gap	pro
Yoruba	gap	gap	0	pro
Malay	gap	gap	RP	pro
Javanese	gap	*	*	pro
Japanese	gap	gap	gap	gap/pro
Gaps =	24 [100%]	17 [65%]	6 [25%]	1 [4%]
Pros =	0 [0%]	9 [35%]	18 [75%]	24 [96%]

Key: gap = [–Case] strategy
pro = copy pronoun retained (as a subinstance of [+Case])
* = obligatory passivization to a higher position prior to relativization
0 = position does not exist as such
RP = relative pronoun plus gap (as a subinstance of [+Case])
[–Case] gap languages may employ a general subordination marker within the relative clause, no subordination marking, a participial verb form, or a fronted case-invariant relative pronoun. For Tongan, an ergative language, the top two positions of AH are absolutive and ergative respectively, not SU and DO; cf. Primus (1999).

sample that have both. Gaps decline down the AH, 100 percent to 65 percent to 25 percent to 4 percent, while pronouns increase (0 percent to 35 percent to 75 percent to 96 percent).

Gaps are associated with simpler and smaller 'filler-gap domains' such as SU and DO and extend to lower positions only if all higher AH positions also permit a gap. Conversely pronouns favor more complex environments (such as GEN and OBL) and extend to simpler ones only if the complex positions also permit a pronoun. For a full and recent literature review of experimental work in psycholinguistics testing for ease and difficulty in relative clause processing, especially with respect to SU and DO positions of the AH, see Kwon et al. (2010). For a review of the performance support for all positions, see Hawkins (2004: 180–5).

The distribution of gaps in Table 2.1 accords well with Minimize Forms and (2.9a). Gaps are found in easier-to-process environments, whereas the more explicit resumptive pronouns that actually encode the position relativized on are found in more complex relatives. I have argued (Hawkins 1999, 2004: 180–5, following Keenan and Comrie's intuition) that resumptive pronouns aid processing by providing an explicit signal, or flag, for understanding the head's role within the relative clause. But in addition resumptive pronouns minimize the domains for processing co-occurrence relations within the relative clause itself. In (2.16b) the pronoun *ota* provides a local argument for processing the lexical co-occurrences of the verb *natan* (loves) and the search for this argument does not need to extend to the head of the relative itself (*isha*), making the lexical domain for the verb's arguments non-minimal. Only coindexing need apply non-locally linking $isha_i$ and la_i. The resumptive pronoun eases overall processing load, therefore, by minimizing some of the domains in relative clause processing, but at the expense of processing an additional form (the pronoun). We see here increasing competition between Minimize Domains (2.1) and Minimize Forms (2.8) down the AH. The minimal forms (the gaps) are preferred in relativization domains that are easier to process and structurally more minimal, but as these domains become more complex there is an increasing tension between the benefits of less form processing and the expanding environments for processing both the filler-gap relationship and the lexical relations between the gap and its local subcategorizer—i.e., there is an increasing tension between Minimize Forms (2.8) and Minimize Domains (2.1).

Corpus data from Hebrew (given in Ariel 1999) provide performance support for the grammatical pattern involving gaps versus pronouns given in Table 2.1. Ariel shows that the Hebrew gap is favored with smaller distances between filler and gap. For example, (2.16a), with a minimal distance between filler and gap, is significantly preferred over (2.16b) with a resumptive

pronoun. The resumptive pronoun becomes productive when filler-gap domains are larger, as in (2.18).

(2.18) Shoshana hi ha-isha*i* [she-dani siper she-moshe rixel
Shoshana is the-woman that-Danny said that-Moshe gossiped
she-nili ohevet ota*i*]
that-Nili loves her

Similarly, numerous grammatical rules in particular languages confirm the pattern of gaps in smaller relativization domains and pronouns in larger ones, e.g., in Cantonese. The pronoun is ungrammatical in the simple relative (2.19b) but grammatical in (2.20), in which there is a bigger distance between coindexed pronoun and relative clause head, i.e., a more complex relativization domain (Matthews and Yip 2003):

(2.19) a. [Ngo5 ceng2 0*i*] go2 di1 pang4jau5*i* (Cantonese)
I invite those CL friend
'friends that I invite'

b. *[Ngo5 ceng2*i* keoi5dei6*i*)] go2 di1 pang4jau5*i*
I invite them those CL friend

(2.20) [Ngo5 ceng2 (keoi5dei6*i*) sik6-faan6] go2 di1 pang4jau5*i*
I invite (them) eat-rice those CL friend
'friends that I invite to have dinner'

In Hawkins (1999, 2004: 186–90) I defined this relative complexity down the AH in terms of expanding syntactic domains for relative clause processing. I also showed how the explicit resumptive pronoun reduces domain sizes. The increasing strain on Minimize Domains (2.1) and the competition between it and Minimize Forms (2.8) can be quantified in this way. Other factors apart from the syntactic size of a relativization domain will also be relevant for an assessment of processing ease and for the occurrence of gaps, just as they are for relativizer omission in §2.2.2, since greater processing ease can have different causes. Ariel (1999) points out that there are more resumptive pronouns in adjunct rather than argument positions in Hebrew. The greater preference for gaps in argument positions could be on account of their greater predictability following the verb, if we think in terms of the predictability theories of Levy (2008) and Jaeger (2006). Similarly parallel function effects, whereby the head of the relative and the position relativized on are both subjects, or both direct objects, etc., make processing easier and can extend gap strategies in certain languages beyond their normal constraints. Parallel function has traditionally been shown to facilitate both acquisition and

processing (Sheldon 1974; MacWhinney 1982; Clancy et al. 1986). Aldai (2003) points out its relevance for grammars as well.

For example, relativization in Basque is fully productive on SU, DO, and IO positions using a gap strategy as shown in (2.21a,b,c), from Aldai (2003).

(2.21) a. [0*i* emakumea-ri liburua eman dio-n] gizona*i* (Basque)
woman-DAT book-ABS given AUX-REL man-ABS
'the man who has given the book to the woman'

b. [gizona-k 0*i* emakumea-ri eman dio-n] liburua*i*
man-ERG woman-DAT given AUX-REL book-ABS
'the book which the man has given to the woman'

c. [gizona-k liburua 0*i* eman*i* dio-n] emakumea*i*
man-ERG book-ABS given AUX-REL woman-ABS
'the woman to whom the man has given the book'

Relativization on an oblique (OBL) phrase such as the locative in (2.22) is not normally possible:

(2.22) *[mendi-an 0*i* ibili naiz-en] emakumea*i* ederra da (Basque)
mountain-LOC walked I-am-REL woman beautiful is
'the woman (with whom) I have walked in the mountains is beautiful'

But when there is parallel function between the head of the relative and the position relativized on, the productivity of the gap strategy is increased and it is systematically extended to OBL as well:

(2.23) [mendi-an 0*i* ibili naiz-en] emakumea-rekin ezkondu
mountain-LOC walked I-am-REL woman-COMIT marry
nahi dat (Basque)
desire I-have
'I want to marry (with) the woman (with whom) I have walked in the mountains.'

Kirby (1999) made the claim that certain processing pressures have been conventionalized in grammars (like Minimize Domains, formerly EIC), whereas others have not, and he included parallel function effects in this latter category. Yet Basque appears to be a language in which the grammar has conventionalized additional relative clause formation possibilities in parallel function environments.

Similarly in Hebrew, relativization on IO and OBL generally requires a resumptive pronoun as shown in (2.24). Parallel function extends the well-formedness of the gap strategy from DO as in (2.16a) to IO and OBL as shown in (2.25), as pointed out by Cole (1976: 244–5):

(2.24) ha-isha*i* [she-dani natan la*i* et ha-sefer] (Hebrew)
the-woman that-Danny gave to-her ACC the-book
'the woman that Danny gave the book to'

(2.25) a. natati sefer le oto yeled*i* [she-Miriam natna 0*i* sefer] (Hebrew)
I-gave book DAT same boy that Mary gave book
'I gave a book to the very boy to whom Mary gave a book'

b. yashavta al kise*i* [she-Ben-Gurion yashav 0*i*]
You-sat upon chair that Ben-Gurion sat
'You sat on a chair on which Ben-Gurion sat.'

Another factor of clear relevance for ease and difficulty of processing is frequency of occurrence. There are frequency correlations for relativizations on the different positions of the Accessibility Hierarchy (see Hawkins and Buttery 2010 for sample data from the British National Corpus supporting this and Keenan 1975 for some early corpus data). Since these frequencies correlate with structural complexity and with the metrics defined in Hawkins (2004: 177–80) it is hard to tease apart whether structure or frequency is the greater contributor to ease of processing or whether both are relevant. In German, for example, Diessel and Tomasello (2006) point out that ditransitive verbs will have the same equally complex relativization domain as defined in Hawkins (2004: 177–80) whether the gap is SU, DO, or IO, because of the final position of the verb and the initial position of the head noun in this NRel structure. This is correct, but I pointed out (Hawkins 2004: 180) that my complexity ranking derived from AH still holds across different (intransitive/transitive/ditransitive) verbs in German, though not for relativizations on arguments of one and the same verb. The precise causality of greater complexity in this typologically mixed language appears to be a subtle mix of structural and frequency factors, therefore.

Also relevant in this context are violations of various island constraints that are possible for relatives containing gaps when a pragmatic context is set up that is rich enough to make them easy to process (see Erteschik-Shir and Lappin 1979; Erteschik-Shir 1992).

The preference for gaps in easy-to-process environments is clear from all these examples and supports Minimize Forms and (2.9a). Resumptive pronouns are found in more complex environments and in a subset of these the increasing complexity can be clearly defined and quantified in terms of non-minimal domains for the processing of items within the relative clause. The variation between gaps and resumptive pronouns in Table 2.1 can be seen in terms of cooperation and competition between Minimize Forms (2.8) and Minimize Domains (2.1), therefore. At the top end (SU), there are 100 percent

gaps: both forms and processing domains are minimal here. At the bottom end (GEN) almost all languages choose resumptive pronouns, providing an explicit pronoun for local processing of the genitive head relation within a minimal domain for the containing NP, i.e., Minimize Domains wins. For DO relatives there are 35 percent resumptive pronouns, and for IO/OBL 75 percent. Minimize Forms wins the former (by 65 to 35 percent), while Minimize Domains wins the latter (75 to 25 percent). I shall return to consider this competition in a broader context in §9.3.

2.3 Efficiency principle 3: Maximize Online Processing

The third principle proposed in Hawkins (2004) is (2.26):

(2.26) *Maximize Online Processing* (MaOP)
The human processor prefers to maximize the set of properties that are assignable to each item X as X is processed, thereby increasing O(nline) P(roperty) to U(ltimate) P(roperty) ratios. The maximization difference between competing orders and structures will be a function of the number of properties that are unassigned or misassigned to X in a structure/sequence S, compared with the number in an alternative.

This principle involves the timing with which linguistic properties are introduced in online processing. There is a preference for selecting and arranging linguistic forms so as to provide the earliest possible access to as much of the ultimate syntactic and semantic representation as possible. What is dispreferred is, first, any significant delay or 'look ahead' (Marcus 1980) in online property assignments, referred to as 'unassignments' in Hawkins (2004: 51–2); and second a 'misassignment' of properties online (Hawkins 2004: 53–5). Misassignments result in so-called garden path effects whereby one analysis is chosen online and is then subsequently corrected in favor of a different analysis when more material has been processed. A famous example is *the horse raced past the barn fell* which is first assigned a main clause reading and then a reduced relative reading when the (matrix verb) *fell* is encountered (see MacDonald et al. 1994). Such backtracking is difficult for the processor, but it is also inefficient since initial property assignments are wasted and make no contribution to the ultimate syntactic and semantic representation of the sentence. Unassignments are also inefficient since they delay the assignment of certain properties to forms in online processing. See Hawkins (2004: 55–8) for a metric that calculates the relative severity of garden paths (misassignments) and of competing orders and structures with respect to their unassignments, using 'OP-to-UP ratios.'

Clear evidence for (2.26) comes from a number of patterns across languages that involve asymmetrical orderings between two categories A and B. Ordering A before B maximizes online processing, since the reverse would involve significant unassignments or misassignments, and MaOP provides a plausible explanation for these conventionalized asymmetries. A sample is given in (2.27), together with my best estimate of the level of quantitative support for each preference.

(2.27) a. Displaced Wh preposed to the left of its (gap-containing) clause [almost exceptionless; cf. Hawkins 1999, 2004; Polinsky 2002]
Who*i* [did you say 0*i* came to the party]

b. Topic to the left of a dependent predication [exceptionless for some dependencies, highly preferred for others, cf. below and Hawkins 2004]
E.g., Japanese: *John wa gakusei desu* 'Speaking of John, he is a student' (Kuno 1973)

c. Head noun (filler) to the left of its (gap-containing) relative clause
E.g., the students*i* [that I teach 0*i*]
If a language has basic VO, then N+Relative [exceptions = rare; see §2.5.2, §7.3, and Hawkins 1983, 2004]

VO	*OV*
NRel (English)	NRel (Persian)
*RelN	RelN (Japanese)

d. Antecedent precedes anaphor [highly preferred cross-linguistically]
E.g., ***John*** *washed* ***himself*** (SVO), *Washed* ***John himself*** (VSO), ***John himself*** *washed* (SOV) = highly preferred over, e.g., *Washed* ***himself John*** (VOS)

e. Wide-scope quantifier/operator precedes narrow-scope Q/O [highly preferred]
E.g., ***Every student a book*** *read* (SOV languages) ∀∃ preferred ***A book every student*** *read* (OSV orders in SOV languages) ∃∀ preferred

f. Restrictive relative precedes appositive relative (Hawkins 2002, 2004)
If N+Relative, then restrictive before appositive relative [exceptionless?]
E.g., *Students that major in mathematics, who must work very hard* (R+A)
**Students, who must work very hard, that major in mathematics* (A+R)

In these asymmetric orders there is generally an asymmetric dependency of B on A: the gap is dependent on the filler (for gap-filling), the anaphor on its antecedent (for coindexation), the predication on a topic (for, e.g., argument assignment), the narrow-scope quantifier on the wide-scope quantifier (the number of books read depends on the quantifier in the subject NP in *Every student read a book*/*Many students read a book*/*Three students read a book*), and so on. The assignment of dependent properties to B is more efficient when A precedes, since these properties can then be assigned immediately in online processing to B. In the reverse B + A there would be delays in property assignments to the dependent B online (unassignments) or misanalyses (misassignments).

2.3.1 Fillers First

For example, if the gap were to precede the 'filler' Wh-word in a structure like [*you said* 0_i *came to the party*] *who*$_i$, there would be a delay in assigning the subject argument to *came*. Gaps are dependent on their fillers for coindexation and coreference, and also for recognizing the position to be filled (in conjunction with access to the subcategorizor, if there is one), whereas fillers are not so dependent on their gaps. This results in the preference for fillers before gaps or Fillers First (2.27a,c); see Hawkins (1999, 2004) and J. D. Fodor (1983). When the gap follows the filler, the filler can be fully processed online, and the properties of the gap that are assigned by reference to the filler can be immediately assigned to the gap, resulting in a more efficient distribution of property assignments throughout the sentence. If the gap precedes, its full properties can only be assigned retrospectively when the filler is encountered, resulting in a processing delay and in frequent garden path effects as matrix and subordinate clause arguments are redistributed to take account of a gap that is activated by late processing of the filler (see Antinucci et al. 1979; Clancy et al. 1986; Phillips 1996). Fillers First maximizes online property assignments, therefore.

When the filler is a Wh-word in a Wh-question (2.27a) there is unambiguous empirical support for Fillers First: almost all languages that move a Wh-word to clause-peripheral position move it to the left, not to the right (Hawkins 2002, 2004: 190–2 and §7.9 below). In relative clauses (3.27c) there is also clear support, but Fillers First is now in partial conflict with a Minimize Domains preference for noun-final NPs in head-final languages (Hawkins 2004: 205–10, §2.5.2, and §7.3). Head-initial languages have consistently right-branching relatives (e.g., [V [N S]]), which are motivated both by MiD and by Fillers First. But head-final languages have either left-branching relatives ([[S N] V]), which are good for MiD but which position the gap before the filler,

or right-branching relatives ([[N S] V]), which are good for Fillers First but which create non-adjacency between heads and make domains for phrasal processing longer. The variation here points to the existence of two preferences, whose predictions overlap in one language type but conflict in the other (see §9.3).

The head-final languages that prefer left-branching relatives appear to be the *rigid* ones like Japanese, in which there are more containing head-final phrases (such as V-final VPs) that prefer the head of NP to be final as well (by MiD). *Non-rigid* head-final languages have fewer containing phrases that are head-final and so define a weaker preference for noun finality, allowing Fillers First to assert itself more, which results in more right-branching relatives; cf. Lehmann (1984) for numerous exemplifying languages and §7.3 in the present volume.

2.3.2 Topics First

A related structure involves topicalized XPs with gaps in a sister S. These generally precede the S across languages (Gundel 1988; Primus 1999). The reverse ordering would be just as good for indicating that the sister S and its constituents are within the scope of the topic XP, but this reverse ordering is either ungrammatical or dispreferred, which provides further evidence for MaOP. The asymmetry disappears when a coindexed pronoun replaces the gap, resulting in left- or right-dislocation structures (e.g., *my brother*i*, he*i *is a nice guy* versus *he*i *is a nice guy, my brother*i), suggesting that it is the gap that contributes substantially to the linear precedence asymmetry here. The preference for Topics First is motivated by the dependence of the gap on the filler for gap identification and filling, as before. In addition, the 'aboutness' relation between the predication and the topic (Reinhart 1982; Jacobs 2001), coupled with the regular referential independence or givenness of the topic, means that semantic processing of the predication is often incomplete without prior access to the topic, whereas the topic can be processed independently of the predication. For example, Tsao (1978) gives numerous examples from Mandarin Chinese of a topic phrase providing information that is required for interpretation of the predication, making these predications *dependent* on the topic as this term is defined here (see (2.1) which follows Hawkins 2004: 18–25). These examples include:

argument assignment to the subcategorizor in the predication (Tsao's transcription of Mandarin Chinese is preserved here):

(2.28) **Jang San** (**a**), dzwo-tyan lai kan wo. (argument assignment)
Jang San (TOPIC PART), yesterday come see me
'Yesterday Jang San came to see me.'

various *argument enrichments* whereby the topic provides a *possessor* (2.29), *class* (2.30), *set* (2.31), or *restrictive adjunct* (2.32) relative to which an argument in the predication is interpreted:

(2.29) **Jei-ge ren** (**a**), tounau jyandan. (argument enrichment:
This-CLASSIF man (TOPIC PART), mind simple possessor-possessed)
'This man's mind is simple.'

(2.30) **Wu-ge pinggwo** (**a**), lyang-ge hwai-le. (argument enrichment:
Rice-CLASSIF apples (TOPIC PART), two-CLASSIF spoiled class member)
'Two rice apples are spoiled.'

(2.31) **Ta-de san-ge haidz** (**a**), yi-ge dang lyushr (argument
His three-CLASSIF children (TOPIC PART), one-CLASSIF serve-as lawyer enrichment:
'One of his three children serves as a lawyer.' set member)

(2.32) **Jei-jyan shr** (**a**), wo-de jingyan tai dwo-le. (argument
This-CLASSIF matter (TOPIC PART), my experience too many enrichment:
restrictive adjunct)
'My experience in this matter is too considerable.'

various *predicate enrichments* whereby the topic provides a *location* (2.33), *time* (2.34), or *cause* (2.35) adjunct, or a domain for a *superlative* (2.36) interpretation relative to which the predication is interpreted.

(2.33) **Nei kwai tyan** (**a**), daudz jang de hen da. (predicate enrichment:
That piece land (TOPIC PART), rice grows PART very big location)
'Rice grows very big in that piece of land.'

(2.34) **Dzwo-tyan** (**a**) Jang San lai kan wo. (predicate enrichment:
Yesterday (TOPIC PART), Jang San come see me time)
'Yesterday Jang San came to see me.'

(2.35) **Weile jei-ge haidz**, wo bu je chr-le dwoshau ku. (predicate enrichment:
For this-CLASSIF child, I not Know eat-Asp how much hardship cause)
'I have endured much hardship on account of this child.'

(2.36) **Yu** (**a**), wei-yu syandzai dzwei gwei. (predicate enrichment:
Fish (TOPIC PART), tuna now most expensive superlative domain)
'Tuna is now the most expensive type of fish.'

If predication and topic were reversed in these examples, there would be little impact on the online processing of the topic, but significant aspects of the interpretation of the predication would be delayed, i.e., there would be online unassignments and misassignments. For example, in (2.29) it would be unclear whose mind was intended, in (2.32) the absence of the restriction imposed by the topic would lead to an overly general interpretation online

that could be untrue (namely my experience in general vs my experience in this matter), in (2.36) the expensiveness of tuna needs to be interpreted relative to fish, rather than, say, food in general, and unless this restriction is contextually given it cannot be assigned online to the predication if the topic, fish, follows.

These asymmetries predict a topic + predication ordering preference, avoiding temporary unassignments or property misassignments online. Across languages, argument enrichments and predicate enrichments (i.e., with fully asymmetric dependencies) appear to be uniformly ordered this way (Gundel 1988), i.e., for gap-containing non-dislocation predications. Argument assignment dependencies like (2.28) (which are predominantly but not fully asymmetric since a topic can also be dependent on the predication for thematic role assignment) are preferably topic + predication (Hawkins 2004: 238–40).

2.3.3 Other linear precedence asymmetries

Further ordering asymmetries that are plausibly motivated by MaOP include the preference for antecedents before their anaphors (dependent on the former for coindexing and coreference (2.27d)), and wide-scope before narrow-scope operators and quantifiers (2.27e). Positioning the wide-scope item first permits immediate assignment of the appropriate interpretation to the narrow-scope item, by reference to the already processed wide-scope item, and avoids un-/misassignments online. Compare the different interpretations of the indefinite singular *a book* in *All the students read a book/Some students read a book/Three students read a book*. When *a book* precedes (*A book all the students read*, etc.) there is no higher-scope element in working memory relative to which a narrow-scope interpretation can be assigned, and the preferred interpretation shifts to wide scope.

Also relevant here is the preference for restrictive ('R') before appositive ('A') relatives exemplified by (2.37) in English (cf. (2.27f)):

(2.37) a. Students that major in mathematics, who must of course work hard, . . . R + A

b. *Students, who must of course work hard, that major in mathematics, . . . A + R

In the online processing of (2.37b) there would always be a semantic garden path. The appositive relative would first be predicated of all students, and would then be revised to a predication about that subset only of which the restrictive relative was true, once the latter was encountered and processed. The ordering of (2.37a) avoids the regular garden path by placing together all

the items that determine the reference set of which the appositive clause is predicated, positioning them before the appositive claim in surface syntax. R+A appears to be widespread in head-initial languages. For head-final languages, see Hawkins (2004: 241) and Lehmann (1984: 277–80).

Notice finally that in contrast to the asymmetrical dependencies of (2.27), dependencies between a verb and an NP direct object are broadly symmetrical. The NP depends on V for case- and theta-role assignment and also for mother node recognition (VP) and attachment (Hawkins 1994), while V depends on NP for selection of the intended syntactic and semantic co-occurrence frame (e.g., transitive vs intransitive *run* [*John ran*/*John ran the race*]), and for the intended semantics of V from among ambiguous or polysemous alternatives (*ran the race*/*the water*/*the advertisement*/*his hand through his hair*; Keenan 1979). These symmetrical dependencies are matched by symmetrical ordering patterns across languages (A+B/B+A), e.g. VO & OV. Asymmetrical orderings therefore appear to involve strong asymmetries in dependency (as defined here in processing terms, see (2.1) above), whereas symmetrical dependencies result in symmetrical orderings (Hawkins 2004: ch. 8).

2.4 The relationship between complexity and efficiency

There is a clear relationship between simplicity/complexity on the one hand and efficiency/inefficiency on the other. If the speaker can refer to a given entity or event with a simpler and shorter form or structure then this is generally more efficient than using a longer and more complex one. The pronoun *he* is formally much simpler than a full and complex referential phrase like *the professor who was teaching my linguistics class this morning*. But in the event that speaker and hearer do not share the rich knowledge base that permits the simpler *he* to be successful in a particular discourse then the more complex expression is required. On these occasions it is not efficient for the speaker to use *he* since this does not achieve the communicative goal, namely successful reference, whereas the more complex expression does.

What this example shows is that simplicity/complexity and efficiency/inefficiency are not always aligned. More generally what it shows is that efficiency relates to the basic function of language, which is to communicate information from the speaker (S) to the hearer (H), and to the manner in which this function is performed. I propose the definition in (2.38):

(2.38) Communication is efficient when the message intended by S is delivered to H in rapid time and with the most minimal processing effort that can achieve this communicative goal.

and the following hypothesis:

(2.39) Acts of communication between S and H are generally optimally efficient; those that are not occur in proportion to their degree of efficiency.

Complexity metrics, by contrast, are defined on the grammar and structure of language itself. An important component of efficiency often involves structural and grammatical simplicity—for example, when minimizing the domains for processing phrase structure groupings or filler-gap dependencies (see §2.1). But sometimes efficiency results in greater complexity, as we have seen. And it also involves additional factors that determine the speaker's selections of forms and structures, leading to the observed preferences of performance, including: fine-tuning to frequency of occurrence, accessibility, and inference (see the principle of MiF, §2.2); speed in delivering linguistic properties in online processing (see the principle of MaOP, §2.3); and reduced unassignments and misassignments (garden paths) in online processing (see also MaOP, §2.3).

These factors interact, sometimes reinforcing, sometimes opposing one another. In my earlier books (Hawkins 1994, 2004) I presented evidence that grammatical conventions across languages have conventionalized these performance factors and reveal a similar interaction and competition between them. Comparing grammars in terms of efficiency rather than complexity alone gives us a fuller and more accurate picture of the forces that underlie cross-linguistic variation. The variation patterns also provide quantitative data that can be used to determine the relative strength of different factors and the manner of their interaction (see Chapter 9).

Analysis of complexity alone without regard for broader communicative goals has also, I would argue, led to some unresolvable problems. Many of these either do not arise or can be solved within the broader efficiency perspective proposed here.

Theories of linguistic complexity share the guiding intuition that 'more [structural units/rules/representations] means more complexity.' This intuition has actually proved hard to define. See, for example, the lively debate in the papers of *Linguistic Typology* 5 (2/3) (2001), responding to McWhorter's (2001) claim that "the world's simplest grammars are creole grammars." The intuition works well enough when we compare one and the same grammatical area across languages, like different structural distances between fillers and gaps (Hawkins 1999, 2004: ch. 7). We can also measure consonant inventory sizes and rank languages according to the number and nature of their consonant phonemes (Lindblom and Maddieson 1988; Maddieson 2005). But complexity claims that go beyond this and that attempt to say something

about a given language type or about a whole component of the grammar, like the syntax, soon become unworkable and vacuous.

The problems include:

Trade-offs: simplicity in one part of the grammar often results in complexity in another; see the illustrations given in the examples of (2.40)–(2.45) and in §§2.5.1, 2.5.2, and 9.3.

Overall complexity: the trade-offs make it difficult to give an overall assessment of complexity, resulting in unresolvable debates over whether some grammars are more complex than others, when there is no clear metric of overall complexity for deciding the matter.

Defining grammatical properties: the (smaller) structural units of grammars are often clearly definable, but the rules and representations that refer to them are anything but and theories differ over whether they assume 'simplicity' in their surface syntactic structures (see, e.g., Culicover and Jackendoff 2005) or in derivational principles (as in Chomsky 1995), making quantification of complexity difficult in the absence of agreement about what to quantify over.

Defining complexity itself: should the definition be stated in terms of rules or principles that generate the structures of each grammatical area (i.e., in terms of the 'length' of the description or grammar, as discussed in Dahl 2004), or in terms of the structures themselves (the outputs of the grammar)? Definitions of this latter kind are inherent in metrics like Miller and Chomsky's (1963) that use non-terminal to terminal node ratios, and in Frazier's (1985), Hawkins's (1994), and Gibson's (1998) later metrics. Rule-based and structure-based definitions will not always give the same complexity ranking.

Consider an example of a trade-off that is generally not considered in the literature on creoles and other simple-looking languages that have SVO word order and little or no inflectional morphology. This trade-off, as we shall see, complicates any claim about the overall simplicity of these systems. Take NP–V–NP sequences in English. They must often be mapped onto complex argument structures in ways that many (often most) languages do not permit, even the closely related German. Some less common thematic role assignments to transitive subjects that are grammatical in English but ungrammatical in German and that depart from the Agent–Verb–Patient prototype of (2.40) are illustrated in (2.41)–(2.43) (Rohdenburg 1974; Hawkins 1986; Müller-Gotama 1994; König and Gast 2009).

(2.40) a. *The professor* wrote an important book. [Agent]
b. *Der Professor* hat ein wichtiges Buch geschrieben.

(2.41) a. A few years ago *a penny* would buy two to three pins. [Instrument]
b. *Vor einigen Jahren kaufte *ein Pfennig* zwei bis drei Stecknadeln.

(2.42) a. *This tent* sleeps four. [Location]
b. **Dieses Zelt* schläft vier.

(2.43) a. *The book* sold 10,000 copies. [Theme]
b. **Das Buch* verkaufte 10,000 Exemplare.

German regularly uses semantically more transparent prepositional phrases for these non-agentive subjects of English, resulting in translations such as (2.45a–i) for English (2.44a–i):

(2.44) a. *This advert* will sell us a lot of dog food.
b. *Money* can't buy everything.
c. *This statement* overlooks the fact that the situation has changed.
d. *This* loses us our best midfield player.
e. *The lake* prohibits motor boats.
f. *The Langdon (river)* could and frequently did drown people.
g. *My guitar* broke a string mid-song.
h. *The latest edition of the book* has added a chapter.
i. ... Philips, *who* was streaming blood, ...

(2.45) a. *Mit dieser Werbung* werden wir viel Hundefutter verkaufen.
b. *Mit Geld* kann man nicht alles kaufen.
c. *Mit dieser Aussage* übersieht der Autor, dass ...
d. *Damit* verlieren wir unseren besten Mittelfeldspieler.
e. *Auf dem See* sind Motorboote nicht zugelassen.
f. *Im Langdon* konnte man ertrinken, was übrigens häufig genug vorkam.
g. *An meiner Gitarre* riss mitten im Lied eine Saite.
h. *Zur letzten Ausgabe* des Buches wurde ein Kapitel angefügt.
i. ... Philips, *von dem* Blut herunter strömte, ...

English transitive and ditransitive clauses are syntactically and morphologically simple. The NP–V–(NP)–NP surface structure is minimal and contains fewer syntactic categories than its German counterpart with both NPs and PPs. English also lacks the case morphology of German, except residually on personal pronouns. The rules that generate these English surface forms are also, arguably, more minimal than their German counterparts. Yet in any formalization of the mappings from surface forms to argument structures and semantic representations, the set of English mappings must be regarded, by anyone's metric, as more complex than the German set. There are more argument structure types linked to the NP–V–(NP)–NP surface structure than to the corresponding transitive and ditransitive structures of German

(containing NPs only and without PPs), as seen in (2.41)–(2.43); see Hawkins (1986) for details. This adds 'length' and complexity to the grammar of English. It also makes processing more complex. The assignment of an instrument to the subject NP in (2.41a) and of a locative in (2.42a) requires crucial access by the processor to the verbs in these sentences, to their lexical semantics and co-occurrence structure, and also possibly to the postverbal NP. More disambiguation needs to be carried out by the English processor, and greater access is needed to more of the surface structure for this disambiguation, i.e., the processing domains for thematic role assignments in English are less minimal (recall §2.1).

The claim that these SVO orders, which are typical of so many inflectionally impoverished languages and of creoles, are simple is not readily defensible, therefore. The rules mapping these forms onto meanings have been conventionalized in English, i.e., they constitute conventions of grammar, and these conventions have changed in the recorded history of English: Old English used to be much more like German (Rohdenburg 1974).

It is no accident that German has richly case-marked NPs, whereas Modern English does not. In Müller-Gotama's (1994: 143) sample of fifteen languages comparing thematic role assignments to NPs, it was the languages with the most minimal surface structures (Chinese, Indonesian, and English) that had the widest and least semantically transparent mappings onto argument structures. Other richly case-marked languages (Korean, Japanese, Russian, Malayalam) behaved like German. The precise typology here and its causes needs to be investigated further (see §7.2). The point in this context is that while claims of simplicity or complexity can be made for individual structures (or for morpheme paradigms and phoneme inventories), larger claims about language types or whole grammars are fraught with difficulty, even when they make reference to properties like SVO in uninflected languages whose simplicity seems self-evident. It is not.

2.5 Competing efficiencies in variation data

The efficiency principles introduced in §§2.1–2.3 can subsume some of the simplicity/complexity metrics and ideas that have been proposed in the literature hitherto but they add to them an efficiency perspective as defined in MiD, MiF, and MaOP. Variation data can now be better understood in terms of cooperating and competing efficiencies between principles. In this section I give some brief examples that support this point.

2.5.1 Extraposition: Good for some phrases, often bad for others

One structure that has figured prominently in psycholinguistic metrics of complexity is Extraposition (Miller and Chomsky 1963; Frazier 1985; Hawkins 1994; Gibson 1998). An English clause with a sentential subject, *that their time should not be wasted is important*, would be (just slightly) more complex, according to Miller and Chomsky's non-terminal to terminal node ratio, than its extraposed counterpart *it is important that their time should not be wasted*. This latter has one additional terminal element (*it*), but the same amount of higher non-terminal structure, resulting in a lower non-terminal to terminal node ratio. In general, complexity increases by their metric in proportion to the amount of higher structure associated with the words of a sentence.

The more local metric of Hawkins (1994, 2004), stated in terms of Phrasal Combination Domains (PCDs) and the principle of Minimize Domains (2.1) and Early Immediate Constituents (2.4), defines complexity differently and from an online processing and efficiency perspective. These two sentences would be compared as follows, with higher overall percentages corresponding to more efficient structures for parsing:

(2.46) a. [$_S$ [that their time should not be wasted] [$_{VP}$ is important]]
PCD:$_S$ 1 2 3 4 5 6 7 8 2/8 = 25%
PCD:$_{VP}$ 1 2 2/2 = 100%

b. [$_S$ it [$_{VP}$ is important [that their time should not be wasted]]]
PCD:$_S$ 1 2 2/2 = 100%
PCD:$_{VP}$ 1 2 3 3/3 = 100%

The sentential subject in (2.46a) results in a complex PCD for the matrix S. Eight words must be processed for the recognition of two immediate constituents (resulting in a 2/8 IC-to-word ratio, or 25 percent). The PCD for the VP has just two words constructing two ICs (2/2 = 100 percent IC-to-word ratio). By extraposing in (2.46b), both S and VP domains become optimally efficient. The written corpus of Erdmann (1988) provides empirical support for these ratios: structures like (2.46b) are massively preferred over (2.46a) by 95 to 5 percent.

But Erdmann's corpus data also point to a trade-off in English performance. In some cases adding the sentential subject to the VP makes its PCD less efficient. Compare (2.47a) and (2.47b) in which the VP has a PP *for us all* to the right of *important*. This PP can be parsed and projected from the preposition *for*. The three ICs of the VP can accordingly be recognized by three adjacent words in (2.47a), *is important for*, without the parser having to access the remainder of the PP (IC-to-word ratio 3/3 = 100 percent). But when the sentential subject is added to the VP in (2.47b), becoming a fourth IC in

the VP, the added clause reduces the efficiency of VP. Six adjacent words must now be processed, *is important for us all that*, in order to construct the four ICs of VP, which lowers the IC-to-word ratio from 100 percent in (2.47a) to 67 percent in (2.47b):

(2.47) a. [$_S$ [that their time should not be wasted] [$_{VP}$ is important [for us all]]]

PCD:$_S$ 1 2 3 4 5 6 7 8 2/8 = 25%

PCD:$_{VP}$ 1 2 3 3/3 = 100%

b. [$_S$ it [$_{VP}$ is important [for us all] [that their time should not be wasted]]]

PCD:$_S$ 1 2 2/2 = 100%

PCD:$_{VP}$ 1 2 3 4 5 6 4/6 = 67%

The overall benefit from extraposing is now less than in (2.46), and Erdmann's corpus reflects this. There are almost five times as many sentential subjects in structures corresponding to (2.47a) (namely 24 percent) compared with (2.47b), than there are for (2.46a) (5 percent) compared with (2.46b). This follows from the efficiency approach to complexity given here. The extraposed alternative in (2.47b) has a ratio that is less than that of (2.46b), hence the sentential subject in (2.47a) can be more efficient relative to its extraposed counterpart on more occasions—for example, in sentences with a short sentential subject and a long PP like [*that he should succeed*][*is important* [*for all the good people who have invested in him*]].

We can use these efficiency ratios to make fine-tuned predictions for structural selections in performance, on the assumption that there are trade-offs in these different word orders. One clear case that has been investigated in detail involves extraposition of a relative clause from an NP in German (see §3.3), illustrated here by (2.48b) which alternates with the unextraposed relative in (2.48a). The idiomatic English translation would be 'Yesterday he read the book that the professor wrote.'

'He has the book that the professor written has yesterday read'

(2.48) a. Er hat [$_{VP}$ [$_{NP}$ das Buch das der Professor geschrieben hat]

PCD:$_{VP}$ 1 2 3 4 5 6 7 [$_{XP}$ gestern] gelesen] 8 9 3/9 = 33%

PCD:$_{NP}$ 1 2 3 3/3 = 100%

b. Er hat [$_{VP}$ [$_{NP}$ das Buch] [$_{XP}$ gestern] gelesen] [$_{NP}$ das der Prof. geschrieben hat]

PCD:$_{VP}$ 1 2 3 4 3/4 = 75%

PCD:$_{NP}$ 1 2 3 4 5 3/5 = 60%

(2.48a) has an efficient NP in which three adjacent words *das Buch das* suffice to construct three ICs (the determiner, the head noun, and the relative clause constructed by the relative pronoun *das*), i.e., 3/3 = 100 percent, but at the expense of a lengthened VP whose processing domain proceeds from *das Buch* through *gelesen*, i.e., 3/9 = 33 percent. (2.48b) with Extraposition from NP, has the reverse costs and benefits: a less efficient NP with separation of *das Buch* from the relative pronoun *das*, i.e., 3/5 = 60%, but a shorter and more efficient VP resulting from the extraposition, 3/4 = 75%.

By quantifying these trade-offs we can predict which variant will be preferred in performance, on the assumption that structures selected will be those that are most efficient overall. For example, depending on the length of the XP preceding V in (2.48) one soon reaches the point (after only a two-word XP, in fact) when benefits for the VP from extraposing are outweighed by disadvantages for the NP, and extraposition is predicted to be disfavored (Hawkins 2004: 142–6). Uszkoreit et al. (1988) tested this prediction, and the following data from their German corpus support it. Extraposition from NP is favored with a one-word intervening XP, but disfavored with a 3+ word XP, while 2-word XPs are transitional, as shown in (2.49). (The frequency of Extraposition from NP over V alone was 95 percent.)

(2.49) Uszkoreitet al.'s (1988) corpus data: Extraposition from NP frequencies
XP(1 word)+V = 77%, XP(2)+V = 35%, XP(3–4)+V = 8%, XP(5+)+V = 7%

What we see here is a clear trade-off between keeping a phrasal constituent in situ or moving it, as a function of the size of potentially intervening constituents (XP and V in (2.48)). This can be analyzed in terms of competing efficiency benefits for NP versus VP phrases. The Minimize Domain principle (2.1) and specifically EIC (2.4) give us a ready means of quantifying these competing preferences. (2.47a) and (2.47b) showed that a minimal domain for the processing of S can make the VP less minimal, and vice versa. (2.48a) and (2.48b) show that the benefits for the VP can be at the expense of the NP, and vice versa. This approach enables us to calculate an overall benefit for the sentence as a whole resulting from the use of one variant versus another and to predict the selection of different variants in performance.

2.5.2 Competing head orderings for complex phrases

The typology of relative clause ordering points to a trade-off between competing forces in head-final (OV) languages. VO languages are almost exceptionlessly head noun before relative clause (NRel), as in English, *the students*i [*that I teach* 0i]. OV languages are mixed, either NRel (Persian) or RelN

(Japanese). The distribution of NRel to RelN appears to correlate with the degree of head finality: the more rigidly verb-final languages like Japanese prefer RelN, the less rigid ones have NRel. This suggests that consistent head ordering across phrases is one factor influencing the typology of relative clause ordering. A competing factor appears to be Fillers First (§2.3.1), the preference for a head noun (filler) to occur to the left of its gap or resumptive pronoun within the relative. Both preferences can be realized in VO languages, but at most one in OV languages, resulting in consistent and almost exceptionless VO and inconsistent OV languages, with these latter selecting RelN in proportion to their degree of head finality elsewhere (Hawkins 2004: 208):

(2.50)

VO	**OV**
NRel (English)	NRel (Persian)
*RelN	RelN (Japanese)

I return to this distribution in §7.3 and §9.3.

2.5.3 Complex inflections and functional categories benefit online processing

Complex inflectional marking, e.g., nominal cases, and functional categories like definite articles have been much discussed in the literature on complexity: cf., e.g., *Linguistic Typology* 5 (2/3) (2001). It is commonly observed that these morphological and syntactic categories arise as a result of lengthy grammaticalization processes and are not characteristic of the earliest and simplest systems, such as creoles. I argue in Chapter 4 that grammaticalization makes processing of the relevant grammars more efficient. If true, this means that there are benefits that compensate for added morphological and syntactic complexity: complexity in form processing is matched by simplicity with respect to the processing functions that rich case marking and definite articles perform.

Japanese has 'rich case marking' (R-case) which is defined in Hawkins (2002, 2004) to mean that a language has explicit marking for at least two arguments of the verb:

(2.51)

John	**ga**	tegami	**o**	yonda	(Japanese)
John	NOM	letter	ACC	wrote	

I mentioned in §2.4 that such languages generally have transparent and consistent mappings of cases onto theta-roles, avoiding the one-to-many mappings of English. The result is especially favorable for verb-final languages, since this permits thematic roles to be assigned early in parsing prior to the verb (see Bornkessel 2002 and §7.2). The clear productivity of

case marking in these languages can be seen in the language samples referenced in (2.52) (see Hawkins 2004:249):

(2.52) V-final with R-case: 89% (Nichols 1986), 62–72% (Dryer 2002)
V-initial with R-case: 38%, 41–47% respectively
SVO with R-case: 14–20% (Dryer 2002)

Definite articles generally arise historically out of demonstrative determiners (C. Lyons 1999). Their evolution is especially favored, in the analysis of Hawkins (2004: 82–93), in languages that have VO or other head-initial phrases (e.g., in OVX languages that combine head-final with head-initial orders; see §5.8.1) and in which the article can serve an NP 'construction' function; see §4.2 and §6.2.3 for data and discussion.

There is a trade-off here in both complex morphology and definite articles between additional form processing, which is disfavored by MiF, and certain complementary benefits resulting from the presence of these forms, as defined by MaOP and MiD. Complex morphology in the form of rich case marking provides early online information about morphosyntactic case and thematic role in advance of the verb, which is favored by MaOP, and which is especially beneficial in V-final languages; see §7.2. Definite articles separate from demonstratives result in less minimal forms for NP than bare NPs but will be argued in §6.2.3 to have processing advantages, especially in head-initial languages. These advantages include early construction of mother NP nodes, favored by both MiD and MaOP. Hence MiF competes with both MaOP and MiD here and this is reflected in the variation data.

An interesting example of cooperation and competition between MiF and MiD can be seen in the retention versus removal of explicit case marking in languages that have it, which is reminiscent of the distribution of gaps and resumptive pronouns in relative clauses discussed in §2.2.3. Consider Kannada, a Dravidian SOV language with morphological case marking on a nominative–accusative basis. Bhat (1991) points out that an accusative-marked *pustaka-vannu* ('book-ACC') can occur without the case marking when it stands adjacent to a transitive verb, giving the structural alternation exemplified in (2.53a) and (2.53b). But there is no zero counterpart to an explicitly marked accusative if the relevant NP has been "shifted to a position other than the one immediately to the left of the verb" (Bhat 1991: 35), as in (2.53c):

(2.53) a. Avanu ondu pustaka-vannu bareda. (Kannada)
He (NOM) one book-ACC wrote
'He wrote a book.'

b. Avanu ondu pustaka bareda.
he (NOM) one book wrote

c. A: pustaka-vannu avanu bareda.
that book-ACC he (NOM) wrote
'That book he wrote.'

An accusative NP without explicit case marking becomes dependent on the verb for the assignment of its case and for its thematic role, just as non-case-marked NPs in languages like English are dependent on their verbs for the assignment of these properties (see the discussion of dependency in processing in Hawkins 2004: 18–25). These additional dependencies create stronger pressure for proximity between items, on this occasion adjacency, by Minimize Domains (2.1). Minimize Forms (2.8) favors zero case marking which sets up these additional dependencies, and so (2.53b) satisfies both MiF and MiD. But if the case marking were absent in (2.53c) or in other non-adjacent environments involving a verb and its governed NP, then there would be less minimal domains for more processing relations, which is at variance with MiD at the same time that the NP retained a minimal form. The cooperation between MiD and MiF under adjacency becomes competition, therefore, in less minimal environments like (2.53c), and the grammar of Kannada has conventionalized what we can assume to have been a performance preference for the explicit form in these non-minimal domains prior to conventionalization (see §4.4). Similarly, gaps give way to resumptive pronouns in relative clauses that involve increasingly non-minimal and complex relativization domains (§2.2.3).

2.5.4 Interim conclusions

The trade-offs in this section are the result of interacting efficiency factors that sometimes reinforce each other but sometimes compete in the relevant structures and languages. These competitions can be summarized in (2.54):

(2.54)
1. §2.5.1 competing MiD preferences in different phrases (S vs VP, or VP vs NP)
2. §2.5.2 competing head orders in complex phrases (MaOP vs MiD advantages)
3. §2.5.3 complex morphology and definite articles (MiF vs MiD & MaOP advantages)

Variation results when factors compete or when different structures satisfy one and the same factor (e.g., Minimize Domains) in different ways, as in the extraposition data (§2.5.1). Performance preferences are motivated by a number of factors, therefore, of which relative complexity measured by size of domain and amount of structure is just one. Cross-linguistic grammatical conventions point to a similar interaction and competition between multiple

factors. According to the Performance-Grammar Correspondence Hypothesis ((1.1) in §1.1) the more efficiency there is in a given structure, the more grammars incorporate it as a convention. Recall the wording 'Grammars have conventionalized syntactic structures in proportion to their degree of preference in performance.'

Ranking grammars according to overall complexity is problematic for the reasons given in §2.4, namely the trade-offs, unclarity over the measurement of overall complexity, unclarity over how best to describe many grammatical properties, and different (rule- or structure-based) definitions of complexity itself. The most we can do, I have suggested, is to say that languages are more or less complex in certain individual areas of grammar.

On the other hand, comparing grammars in terms of efficiency in communication as defined in §2.4 enables us to model more of the factors that ultimately determine the preferences of performance and of grammars, in addition to complexity itself, namely speed in the delivery of linguistic properties, fine-tuning of structural selections to frequency, accessibility, and inference, and online delays in property assignments and online error avoidance (preventing unassignments and misassignments). This provides a more general theory of performance and of grammars, it puts structural complexity in its proper context, and it helps us understand the trade-offs better: preferred structures can be simpler in one respect, more complex in another; and the trade-off may involve simplicity competing with some other efficiency factor, e.g., speed of online property assignments in processing.

Having identified the multiple factors determining efficiency we can then use quantitative data from both performance and grammars to guide our theorizing about the strength of each, about their relative strength in combination, and about their strength of application in different linguistic structures. Performance data such as (2.49) involving Extraposition from NP frequencies, grammatical frequencies such as (2.52), as well as distributional asymmetries such as (2.50) (see §7.3 for precise data), all shed light on this crucial issue: how strong is each factor, intrinsically and relatively? I have argued that the kind of multi-factor model of grammar and language use that we now need (and any serious model of these two domains has to be a multi-factor one) is best characterized in terms of different ways of achieving efficiency in communication, rather than in terms of simplicity and complexity alone.

I return to the question of how multiple principles cooperate and compete in Chapter 9. Before that I must first consider some general issues in psycholinguistics that are raised by this research program (Chapter 3), and in historical linguistics and language change (Chapter 4), and I must examine numerous variation patterns of performance and grammars in greater detail (Chapters 5–8).

3

Some current issues in language processing and the performance–grammar relationship

Any attempt to establish a correspondence between performance data and grammars, as proposed in the last two chapters, immediately raises some basic questions at both ends of the equation. On the one hand, what exactly are the data of performance? How have they been collected? How reliable are they? How are they to be understood and modeled? On the other hand, how are the grammatical generalizations that are claimed to correlate with them best formulated and modeled? In this chapter I focus mainly on the performance side of things, as studied primarily by psychologists. There are different methods of data collection in psycholinguistics, with sometimes very different findings, and there are different theories about the underlying mechanisms that lead to the observed data, just as there are different descriptions and models of grammar in linguistics.

It is important that we try to rise above these differences and attempt to see the forest rather than the trees, so that we can shed light on the basic question in this context: the relationship between performance and grammars. Sections 3.1–3.4 summarize some current methodological and theoretical issues in psycholinguistics, especially those that arise in connection with data of the kind presented here. These issues need to be noted but they do not, I believe, detract from the basic correctness of the performance–grammar correspondence hypothesized in Chapter 1. Much of this book argues the case for more performance factors in grammatical modeling. The fifth section of this chapter (§3.5) looks at things the other way round and shows how facts about grammars can and should be taken into account by psychologists in their psycholinguistic model-building, precisely because there is a correspondence between performance data and grammars (§1.1). The final section (§3.6) discusses a recent proposal by Mobbs (2008) to incorporate the efficiency

principles proposed here within Chomsky's 'Minimalist Program' (Chomsky 1995, 2004, 2005).

I begin in the first section (§3.1) with a brief discussion of how I see the relationship between ease of processing, a much-studied concept in psycholinguistics, and efficiency, on which I am focusing here.

3.1 Ease of processing in relation to efficiency

The Dependency Locality Theory of Gibson (1998, 2000) is a much-discussed recent version of a working memory model in psychology that builds on the idea of 'constrained capacity' (see, e.g., Frazier and Fodor 1978, Just and Carpenter 1992 for earlier proposals, and Jaeger and Tily 2011 for a full overview of these models). The ceiling effects within this capacity are claimed to make certain sentence types unprocessable beyond the complexity limit. My EIC/MiD theory summarized in §2.1 has been widely linked to this constrained capacity tradition, but I have often stated (most recently in Hawkins 2004: 268–9) that I make no commitment as to what this capacity may ultimately be, stressing only that processing complexity and difficulty increase as the size and complexity of the different processing domains increase (the constituent recognition domains of Hawkins 1994, filler-gap domains in Hawkins 1999, etc.). Lewis and Vasishth (2005) and Vasishth and Lewis (2006) have meanwhile developed an alternative working memory model to Gibson's in terms of activation, using primitives taken from cognitive psychology and from mechanisms of neural activation (cf. Anderson 1983).

The key insight to emerge from the working memory literature, I believe, is simply that processing becomes harder, the more items are held and operated on simultaneously when reaching any one parsing decision (e.g., what is the phrase structure for this portion of the sentence, where is the gap for this filler, etc.?). When processing domains relevant to each decision become larger and contain additional items whose properties must also be computed, there is a corresponding strain on processing. Small domains make processing easier. Small domains also contain fewer items that can interfere with current parsing decisions, e.g., fewer potential gap sites for a filler, fewer NP arguments to be linked to a given verb, and so on. Less interference is a key component of ease of processing in the model of Lewis and Vasishth (2005) and Vasishth and Lewis (2006). The advantage of fewer items and less interference in working memory is that less cognitive activity takes place in online processing.

It must also be remembered here that claims about working memory capacity have been derived largely from psycholinguistic studies on English and a handful of other languages. There is a risk that they have defined as

unprocessable certain structures that do occur in other language types (see Hawkins 1994: 266–7, 2007 for illustration of this point, and Jaeger and Norcliffe 2009 on the urgent need for more cross-linguistic work in sentence processing). A more relative approach to working memory that avoids this commitment to an absolute ceiling reduces this problem.

In addition to the size of processing domains and avoidance of interfering items, there is a third insight about ease of processing that has increasingly gained ground in the psycholinguistic literature. Linguistic items are easier to process when they are more predictable and expected and less surprising. Levy (2008), Jaeger (2006), and Jaeger and Wasow (2008) define and quantify this idea while controlling for numerous alternative explanatory possibilities. Recall the discussion of relativizer omission in §2.2.2 based on data from Wasow et al. (2011): zero relativizers are more frequent in environments where a restrictive relative clause is more predictable and frequently occurring (compare *the man I know* with *the man who(m)/that I know*).

The concept of efficiency, on which I shall focus in this book, is more goal-oriented than grammatical simplicity and complexity and their processing counterparts, namely ease vs difficulty in processing. As discussed in §2.4, efficiency relates to the basic goal of communication, which is, as I see it, to convey information from a speaker to a hearer in rapid time and with minimal processing effort. It is often the case that communicative goals are best met by using simpler/easy-to-process structures, but sometimes this is not the case. The hearer's uptake may require more complex and harder-to-process alternatives, in order to understand what the speaker intends.

More generally, I believe that many phenomena that have been attributed to working memory capacity are really just the product of 'least-effort' processing (of the kind discussed in Zipf 1949) and of efficiency as defined here (in Chapter 2) and in Hawkins (2004) and (1994). Speed in communicating the intended message from speaker to hearer and minimal processing effort in doing so are the two driving forces of efficiency that I propose. The former is captured in the principle of Maximize Online Processing (MaOP) which defines a preference for selecting and arranging linguistic forms so as to provide the earliest possible access to as much of the ultimate syntactic and semantic representation as possible. In the process MaOP defines a dispreference for online 'misassignments' or garden paths, and for 'unassignments,' in proportion to their quantifiable severity (Hawkins 2004: 51–8). See §2.3 here for more detail. The latter (minimal processing effort) is captured in Minimize Domains (MiD) and in a second minimization principle, Minimize Forms (MiF).

MiD explains, for example, why corpora reveal such clean, gradient weight effects, whereby the selection of short before long phrases in English-type languages is in direct proportion to the degree of weight difference between them, reflecting the increasing inefficiency of long before short orders when there are larger weight differentials (see §2.1). These efficiency preferences are visible well within the capacity constraints of any working memory theory and are not explained by them. Nor do they appear to be explainable by activation-based models of memory, which predict numerous non-minimal and antilocal preferences at variance with EIC/MiD's efficiencies (Vasishth and Lewis 2006). See §3.4 below.

Minimize Forms (MiF) defines a preference for minimizing the formal complexity of linguistic forms (phonemes, morphemes, etc.) and the number of forms with unique conventionalized property assignments, in proportion to the ease with which grammatical and lexical properties can be assigned to minimally specified forms. See §2.2 for further details. MiF explains many of the morphological patterns in Greenberg's markedness hierarchies, as well as structural alternations such as the presence versus absence of the relativizer in §2.2.2. Relevant corpus data in all these cases suggest that the zero alternate is preferred in environments where the morphological and syntactic categories in question are more expected and predictable, making processing more efficient overall. Such efficiencies as well as those that result from minimizing domains ultimately involve a minimization of the amount of cognitive activity that takes place within relevant processing domains.

There are other recent approaches to defining efficiency in psycholinguistics which are very much in the spirit of the research program of this book and earlier books and which add useful dimensions to it. For example, sophisticated measures have been developed to test the idea that expectedness and predictability are reflected in ease of processing on a word-by-word basis in a sentence (see Jurafsky 1996; Hale 2003; Levy 2008). It has been proposed that efficiency is achieved through uniform information density over the linguistic signal and over time (Jaeger 2006). These proposals, which are empirically well supported, provide an additional online component to the measurement of efficiency given here, with reflexes in syntax, morphology, phonology, and prosody (see Jaeger and Tily 2011 for a summary). They also build on the mathematical theory of information given in Shannon (1948), which has been neglected for far too long and whose significance is now leading to exciting new insights and research paradigms for the future (see again Jaeger and Tily 2011).

3.2 Production versus comprehension

I have traditionally defined EIC and MiD and the processing domains to which they apply primarily from the hearer's rather than the speaker's perspective. Wasow (1997, 2002) has pointed out that the data that motivate them include significant amounts of corpus data, i.e., data from language production, and he has stressed the speaker's perspective when accounting for 'end weight' in English. I agree with him with respect to English. It needs to be pointed out, however, that my focus on parsing has been largely for practical reasons, not theoretical ones. More work has traditionally been done on the parsing of syntax than on its production, and more is known about parsing. The precise relationship between production and comprehension models is also currently a matter of intense debate (cf., e.g., Kempen 2000, 2003 for an attempted synthesis). In Hawkins (1998: 762–4) I advocated a position in which the domain minimization benefits of EIC/MiD apply to both production and comprehension. See O'Grady (2005: 7) for a similar position and a review of supporting psycholinguistic evidence, and MacDonald (1999) for discussion of numerous production–comprehension parallels, including weight effects.

One challenge for a unification of production and comprehension models comes from head-final languages and from head-final structures generally (like subordinate clauses with final complementizers or prenominal relative clauses before a head noun), whether they occur in consistently head-final languages (Japanese and Korean) or in a mixed type like Chinese. For example, De Smedt's (1994) Incremental Parallel Formulator builds on Levelt's (1989) Production Model and has a ready explanation for the postposing of heavy constituents, as in Extraposition from NP sentences like (3.2b) in the next section and Heavy NP Shift sentences like (3.4b) in §3.5. Syntactic constituents are assembled incrementally in this model in the Formulator, following message generation within a Conceptualizer. The relative ordering of constituents can reflect both the original order of conceptualization and the processing time required by the Formulator for more complex constituents. The late occurrence of heavy constituents is explained in De Smedt's model in terms of the greater speed with which short constituents can be formulated in the race between parallel processes in sentence generation.

For head-final structures, however, there is growing evidence for the prediction made back in Hawkins (1994) for a heavy-first *pre*posing preference, i.e., for the mirror image of English *post*posing rules; see Yamashita and Chang (2001, 2006) for Japanese, Choi (2007) for Korean, and Matthews and Yeung (2001) for Cantonese (a comprehension study). Preposing appears to be in direct conflict with the predictions of De Smedt's

Incremental Parallel Formulator (see §3.5 below). An alternative way of capturing domain minimization benefits in a production model is outlined in Hawkins (1998: 763–4), where it is assumed that Constituent Recognition Domains and Constituent Production Domains are closely aligned. The interesting consequence of this alignment for head-final languages is that the Conceptualizer could be cognitively aware that a clause currently being processed is a subordinate one within a matrix S, at the same time that its syntactic status as a preposed subordinate clause within a matrix will not yet have been constructed by the Formulator until the right-peripheral complementizer is reached, which projects to its dominating mother and grandmother nodes (by the kinds of node construction principles proposed in Hawkins 1994: ch. 6).

The efficiency of head-final languages derives from their adjacent positioning of heads on the right periphery of their respective phrases; see §5.3. It is this consistent right-adjacency that minimizes phrasal processing domains, corresponding to the consistently left peripheral positioning of heads in head-initial languages. This also results in frequent bottom-up processing effects, including certain late syntactic property assignments like the one just mentioned when a subordinate clause precedes its matrix. This result does not necessarily argue against the alignment of production and comprehension strategies, however, and hence it does not argue against a possible production theory of mirror-image weight effects in head-initial and head-final languages (see §7.1). The adjacency of heads, through consistent left- or consistent right-positioning, will minimize domains for phrase structure processing and can be efficient for both comprehension (in a syntactic parser) and for production (within a Levelt 1989-type Formulator).

3.3 Online versus acceptability versus corpus data

Different methods of data collection in psycholinguistics can sometimes lead to different patterns of results, which can then lead to different theories of ease versus difficulty in processing. A striking example can be seen in the landmark study by Lars Konieczny and his colleagues on extraposed and unextraposed relative clauses in German (Konieczny 2000; see also Uszkoreit et al. 1998). The German relative clauses were examined in corpora, in an offline acceptability judgment task and in an online self-paced reading experiment. Examples of these structures are given in (3.1a) and (3.1b). In the former, the relative clause is adjacent to its nominal head; in the latter it is extraposed:

(3.1) a. Er hat die Rose [die wunderschön war] hingelegt . . .
(Rel clause adjacent)
He has the rose [that beautiful was] laid-down

b. Er hat die Rose hingelegt [die wunderschön war] . . .
(Rel clause extraposed)
He has the rose laid-down [that beautiful was] . . .
'He has laid down the rose that was beautiful.'

Konieczny found different patterns of preference in different types of data and in response to different tasks, and this led him and has subsequently led others to rethink earlier findings in psycholinguistics about capacity constraints in working memory, about locality versus antilocality preferences, and about the production–comprehension relationship. Let us first establish what the different empirical patterns were in Konieczny's study.

The corpus study by Uszkoreit et al. (1998), in which Konieczny participated, tested some predictions that were made for the alternation in (3.1) by the Early Immediate Constituents (EIC) principle that I proposed in Hawkins (1994). EIC and its more recent version, Minimize Domains (MiD), predict that the extraposed (3.1b) will be preferred in proportion to the *in*creasing weight and complexity of the relative clause and the *de*creasing potential distance between head noun (*Rose*) and relative when items can intervene. The adjacent (3.1a) will be preferred over (3.1b) in proportion to the *de*creasing weight and complexity of the relative clause itself and the *in*creasing potential distance between head noun and relative. See §2.1 and §§5.1–5.2 in the present book for definitions of EIC and MiD following Hawkins (1994: 69–83) and (2004: 31), and §2.5.1 for precise predictions for this relative clause alternation following Hawkins (1994: 198–210) and (2004: 142–6).

These patterns of preference were confirmed in Uszkoreit et al.'s corpus data. The predicted increase in the frequency of extraposition as a function of the increasing weight of the relative was supported (see Hawkins 2004: 146, ex. (5.32) for a summary of Uszkoreit et al.'s figures). Figure 3.1 here shows their results for the potential distance prediction. It reproduces the corpus selections for extraposed versus adjacent relatives, expressed as a percentage (as given in their Abbildung 2), and reveals a preference for the extraposed version over short distances (1–3 words) but a strong preference for adjacency of the relative and head over middle (4–6) and long potential distances (7–9) (the percentage of relatives extraposed over a distance of one word was 95.2 percent, over a distance of two words 77.1 percent, and over a distance of three words 34.7 percent). The corpus consisted of written journalistic German from the *Frankfurter Rundschau* which was subjected to both automatic and manual parsing techniques. The various subcorpora analyzed by Uszkoreit

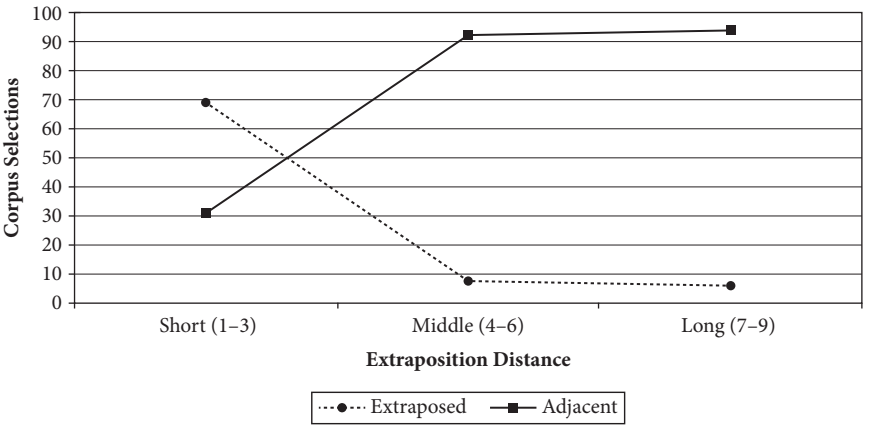


FIGURE 3.1 Mean corpus selection percentages for extraposed vs adjacent relatives at different distances

et al. (see their Tabelle 1 as well as Abbildung 2) yielded similar results. Figure 3.1 shows a clear crossing and complementary pattern of preferences that follow the predictions of EIC/MiD closely and that support the kind of 'locality' principle underlying EIC/MiD and also much related work reported in Gibson (1998, 2000).

Konieczny (2000) and Uszkoreit et al. (1998) also gave data from an offline psycholinguistic study measuring the relative acceptability of adjacent and extraposed relative clauses and using the magnitude estimation technique of Bard, Robertson, and Sorace (1996). Participants provided a score indicating how much better or worse the current sentence was compared to a given reference sentence (for example, if the reference sentence was assigned 100 and the target sentence was judged five times better it received 500). Figure 3.2 shows the mean acceptability scores for the two relative clause possibilities over a short distance (one word), a middle distance (3–4 words), and a long distance (5–6 words) (following Table IV and Figure 3 in Konieczny 2000 and Abbildung 5 in Uszkoreit et al. 1998). As in the pattern of Figure 3.1, extraposition is judged worse in Figure 3.2 when there are increasing distances between head and relative. Adjacency is also less preferred in their acceptability data for relative clauses of increasing weight and complexity (see again Table IV and Figure 3 in Konieczny 2000 and Abbildung 5 in Uszkoreit et al. 1998). But in contrast to the corpus data, these acceptability data show a strong preference overall for adjacency over extraposition. Only in the case of the longest possible relative clause of 9–11 words separated by its head from one word was extraposition actually preferred in the acceptability task (this is not visible in

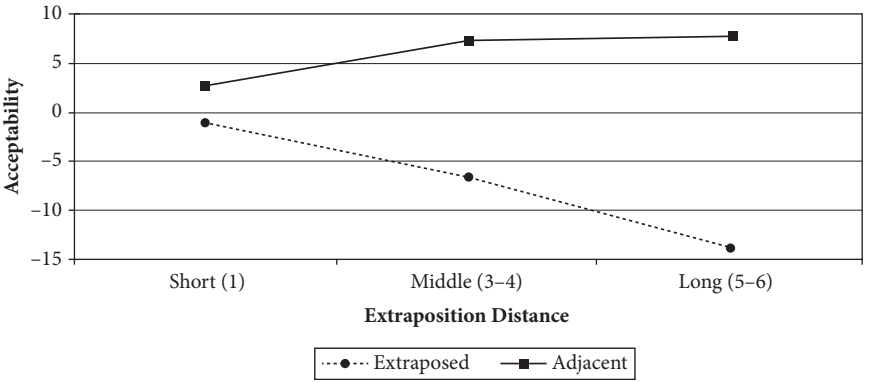


FIGURE 3.2 Mean acceptability scores for extraposed vs adjacent relatives at different distances

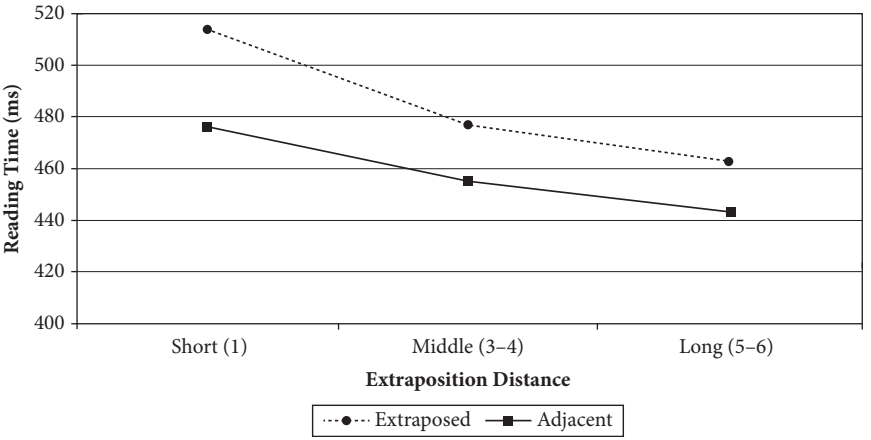


FIGURE 3.3 Mean reading times at the clause-final verb for extraposed vs adjacent relatives at different distances

Figure 3.2 since this extreme was averaged out with relative clauses of shorter weights in the one-word condition). Whereas the patterns of Figure 3.1 are predicted by EIC/MiD, therefore, the overall result of Figure 3.2 is not, though similar tendencies are evident in both.

In addition to the offline experiment of Figure 3.2 Konieczny (2000) conducted an online self-paced reading experiment using the same materials as in the offline acceptability task and measuring reading times at the clause-final matrix verb (*hingelegt*). The experiment revealed that the verb was read systematically faster when the relative clause preceded it, i.e., when the relative clause was adjacent to the head noun, as shown in Figure 3.3 (which

reproduces Konieczny's Table V and Figure 5). There was also a main effect of distance between head noun and relative in the online experiment, but not of relative clause weight and complexity.

Konieczny concluded that online reading times do not support locality-based predictions. I shall return to this in the next section, but notice first the important methodological point. Different types of data and different methods of collecting them can give different patterns of results. Corpus selections show strong EIC/MiD effects in these German structures, with preferences for the extraposed or the adjacent relative depending on the potential distance between head and relative (see Figure 3.1) and on the size and complexity of the relative clause itself. Online data show a consistent preference for the adjacent relative clause (see Figure 3.3), with a subsidiary distance effect and no relative clause weight effect. Offline acceptability data are in between the two, with an overall adjacency preference (see Figure 3.2) but with both distance and weight effects evident in the data.

A similar set of differences has been established for the corresponding Extraposition from NP alternation in English (involving preverbal subject NPs) by Francis (2010) as exemplified in (3.2):

(3.2) a. New sets [that were able to receive all the TV channels] soon appeared.
b. New sets soon appeared [that were able to receive all the TV channels].

Her corpus findings fully support EIC and MiD predictions in terms of relative clause weight and complexity and in terms of the potential size of the distance between nominal head and relative. But a self-paced reading study provided only partial support (extraposed sentences showed an advantage when the relative clause was heavy), while an acceptability judgment task showed the same preference for adjacency of the relative clause as in German, leading Francis to speculate that a prescriptive rule banning displaced modifiers might be impacting native speakers' judgments here.

We have to ask where these different patterns of results come from and how we can account for them. EIC/MiD is a principle that assigns a quantified and gradient preference to one structure over another among alternative, truth-conditionally equivalent sentences. Corpus data that involve a structural discontinuity between relative clause and head noun with competing demands on language processing provide a challenging test for it and have shown strong support. This suggests, at the very least, that this is a principle that guides structural *selections* in language production, supporting the point made in Wasow (1997, 2002) that we need to look at weight effects from the speaker's perspective. Online self-paced reading data like those of Konieczny (2000), on the other hand, reflect processing ease and speed at certain temporal points in competing structures. These data do not implicate a principle that

is relevant for structural selection, since if they did the adjacent structures would be selected almost all the time. Faster processing at the matrix verb does not mean, therefore, that the relevant structure is necessarily easier to process as a whole nor that it will be chosen more frequently. In fact, the corpus selections suggest that overall structural complexity can be quite independent of ease of processing measured at the verb (compare Figures 3.1 and 3.3). Whether the relevant processing facilitations that these online data do reveal at the verb are unique to comprehension, or compatible with both production and comprehension models, seems less clear (see §3.2). Language producers could experience points of high and low processing load at different points in their online structure generation just like comprehenders do, while still selecting one structure over an alternative based on a global measure of processing ease and minimality for the whole sentence, which is what EIC/MiD is. There is nothing contradictory about this and both the online and the corpus data are presumably tapping into different and real aspects of processing ease. The acceptability ratings, on the other hand, suggest a possible confound resulting from a normative bias.

3.4 Locality versus antilocality effects

Gibson's (1998, 2000) Dependency Locality Theory is similar in spirit to EIC/MiD (Hawkins 1994, 2004) and defines many similar preferences, but differs from it in certain ways. It is an online measure of processing complexity, whereas EIC/MiD calculates the complexity of each processing domain of a sentence globally in terms of its overall size and in terms of the quantity of syntactic, semantic, and other properties that are assigned within it, without regard for the moment-by-moment levels of processing difficulty. The ultimate units of processing cost for Gibson are measured in terms of new discourse referents within processing domains, whereas the ultimate units for EIC/MiD are word quantities, though this may just be a convenient and readily measurable shorthand for the full set of syntactic, semantic, and other properties that are assigned within each processing domain, with which word quantity totals are claimed to correlate (see Wasow 2002 and also Szmrecsanyi 2004 for a demonstration of the correlations between word quantities and other syntactic complexity measures). This particular difference between the Dependency Locality Theory and EIC/MiD may not be an essential one, therefore (Ted Gibson, personal communication).

Konieczny (2000) contrasted the 'antilocality' effects in his study with the predictions of both dependency locality and EIC/MiD and this has since led to a number of other experimental findings supporting 'antilocality.' The result is a major current debate in the field involving when exactly locality

effects will be found, and when not. For example, Vasishth and Lewis (2006) report that clause-final verbs are read faster in Hindi when processing domains are increased through the addition of relative clause modifiers, PPs, and adverbs. At the same time they recognize the existence of locality effects. They propose an activation-based model of sentence processing using the ACT-R architecture that combines activation decay and interference (accounting for locality) with the possibility of reactivating dependents, which is argued to account for antilocality effects. Konieczny (2000) and a number of other studies have appealed instead to the facilitating role that longer processing domains can have in predicting the upcoming matrix verb, making it faster to recognize and read. See again Levy (2008) and Jaeger (2006) for recent developments of this prediction- or 'expectation-'based approach to processing, controlling for numerous factors and alternative explanatory possibilities.

What is interesting about the current locality versus antilocality debate, and about the different data summarized in §3.3, is that corpora seem to strongly and consistently support locality, when there are alternating pairs of structures to choose from. Antilocality is supported by a subset of online experimental measures, which, I would argue, ultimately reflect the ease or difficulty of processing at certain temporal points online, especially at the verb, rather than the complexity of structures as a whole. Frequencies of selection in corpora appear, from these data, to reflect the overall complexity of sentences, as evidenced by frequency patterns in which there are competing complexity demands in different portions of the structure depending on whether the relative clause is adjacent to its head or extraposed from it.

3.5 The relevance of grammatical data for psycholinguistic models

The potential interest of the Performance–Grammar Correspondence Hypothesis for psychologists is that grammars are viewed as conventionalizations of performance preferences, by grammaticalization principles of the type summarized in the next chapter. As a result grammars can provide a new set of data that psychologists can consult, in addition to their experimental and corpus data from more familiar (generally European) languages, when setting up psychological models of production and comprehension. These data are especially valuable when they involve languages and structures that are rather different from English and other similar languages (see Jaeger and Norcliffe 2009).

I referred in §3.2 and §3.3 to the fact that languages like German and English postpose heavy constituents. Further examples (from English) involving

Extraposition of a sentential subject and Heavy NP Shift are given in (3.3b) and (3.4b) respectively.

(3.3) a. [$_S$ That John got married yesterday] surprised Mary
b. It surprised Mary [$_S$ that John got married yesterday]

(3.4) a. I gave [$_{NP}$ the antique that was extremely valuable and expensive] [to Mary]
b. I gave [to Mary] [$_{NP}$ the antique that was extremely valuable and expensive]

Japanese does things differently, *pre*posing heavy sentential phrases as shown in (3.5b).

(3.5) a. Mary ga [[kinoo John ga kekkonsi-ta to $_S$]
Mary NOM yesterday John NOM married that
it-ta $_{VP}$] (Japanese)
said
'Mary said that John got married yesterday.'

b. [kinoo John ga kekkonsi-ta to $_S$] Mary ga [it-ta $_{VP}$] (Heavy First)

In English corpora and in experimentally based processing preferences, heavy and complex clauses and NPs, etc., are regularly positioned at the end of their immediately containing clause or phrase, following lighter phrases (Erdmann 1988; Stallings 1998; Wasow 1997, 2002); see §2.5.1 and §5.2. This is a general characteristic of languages that are 'head-initial,' i.e., verbs occur initially in the verb phrase (VP) and early in the clause, nouns are initial or early in the NP, prepositions are initial in the PP, complementizers are initial in subordinate clauses, and so on. In other words, it is characteristic of languages that have VO. But in Japanese, heavy phrases (clauses, NPs, etc.) are preferably fronted in corpus data and in online experiments (Yamashita and Chang 2001, 2006; Hawkins 1994); see §5.3. This preference appears to be systematic in Japanese and in other strongly 'head-final' and OV languages as well; see Hawkins (1994) and Choi (2007) for similar data from Korean.

I argued in §3.2 that Heavy First provides counterevidence to the speed of formulation idea proposed by De Smedt (1994). If short phrases systematically win the competition for early production in English, they should do so in Japanese as well. To the extent that they do not, then either the proposed architecture must be abandoned, or some competing mechanism needs to be proposed explaining why this outcome of the proposed universal processing architecture is no longer an outcome in these languages.

There is another complication posed by these weight effects for psycholinguistic models. They are gradient. The preference for heavy last or heavy first

depends on the degree of difference in weight between the heavy phrase and its lighter sisters, not on some absolute size that might exceed the capacity constraints of models like Frazier's (1979, 1985), Just and Carpenter's (1992), or Gibson's (1998, 2000). The bigger the weight difference, the greater is the proportion of structures with heavy last or heavy first. For example, Heavy NP Shift in English (3.4b) applies in proportion to the weight difference between the NP and the PP (Hawkins 1994: 182–7, 2004: 111–12; Stallings 1998; Stallings and MacDonald 2011).

This complicates the explanation for weight effects in terms of capacity constraints in working memory load (see §3.1) since weight effects are visible among phrases with quite small weight differences (see the data in §2.1). The observed patterns of preference in these cases fall within the proposed capacity constraints, i.e., the constraints are too weak to account for weight effects in the data. They are also too strong since there are some language types with structures that will often exceed the proposed capacity constraints (Hawkins 1994: 266–8).

Another linear ordering effect that has been discussed in psychological models without full awareness of the relevant cross-linguistic differences and details involves the relative positioning of a subject and a direct object of a verb. It has been observed that subjects are preferably positioned before objects in languages like German and Finnish in which both SO and OS occur, as in the following examples from German:

(3.6) a. Die kleine blonde Frau küsste den roten Bären. (German)
The small blond lady kissed the red bear

b. Den roten Bären küsste die kleine blonde Frau.

The explanation proposed for this in Gibson (1998) appeals to 'memory cost' in online processing and uses an idea that goes back at least to Yngve (1960). According to Gibson (1998: 59) object-first orders such as (3.6b) are more complex at the initial noun phrases because it is necessary to retain the prediction of a subject at this location. A direct object always co-occurs with a subject, and a clearly and unambiguously marked initial accusative object as in (3.6b) will activate the prediction that a (nominative) subject follows. Subject-first noun phrases do not activate the prediction of an object (clauses may be transitive or intransitive) and hence, according to Gibson, they require less working memory in online production and comprehension and can be expected to be easier to process and more frequent.

At first sight, Gibson's explanation for subject-before-object ordering appears to be supported cross-linguistically in both performance and grammars. For example, only 4–5 percent of Japanese sentences in the corpus of Hawkins

(1994: 145) had OSV as opposed to SOV (see Yamashita and Chang 2001, 2006 for similar statistics). Basic orders across grammars also appear to support it, as shown in Tomlin's (1986) language sample reproduced in (3.7). Only 4 percent of these grammars have O before S (namely VOS and OVS orders).

(3.7) *Relative frequencies of basic word orders in Tomlin's (1986) sample* (402 lgs)

SOV (168)	VSO (37)	VOS (12)	OVS (5)
SVO (180)			OSV(0)
87%	9%	3%	1%

But when we look more closely at the morphological and syntactic properties of some of these languages, we see that this explanation in terms of the added memory cost of a predicting O before a predicted S cannot be correct, at least not as currently formulated. There may, however, be a minor explanatory role for it in a subset of typological data.

First, a significant number of SOV and VSO languages have ergatively case-marked transitive subjects (Comrie 1978; Primus 1999). In these systems the (ergative) subject receives special and distinctive marking signaling a transitive agent acting on a patient. This marking predicts the co-occurrence of a required object. The (absolutive) object morphology, by contrast, does not predict a second (subject) argument since the absolutive case is also the case found on intransitive subjects. An initial absolutive NP could be followed by an intransitive verb, therefore, or by an ergatively marked subject and a transitive verb. An initial absolutive in these languages is like an initial nominative in nominative–accusative systems: it does not predict a second following case-marked NP. The important point about these ergative–absolutive systems is that, in the great majority, the (predicting) ergative subject precedes the object in the most frequent and often basic order, despite the added memory cost of positioning the subject first. The North-East Caucasian language Avar is typical. (3.8) is a transitive clause with the most common ordering of S and O, and (3.9) an intransitive. The absolutive case is zero-marked in this and many other such languages (see Anderson 1976).

(3.8) Vas-as: jas j-ec:ula. (Avar)
boy-ERG girl-ABS SG.FEM.ABS-praise
'The boy praises the girl.'

(3.9) Jas j-ekerula.
girl-ABS SG.FEM.ABS-run
'The girl runs.'

A second problem for the memory cost explanation is that it is not the case that X before Y is always preferred over Y before X across languages when

Y predicts X and not vice versa. Sometimes the non-predicting X is indeed initial (English SO). Sometimes the predicting Y is initial, as in (3.8). Topic-marked phrases (e.g., Japanese *wa*) predict a predication and are also generally initial (Hawkins 2004: 235–40 and §2.3.2 here). Displaced Wh-words predict a gap and are almost invariably initial (Hawkins 2004: ch. 7 and §7.9 here). In all these cases the structures with greater memory cost are systematically preferred over those with less. Sometimes neither X nor Y predicts the other in these asymmetries. There is no consistent correlation, therefore, between ordering asymmetries across languages and memory cost (see Hawkins 2002, 2004: ch. 8 for further details).

There are further complications for a prediction-based explanation for ordering. One problem is that online predictions can both help processing, by activating co-occurrences at least one of which will be realized, but they can also hinder it by adding to working memory. It is not clear, therefore, what ordering one should expect when an online prediction is made by one category for the co-occurrence of another. Empirically the predicting category is sometimes first as we have seen (Avar (3.8) and Japanese topics, etc.) and sometimes last (accusatives following nominatives). The benefits and costs of an online prediction may simply be canceling each other out in these cases, making it an irrelevant consideration for ordering.

There is one set of ergative languages, however, that may provide some evidence for Gibson's memory cost ordering explanation, in redefined form, namely those that are often classified as OSV (e.g., Dyirbal, Kabardian) and OVS (e.g., Hishkaryana). There is evidently a competing motivation in these languages between a strong universal preference to position thematic agents before patients, and a second principle that favors positioning the (unpredictive) absolutive case first—i.e., the case which is found in both intransitive and transitive clauses, before the ergative (which is characteristic of transitives only and which predicts a co-occurring absolutive). This tension is discussed in detail in Primus (1999, 2002) and §8.2 below. The added memory cost of an initial ergative in the SOV order may be the explanation for this weak counter-principle to agents first. Hence a variant of Gibson's idea may be applicable not to explain the ordering of subjects before objects in all languages, but to actually explain the reverse object-before-subject ordering in a small subset of ergative languages.

What we see in these brief examples is that psycholinguistic models have been set up on the basis of familiar European languages, with insufficient attention to cross-linguistic variation. There is a welcome trend towards experimental work on different language types (cf., e.g., Hsiao and Gibson 2003; Matthews and Yeung 2001; Yamashita and Chang 2001; Saah and Goodluck 1995, to cite just a handful of studies). But there needs to be much

more awareness of cross-linguistic variation in psycholinguistics at this point. And since the study of diversity in grammars is decades ahead of studies of performance for relevant structures in more exotic languages, and since a strong correlation has now been shown between grammatical conventions and performance preferences in languages with variation (recall §1.2), it is useful for psycholinguists to examine grammars from a processing point of view before performance studies are undertaken in these languages. Interesting details about these grammars that deviate substantially from European norms, like heavy-first weight effects, and orderings of subjects and objects in ergative languages, can then suggest critical areas of performance to examine, both experimentally and in corpora as these become available for these languages.

3.6 Efficiency in Chomsky's Minimalist Program

A proposal has recently been made by Mobbs (2008) to incorporate the three efficiency principles of Chapter 2 into Chomsky's Minimalist Program (Chomsky 1995, 2004, 2005) and into the general picture of the human language faculty that is emerging within that program. A useful overview of this general picture can be found in Berwick and Chomsky (2011).

The Minimalist Program sees three factors as significant in the evolution of the language faculty. The first comprises general properties of the innate U(niversal) G(rammar). The second involves experience and the primary linguistic data that must be converted into particular grammars. And the third refers to more general cognitive principles not specific to the faculty of language that are shared with other cognitive systems and that apply to language as well. These third factors include 'principles of data analysis that might be used in acquisition,' 'principles of structural architecture and developmental constraints' (Chomsky 2005: 6), and also properties of the human brain that determine what cognitive systems can exist (Mobbs 2008: 9). Berwick and Chomsky (2011) propose that cognitive systems operate quite generally using principles of computational efficiency and economy in integrating perceptual and conceptual information. In the case of human language this integration is ultimately between linguistic sounds or forms on the one hand and meanings on the other, with interfaces to perception and articulation in the former case and to conception and intention on the other. The origin of these principles is seen not in the advantages of, and constraints on, externalized communication as such, but rather in the advantages/constraints pertaining to the internalized mental processes of sound–meaning mappings and their interfaces.

Mobbs (2008) proposes incorporating the three principles developed here and in Hawkins (2004) into the Minimalist Program, thereby adding some specificity to the rather programmatic remarks on efficiency in Minimalist writings hitherto. Hawkins's principles are seen now not as principles of performance but as general computational constraints and "'efficiency-oriented' third factors" (Mobbs 2008: 12) in the internalized mental processes for language. He (re)defines them as follows:

(3.10) *Minimize Domains*
Human computation minimizes the connected sequences of forms and associated properties in which relations of combination and/or dependency are considered.

(3.11) *Minimize Forms*
Human computation minimizes the formal complexity of each form F and the number of forms with unique property assignments, thereby assigning more properties to fewer forms.

(3.12) *Maximize Online Processing*
Human computation maximizes the set of properties that are assignable to each item X as it is considered, thereby increasing I(nitial) P(roperty) to U(ltimate) P(roperty) ratios.

3.6.1 Internal computations versus performance

There are two differences from the original definitions of Minimize Domains (2.1), Minimize Forms (2.8), and Maximize Online Processing (2.26) in Mobbs's reformulations: 'Human computation' has replaced 'The human processor' at the beginning of each; and reference to specifically linguistic entities has been removed since third factors are held to be general principles of cognition that apply to language as well as other cognitive systems. Apart from this, the essential details of (3.10)–(3.12) remain exactly as they were. By incorporating these changes, Mobbs sees the efficiency principles as being "radically reconceived" (p. 12) and he succeeds in "finding a use" (p. 11) for them within the Minimalist Program.

There is certainly some reconceptualization here at a general level. But the fact that these principles derived in large part from performance data can be so readily recategorized, preserving essential details, as computational principles regulating the internal mental processes that map forms onto meanings raises some deeper questions that Mobbs does not consider. It is interesting, though, that he recognizes and agrees with the role of efficiency in constraining grammars and cross-linguistic variation (Mobbs 2008: 16–42). One of the strong points of his thesis, in fact, is that for each principle he shows its

applicability to a number of syntactic and phonological phenomena discussed in the generative literature that go beyond the grammatical data of Hawkins (2004). For example, the Agree relation between a head and a specifier (Chomsky 1995; Rizzi 1991) is subject to a strict locality constraint, the 'Minimal Link Condition,' providing further support for Minimize Domains (Mobbs 2008: 29).

Another interesting feature of Mobbs (2008) is the explicit link that he proposes from this theory of language to principles of computational science and his proposal to ground and explain efficiency in terms of complexity theory, specifically time and space complexity. Citing Manber (1989) he writes (p. 13): "The time complexity of an algorithm is the number of operations it must perform to complete its task. The space complexity of an algorithm is the amount of working memory it requires while running." Computational science thus provides the general apparatus for characterizing the computations that operate within the language faculty. Processing, performance, and communication, on the other hand, involve externalized products of the language faculty and, following Chomsky, he argues that these are not ontologically prior: they did not provide a motivation for the evolution of human language and of the language faculty (cf. Berwick and Chomsky 2011); and instead of viewing performance and use as the ultimate explanation for principles and constraints in grammars (as argued by Hawkins 2004), he sees grammars as ontologically prior and performance as derived. In this way the Minimalist Program continues to assert the fundamental asymmetry between competence and performance that Chomsky proposed back in 1965 in *Aspects of the Theory of Syntax*, according to which principles of the competence grammar are constantly accessed in language production and comprehension—i.e., they shape and determine the data of performance, but performance did not shape the rules, principles, and constraints of grammars (recall §1.2 and see Hawkins 1994: 3–24 for a more detailed summary of the issues here and a review of relevant literature). Similarly Mobbs (2008) proposes that principles of (internalized) computational efficiency can significantly impact grammars and constraints on cross-linguistic variation, but (externalized) principles of processing cannot.

These Minimalist reflections are clearly relevant for the efficiency proposals of this book, given that they have now been incorporated and recast as so-called third factors and given the explicit rejection of the relevance and causality of performance. I find these proposals interesting in some respects, but unclear and questionable in others.

It is important to draw attention to the internalized computations that link sounds and meanings and their respective interfaces, and to try to make explicit the general cognitive principles, of efficiency or whatever they might

be, that apply to these mappings in addition to the constraints of UG. It is also interesting and provocative to suggest, as Berwick and Chomsky do, that human language evolved not in order to externalize these internal computations for the purpose of actual communication and use, but rather to have an improved internal language of thought for enhanced understanding, conception, problem-solving, etc. It is certainly possible that the route through which the above principles of efficiency entered into grammatical systems and their parameters was from these internalized computations and not from their externalized counterparts in real-time language processing and performance, as Mobbs argues.

But here is the basic problem, as I see it. What do minimalists think the relationship is between these internalized computations and the externalized principles of real-time processing that we can actually observe and measure (in controlled experiments and corpus productions)? Are they the same or different? And how could we find out? In short, what is the evidence for these internalized computations, if different from the externalized processes that we can actually study? For the rules and conventions of the competence grammar we have traditionally had the evidence of grammaticality judgments by native speakers. These judgments provide the data from which internalized grammatical conventions and properties can be extrapolated. But what are the computations that operate on these conventions and properties of particular languages to create individual meaningful sentences and internalized thoughts in our heads, if different from those that we activate in actual performance?

This is the first unclarity. It may well be, as far as I'm concerned, that the internal computations are different, in many or some respects, from the mechanisms that psycholinguists have proposed by which sentences and their meanings are actually produced and comprehended in real time. I am also open to the possibility that efficiency entered into grammatical conventions and parameters through 'pressure' from these internal computations and not from performance as such. But if there is no proposed way of finding out, and if the only evidence for the nature of the computations comes from externalized performance data, then the claimed separateness of the former, and the dismissal of performance, makes this third factor proposal vacuous and untestable.

The ease with which Mobbs (2008) converts the efficiency principles of Hawkins (2004) into internalized computational principles supports this criticism. These three principles have been supported by two types of data, performance data from individual languages and conventions of different grammars as reflected in current language samples, and it has been claimed that there is a correspondence between them. Hence observations from actual performance contributed crucially to these principles (in (2.1), (2.8), and

(2.26)). Mobbs, in effect, takes these same principles and relabels them, with minor changes but identical details ((3.10)–(3.12)) as internalized computational principles that apply to human language (inter alia). Before, he claims, they could not explain grammars. Now they can, in fact they can explain even more properties of grammars than I enumerated, just as long as we avoid the term 'performance' and preserve the claimed asymmetry between performance and grammar. I do not think I will be the only one who finds this disingenuous. Crucial evidence for these 'non-performance' principles has come from performance itself!

The explanation that Mobbs gives for the efficiencies, from complexity theory, further supports the idea of a strong overlap at least between internal computations and the processes required for externalized performance. Metrics of time complexity and of space complexity are exactly what psycholinguists have proposed in their measurements of online processing load—for example, in proposals for the number of items that can be held and processed at any one time in working memory (recall §3.1). It looks again like performance mechanisms and internal computations share essential details.

Let me stress that I am not arguing against the Minimalist proposal that there may be efficient internal computations of a general nature that apply to language and other cognitive systems. I am also open to the possibility that these contributed to shaping the conventions of grammars. What I am drawing attention to is the need for some clarity in the proposed relationship between internal computations and their externalized counterparts, and for some empirical evidence about the former if they are claimed to be substantively different from the latter. At the moment their supposed separateness lacks credibility since a crucial part of the evidence for the internal computations comes ultimately from performance data.

We can summarize the main issue here as follows. Let G stand for the set of grammatical principles and parameters and P the set of performance efficiency principles that Hawkins (2004) sees reflected both in G and in performance data supporting P. Let us assume now a set I of internalized computational principles operating in the manner of Chomsky. The question then becomes what is the relationship between I and P, and between I and G? There is empirical evidence for P from performance data. There is also empirical evidence for G from the grammaticality judgments of native speakers. But if claims about I are not testable independently of the data supporting P, or if I shares essential details with P, then the claim that I and not P shaped the conventions of G is either unsupported or unclear, motivated no doubt by earlier entrenched assumptions about performance not shaping the competence grammar (cf. §1.2).

3.6.2 Further issues

There are some further issues raised by these Minimalist reflections. Mobbs criticizes me for failing to see that the correspondence between performance and grammars can derive equally from the primacy of grammars shaping performance and not vice versa, i.e., from a Grammar–Performance Correspondence Hypothesis (Mobbs 2008: 6–7). In fact, I have consistently adopted a fairly conservative stance in my writings with respect to Chomsky's (1965) position that grammars of particular languages exist as internalized conventionalized mappings from sound to meaning and that they are constantly accessed in the performance of these languages, among the other factors that impact performance. Where I disagree with Chomsky is over the question of how these grammatical conventions actually got to be the way they are, and whether through constantly being accessed in performance they were not also influenced by performance pressures. And the way I have argued for this pressure has been by looking at structural variants such as 'free' word orders and their preferred selections in performance in languages permitting choices and comparing them with grammatical data involving 'fixed' ordering conventions, for example (see Hawkins 1994: 10–24 and Hawkins 2004: 5–7 for the logic of the argument here and §1.3 above).

There is no disagreement, therefore, that grammars shape the observable data of performance, and fundamentally so. That much is trivial. Two issues are not trivial, however. One is the question of whether, as a result of the correspondence between performance and grammars, there is any separate reality to grammars or whether they are simply routines for using and processing language (see O'Grady 2005). Here I actually side with Chomsky in assuming a conventionalized competence grammar separate from performance procedures, while agreeing with O'Grady philosophically on the profound influence of performance on grammars. The second issue is whether performance considerations can be said to explain grammatical conventions in the areas where they correspond, or whether there is some independent explanation for the grammatical conventions which then gets reflected in performance as well, as a result of grammars being accessed in performance. This is the direction of causality issue.

I am aware that a correspondence or a correlation between two sets of data is not an argument for the causality of one with respect to the other. More is clearly required. I address this, however, throughout my writings, by arguing that there is no purely grammatical explanation, except for stipulation, for many of the grammatical principles I discuss, whereas motivated explanations can be given from usage considerations (from frequency of occurrence, ease of processing, etc.) for properties of grammars. I see the distribution of gaps to

resumptive pronouns in relative clauses as reflecting the processing ease versus greater difficulty of identifying the position relativized on, zero allomorphs in morphological paradigms as correlating with frequencies of occurrence for the categories in question permitting correct default value assignments to zero forms, exceptions to certain word order parameters as involving linear orderings that are less than optimal but still relatively easy to process, and so on. Not only is there no independent grammatical explanation for these phenomena (they are either stipulated or exceptional), but the performance patterns are correlated with causative factors that are either unique to performance, such as frequency of occurrence, or they involve notions of complexity that possibly impact performance and internal computations equally, but for which we only have evidence from performance. Either way, the direction of causality seems to clearly favor the primacy of performance.

Notice that this issue of explaining the Performance–Grammar Correspondence Hypothesis (1.1) is separate from the question of how and why human language evolved, as discussed in Berwick and Chomsky (2011). It may well be, as they suggest drawing on a distinguished list of prior opinions, that language evolved as an instrument of thought rather than as a means of communication. This could well have been the ultimate selective advantage, as far as we know. But the cross-linguistic variants that sprang up from the prototypical language of the earliest homo sapiens sapiens are ultimately the variants that we observe today, many of which are summarized in this book. And for these, I have argued, there is a profound correspondence between grammatical principles and performance principles of efficiency, with only the latter being able to explain the former in an independently motivated way. If the impact of these latter was mediated through internal computations, then so be it. In this case, mechanisms tied clearly to performance (P above) will have shaped the internal computations of thinking (I), and P or I or both would then have shaped the grammatical conventions and parameters G. It may well be that P and I influenced each other and became profoundly interconnected and indistinguishable. Language had to be both thinkable and usable. This is plausible but highly speculative in the absence of details. What cannot be reasonably denied is that crucial empirical evidence for I comes from P, graphically so in Mobbs's minor adaptation of my efficiency principles, and also the crucial causal role of P in shaping G, possibly mediated through I. Chomsky's dismissal of the role of P in G is not well-founded.

Mobbs (2008: 2–4) also criticizes Hawkins (2004) for failing to "diagnose a real problem with a UG-based approach to learnability, or to offer a viable alternative" (Mobbs 2008: 7). In fact, I have been at pains to point out that

neither I, nor anyone else as far as I am aware, has a real solution to the 'negative evidence problem' (see Bowerman 1988 for one of the clearest statements of the issues here). This is the problem whereby the child hears only a small subset of the grammatical sentences of a language and does not receive instruction in ungrammaticality. The sentences that are not heard are either grammatical or ungrammatical therefore, and yet the child manages in a short time to learn the limits of grammaticality from the small set to which he or she is exposed. This ability is indeed a remarkable one and Chomsky was right to draw attention to it from his earliest writings.

There are three ways to attack the problem, as far as I can see. The first is to propose that, contrary to Chomsky, the positive data are sufficient to provide information about what is ungrammatical and absent in principle from the data, in addition to giving obvious evidence about the grammaticality of occurring sentences. So, for example, if the English-learning child hears only sentences with verb before object, and no Japanese-style sentences with the reverse, then the grammaticality of the former and the ungrammaticality of the latter can be safely inferred at an early age. More generally, if learners consistently hear structures of type X, and never hear some alternative to X, Y, then inferring the ungrammaticality of Y seems a safe bet. This is in essence what Stefanowitsch (2008) argues.

The second and related approach is to provide a model of learning that can operate on the data of experience and derive regularities and patterns from it that simulate the acquisition stages through which children progress. The general learning procedures discussed by Braine (1971), Berman (1984), Bowerman (1988), and Slobin (1973) were early examples of this type. Connectionist models are more recent influenced ultimately by attempts to capture the neural circuitry of the brain (cf., e.g., Rumelhart et al. 1986).

The third, Chomskyan approach is to propose certain innate predispositions or a Universal Grammar that can compensate for the claimed poverty of the stimulus, with the result that the relevant grammaticality distinctions do not need to be learned. One of the clearest statements of this position can be found in Hoekstra and Kooij's (1988) paper discussing subjacency violations, for example, which they claim are not learnable from the positive data to which the child is exposed.

There is surely something right about each of these approaches. If learners consistently hear structures of type X, and never hear some alternative to X, Y, then inferring the ungrammaticality of Y seems entirely reasonable. The issue here is: how many cases of ungrammaticality actually work like this such that positive data are sufficient for learning them? Stefanowitsch (2008) argues that these data are indeed sufficient for learning in general. At the very least we can conclude from this that ungrammatical sentences are

not always unlearnable from positive data. The pattern-finding capabilities of connectionist models are also remarkable and their ability to match the observed stages of language acquisition in many cases suggests that human learning has similar capabilities (Rumelhart et al. 1986).

With respect to innateness, the discussion in Hawkins (2004) which Mobbs (2008) criticizes does not deny that there is a strong innate component to language that distinguishes us from other species and which learners bring to the task of acquiring each particular language. The issue is: what aspects of language command does it apply to exactly? Principles and parameters of the grammar—if so which? Learning procedures? Processing mechanisms? Hawkins's (2004) main point about learnability was twofold. First, if many cross-linguistic universals are explainable through performance and are not the result of innate grammatical parameter setting, then they will not be available to the learner to solve learnability problems. Work in the Minimalist framework has now in any case greatly scaled back claims about the contents of an innate UG, while Mobbs (2008) is explicitly proposing efficiency explanations for grammars in the spirit of Hawkins (2004), so there is much less disagreement than there was. The second point that I drew attention to (Hawkins 2004: 10–12, 272–6) involved a critique of learnability arguments for the innateness of certain grammatical details. My point was, and is, that these arguments are hopelessly weak and prove nothing, irrespective of whether my performance-based explanation for grammars is correct or not.

A typical example can be found in Hoekstra and Kooij's (1988: 38) discussion of the pair of sentences given here as (3.13):

(3.13) a. Where did John say that we had to get off the bus?
b. Where did John ask whether we had to get off the bus?

They note that (3.13a) is ambiguous with *where* having matrix or embedded scope, i.e., the place that is being questioned is either where John said something or where we had to get off the bus. (3.13b) is not ambiguous and has only matrix scope, i.e., the place that is being questioned is where John asked something. Hoekstra and Kooij argue that "This piece of knowledge is shared by all native speakers, but it can hardly have been established on the basis of induction, simply because there are no data from which induction could conceivably proceed," and they add "Data like [(3.13a) and (3.13b), JAH], no matter how carefully 'stored', are in themselves largely insufficient to provide an answer to these questions unless the child also has access to complex principles of UG that determine the (im)possibility of extraction in otherwise similar configurations."

The confidence with which these authors assert the impossibility of learning here is remarkable, given that they cite no relevant empirical evidence for such a strong claim and see no need for it, they give no consideration to whether and to what extent ungrammaticalities might be derived from positive evidence in general, and they have no discussion of learning procedures and strategies that were being pursued in language acquisition studies at the time. It is certainly possible to imagine that learning might proceed along the lines outlined by Stefanowitsch (2008), by the child remembering that sentences of type (3.13b), with an initial Wh-word in the embedded clause, always have matrix scope, whereas those like (3.13a) with the complementizer *that* have both matrix and embedded-scope possibilities. This comparison across similar structures might then lead to a correct inference of ungrammaticality for certain syntactic movements over certain barriers. Whatever the details of this solution to the learnability problem turn out to be, it is legitimate to criticize proposals such as Hoekstra and Kooij's (1988) on the grounds that they are unempirical, theoretically uninformed, and unimaginative. They tell us nothing about learning and even less about a supposedly innate UG.

The logic of Hoekstra and Kooij's argument is further weakened by the argument in Hawkins (2004) that learners solve parallel learnability problems involving idiosyncrasies of particular languages that have nothing to do with UG and that cannot possibly reveal properties of the innate grammar or language faculty. This argument is also supported by the extensive data of Culicover (1999) on what he calls "syntactic nuts." Hawkins (2004: 10–11) captures this problem as follows:

> Consider, for example, the contrast between grammatical and ungrammatical subject-to-subject raisings in English, e.g., *John is likely to pass the exam* versus **John is probable to pass the exam*. It poses a learning problem that involves idiosyncratic facts of English, rather than principles of UG. Some adjectives (and verbs) behave like *likely* and permit raising, others behave like *probable* and do not. The child hears the former but not the latter and evidently succeeds in learning the grammaticality distinction based on positive evidence alone. The negative evidence problem here is so similar to that for which UG has been offered as a solution (e.g., for subjacency violations in Hoekstra & Kooij 1988), and the number of negative evidence problems that reduce to language-particular idiosyncrasies is so overwhelming (see Culicover 1999), that the whole relevance of UG to learnability must be considered moot. There is a real issue here over how the child manages to infer ungrammaticality from the absence of certain linguistic data, while not doing so for others, i.e., there is a negative evidence problem, as Bowerman (1988) points out. But UG cannot claim to be the solution.

In summary, Mobbs's (2008) incorporation of Hawkins's (2004) efficiency principles into the Minimalist Program is a welcome development that helps

to further integrate the discussion of efficiency and processing, on the one hand, and grammars on the other. His proposals, and the Minimalist Program itself, raise some issues in this regard to which I have drawn attention in this section. I return now to a focus on grammars and language typology and consider questions of language change. Specifically how do the efficiencies and processing preferences of the kind summarized in the first three chapters of this book enter into grammatical systems resulting in grammatical variation across languages?

4

The conventionalization of processing efficiency

According to the PGCH (1.1), efficiency principles of the kind proposed in Chapter 2 bring about variation patterns across grammars when they are 'conventionalized.' In other words, the forms and rules of grammars can become 'fixed' in favor of one or other performance alternative, in ways that reflect efficient processing. In this chapter I consider how such conventions come about. A key notion in this diachronic context is that of 'grammaticalization.' This term, as standardly used and understood (see §4.1 and §4.2), must be slightly extended to include the grammaticalization of syntactic rules within a formal grammar (§4.3). The 'adaptive mechanism' by which speakers implement new conventions in their grammars will be discussed briefly in §4.4.

4.1 Grammaticalization and processing

The term 'grammaticalization' was first introduced by Meillet (1912) to refer to the diachronic progression from a lexical item to a grammatical form, or from one grammatical form to another. Typical examples are future *will* in English which developed out of a corresponding verb meaning 'want' (cf. German *will* '(he/she) wants'), and the English definite article *the* which emerged from a demonstrative. There has been much research into grammaticalization phenomena in different languages in recent years, whose fruits are summarized in, e.g., Heine and Reh (1984), Traugott and Heine (1991), Hopper and Traugott (1993), Lehmann (1995), Heine and Kuteva (2002), and in historical linguistics textbooks such as Campbell (2004).

Grammaticalization has been described as a 'weakening' in both form and meaning. Heine and Reh (1984: 15) define it as 'an evolution whereby linguistic units lose in semantic complexity, pragmatic significance, syntactic freedom, and phonetic substance.' With respect to form, grammaticalization does very often result in a reduction in phonological units and/or in morphological and syntactic structure—for example, when independent words lose certain

sounds and become cliticized or affixed. This is exemplified by English *will* which can attach to a preceding subject, *I'll*, *he'll*, etc., and which loses its first two phonemes when it does so. Similarly *the* is phonologically reduced compared with *that* and *this*, etc., both in segments and in stress, and *a(n)* is reduced compared with the numeral *one* from which it developed.

As for the properties that are assigned to grammaticalized forms, they always play a productive role in the grammars of the relevant languages. *Will* joins the system of finiteness marking in English syntax and the tense system of English semantics. The definite article becomes an important part of noun phrase (or determiner phrase; cf. Abney 1987) syntax and semantics. The precise relationship between the 'donor' or source property P1 and the grammaticalized property P2 can vary, but there is always some natural and plausible link between P1 and P2 such that the one can merge gradually into the other over time. One's wishes and wants generally relate to the future, so future tense is a plausible semantic extension from wishing and wanting. The definite article also emerges via gradual and plausible stages from a demonstrative; see Hawkins (2004: 84–6). In fact, most of the research on grammaticalization has been devoted to these P1 → P2 pathways or 'clines' and to a linguistic analysis of the different relations between P1 and P2 and the types of semantic and syntactic changes that they exemplify (see Traugott and König 1991 for a summary and discussion).

But there is also a performance and processing aspect to grammaticalization. Campbell (2004: 294–6) summarizes forty grammaticalization clines, in all of which the P2 property appears to be more productive in grammars containing them than a lexical P1 would be, with the result that these F2:P2 forms will be used more frequently than those with F1:P1. The first ten of these clines are: (1) auxiliary < main verb; (2) case suffixes < postpositions; (3) case marking < serial verbs; (4) causatives < causal verb; (5) complementizer/subordinate conjunction < 'say'; (6) coordinate conjunction ('and') < 'with'; (7) copula ('to be') < positional verbs 'stand,' 'sit,' or 'give,' 'exist'; (8) dative case marker < 'give'; (9) definite article < demonstrative pronoun; (10) direct object case markers < locatives or prepositions.

The greater productivity of the grammaticalized form can be seen straightforwardly when a grammatical item develops out of a lexical one, e.g., when the verb 'give' provides the source for a dative case marker, since the lexical semantics of 'give' limits its applicability to a narrower range of sentences than does the corresponding syntax and semantics of dative case. But there appear to be regular expansions as well when one grammatical form (e.g., a demonstrative) gives rise to another (*the*). English frequency counts such as Francis and Kučera (1982) show that the definite article is used more frequently than demonstrative determiners. What this means is

that grammaticalization becomes relevant to the form minimization predictions of (2.8) and (2.9) in §2.2. See Bybee (2001) for a useful discussion of the general role of frequency in grammaticalization.

Greater frequency of use predicts reductions of form in grammaticalized F2:P2 pairs by (2.9a). This is not something that is unique to grammaticalization: it follows from Minimize Forms (2.8), and from the general principle of least effort that motivates it. Moreover, the greater productivity and frequency of F2:P2 is in accordance with the generalization captured in (2.9b) to the effect that form–property pairings are preferred in proportion to their potential for use. These grammaticalization clines are conventionalizing more frequently usable, not less frequently usable, and more preferred properties at successive historical stages.

I see three general issues that are raised by the grammaticalization literature when viewed from a processing perspective, and I believe that processing can contribute to their resolution.

First, why do we find *both* reductions in form *and* expansions in F2:P2 usage relative to F1:P1? Why not just P2 expansions without the formal reductions? Why aren't there *increases* in formal complexity co-occurring with property expansions? MiF and (2.9a) provide a general answer.

Second, can we predict and limit the set of regular grammaticalization pathways that we find evidence for in language history? The performance generalization which appears to be consistent with the literature is that the changes must result in an *increase* in frequency of usage. You cannot have P1 being more preferred and more frequent than P2 when P1 → P2. This generalization follows from (2.9b), since if universal diachronic changes were in the direction of less frequent and less preferred properties, or were indifferent to frequency and preference, then we would not see the kinds of priorities in favor of high frequency and preferred properties that we do see in cross-linguistic universals like the markedness hierarchies of §2.2.1. For example, we would not see the higher positions on these hierarchies being both more frequent in usage and more frequently conventionalized as grammatical forms in synchronic language samples now. It remains to be seen whether this increasing frequency prediction is always correct. It also remains to be seen whether the set of grammaticalization clines currently proposed in the literature is well defined. Some grammatical forms will lose in productivity as others gain. If a definite article develops out of a demonstrative, other forms of definiteness marking in a language may decline and the demonstrative determiners that overlap in usage with the definite article will also decline. The point about grammaticalization pathways is that they have the status of diachronic universals and laws, which should hold regardless of language type and regardless of the idiosyncrasies of particular languages.

The processing approach advocated here incorporates a hypothesis for constraining these universals of change and for making predictions about productive diachronic changes.

There is a third puzzle about grammaticalization which has not received the attention it deserves. Why is it that grammaticalization clines are set in motion in some languages and not in others, or set in motion at some stage of a language and not at another? Why or when does 'want' get recruited for a future tense marker, if at all? Why or when does 'give' become a dative marker, or a demonstrative a definite article, again if at all? The intuition that underlies discussions of this point, when they are discussed, is that there is some structural pressure or need for the relevant grammatical category (F2: P2) and that this motivates its expansion out of an earlier F1:P1. It is difficult to give precision and content to this claim. But a processing approach to grammar can clarify some of the intuitions here and come up with new generalizations that help us explain these expanding P1 → P2 properties. We need to look not just at the semantics/pragmatics but also at the syntax of each historical stage, and we need to consider how the relevant properties are processed, in order to formulate these generalizations.

These issues can be made clearer on the basis of an example, involving definiteness marking.

4.2 The grammaticalization of definiteness marking

The definite article, in languages that have one, is almost always descended from a demonstrative morpheme with deictic meaning (C. Lyons 1999: 331) and specifically from a demonstrative determiner (Diessel 1999a,b). The precise nature of this deictic meaning is discussed at length in J. Lyons (1977: vol.2) and can be roughly paraphrased as: *that book = the book (which is) there*; *this book = the book (which is) here* (J. Lyons 1977: 646). Anderson and Keenan (1985) survey demonstrative systems in a broad range of languages and they point out that they almost always involve a deictic contrast between at least two locations defined either with respect to the speaker alone or with respect to the speaker and the hearer. Their meaning involves a restrictive modification of a pragmatic nature that identifies which referent of the head noun is intended, and it is this restriction that is made explicit in John Lyons's paraphrases. The deictic restriction identifies the referent for the hearer by indicating that it exists and is unique within the appropriate subset of entities in the visible situation or previous discourse. By searching in this subset, the hearer can access it.

What we see in the evolution of definite determiners out of demonstratives, and in their further syntacticization, is a set of changes that are directly relevant for the three points raised in §4.1.

First, definite articles involve either less or equal formal marking compared with demonstrative determiners, not more: *the* has fewer and more reduced segments in English (CV rather than CVC, and a schwa vowel) and less stress. There are languages in which the erstwhile demonstrative and definite forms have evolved further into cliticized NP particles compatible with indefiniteness as well as definiteness—e.g., Tongan *e* (see Broschart 1997) and Maori *te* (see Bauer 1993). Such items have also been reconstructed as the source of affixal agreement markers on adjectives (in Lithuanian: C. Lyons 1999: 83), as the source of noun class markers (in Bantu: Greenberg 1978), gender markers (in Chadic: Schuh 1983), and gender and case markers (in Albanian: C. Lyons 1999: 71). Formal reduction and loss of syntactic and morphological independence go hand in hand here with increasing grammaticalization.

Second, the properties that are associated with later forms in this cline involve a gradual expansion in the set of NPs that are compatible with the erstwhile demonstrative marking. Definite articles have semantic/pragmatic and syntactic properties that permit their attachment to more NPs than do demonstrative determiners and they are used more frequently. Expansion to indefiniteness, as in Polynesian, permits more attachments still and even greater frequency of usage. Further syntacticization as agreement, noun class, and gender markers, etc. (Greenberg 1978) permits usage of the relevant cliticized or affixal forms regardless of their original semantic/pragmatic properties. We can hypothesize that these final stages of grammaticalization will also be accompanied by increasing frequency of usage, until the cycle is complete. They can then be removed from the language altogether, as the reverse process of morphological reduction and syncretism sets in of the kind we have witnessed in the once richly inflected Indo-European. As this happens, new grammaticalization clines are set in motion, such as demonstrative → definite article in the recorded Germanic and Romance languages.

Third, a processing approach to these historical expansions leads to new hypotheses about why these originally demonstrative forms should be recruited for property extensions. It has been commonplace to focus on the expanded semantics and pragmatics of the definite article compared with the demonstrative (e.g., Hodler 1954). But this is a consequence, as I see it, not a cause of the grammaticalization. There is no compelling semantic/pragmatic reason why the definite article should emerge out of a demonstrative to express meanings that are perfectly expressible in languages without definite articles. But there are some compelling reasons, involving the processing of grammar, that can motivate an expansion of the determiner

category, and that can motivate it in language types and at historical stages at which it appears to happen. And these reasons can also make sense of a number of properties in the grammar of definiteness that would otherwise be mysterious.

These processing motivations for the evolution of definiteness marking were considered in detail in Hawkins (2004: 84–93), to which the reader is referred. A more general discussion of noun phrase syntax and morphosyntax from a processing perspective is provided in Chapter 6 of the present book, in which the kinds of grammaticalization motivations for definiteness discussed in Hawkins (2004: 84–93) are extended to many more cross-linguistic patterns in noun phrase typology.

4.3 The grammaticalization of syntactic rules

For many of the conventionalized efficiencies presented in this book we see evidence not for the diachronic progression from a lexical to a grammatical form, or from one grammatical form to another, but for the emergence of a new syntactic rule in response to processing. In Hawkins (1994: 19–24) I laid out several basic ways in which syntactic rules can be grammaticalized, and these can be usefully summarized here.

First, syntactic rules or principles or constraints may select some categories rather than others and apply selectively to the relevant categories, in a way that reflects processing efficiency. For example, the grammar may block a center-embedded clause (S) in structures like (4.1) in English, while tolerating a center-embedded NP in this environment as in (4.2):

(4.1) *Did [$_{S}$ that the boy failed his exam] surprise Mary?

(4.2) Did [$_{NP}$ the boy] surprise Mary?

Embedded clauses are typically much longer than NPs in performance, which makes the Phrasal Combination Domain (2.3) for the matrix clause longer and more complex in (4.1) than in (4.2). The Minimize Domains preference of (2.1) has been grammaticalized here in the selective blocking of (4.1) and is defined in terms of a center-embedded S, assuming that this is indeed an ungrammatical sentence in English and that its blocking is a property of the grammar rather than just a performance dispreference. The important point here is that different phrasal types are available to a grammar, S versus NP, and that these *can* be selected by syntactic rules in particular languages to define as ungrammatical certain sentence types that are systematically bad for processing, due to the average weights of the phrases in question in performance.

Correspondingly, extraposition rules may apply to an S in English but not to NP moving S to the right within its clause and leaving a place-holder *it* in its original position as in (4.1'):

(4.1') Did it surprise Mary [$_S$ that the boy failed his exam]?

(4.2') *Did it surprise Mary [$_{NP}$ the boy]?

The improvement in IC-to-word ratios from (4.1) to (4.1') is considerable; see §2.5.1 and Hawkins (1994: 190–6) for precise quantification of the effects of extraposition. Extraposition of a clause S is accordingly highly beneficial for a language like English, which motivates its 'grammaticalization' in the form of a syntactic rule and which makes structure (4.1') available as a solution to the center-embedding constraint seen in (4.1). Extraposition of an NP as in (4.2') is not similarly motivated by Minimize Domains, though right-dislocation structures with referential pronoun copies as in (4.2") may also be motivated by domain minimization in addition to whatever discourse properties and motivations they have (see, e.g., Gundel 1988):

(4.2") Did he*i* surprise Mary [$_{NP}$ the boy who failed his exam]*i*?

Notice an interesting consequence of the grammar's conventionalization of the well-formedness of (4.2), however. Syntactic rules and constraints apply to phrases and categories, not to terminal strings. So on those occasions in performance when an NP happens to be longer than an embedded clause normally is, the grammaticality facts remain the same. Compare (4.1) and (4.2) with (4.3):

(4.3) Did [$_{NP}$ the failure of the boy to pass his exam] surprise Mary?

The long and complex center-embedded NP of (4.3) is grammatical, just as (4.2) is grammatical, and provides a structure in which English speakers can express the meaning intended by the ungrammatical (4.1). The grammar is sensitive to the categorial status of these word groupings, not to their length on particular occasions—i.e., because the length and complexity of NP is typically less than that of an embedded clause, English has grammaticalized a distinction between them which permits (4.2) to be generated as grammatical. As a result, the long center-embedded NP of (4.3) can survive as grammatical to express the meaning of (4.1) despite its unusual length and complexity. The typical and average weights of these phrases are arguably what led the grammar to make the distinction between center-embedded S and NP in the first place.

Second, processing efficiency can determine the relative ordering of categories. English has grammaticalized [$_{VP}$ V NP PP] within its VP, and

[PP P NP] within PP because these are orders that are most efficient for Minimize Domains in (2.1). V is typically a one-word item, and NP is on average significantly shorter than PP but longer than one word (see §7.8.2), which means that this relative ordering conforms to the short-before-long preference of head-initial languages (§5.2). A direct object also contracts more syntactic and semantic relations with the verb than a PP (see §7.8), which also favors the adjacency of V and NP by Minimize Domains. When an NP is exceptionally longer and more structurally complex than a PP, these are the circumstances in which 'Heavy NP Shift' (Ross 1967) applies to rearrange [VP V NP PP] into [VP V PP NP] (as in *I* [*gave* [*to my uncle*] [*the ideal gift which I had been looking for for many years*]] see Hawkins 1994: 182–4), thereby returning these VPs to the short-before-long regularity of the basic order. The shift to V-PP adjacency is also highly frequent when V and PP contract a lexical relation or a 'collocation' in Wasow's (2002) terms, i.e., when the V and PP constitute a processing unit similar to that which normally holds between V and a direct object NP; see §7.8.2.

For the grammaticalization of short-before-long basic orders in other major syntactic phrases of English, the reader is referred to Hawkins (1994: 282–96). The directionality of rearrangements such as Heavy NP Shift and Extraposition in English, namely to the right, is also predicted by Minimize Domains. Such rearrangements will always be to the right in a head-initial language, and will be either to the left or to the right in a head-final language depending on whether the moved phrase is constructed on its right periphery, as in rigidly head-final Japanese (see §5.3), or on its left periphery as in non-rigid OV languages such as Persian and German (see §5.8.1).

Third, many mother–daughter relations in phrase structure rules are driven by processing. If an abstract phrase structure tree is to be clearly recognized and derived from a rapid linear string of words, this imposes certain constraints. Some lower categories and parts of speech must provide evidence about higher phrases and about the phrasal groupings of words. I have argued (Hawkins 1993, 1994: 343–58) that this is what ultimately underlies and motivates the 'head of phrase' generalization. It accounts in a straightforward way for the many properties of heads (Zwicky 1985; Hudson 1987), such as their obligatory presence and the distributional equivalence of heads to their mothers. It also accounts for why heads are the morphosyntactic locus for inflections, and for the constituent domains within which relations such as government and subcategorization can hold. All of these points are argued in detail in Hawkins (1993, 1994: 343–58). This processing perspective on phrase structure also makes sense of facts that are not so readily handled by purely grammatical approaches to the head of phrase generalization (cf. Corbett et al. 1993 for summaries of several such approaches) and that

involve alternations within and across languages between different categories (not all of them heads) that can construct higher nodes. Many examples of such alternations are given in the present book (Chapter 6) as a way of explaining morphosyntactic differences between noun phrases across languages, and as grammaticalizations of different categories that can construct noun phrases.

Fourth, processing efficiency leads to constraints and exceptions on independently motivated rules and principles. These latter can be quite general, involving say Wh-movement, or more language-particular. In either case certain unexpected departures from an independently motivated and general grammatical pattern can be motivated in terms of processing. This was exemplified in detail in Hawkins (2004: ch. 7) for 'subjacency'-type restrictions on otherwise productive filler-gap dependencies. I argued there that several hierarchies of Wh-movement and relativization, with their systematic cut-off points in different languages and alternations between gaps and resumptive pronouns in easier and more difficult environments respectively, could be readily explained in terms of increasing processing complexity down the relevant hierarchies (see especially Hawkins 2004: 169–202 and also §2.2.3 here).

Numerous cases of processing efficiency motivating exceptions to independently motivated language-particular rule formulations are also relevant here. Adjective phrases in English provide a case in point. English allows single-word adjectives in prenominal position (4.4), but not adjectives with post-modified phrases as in (4.5a,b):

(4.4) a yellow book

(4.5) a. *a [yellow with age] book
b. *a [with age yellow] book
c. a book [yellow with age]

It does allow adverbial modifiers to precede adjectives, however, in both prenominal and postnominal position:

(4.6) a. a [very yellow] book
b. a book [very yellow with age]

By standard assumptions (see, e.g., Jackendoff 1977), an adjective projects to an AdjP and hence all of these examples have an AdjP sister to an N, either before it or after it. Any attempt to argue that single-word adjectives such as *yellow* are not adjective phrases, and to formulate a separate positioning rule for single-word categories versus phrases is complicated by (4.6a) in which the addition of the adverb makes [*very yellow*] an AdjP, just as [*yellow*

with age] is an AdjP. So if AdjP can occur before N within NP, then all lower material dominated by AdjP should also be able to occur wherever AdjP can occur. But it can't—see (4.5a,b). So what is wrong with these phrases?

Processing efficiency suggests an answer. A left-branching modifier of N is at variance with the general head-initial syntax of English. Dryer (1992) has found that departures from consistent head ordering like this are typically limited to single-word non-heads or modifiers. The explanation proposed in the next chapter (see esp. §5.5) is that single-word non-heads intervening between heads make only a limited addition to phrasal combination domains (2.3) and are accordingly tolerated by Minimize Domains (2.1). Phrasal non-heads with right-branching modifiers as in (4.5a,b) add significantly to phrasal combination domains within the NP (e.g., (4.5a,b) have a 3/5 = 60% IC-to-word ratio for NP compared to 3/3 = 100% for (4.5c)). They also add significantly to phrasal combination domains within other higher phrases like VP in the event that there is no initial article to construct NP (cf., e.g., **I read yellow with age books*, in which the PCD for VP proceeds from the verb *read* to the head of the NP, *books*, making a ratio of 2/5 = 40% compared to 2/2 = 100% for *I read books yellow with age*).

The pre-adjectival adverb in [*very yellow*] is not blocked because this is typically, again, a single-word premodifier in performance which adds little to the inefficiency of phrasal combinations domains.

Consider another syntactic idiosyncrasy that follows readily from efficiency. The rule placing a *that*-complementizer in a subordinate clause would normally be described as 'optional,' since such clauses are grammatical both with and without them (*I believe (that) he is sick*). In some environments, however, *that*-placement is obligatory:

(4.7) a. *He is sick surprised Mary.
b. That he is sick surprised Mary.

This example was used in early discussions of the influence of processing on grammars (e.g., Bever 1970; Frazier 1985). For example, Frazier (1985) attributed its ungrammaticality to the Impermissible Ambiguity Constraint and argued that structures that garden-path the hearer on every occasion of use, as (4.7a) does (*He is sick* will always first be assigned a main clause interpretation), will be blocked by the grammar. The impossibility of deleting subject relative pronouns (*the professor *(who) wrote the book gave a lecture*) versus the optionality of non-subject relative pronouns (*the professor (who(m)/that) I know gave the lecture*) also fits in here since omitted subject relative pronouns will generally result in a misassigned main clause interpretation, *the professor wrote the book* in place of *the professor who wrote the book*. These garden paths are very inefficient by the misassignment criteria of Hawkins

(2004: 51–8) and we can accept Frazier's explanation for their grammatical blocking, though I would argue that the impact of processing efficiency on individual grammars and on typological variation has been much more pervasive than was recognized in these early discussions, as argued throughout this book.

Recall in this context the surprising contrast between (2.19b) and (2.20) in Cantonese repeated here as (4.8b) and (4.9):

(4.8) a. [Ngo5 ceng2 0*i*] go2 di1 pang4jau5*i* (Cantonese)
I invite those CL friend
'friends that I invite'

b. *[Ngo5 ceng2*i* keoi5dei6*i*)] go2 di1 pang4jau5*i*
I invite them those CL friend

(4.9) [Ngo5 ceng2 (keoi5dei6*i*) sik6-faan6] go2 di1 pang4jau5*i*
I invite (them) eat-rice those CL friend
'friends that I invite to have dinner'

The relative clause with a resumptive pronoun is ungrammatical in the simple environment of (4.8b) but grammatical in the more complex (4.9) with a non-minimal filler-gap domain. This exception to the normal optionality of both gaps and pronouns when relativizing on a DO in Chinese can be seen as conventionalized efficiency. The pronoun is not necessary in the easy-to-process (4.8) and merely adds an additional word to the relativization domain. The gap structure is therefore good for both Minimize Forms (2.8) and Minimize Domains (2.1); see §2.2.3. In the less easy-to-process (4.9), however, the pronoun is helpful as a way of identifying the position relativized on and it provides a local lexical domain for processing the arguments of the verb *ceng2* ('invite'). The gap involves less form processing, the pronoun helps with relative clause processing, hence both strategies are grammatical in (4.9).

The reader is also referred here to the explanation given in Hawkins (2004: 215–18) for the so-called 'that-trace' effect in English (Chomsky 1973) and in certain other languages. In contrast to Impermissible Ambiguity examples like (4.7a) in which the presence of *that* is obligatory, its occurrence is blocked, surprisingly, immediately before a gap or trace, as in (4.10a):

(4.10) a. *the person*i* who*i* you think [$_S$ that 0*i* committed the crime]
b. the person*i* who*i* you think [$_S$ 0*i* committed the crime]

The argument that I gave in Hawkins (2004: 215–18) is that (4.10b) is the more efficient structure overall when total domain sizes for all processing operations are calculated that apply to this sentence. The complementizer *that*

adds an extra word that is unnecessary in this environment since the very next word, the finite verb *committed*, will project to and construct the subordinate clause just as effectively as *that* does. English has conventionalized the more efficient option here.

Matters are actually more subtle since there are also instances of this structure in which the presence of *that* can be more efficient, contrary to the normal case, namely when the finite verb is not initial in S and when material has been fronted to its left. Culicover (1993) drew attention to examples like (4.11) of exactly this type, which appear to be grammatical:

(4.11) John is the kind of person*i* who*i* I suspect [$_S$ that after a few drinks 0*i* would be unable to walk straight]

It may be that English grammar blocks *that* only when it is immediately before a gap, therefore, and not when there is intervening material. Hawkins's (2004) efficiency principles Minimize Domains and Maximize Online Processing (see §§2.2, 2.3) predict this preference for a retained *that* over its absence here, since *that* is useful as a left-edge constructor of S when it does not immediately precede the finite verb *would*. Whatever the best grammatical formulation is for the 'that-trace' principle, well-formedness judgments appear to reflect processing efficiency, and this latter can explain when certain rules may, must, or cannot apply.

In all of these examples the exceptions and constraints on rules are hard to motivate in purely grammatical terms. They make sense in efficiency terms, and grammars appear to have conventionalized the most efficient options in these cases.

Fifth, processing efficiency can be seen to constrain general patterns of variation throughout a whole grammar. For example, it determines whether a language will employ raising rules or Wh-fronting in its syntax, whether it will assign a productive set of non-agentive thematic roles to grammatical subjects, whether it will have rich case morphology, and so on. In Chapter 7 of the present book I examine a number of correlations with different verb positions across languages and argue that a large number of structural differences between VO and OV languages have arisen as a consequence of their ease and efficiency in processing.

Finally in this section notice that whereas conventions of syntax can emerge that grammaticalize efficiency in the ways we have seen (by distinguishing between different phrasal categories in their inputs and outputs, imposing orders on configurations, limiting the domains for government and subcategorization, making rules apply obligatorily or optionally or not at all, and so on), these conventions have to conform to certain formal conventions that hold for grammars in general. So, rules apply to phrases like NP to define its

immediate constituents or they apply to effect a certain change, of movement, deletion, etc. Rules do not in general apply to terminal elements or to individual lexical items (though there may be certain lexical exceptions to otherwise productive rules; cf., e.g., Culicover 1999). So a putative 'rule' of Heavy NP Shift (J. R. Ross 1967) which relies crucially for its application on the length and complexity of terminal elements dominated by NP is not a well-defined rule, since its application is vague and indeterminate. Not only is it unclear which NPs it would apply to, but the notion of heaviness that is relevant here, and that can be shown empirically to result in the shifting of NPs from their immediate postverbal position, depends crucially on the relative weight of NPs in comparison with whatever sisters happen to be present, and not on any absolute weight in the NP itself (as shown in Hawkins 1994: 182–8; Stallings 1998; Stallings and McDonald 2011). It is impossible for the grammar to conventionalize a response in this case in the form of standard phrasal categories and rules. Similarly, the (gradient) applicability of relativizer deletion in non-subject relative clauses in English and the sensitivity of these deletions to lexical choices and predictability within an NP (see Wasow et al. 2011 and §2.2.2) does not lend itself readily to conventionalization within a syntactic rule. These cases differ, therefore, from the productive cases of grammaticalization that we have seen in this section, and also from those we have seen in §§4.1–4.2 in which erstwhile lexical items actually changed their status to become productive parts of the syntax, this being a prerequisite for grammaticalization (C. Lehmann 1995).

4.4 The mechanisms of change

Having laid out the lexical and grammatical sources for new grammatical words (§§4.1–4.2) and the kinds of syntactic changes that grammars can conventionalize (§4.3), all of these historical changes being driven ultimately in crucial respects, as I have argued, by processing efficiency, we now need some account of the 'adaptive mechanisms.' This was the term used by Kirby (1999) to describe the mechanisms by which successive generations of language learners and users actually develop new grammars that incorporate efficiency principles like Early Immediate Constituents (now MiD (2.1); see §2.1), resulting in the observed typological patterns. Kirby's computer simulation was a novel and interesting attempt to model the way in which new grammaticalized linear orderings evolve. But more generally, how do new grammatical forms and rules actually get into the grammar from a preceding stage of the language in which a grammatical convention does not exist?

A number of linguists have been concerned with this question and have clarified these adaptive mechanisms in recent years, often with insightful

comparisons between evolutionary biology and historical linguistics; see especially Haspelmath (1999a), Croft (1996, 2000), Bybee (1988, 2001), Bybee and Hopper (2001), and Heine (2008). Haspelmath (1999a) summarizes three crucial ingredients for the evolution of new linguistic conventions. First, there has to be variation among structural alternatives at some stage of a language, e.g., alternative orderings of the same phrases (see §2.1 and §§5.2–5.3), or both gaps and resumptive pronouns when relativizing on the same position (see §2.2.3), reduced and less reduced morphological forms (§2.2.1), phonological variants, and so on. Second, principles of 'user optimality' apply to these variants at each successive stage to select some rather than others, resulting in different frequencies of use in performance. In the present context, these principles would include the efficiency principles developed in this book. And third, the preferred and most frequently used variants of performance may become obligatory and categorical, with the less frequent alternatives being lost altogether.

Haspelmath illustrates these points in detail, choosing the constraints of Prince and Smolensky's (1993) Optimality Theory as the conventions of grammar that have arisen in response to user optimality, and discussing how these constraints are responses to user optimality and how they could evolve from one generation of learners and users to the next out of the structural variants that preceded them. The reader is referred to Haspelmath's paper and to the other works cited here for further elaboration on these adaptive mechanisms. In the remainder of this chapter I shall just mention some further points of direct relevance to the efficiency principles developed here.

The relative sequencing of changes, and the innovation of new variants to the current grammatical state of the language, proceeds gradually, as historical linguistics textbooks point out. A language will change the ordering of single-word adjectives before it reorders adjective phrases, i.e., it will develop variant single-word orders first before phrasal variants. This is the clear diachronic basis for the synchronic typological pattern that Dryer (1992) documents, whereby single-word modifiers of heads are often typologically inconsistent with the predominant head ordering of the grammar. A change involving these word orders also proceeds via a 'word order doubling' stage (Hawkins 1983) in which both orders (e.g., AdjN and NAdj) exist at the same time, as in the Romance languages. I pointed out in Hawkins (1983: 213–15) that these word order doublets arise at the points of transition between pre-posed and postposed modifiers on various word order hierarchies (such as the Prepositional Noun Modifier Hierarchy summarized in (5.24) of §5.5 which is structured by phrasal complexity).

In relativization hierarchies, like the Keenan–Comrie (1977) Accessibility Hierarchy (2.17) (see §2.2.3), gaps or pronouns are innovated again as competing variants or doublets, as in the Hebrew direct object relatives (cf. (2.16a,b)), and again at the points of transition between these strategies on that hierarchy. See the data of Table 2.1 for confirmation of this point. The gap/pro structures are transitional between the higher gap positions and the lower pro positions in the relevant languages. The competing gap/pro positions are not at points that are distant from either strategy. For example, the SU gap is extended first to the DO and only later to the IO or OBL.

From the perspective of the PGCH (1.1) this means that small conventionalized processing load departures from the current grammar precede bigger ones. If you are going to change your word order, change the single-word categories first that have minimal impact on the processing of phrase structure before you change the phrasal categories that have a bigger impact. Languages, dialects, and sociolects will therefore innovate certain grammatical changes before others and we accordingly see variation patterns arising in different social and regional speech communities in certain parts of the grammar first and not in others.

The ultimate triggers or causes for these shifts that motivate speakers to change their speech habits can be various, as historical textbooks point out (see, e.g., Campbell 2004; McMahon 1994). Increasingly, historical linguists are coming to recognize the profound role of contact between languages (cf. Thomason and Kaufman 1988; Thomason 2001) and of different types of bilingualism between speakers that result in transfers from one language to another (see Hawkins and Filipovic 2012 for discussion of transfer from different L1s into L2 English). Trudgill (2011) distinguishes between contact situations in which there are many non-native adult learners of the L2 versus those with genuinely bilingual native speakers of the two languages. The contact-induced changes that result from the former are different, he argues, from those in the latter and include morphological simplifications as a consequence of imperfect learning. Native-speaking bilinguals, by contrast, can serve as a conduit for the transfer of more complex linguistic features between languages.

It may not be the bilingual learners themselves who innovate the grammatical changes, i.e., children, but rather adolescent bilinguals or other social groups. Trudgill's (2011) perspective provides a welcome antidote to what seems to have become an unquestioned assumption in much of linguistics, to the effect that it is necessarily children who innovate language change (cf., e.g., King 1969; Lightfoot 1991; Kirby 1999). First of all, to the extent that learners do change features of the language they are learning that subsequently spread to the rest of the speech community, they may be second-language learners

who are present in sufficient numbers to have an impact. Second, the language changers may not be children at all (see further the work of Labov 1972, 1994, 2007 in this connection).

Contact and bilingualism provide the social conditions for the spread of head-initial and head-final word orders in strikingly contiguous areas of the globe, as seen in *WALS* maps such as 83 and 85 (Dryer 2005a,b) involving VO versus OV and prepositions versus postpositions respectively. This explains why the only Austronesian languages to depart from their VO syntax and change to OV are those that are in New Guinea and in contact with indigenous OV languages. See M. D. Ross (1996, 2001) for discussion of a particularly extreme case of this type involving the development of OV patterns in the Western Oceanic language Takia under the influence of Papuan Waskia on the north coast of Papua New Guinea. In Mesoamerica we see the reverse shift of OV to VO among Uto-Aztecan languages like Nahuatl in contact with VO Mesoamerican languages (see Gast 2007). Heine (2008) gives an insightful discussion of how new word orders arise diachronically in these kinds of contact situations by building on and extending minority orders that existed before and by adapting already existing structures in the recipient language. In addition to the important role played by non-native speaking adults versus genuine bilinguals in these contact situations Trudgill (2009, 2011) draws attention to the denseness versus looseness of social networks and to community size as major forces that impact the spread of linguistic features, and he gives intriguing correlations between these social variables and the evolution of simpler or more complex variants. Lupyan and Dale (2010) provide further evidence for social correlations with structural simplicity and complexity in grammars. For a summary of the various causes that have been proposed for major typological changes in the history of English, see Hawkins (2012).

Processing efficiency adds three things to these other, primarily external, causes of change. First, it modulates their sequencing and relative timing, as we have seen, with smaller changes in processing load preceding larger ones. This constrains the types of variation one finds synchronically, e.g., word order doublets for single-word modifiers of heads before doublets are found with phrasal modifiers, and so on. Second, processing can cause internal adjustments and a ripple effect throughout the grammar in the event that, e.g., a head-final OV language changes its verb position to VO. Other phrases will then change their head ordering in response to ease and efficiency of use as the domains for phrase structure processing are gradually minimized throughout the whole grammar, with smaller adjustments again preceding larger ones. Third, different efficiency principles sometimes cooperate and sometimes compete within languages and structures, as we will see throughout this book

(see Chapter 9 for a summary and some proposed patterns of interaction). It is not possible for each language type to satisfy all the preferences all the time, and this is the essential reason, I believe, why we see the variation that we do. The most common language types appear to be good for most principles, albeit not for all, and the efficiencies that cannot be satisfied in a given language type create an internal tension for change, and a readiness for change in the event that external factors such as language contact and language transfer come to favor it. I drew attention to some instances of this tension in §2.5 and will discuss more of them in §7.2, §7.6, and §9.3.

Processing efficiency therefore helps us understand which structures will change, when they will change, and the directionality of the changes, i.e., what they will change into. Efficiency is a hitherto relatively neglected cause of language change and it motivates many changes that are often discussed as grammatical or 'language-internal' in historical linguistics, such as head ordering. The variation patterns discussed in this book have ultimately arisen, I argue, not because of some mechanism internal to the grammar, nor as a consequence of innate parameters triggered by the data of experience, and not necessarily because of changes introduced by child language learners, but as a result of processing efficiency acting within various social groups of language users to select and conventionalize certain variants from among competing options while gradually eliminating others. Efficiency can be added to the language-external and -internal factors that are routinely recognized and investigated in historical linguistics and sociolinguistics, to give us a more complete model of language change, variation, and mixing.

In the chapters that follow I consider in detail the kinds of typological variation patterns that have emerged from these efficiency-determined grammaticalizations in a number of areas, starting with word order.

5

Word order patterns: Head ordering and (dis)harmony

5.1 Head ordering and adjacency in syntax

An important set of generalizations about word order in both cross-linguistic typology and formal grammar has been captured in terms of 'head ordering' (Greenberg 1963; Hawkins 1983; Dryer 1992; Corbett et al. 1993; Newmeyer 2005). When a phrase XP immediately contains another phrase YP we have the following four ordering possibilities for the heads X and Y within each:

(5.1) **XP** (5.2) **XP** (5.3) **XP** (5.4) **XP**

X YP	**YP X**	**X YP**	**YP X**
Y ZP	**ZP Y**	**ZP Y**	**Y ZP**
Head-initial	Head-final	Mixed	Mixed

(5.1) and (5.2) are consistently and harmonically head-initial and head-final respectively. (5.3) and (5.4) are 'inconsistent' or 'disharmonic' word orders.

Types (5.1) and (5.2) are almost always fully productive across languages. This can be seen in the 'Greenbergian' correlations, which I discussed briefly in §2.1 and to which I now return. Consider the structure in which a prepositional or postpositional phrase (PP) is immediately dominated by a verb phrase (VP). The four ordering possibilities are exemplified using English words in (5.5). Language frequencies for each are shown in (5.6), using data from Hawkins (1994: 257) provided by Matthew Dryer and taken from his (1992) typological sample (measuring languages rather than groups of related languages or 'genera'):

(5.5) a. [$_{VP}$ went [$_{PP}$ to the movies]] (5.1) b. [[the movies to $_{PP}$] went $_{VP}$] (5.2)
c. [$_{VP}$ went [the movies to $_{PP}$]] (5.3) d. [[$_{PP}$ to the movies] went $_{VP}$] (5.4)

(5.6) a. [$_{VP}$ V [$_{PP}$ P NP]] = 161 (41%) b. [[NP P $_{PP}$] V $_{VP}$] = 204 (52%)
c. [$_{VP}$ V [NP P $_{PP}$]] = 18 (5%) d. [[$_{PP}$ P NP] V $_{VP}$] = 6 (2%)
(5.6a)+(b) = 365/389 (94%)

The consistently head-initial and head-final types, found in English and Japanese respectively, account for the vast majority of languages, 94%. The disharmonic (5.3) (exemplified by Kru) is found in 5%, and disharmonic (5.4) (exemplified by Persian) in 2%.

A very similar distribution can be seen in the other Greenbergian correlations, for example when an NP containing a head noun and a possessive phrase sister is contained within a PP in phrases corresponding to {$_{PP}$ with {$_{NP}$ soldiers {$_{POSSP}$ of [the king]}}}:

(5.7) a. [$_{PP}$ P [$_{NP}$ N Possp]] = 134 (40%) (5.1)
b. [[Possp N $_{NP}$] P $_{PP}$] = 177 (53%) (5.2)
c. [$_{PP}$ P [Possp N $_{NP}$]] = 14 (4%) (5.3)
d. [[$_{NP}$ N Possp] P $_{PP}$] = 11 (3%) (5.4)
(5.7a) + (b) = 311/336 (93%) (data from Hawkins 1983)

Notice, as a preliminary to our head-ordering discussion in this chapter, that grammatical heads are a subset of what are called 'mother node constructing categories' in Hawkins (1994: ch. 6). The parsing principle of Mother Node Construction applies to such categories and builds on Kimball's (1973) New Nodes:

(5.8) *Mother Node Construction* (Hawkins 1994: 62)
In the left-to-right parsing of a sentence, if any word of syntactic category C uniquely determines a phrasal mother node M, in accordance with the PS rules of the grammar, then M is immediately constructed over C.

In the terminology of generative theories of syntax category C 'projects' to M (Radford 1997; Corbett et al. 1993).

A second parsing principle, proposed for non-heads in Hawkins (1994: ch. 6), is Immediate Constituent Attachment:

(5.9) *Immediate Constituent Attachment* (Hawkins 1994: 62)
In the left-to-right parsing of a sentence, if an IC does not construct, but can be attached to, a given mother node M, in accordance with the PS rules of the grammar, then attach it, as rapidly as possible. Such ICs may be encountered *after* the category that constructs M, or *before* it, in which case they are placed in a look-ahead buffer.

Why is it that the head-adjacent orders are preferred over the others? In the theory of Hawkins (1994, 2004) this is because there are principles of processing efficiency that favor the preferred orders and that have become conventionalized in the syntax of most languages. For example, the adjacency

of V and P in (5.6a,b) guarantees the smallest possible string of words for the construction of VP and PP, and for the attachment of V and PP to VP as sister ICs. Non-adjacency of heads in (5.6c,d) is less efficient for phrase structure processing.

Specifically I have argued that the smallest possible string of words is preferred for the construction of phrases and for recognition of the combinatorial and dependency relations that hold within them. This was referred to as the principle of Early Immediate Constituents (EIC) in Hawkins (1994), which was then generalized in Hawkins (2004) to a similar preference for 'minimal domains' in the processing of all syntactic and semantic relations; cf. also Gibson's (1998) similar 'locality' principle. Minimize Domains was defined in (2.1) in §2.1 and is repeated here (without the definitions for combination and dependency):

(5.10) *Minimize Domains* (MiD) (Hawkins 2004: 31)
The human processor prefers to minimize the connected sequences of linguistic forms and their conventionally associated syntactic and semantic properties in which relations of combination and/or dependency are processed. The degree of this preference is proportional to the number of relations whose domains can be minimized in competing sequences or structures, and to the extent of the minimization difference in each domain.

The consistently head-initial and head-final orders (5.1) and (5.2) are optimal by this principle: two adjacent words suffice for the construction of the mother XP (projected from X) and for construction of the mother YP (projected from Y) and for its attachment to XP as a sister of X. Structures (5.3) and (5.4) are less efficient: more words need to be processed for construction and attachment because of the intervening ZP.

MiD can be argued to motivate the grammatical principle of Head Adjacency and the Head Ordering Parameter in generative grammar (Newmeyer 2005). MiD can also explain why there are two highly productive mirror-image language types, the head-initial and the head-final one. The basic reason would appear to be that heads can be adjacent in two logically possible orders and the processing domains for phrase structure recognition and production can be minimal in both—i.e., these two types are equally efficient. Structures (5.3) and (5.4) are not as efficient and both are significantly less productive.

In order to support this explanation it needs to be shown, of course, that Minimize Domains applies to performance data as well as grammars. In the sections that follow, I shall summarize some variation data from English and

Japanese in which users have a choice between the adjacency or non-adjacency of certain categories to their heads. It turns out that there are systematic preferences in performance, mirror-image ones interestingly between these languages, as predicted by MiD. Many further datasets from different languages are given in Hawkins (1994) and (2004). I will then show that this same principle can be found in the fixed conventions of grammars in languages with fewer options. Specifically, this principle gives us an explanation, derived from language use and processing, for general patterns in grammars, but also as we shall see for otherwise puzzling exceptions to these patterns, and for grammatically unpredicted datasets involving, e.g., hierarchies.

5.2 MiD effects in the performance of head-initial languages

Words have to be assembled in comprehension and production into the kinds of phrasal groupings that are represented by tree structure diagrams. The recognition of these phrasal groupings and word combinations can typically be accomplished on the basis of less than all the words dominated by each phrase. Some orderings reduce the number of words needed to recognize a mother phrase M and its immediate constituent daughters (ICs), making phrasal combination faster. Compare (5.11a) and (5.11b), which were discussed briefly in §2.1 and which are repeated here:

(5.11) a. The man [$_{VP}$ looked [$_{PP_1}$ for his son] [$_{PP_2}$ in the dark and quite derelict building]]
1 2 3 4 5

b. The man [$_{VP}$ looked [$_{PP_2}$ in the dark and quite derelict building] [$_{PP_1}$ for his son]]
1 2 3 4 5 6 7 8 9

Five words suffice for recognition of the three items, V, PP1, PP2 in (5.11a), compared with nine in (5.11b), assuming that head categories such as P immediately project to mother nodes such as PP, enabling the parser to construct and recognize them online. For comparable benefits within a production model, cf. Hawkins (2004: 106).

Minimize Domains predicts that Phrasal Combination Domains (PCDs) should always be as short as possible, and that the degree of this preference should be proportional to the minimization difference between competing orderings. This principle (a particular instance of Minimize Domains) is called Early Immediate Constituents (EIC), again repeated from §2.1:

(5.12) *Phrasal Combination Domain* (PCD)
The PCD for a mother node M and its I(mmediate) C(onstituent)s consists of the smallest string of terminal elements (plus all M-dominated non-terminals over the terminals) on the basis of which the processor can construct M and its ICs.

(5.13) *Early Immediate Constituents* (EIC) [Hawkins 1994: 69–83]
The human processor prefers linear orders that minimize PCDs (by maximizing their IC-to-word ratios), in proportion to the minimization difference between competing orders.

EIC prefers short before long phrases in head-initial structures like those of English, i.e., the short before the long PP in (5.11a). These orders have higher 'IC-to-word ratios' and permit more ICs to be recognized on the basis of fewer words in the terminal string. The IC-to-word ratio for the VP in (5.11a) is 3/5 or 60 percent (five words required for the recognition of three ICs), whereas that of (5.11b) is 3/9 or 33 percent (nine words required for the same three ICs).

The empirical results of a corpus search reported in Hawkins (2000, 2001) were summarized in §2.1. Relevant structures were selected on the basis of a permutation test: two postverbal PPs had to be permutable with truth-conditional equivalence (i.e., the speaker had to have a choice). Only 15 percent (58/394) of these English sequences had long before short. Among those with at least a one-word weight difference (excluding 71 with equal weight), 82 percent had short before long, and there was a gradual reduction in the long-before-short orders, the bigger the weight difference as shown in (5.14) (PPS = shorter PP, PPL = longer PP):

(5.14)

n = 323	PPL > PPS by 1 word	by 2–4	by 5–6	by 7+
[V PPS PPL]	60% (58)	86% (108)	94% (31)	99% (68)
[V PPL PPS]	40% (38)	14% (17)	6% (2)	1% (1)

A PCD as defined in (5.12) is a domain for processing the syntactic relation of phrasal combination or sisterhood. Some of these sisters contract additional syntactic–semantic relations with the verb of the kind that grammatical models try to capture in terms of verb–complement (rather than verb–adjunct) relations, e.g., *count on your father* versus *play in the playground* (place adjunct). Complements are listed in the lexical entry for each head, and the processing of verb–complement relations should also prefer minimal lexical domains, by MiD (5.10).

(5.15) *Lexical Domain* (LD)
The LD for assignment of a lexically listed property P to a lexical item L consists of the smallest possible string of terminal elements (plus their associated syntactic and semantic properties) on the basis of which the processor can assign P to L.

One practical problem here is that the complement/adjunct distinction is a multi-factor one covering different types of combinatorial and dependency relations, obligatoriness versus optionality, semantic dependence versus independence, etc. It is not always straightforward to convert what is not actually a binary distinction into a binary one (cf. Schütze and Gibson 1999). Hawkins (2000, 2001) proposes the following entailment tests for defining PPs that are lexically listed, thereby capturing what seems to be the essential intuition behind the complement relation:

(5.16) **Verb Entailment Test**: does [V, {PP1, PP2}] entail V alone or does V have a meaning dependent on either PP1 or PP2? E.g., *the man waited for his son in the early morning* entails *the man waited*; *the man counted on his son in his old age* does not entail *the man counted.*

(5.17) **Pro-Verb Entailment Test**: can V be replaced by some general Pro-Verb or does one of the PPs require that particular V for its interpretation? E.g., *the boy played in the playground* entails *the boy did something in the playground*, but *the boy depended on his father* does not entail *the boy did something on his father*!

If V or P is dependent on the other by these tests, then the PP is lexically listed, i.e., dependency is used here as a sufficient condition for complementhood and lexical listing. The PPs classified as independent are (mostly) adjuncts or unclear cases.

When there was a dependency between V and just one of the PPs, then 73 percent (151/206) had the interdependent PP (Pd) adjacent to V in the corpus of Hawkins (2000, 2001), i.e., their LDs were minimal. Recall that 82 percent had a short PP adjacent to V preceding a longer one in (5.14), i.e., their PCDs were minimal. For PPs that were *both* shorter *and* lexically dependent, the adjacency rate to V was 96 percent, which was (statistically) significantly higher than for each factor alone.

We can conclude that the more syntactic and semantic relations whose domains are minimized in a given order, the greater the preference is for that order: multiple preferences result in a stronger adjacency effect when they reinforce each other, as predicted by MiD (5.10). MiD also predicts a stronger adjacency preference within each processing domain in proportion to the minimization difference between competing sequences. For PCDs this difference is a function of the relative weights of the sisters; see (5.14). For LDs it is a

function of the absolute size of any independent PP (Pi) that could intervene between the verb and the interdependent PP (Pd) by the entailment tests, thereby delaying the processing of the lexical co-occurrence. This is shown in (5.18) in which the increasing size of Pi results in increasingly fewer separations of Pd from V:

(5.18)	n = 206	Pi = 2–3 words	:4–5	:6–7	:8+
	[V Pd Pi]	59% (54)	71% (39)	93% (26)	100% (32)
	[V Pi Pd]	41% (37)	29% (16)	7% (2)	0% (0)

Multiple preferences have an additive adjacency effect when they work together, therefore (see §9.2), but they result in exceptions to each when they pull in different directions (see §9.3). Most of the 58 long-before-short sequences in (5.14) involved some form of lexical dependency between V and the longer PP (Hawkins 2000). Conversely, V and Pd can be pulled apart by EIC and MiD in proportion to the weight difference between Pd and Pi (see Hawkins 2004: 116). I return to this competition in §9.3.1, within the context of a broader discussion of competing principles and their relative strengths. In that section I also summarize results from a valuable recent study by Wiechmann and Lohmann (2013) which uses a larger database of postverbal PPs in English and reassesses the findings of Hawkins (2000).

5.3 MiD effects in head-final languages

In head-final languages the categories that construct mother nodes (V, P, Comp, case particles, etc.) are on the right of their respective sisters: recall (5.2) above. Long-before-short orders now provide minimal PCDs in structures containing a plurality of phrases or a plurality of sisters or both. For example, if the direct object of a Japanese verb is a complement clause headed by a complementizer *to*, as in (5.19), the distance between this complementizer and other constituents of the matrix clause in (5.19b), the subject *Mary ga* and the verb *it-ta*, is short, just as short in fact as it is in the mirror-image English translation *Mary said that*.... Hence the Phrasal Combination Domain for the matrix clause in (5.19b) is minimal. In (5.19a), by contrast, with a center-embedded complement clause, this PCD for the matrix clause proceeds all the way from *Mary ga* to *it-ta*, and is much longer.

(5.19) a. Mary ga [[kinoo John ga kekkonsi-ta to $_S$]
Mary NOM yesterday John NOM married that
it-ta $_{VP}$] (Japanese)
said
'Mary said that John got married yesterday.'

b. [kinoo John ga kekkonsi-ta to $_S$] Mary ga [it-ta $_{VP}$]

Calculating the PCD for the VP depends on the precise syntactic analysis for (5.19b). If the complement clause is still discontinuously attached to the VP, then the subject *Mary ga* will intervene. If it is extracted into the highest S as a sister to the remaining clause *Mary ga it-ta* then either the nominative case marker *ga* or the verb will plausibly project to this clause (see Hawkins 1994: ch. 6 for relevant constructing principles) and again there will be some intervening material between the fronted complement clause and the category that constructs its sister clause. Either way a preference for (5.19b) is predicted in proportion to the relative weight difference between subject and object phrases (recall the similar discussion of EIC trade-offs in §2.5.1 between phrases in the rightward-moving extraposition structures of English and German).

A long-before-short preference is also predicted for [{NPo, PPm} V] structures in Japanese, in alternations such as (5.20) (with *-o* standing for the accusative case particle, and PPm for a postpositional phrase with a head-final postposition):

(5.20) a. (Tanaka ga) [[Hanako kara $_{PP}$] [sono hon o $_{NP}$] kata $_{VP}$] (Japanese)
Tanaka NOM Hanako from that book ACC bought
'Tanako bought that book from Hanako'

b. (Tanaka ga) [[sono hon o $_{NP}$] [Hanako kara $_{PP}$] kata $_{VP}$]

Relevant corpus data were collected by Kaoru Horie and are reported in Hawkins (1994: 152). Letting ICS and ICL stand for the shorter and longer immediate constituent respectively (i.e., with weight as the crucial distinction rather than phrasal type), these data are summarized in (5.21) (excluding the phrases with equal weights):

(5.21)

n = 153	ICL>ICS by 1–2 words	by 3–4	by 5–8	by 9+
[ICS ICL V]	34% (30)	28% (8)	17% (4)	9% (1)
[ICL ICS V]	66% (59)	72% (21)	83% (20)	91% (10)

These data are the mirror image of those in (5.14): the longer IC is increasingly preferred to the left in the Japanese clause, whereas it is increasingly preferred to the right in English. This pattern has been corroborated in experimental and further corpus data by Yamashita and Chang (2001, 2006) and Yamashita (2002), and it underscores an important principle for psycholinguistic models. The directionality of weight effects depends on the language type. Heavy phrases shift to the right in English-type (head-initial) structures, but to the left in Japanese-type (head-final) structures; see Hawkins (1994, 2004) for extensive illustration and discussion.

In (5.22) the data of (5.21) are presented by phrasal type (NPo versus PP) and relative weight:

(5.22)

	NPo>PPm by			NPo=PPm	PPm>NPo by		
n = 244	5+	3–4	1–2		1–2	3–8	9+
[PPm NPo V]	21% (3)	50% (5)	62% (18)	66% (60)	80% (48)	84% (26)	100% (9)
[NPo PPm V]	79% (11)	50% (5)	38% (11)	34% (31)	20% (12)	16% (5)	0% (0)

We see preferred adjacency for a direct object complement to a verb here (i.e., [PPm NPo V]) when weight differences are equal or small, namely 66 percent to 34 percent for equal weights, and 62 percent to 38 percent when NPo is longer than PPm by 1–2 words. Long-before-short and verb–direct object adjacency are equally balanced when weight differences are larger, at 3–4 words, while long before short wins impressively by 79 percent to 21 percent when weight differences are large (5+) in the leftmost column. This is plausibly a consequence of the fact that NPs are generally complements and in a lexical combination with V, whereas PPs are either adjuncts or complements, mostly the former (recall the data for English PPs in the last section that were in a lexical complement relation to the verb and that showed tighter adjacency to V as a result). The data of (5.21) show clear progressive evidence for the long-before-short preference in Japanese, increasing with increasing weight differences, while (5.22) shows that many of the exceptional short-before-long orders of Japanese involve (longer) direct object NPs that follow the competing preference for short lexical domains for verb and complement processing.

The patterns of preference seen in these data samples involving freely ordered NPs and PPs from English and Japanese are increasingly being replicated in further findings from these and other languages, using both corpus and experimental methodologies; see, e.g., Yamashita (2002) and Yamashita and Chang (2001, 2006) for Japanese, Choi (2007) for Korean, Matthews and Yeung (2001) for Cantonese, Uszkoreit et al. (1998) for German, Kizach (2010, 2012) for Russian, Hawkins (1994) for a variety of typologically different languages, and Stallings (1998), Stallings and MacDonald (2011), Wasow (2002), Lohse et al. (2004), Marblestone (2007), and Wiechmann and Lohmann (2013) for English. We turn now to grammars and show how these same preferences can be seen in cross-linguistic variation patterns, in accordance with the Performance–Grammar Correspondence Hypothesis (1.1) in §1.1.

5.4 Greenberg's word order correlations and other domain minimizations

Grammatical conventions across languages reveal a clear preference for minimal domains. The relative frequencies of language types reflect the preferences, as do hierarchies of co-occurring word orders. An efficiency approach

can also explain a number of grammatical principles such as consistent head ordering.

Let us return to the implicational universal with which I began this chapter in §5.1. Greenberg (1963) examined alternative verb positions across languages and their correlations with prepositions and postpositions in phrases corresponding to (5.5) repeated here as (5.5′) with the phrasal combination domains for VP (cf. (5.12)) shown by the underlinings:

(5.5′) a. [$_{VP}$ went [$_{PP}$ to the movies]] b. [[the movies to $_{PP}$] went $_{VP}$]
- - - - - - - - - - - -
c. [$_{VP}$ went [the movies to $_{PP}$]] d. [[$_{PP}$ to the movies] went $_{VP}$]
- -

(5.5′a) is the English order, (5.5′b) is the Japanese order, and these two sequences are massively preferred in language samples over the inconsistently ordered heads in (5.5′c) and (5.5′d), as was shown in (5.6):

(5.6) a. [$_{VP}$ V [$_{PP}$ P NP]] = 161 (41%) b. [[NP P $_{PP}$] V $_{VP}$] = 204 (52%)
c. [$_{VP}$ V [NP P $_{PP}$]] = 18 (5%) d. [[$_{PP}$ P NP] V $_{VP}$] = 6 (2%)
(5.6a)+(b) = 365/389 (94%)

The adjacency of V and P guarantees the smallest possible string of words (indicated by the underlinings in (5.5′)) for the recognition and construction of VP and of its two immediate constituents, namely V and PP—i.e., adjacency provides a minimal Phrasal Combination Domain for the construction of VP and its daughters, of the same kind we saw in the performance preferences of §§5.2–5.3.

Consistent head ordering in grammars can be argued to derive from Minimize Domains (5.10), therefore. Conventions of ordering have emerged out of performance preferences, and one and the same principle explains both the preferred conventions of grammar and the preferred structural selections in performance (in languages and structures in which speakers have a choice). MiD can also explain why there are two productive mirror-image types here, head-initial and head-final languages (5.1) and (5.2) respectively: they are equally efficient strategies for phrase structure comprehension and production.

Purely grammatical approaches can also define a head-ordering parameter (cf. Newmeyer 2005: 43 and Haspelmath 2008 for full references), and Svenonius (2000: 7) states that this parameter is 'arguably not a function of processing.' It is certainly possible that this is an autonomous principle of grammar with no basis in performance. But how do we argue for or against this?

A classic method of reasoning in generative grammar has always involved capturing significant generalizations and deriving the greatest number of observations from the smallest number of principles. An autonomous head-ordering principle would fail to capture the generalization that both grammatical and performance data fall under Minimize Domains. The probabilistic and preferential nature of this generalization is also common to both types of data. Moreover, many other ordering universals point to the same preference for small and efficient phrasal combination domains, e.g., in noun-phrase-internal orderings corresponding to [$_{NP}$ bright students [$_{S'}$ that Mary will teach]] in English:

(5.23) [$_{NP}$ Adj N [$_{S'}$ C S]] C = the category that constructs S′: e.g. relative pronoun, complementizer, subordinating affix or particle, participial marking on V, etc. (Hawkins 1994: 387–93)

There are twelve logically possible orderings of Adj, N, and S′ (ordered [C S] or [S C]). Just four of these have minimal PCDs for the NP (100 percent IC-to-word ratios), all of them with adjacent Adj, N, and C, namely [N Adj [C S]] (Romance), [Adj N [C S]] (Germanic), [[S C] N Adj] (Basque), and [[S C] Adj N] (Tamil). These four account for the vast majority of languages, while a small minority of languages are distributed among the remaining eight in proportion to their IC-to-word ratios measured online (Hawkins 1990, 1994, 2004). There appears to be no straightforward grammatical account for this distribution of occurring versus non-occurring and preferred versus less preferred grammars. The distribution does correlate with degrees of efficient processing in NP phrasal combination domains, however.

Consider also the fact that complements prefer adjacency to heads over adjuncts in the basic orderings of numerous phrases in English and other languages and are generated in a position adjacent to their heads in the X-bar theory of Jackendoff (1977) and in Head-driven Phrase Structure Grammar (Pollard and Sag 1987, 1994). Tomlin's (1986) Verb–Object Bonding discussion provides cross-linguistic support for this by pointing to languages in which it is impossible or dispreferred for adjuncts to intervene between a verbal head and its DO complement.

Why should complements prefer adjacency to heads in grammars when there are basic ordering conventions defined on these categories and phrases? The reason, I suggest, is the same as the one I gave for the preferred orderings of complements (Pd and NPo) in performance in §5.2 and §5.3. There are more combinatorial and/or dependency relations linking complements to their heads than linking adjuncts. Complements are listed in a lexical

co-occurrence frame defined by, and activated by, a specific head (e.g., a verb); adjuncts are not so listed and occur in a wide variety of phrases with which they are semantically compatible (Pollard and Sag 1994). The verb is regularly lexically dependent on its DO, not on an adjunct phrase: compare the different senses of 'run' in *run the race/the water/the advertisement/the campaign*, cf. Keenan (1979). A direct object receives a thematic role from V, typically a subtype of Dowty's (1991) Proto-Patient, whereas adjuncts do not receive thematic roles. The DO is also syntactically required by a transitive V, while adjuncts are not syntactically required sisters. Processing these lexical co-occurrence relations favors minimal lexical domains (cf. (5.15)).

5.5 Explaining grammatical exceptions and unpredicted patterns

Further support for a Minimize Domains explanation for head ordering comes from grammars with exceptional head orderings. Dryer (1992) points out that there are systematic exceptions to Greenberg's correlations when the category that modifies a head is a single-word item, e.g., an adjective modifying a noun (*yellow book*). Many otherwise head-initial languages have non-initial heads here (English is a case in point); many otherwise head-final languages have noun before adjective (e.g., Basque). But when the non-head is a branching phrasal category (e.g., an adjective phrase as in English *books yellow with age*) there are good correlations with the predominant head ordering (recall the discussion in §4.3). Why should this be?

When a head category like V has a phrasal sister, e.g., PP in {V, PP}, then the distance from the higher head to the head of the sister will be very long when heads are inconsistently ordered and are separated by a branching phrase, e.g. [$_{VP}$ V [NP P $_{PP}$]] in (5.6c). An intervening phrasal NP between V and P makes the PCD for the mother VP long and inefficient compared with the consistently ordered counterpart (5.6a) [$_{VP}$ V [$_{PP}$ P NP]], in which just two words suffice to recognize the two ICs. But when heads are separated by a non-branching single word, then the difference between, say, [$_{VP}$ V [$_{NP}$ N Adj]] and [$_{VP}$ V [Adj N $_{NP}$]] is short, only one word. Hence the MiD preference for noun initiality (and for noun finality in postpositional languages) is significantly less than it is for intervening branching phrases, and either less head-ordering consistency or no consistency is predicted. When there is just a one-word difference between competing domains in performance—see, e.g., (5.14) in English—we also saw that both ordering options are generally productive, and so too in grammars.

MiD can also explain numerous patterns across grammars that do not follow readily from grammatical principles alone. Hierarchies of permitted

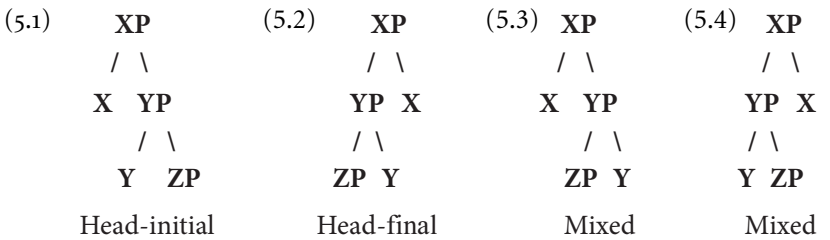
center-embedded phrases provide an example. In the environment [$_{PP}$ P [__ N $_{NP}$]] we have the following center-embedding hierarchy (Hawkins 1983):

(5.24)	Prepositional languages:	AdjN	32%	NAdj	68%
		PosspN	12%	NPossp	88%
		RelN	1%	NRel	99%

As the average size and complexity of nominal modifiers increases (relative clauses exceeding possessive phrases, which in turn exceed single-word adjectives), the distance between P and N increases in the prenominal order and the efficiency of the PCD for PP declines compared with postnominal counterparts. As efficiencies decline, the relative frequencies of prenominal orders in conventionalized grammatical rules decline also.

5.6 Disharmonic word orders

Let us now consider the disharmonic word orders (5.3) and (5.4) in greater detail from this processing perspective. I repeat the four basic options for head ordering:



Within formal grammar, a proposal has recently been developed in a series of publications beginning with Holmberg (2000) for a different partitioning that distinguishes the disharmonic type (5.4) from (5.3) and from the consistent orders as well. It is referred to as the Final-Over-Final Constraint (FOFC) and in Holmberg (2000) it was defined as follows:

(5.25) *The Final-Over-Final Constraint* (FOFC)
If α is a head-initial phrase and β is a phrase immediately dominating α, then β must be head-initial. If α is a head-final phrase, and β is a phrase immediately dominating α, then β can be head-initial or head-final.

This rules out (5.4) and permits all of (5.1)–(5.3). The precise formulation of FOFC given in refinements made over the years (see Biberauer, Holmberg, and Roberts 2007, 2008) incorporates principles of Minimalist Syntax

(Chomsky 1995, 2000; Kayne 1994) and has, in effect, limited the applicability of FOFC to those instances of (5.4) in which YP and XP are of the same or similar syntactic type, e.g. both verbal heads of some kind, as opposed to different syntactic phrases such as a PP within a VP. Only the former are now excluded as impossible. In what follows I shall consider different types of disharmony and various instances of type (5.4), whether they fall under the current limitations of FOFC or not, since the perspective to be pursued here is different and since the patterns to be revealed in this way point to a different explanation.

The alternative proposed here is, of course, a processing one. Formal grammarians view FOFC as a grammatical and ultimately innate principle of U(niversal) G(rammar). They acknowledge (Theresa Biberauer, Ian Roberts, Anders Holmberg, personal communications) that many disharmonic word order data are not covered by FOFC, especially those of type (5.3) and also some instances of type (5.4), for which processing efficiency principles may be applicable. However, they wish to preserve FOFC as an autonomous and grammatical principle and as an exceptionless absolute universal.

The interesting research question then becomes whether the processing principles offered here, which account for many statistical patterns across languages but also as we shall see for some apparently exceptionless ones as well, can also subsume the data for which FOFC has been proposed. Lack of comparability between available typological data and formal syntactic analyses of relevant languages makes it difficult to give a definitive answer to this question at this time. Hence I shall try to make the strongest possible case I can for an efficiency explanation, which I believe does make FOFC redundant as an independent principle of grammar, but I leave it to others to adjudicate whether this is indeed the case.

In the meantime the research done by formal grammarians on disharmonic word order languages is most welcome and the online processing considerations to be given here will hopefully complement their work and allow deeper questions of causality to be resolved in the future. I have argued in Hawkins (1985), and will reiterate here, that typologists and formal grammarians need to work together in areas such as this in order to identify the precise cross-linguistic patterns. The formal grammarian gives depth to the syntactic analysis, the typologist gives breadth, while a psycholinguistic approach to typology gives a new type of explanation that can potentially account for the productive patterns, including absolute universals, and also for the minority types and exceptions.

From a typological perspective the FOFC looks, at first sight, like it is not quite right: languages with (5.4) are generally dispreferred, occasionally unattested (i.e., it appears to be too strong); while languages with (5.3) appear

to be similarly dispreferred and occasionally unattested (i.e., it is too weak). Types (5.1) and (5.2) are almost always fully productive. This supports the traditional generalization whereby (5.1) and (5.2) are preferred over both (5.3) and (5.4). It does not support a generalization in which (5.4) is singled out for special treatment.

5.7 The timing of phrasal constructions and attachments

When parsing principles (5.8) Mother Node Construction and (5.9) Immediate Constituent Attachment apply to the terminal elements of structures (5.1)–(5.4) they result in very different timing patterns online for the construction and attachment of these phrases. I repeat the two principles and also (5.1)–(5.4), for convenience:

(5.8) *Mother Node Construction* (Hawkins 1994: 62)
In the left-to-right parsing of a sentence, if any word of syntactic category C uniquely determines a phrasal mother node M, in accordance with the PS rules of the grammar, then M is immediately constructed over C.

(5.9) *Immediate Constituent Attachment* (Hawkins 1994: 62)
In the left-to-right parsing of a sentence, if an IC does not construct, but can be attached to, a given mother node M, in accordance with the PS rules of the grammar, then attach it, as rapidly as possible. Such ICs may be encountered *after* the category that constructs M, or *before* it, in which case they are placed in a look-ahead buffer.

(5.1) **XP**: **X** **YP**; **YP**: **Y** **ZP** — Head-initial

(5.2) **XP**: **YP** **X**; **YP**: **ZP** **Y** — Head-final

(5.3) **XP**: **X** **YP**; **YP**: **ZP** **Y** — Mixed

(5.4) **XP**: **YP** **X**; **YP**: **Y** **ZP** — Mixed

In (5.1) X first constructs XP, then Y constructs YP at the next word, and YP is immediately attached left as a daughter to the mother XP. The processing of ZP then follows.

In (5.2) ZP is processed first, Y then constructs YP, and X constructs XP at the next word. YP is immediately attached right as daughter to the mother XP. Note that the attachment of YP follows its construction by one word here, a point that will be of some interest and to which I return in section §5.8.3.

In (5.3) X first constructs XP, then after processing ZP Y constructs YP and this YP is attached left to the mother XP, possibly several words after the construction of XP. The result is delayed assignment of the daughter YP to XP.

In (5.4) Y constructs YP first, then after processing ZP X constructs XP and YP is attached right to the mother XP, possibly several words after the construction of YP. This results in delayed assignment of the mother XP to YP.

Structures (5.1) and (5.2) are optimal from the perspective of Minimize Domains (5.10), therefore. Both (5.3) and (5.4) are non-optimal. From the perspective of Maximize Online Processing (see (2.16) in §2.3), which I repeat here for convenience as (5.26), we can note the following.

(5.26) *Maximize Online Processing* (MaOP) (Hawkins 2004: 51)
The human processor prefers to maximize the set of properties that are assignable to each item X as X is processed, thereby increasing O(nline) P(roperty) to U(ltimate) P(roperty) ratios. The maximization difference between competing orders and structures will be a function of the number of properties that are unassigned or misassigned to X in a structure/sequence S, compared with the number in an alternative.

Phrasal construction and attachment proceed on the basis of two adjacent words in (5.1) and (5.2), though I pointed out that both construction and attachment are simultaneous for (5.1), whereas the attachment of YP to XP follows the actual construction of YP by one word in (5.2). (5.3) involves a significant delay in the assignment of the daughter YP to the mother XP following construction of the latter, while (5.4) involves a significant delay in the assignment of the mother XP to the daughter YP following construction of this latter. This is all summarized in (5.27):

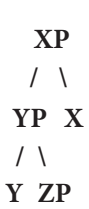
(5.27)

	MiD	MaOP
Structure 5.1	optimal	adjacent words for XP & YP construction & attachments
Structure 5.2	optimal	adjacent words for XP & YP construction & attachments
	(Mother XP assignment to YP delayed by one word)	
Structure 5.3	non-optimal	non-adjacent . . . Delayed Daughter YP assignment to XP
Structure 5.4	non-optimal	non-adjacent . . . Delayed Mother XP assignment to YP

5.8 Predictions for disharmonic word orders

5.8.1 Structure (5.4)

(5.4)



The inefficiency of structure (5.4) is that it involves the delayed assignment of a mother XP to a head-initial daughter YP—i.e., there is no mother to attach YP to for several words of online processing, which goes against both Minimize Domains (5.10) and Maximize Online Processing (5.26).

By the Performance–Grammar Correspondence Hypothesis (1.1) we expect that structures of type (5.4) will be limited as basic word orders in grammars in comparison to structure (5.2) (which keeps X final) and to structure (5.1) (which keeps Y initial). This is shown in (5.28) and (5.29) respectively, using data mainly from Dryer (1992) and counting 'genera' this time rather than languages (i.e., groups of genetically related languages at a time depth corresponding roughly to the subgroupings of Indo-European). The phrases corresponding to YP and XP are (i) an NP within a VP where the NP consists of a head noun plus Possessive Phrase (Dryer's 1992 Noun and Genitive orders, p. 91), (ii) a PP within a VP (Dryer 1992: 83), (iii) a VP within a TP (headed by a Tense or Aspect Auxiliary Verb with a VP sister; Dryer 1992: 100), and (iv) a CP within an NP (where CP is represented by a relative clause structure; cf. Lehmann 1984):

(5.28) *Limited productivity of (5.4) compared with (5.2) as basic orders* (keeping X final)

(i) $[[_{NP}$ N Possp$]$ V $_{VP}]$ vs $[[$Possp N $_{NP}]$ V $_{VP}]$ = 9.7% genera (12/124) Dryer (1992)
(ii) $[[_{PP}$ P NP$]$ V $_{VP}]$ vs $[[$NP P $_{PP}]$ V $_{VP}]$ = 6.1% genera (7/114) Dryer (1992)
(iii) $[[_{VP}$ V NP$]$ T $_{TP}]$ vs $[[$NP V $_{VP}]$ T $_{TP}]$ = 10% genera (4/40) Dryer (1992)
(iv) $[[_{CP}$ C S$]$ N $_{NP}]$ vs $[[$S C $_{CP}]$ N $_{NP}]$ = 0 Lehmann (1984)

(5.29) *Limited productivity of (5.4) compared with (5.1) as basic orders* (keeping Y initial)

(i) $[[_{NP}$ N Possp$]$ V $_{VP}]$ vs $[_{VP}$ V $[_{NP}$ N Possp$]]$ = 16% genera (12/75) Dryer (1992)
(ii) $[[_{PP}$ P NP$]$ V $_{VP}]$ vs $[_{VP}$ V $[_{PP}$ P NP$]]$ = 9.1% genera (7/77) Dryer (1992)
(iii) $[[_{VP}$ V NP$]$ T $_{TP}]$ vs $[_{TP}$ T $[_{VP}$ V NP$]]$ = 12.5% genera (4/32) Dryer (1992)
(iv) $[[_{CP}$ C S$]$ N $_{NP}]$ vs $[_{NP}$ N $[_{CP}$ C S$]]$ = 0 Lehmann (1984)

These figures clearly show that structure (5.4) is unproductive, compared with (5.1) and (5.2). It is not exceptionless for the combinations (i)–(iii), but it is for (iv) involving a CP within an NP. This absolute universal is not predicted under recent formulations of FOFC (Biberauer, Holmberg, and Roberts 2007, 2008), since mother and daughter phrases are of very different types (NP and CP), while structures of type [[$_{VP}$ V NP] T $_{TP}$] which it does exclude seem to be attested. Clearly these potentially offending language types in Dryer's sample need to be investigated more closely, in order to see whether they are genuine violations.

From a processing perspective we make a different set of predictions. First, the more structurally complex YP is, the more it should be dispreferred in (5.4). For example, a CP as YP should be worse than an NP or PP. This appears to be the case. This prediction can be made because domains for phrase structure processing are least minimal when YP is complex, offending Minimize Domains (5.10), and the more they delay the assignment of the mother XP to an already constructed YP, the more they offend Maximize Online Processing (5.26). Whether a more fine-tuned ranking and prediction can be made between the less complex YP categories, NP, PP, and VP based on their average weights and complexities remains to be investigated. This is less straightforward because there are cross-linguistic differences between PPs, for example, and between these and the other phrases that correspond to them semantically. They can involve single-word adpositions or affixes (see Hawkins 2008 and §7.8 below). There are also differences between languages with regard to the very existence of a VP or at least with regard to its productivity in the rules of the grammar. But certainly a sentence-like CP should be more complex on average than these other phrases, and it is the [[$_{CP}$ C S] N $_{NP}$] configuration that is unattested typologically (and not ruled out under the most recent definition of FOFC).

Second, we make predictions for 'rigid' versus 'non-rigid' OV languages. This distinction goes back to Greenberg (1963). Non-rigid OV languages are those with basic OV order that have certain complements and/or adjuncts of the verb positioned after it, i.e., to the right of V. Such languages are predicted here to be those that combine a Y-initial YP with an X-final XP, i.e., languages of type (5.4), and they are further predicted to postpose YP to the right of V, in proportion to the complexity of YP, creating alternations with structure (5.1). This can be seen in the obligatory extraposition rules of Persian, German, and other non-rigid OV languages that convert [[$_{CP}$ C S] V $_{VP}$] into [$_{VP}$ V [$_{CP}$ C S]] (Dryer 1980; Hawkins 1990), as in the following Persian example from Dryer (1980):

(5.30) a. *An zan [$_{CP}$ ke an mard sangi partab kard] mi danat (Persian)
the woman that the man rock threw CONT knows
'The woman knows that the man threw a rock'
b. An zan mi danat [$_{CP}$ ke an mard sangi partab kard]

Data from the *World Atlas of Language Structures* (*WALS*) confirm this prediction (see Dryer and Gensler 2005). 78 percent (7/9) of OV genera in *WALS* with prepositions (rather than postpositions) are non-rigid OV rather than rigid, i.e., these are potential combinations of type (5.4), and PPs regularly follow V in these languages converting (5.4) into (5.1); see (5.29ii) above (see Hawkins 2008). Similarly, 73 percent (8/11) of OV genera in *WALS* with [$_{NP}$ N Possp] (i.e., postnominal rather than prenominal genitives) are non-rigid OV rather than rigid, and NPs regularly follow V in these languages; see (5.29i) (Dryer and Gensler 2005; Hawkins 2008).

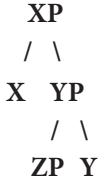
Rigid OV languages, by contrast, are those with basic OV in which V is final in VP and sisters precede. Such languages are predicted here to combine an X-final XP (i.e., OV) with a Y-final YP. And indeed 96 percent (47/49) of rigid OV genera in *WALS* have postpositions (rather than prepositions), i.e., [[NP P $_{PP}$] V $_{VP}$] (Dryer and Gensler 2005; Hawkins 2008). 94 percent (46/49) of rigid OV genera in *WALS* also have [[Possp N $_{NP}$] V $_{VP}$], i.e., prenominal rather than postnominal genitives (Dryer and Gensler 2005; Hawkins 2008). For the testing of more fine-tuned predictions for rigid and non-rigid OV languages, see §7.8.3.

Another solution available to non-rigid OV languages for relieving inefficiency involves keeping YP in situ in (5.4) but shortening it by extraposing items within it. This is very productive in German, for example, in which the initial head noun of an NP remains in situ within a verb-final VP while a relative clause CP is extraposed to the right of V, as shown in (5.31b):

(5.31) a. Ich habe [[$_{NP}$ den Lehrer [$_{CP}$ der das Buch geschrieben
I have the teacher who the book written
hat]] gesehen $_{VP}$]
has seen
'I have seen the teacher who wrote the book'
b. Ich habe [[$_{NP}$ den Lehrer] gesehen $_{VP}$] [$_{CP}$ der das Buch geschrieben hat]

Detailed predictions for when Extraposition from NP will apply, based on the competing efficiencies for NP and VP processing domains, are illustrated and supported in Hawkins (1994: 198–210, 2004: 142–6) using corpus data from Shannon (1992) and in Uszkoreit et al. (1998), and are summarized in §2.5.1 above.

5.8.2 Structure (5.3)

(5.3)


The inefficiency of structure (5.3) is that it involves the delayed assignment of a daughter YP to a constructed mother XP—i.e., no clear daughter can be assigned online to XP for several words of processing during which terminal material is processed that is contained within ZP. This goes against both Minimize Domains (5.10) and Maximize Online Processing (5.26).

As with (5.4), the PGCH (1.1) leads to the prediction that structures of type (5.3) will be limited as basic word orders in grammars in comparison to structure (5.1) (keeping X initial) and to (5.2) (keeping Y final). This is tested in (5.32) and (5.33) respectively using data mainly from Dryer (in terms of genera), but also data from Lehmann (1984) and Hawkins (1983). The phrases corresponding to YP and XP are (i) an NP within a VP where the NP consists of a head noun plus Possessive Phrase (Dryer's 1992 Noun and Genitive orders, p. 91), (ii) a PP within a VP (Dryer 1992: 83), (iii) a VP within a TP (headed by a Tense or Aspect Auxiliary Verb with a VP sister, Dryer 1992: 100), (iv) a CP within an NP (where CP is represented by a relative clause structure; cf. Lehmann 1984); and (v) a CP within a VP where CP is a sentence complement structure with a complementizer C (Hawkins 1990):

(5.32) *Limited productivity of (5.3) compared with (5.1) as basic orders* (keeping X initial)

(i)	$[_{VP}$ V [Possp N $_{NP}]]$ vs $[_{VP}$ V $[_{NP}$ N Possp]]	=	32% (30/93) genera	Dryer (1992)
(ii)	$[_{VP}$ V [NP P $_{PP}]]$ vs $[_{VP}$ V $[_{PP}$ P NP]]	=	14.6% (12/82) genera	Dryer (1992)
(iii)	$[_{TP}$ T [NP V $_{VP}]]$ vs $[_{TP}$ T $[_{VP}$ V NP]]	=	9.7% (3/31) genera	Dryer (1992)
(iv)	$[_{NP}$ N [S C $_{CP}]]$ vs $[_{NP}$ N $[_{CP}$ C S]]	=	v. few, if any	Lehmann (1984)
(v)	$[_{VP}$ V [S C $_{CP}]]$ vs $[_{VP}$ V $[_{CP}$ C S]]	=	0	Hawkins (1990)

(5.33) *Limited productivity of (5.3) compared with (5.2) as basic orders* (keeping Y final)

(i)	$[_{VP}$ V [Possp N $_{NP}]]$ vs [[Possp N $_{NP}]$ V $_{VP}]$	=	21.1% (30/142) genera	Dryer (1992)
(ii)	$[_{VP}$ V [NP P $_{PP}]]$ vs [[NP P $_{PP}]$ V $_{VP}]$	=	10.1% (12/119) genera	Dryer (1992)
(iii)	$[_{TP}$ T [NP V $_{VP}]]$ vs [[NP V $_{VP}]$ T $_{TP}]$	=	7.7% (3/39) genera	Dryer (1992)
(iv)	$[_{NP}$ N [S C $_{CP}]$ vs [[S C $_{CP}]$ N $_{NP}]$	=	v. few, if any	Lehmann (1984)
(v)	$[_{VP}$ V [S C $_{CP}]]$ vs [[S C $_{CP}]$ V $_{VP}]$	=	0	Hawkins (1990)

These figures clearly show that (5.3) is unproductive compared with (5.1) and (5.2). The dispreference figures are not that different from those for structure (5.4) in (5.28) and (5.29), which violate FOFC, supporting the point made in §5.6 that the correct partitioning seems to be between structures (5.1) and (5.2) (productive) on the one hand and (5.3) and (5.4) (unproductive) on the other, and that FOFC seems to be both too weak and too strong. Notice that the combination [$_{VP}$ V [S C $_{CP}$]] in (5.32v) and (5.33v) appears to be unattested, making its absence a possible absolute universal, yet FOFC does not apply to this structure. Structures of type [$_{NP}$ N [S C $_{CP}$]] in (iv) may also be unattested, depending on what counts as the category C across languages.

The general prediction that we make for (5.3) from a processing perspective is that the more complex the (center-embedded) ZP is, the more it will be dispreferred. For example, a center-embedded S in (iv) and (v) in (5.32)–(5.33) is worse than an NP or PossP in (i)–(iii). The typological frequency data support this prediction.

One way in which structures of type (5.3) can be made more efficient is by positioning those items within ZP early that can construct YP by the parsing principle of Grandmother Node Construction (Hawkins 1994: 361). This states that 'if any word of syntactic category C uniquely determines a grandmother node G directly dominating its mother node M . . . then G is immediately constructed over M.' If a word within ZP can construct YP in advance of category Y then the delay in recognizing YP can be avoided, in accordance with Minimize Domains (5.10), and YP can be attached to XP early without having to wait for Y, making the processing domain for XP and its immediate constituents minimal.

This may be at least part of the explanation for why non-nominative case-marked pronouns are preposed in the German VP and for why case-marked full NPs precede PPs and other non-case-marked categories in this language. Non-nominative case marking within an NP can construct a VP by Grandmother Node Construction in structures such as [$_{TP}$ T [NP . . . V $_{VP}$]]; see (5.34) (Hawkins 1994: 393–402):

(5.34) Ich [$_{TP}$ habe [ihn [noch einmal] gesehen $_{VP}$] (German)
I have him (+ACC) once again seen
'I have seen him once again'

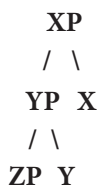
A major source of inefficiency in structure (5.3) involves the potential for online structural misassignments or garden paths (Hawkins 2004), whereby what will eventually be parsed as ZP or a phrase within ZP is initially parsed as YP, the sister of X. Misassignments are inefficient for Maximize Online Processing (5.26). This may explain the non-occurrence of [$_{VP}$ V [S C $_{CP}$]] structures in (v) of (5.32)–(5.33), in addition to its inefficiency by MiD, since

different phrases within S could be attached immediately to VP unless C precedes and constructs CP at the outset, marking a clear clausal boundary for CP-dominated material. When complementizers are optionally deleted in English there are garden paths of this sort in structures like *I believe the clever student wrote*..., which is disambiguated only at *wrote*. There is clear performance evidence in English corpora showing that complementizers are not deleted when the misanalysis would persist over more than a few words, and they are not deleted even when there would be no garden path (e.g., when *realize* replaces *believe* in the example just given) if the 'unassignment' of CP persists for more than a couple of words (see Hawkins 2004: 49–61 for discussion and relevant data).

This potential for misassignments and unassignments may contribute to the explanation for the relative infrequency of [$_{VP}$ V [NP P $_{PP}$]], i.e., (ii) in (5.32)–(5.33), and for [$_{VP}$ V [Possp N $_{NP}$]], i.e., (i). It would be advantageous for grammars to distinguish NP arguments of V from embedded NPs in [NP P $_{PP}$] or [Possp N $_{NP}$] structures in such cases, either through case marking for an NP sister to V and not for an NP within a PP or within another NP, or through different case marking. In the event that the grammar permits a genuine garden path here, as in English *I* [$_{VP}$ *met* [$_{POSSP}$ [*the king's*] *daughter* $_{NP}$]], in which *the king* can first be assigned as a direct object to *met*, we expect certain consequences, such as a limitation on the length of these prenominal genitives in performance. Biber et al. (1999: 597) point out that 70–80 percent of premodified noun phrases in English are limited to single-word premodifiers, including prenominal genitives. The cliticization of the genitive case marking is also interesting since it marks the genitive NP within the genitive phrase itself and without a word boundary, thus distinguishing it more rapidly from other NP types.

5.8.3 Structure (5.2) (head finality)

(5.2)



Structure (5.2), like structure (5.1), is optimal for MiD (5.10) since heads are consistently adjacent. There is one small respect, however, in which it is not optimal for MaOP (5.26). YP is constructed at Y and the parser must then wait one word until X has constructed XP for the attachment of YP to this

latter—i.e., for one word of online processing there is no mother to attach YP to. In the mirror-image (5.1), by contrast, XP is constructed first and YP can then be constructed by Y and attached immediately to XP with no processing delay.

There are some typological patterns in consistently head-final SOV languages that are as yet unexplained grammatically, as far as I am aware, but which suggest that this small processing delay between the construction and attachment of YP does have grammatical consequences. I shall consider two such briefly and then draw attention to a more general characteristic of left-branching structures like (5.2) that is also relevant here.

First there are fewer free-standing X words that follow Y in these languages. When X precedes YP—for example, when a preposition precedes NP as in English—the preposition is typically a free-standing word. But when P follows NP as a postposition there are many fewer free-standing postpositions. More generally, when X follows Y there are many more X affixes on Y, e.g., derivational and inflectional suffixes on nouns and on verbs, and these X affixes can construct YP and XP simultaneously at Y, the former through Mother Node Construction (5.8), the latter through Grandmother Node Construction (Hawkins 1994: 361–6). This is an efficient solution to the one-word processing delay since construction and attachment of YP can now take place at one and the same word, Y, just as they do in the head-initial structure (5.1), making them equivalent in processing efficiency.

Postpositions are not as productive in head-final languages as prepositions are in head-initial. Many languages with strong head-final characteristics have no free-standing postpositions at all, but only suffixes with adposition-type meanings and a larger class of NPs bearing rich case features. Tsunoda, Ueda, and Itoh (1995: 757) attempt to quantify the number of such head-final languages with suffixes and without actual postpositions and in their sample this number is almost 30 percent (19/66). More generally they point out that head-final languages have fewer free-standing postpositions than head-initial languages have free-standing prepositions, since postpositions are more likely to be lost in language change and to develop into affixes. They cite Bybee et al. (1990: 3) who observe that 'postposed grammatical material is more likely to affix than preposed grammatical material.' Hall (1992) gives a detailed processing-motivated account for how this affixing asymmetry comes about diachronically. He argues that morphological integration of a free-standing word into a suffix of the preceding word, often with phonological reduction, can have processing efficiency advantages that are not matched for prefixes.

Consider also sentence complementizers in this context. In head-initial languages they are typically free-standing words that construct subordinate clauses. In head-final languages these correspond more commonly to participial and other subordinate clause indicators affixed to verbs and they are

much less productive as independent words; see Dryer (2009) and §7.6 for details.

Second, consistently head-final SOV languages appear to avoid the grammaticalization of additional constructors of phrasal nodes by categories other than the syntactic head. In the kind of classical syntax model assumed in Hawkins (1994, 2004), which is close in spirit to the Simpler Syntax of Culicover and Jackendoff (2005), trees are flatter than in other models and there can be more than one daughter category that constructs a given phrase (in accordance with Mother Node Construction (5.8)), i.e., the set of constructing categories for phrases is not just limited to syntactic heads (see especially Hawkins 1993, 1994: ch. 6). If we adopt such a perspective, an interesting difference emerges between VO and rigid OV languages. Assume (controversially given the DP theory: cf. Abney 1987; Payne 1993) that definite articles construct NP, just like N or Pro and other categories uniquely dominated by NP (see Hawkins 2004: 82–93 and §6.2.3). If so then either N or Art can construct NP immediately on its left periphery and provide efficient and minimal Phrasal Combination Domains (PCDs) in VO languages (see §3.1 and §3.5.3).

(5.35) $[_{VP}$ V $[_{NP}$ N . . . Art . . .]
$[_{VP}$ V $[_{NP}$ Art . . . N . . .]
|- - - -|

In OV languages on the other hand any additional constructor of NP will lengthen these processing domains, whether it follows or precedes N, by constructing the NP early and extending the processing time from the construction of NP to the processing of V. Additional constructors of NP are therefore inefficient in OV orders, as shown in (5.36):

(5.36) [[. . . N . . . Art $_{NP}$] V $_{VP}$]
[[. . . Art . . . N $_{NP}$] V $_{VP}$]
|- - - - - - - - |

It is significant, therefore, that definite articles, which typically emerge historically out of demonstrative determiners (C. Lyons 1999), are found predominantly in VO rather than OV languages. The data of (5.37) are taken from *WALS* (Dryer 2005c) and compare language quantities for VO and (rigid) OV languages in which there is a separate definite article word distinct from a demonstrative determiner (see §6.2.3 for a more general formulation of this prediction in terms of NP-constructing categories going beyond definite articles, including e.g., other articles and particles within NP):

(5.37) [*WALS* data]

	Def word distinct from Dem	*No definite article*
Rigid OV	19% (6)	81% (26)
VO	58% (62)	42% (44)

This same consideration provides a further motivation for the absence of free-standing complementizers in head-final languages (cf. §7.6). Complementizers can shorten PCDs when they precede V in VO languages, by constructing subordinate clauses on their left peripheries (*John knows* [that *he is sick*]), but they will lengthen PCDs in OV languages, compared with projections from V alone, whether they are clause-initial or clause-final. This all fits in readily with MiD (5.10) and MaOP (5.26).

Finally in this section let me make an observation about a typological asymmetry for which there appears to be no clear grammatical explanation but which may be explainable in terms of MaOP (5.26). Left-branching phrases like YP and ZP in (5.2), i.e., with final heads, are often more reduced and constrained in comparison with their right-branching and head-initial counterparts in (5.1). For example, Lehmann (1984: 168–73) observes that prenominal relative clauses are significantly more restricted in their syntax and semantics than postnominal relatives. They often involve greater nominalization (i.e., more non-sentential properties); restrictions on tense, aspect, and modal forms; more non-finite rather than finite verbs; less syntactic embedding possibilities; the conversion of an underlying subject to a genitive; and less tolerance of appositive interpretations. The effect of these limitations is to make a left-branching relative clause recognizably different from a main clause, thereby signaling its subordinate status early and avoiding a structural misassignment or garden path in online processing, in accordance with MaOP's preference (Hawkins 2004: 205–10).

5.8.4 Conclusions on word order disharmony

A number of fine-tuned predictions for disharmonic words orders have been made by the processing approach advocated here, involving their relative frequencies and structural characteristics, and they appear to be supported. Some predictions have also been made for consistently head-final languages (in §5.8.3). These languages appear to be interestingly different from head-initial languages in the relevant respects (a topic that will be pursued further in Chapter 7). Typological patterns support the general efficiency principles of Minimize Domains (5.10) and Maximize Online Processing (5.26), and particular parsing principles such as Mother Node Construction (5.8) and Immediate Constituent Attachment (5.9).

Purely grammatical principles such as FOFC, on the other hand, do not have the advantage of independent support from performance (see, e.g., the data summarized in §§5.2–5.3) and they are stipulated. The typological patterns presented in this chapter suggest that FOFC (5.25) is not capturing the right generalization: it is too strong (structure (5.4) is generally dispreferred, occasionally unattested), and too weak (structure (5.3) is also dispreferred, occasionally unattested). There are also apparent exceptions in typological samples such as Dryer (1992) to the more recent formulations of FOFC in Biberauer, Holmberg, and Roberts (2007, 2008)—see (5.28iii) and (5.29iii) in §5.8.1—and these need to be investigated further. Above all, there appears to be a broader and more significant generalization for all the patterns we have seen, in terms of processing. The FOFC-violating structures appear to be just a subset of the hard-to-process ordering possibilities for heads, which suggests that they should not be ring-fenced and given a separate explanation in terms of an innate UG.

Typologists, on the other hand, can benefit from the more precise and in-depth descriptions of relevant languages that formal syntax can provide, in order to determine what exactly the cross-linguistic regularities are, how best to formulate them, and what the relevant syntactic categories are that figure in the rules and principles for each language. Apparent exceptions to FOFC may not be genuine counterexamples, depending on the best syntactic analysis.

6

The typology of noun phrase structure

In this chapter I examine one phrasal type in greater detail, the noun phrase, and consider cross-linguistic variation within it from the perspective of efficiency and online processing. I will argue that new descriptive generalizations can be formulated in this way, and new explanations given for why languages exhibit the variants they do.

Two processing axioms will be proposed. First, anything that is an NP must be recognizable as such, i.e., every NP must be **constructible**. This is a basic prerequisite for all successful parsing, as defined by the Axiom of Constructibility in Hawkins (1994: 379), which states that there will always be at least one word of category C dominated by a phrasal node P that can construct P on each occasion of use; see (6.3) in §6.2. Second all the items that belong to a given NP must be **attachable** to it rather than to some other phrase, as defined in the Axiom of Attachability in (6.11). A mechanism for bringing this about is the parsing principle of Immediate Constituent Attachment given in Hawkins (1994: 62): "In the left-to-right parsing of a sentence, if an IC does not construct, but can be attached to, a given mother node M, in accordance with the PS rules of the grammar, then attach it as rapidly as possible." In addition, the amount of syntactic, morphosyntactic, or lexical encoding needed to signal this attachment will be hypothesized here to be in proportion to the processing difficulty of making the attachment; see §6.3. This follows from Minimize Forms (2.8); see especially (2.9a) in §2.2.

Selected predictions following from these axioms and hypotheses derived from them will be defined, tested, and found to be supported in §§6.2–6.3, suggesting that processing has played a significant role in shaping grammars in this area. I begin with an overview of cross-linguistic variation patterns in the noun phrase (§6.1).

6.1 Cross-linguistic variation in NP syntax

There is significant variation in the syntactic and morphosyntactic devices that define the noun phrase across languages (see, e.g., Rijkhoff 2002; Plank 2003). In addition to nouns and pronouns, which are almost universal as separate categories, some languages have definite and/or indefinite articles, some have classifiers, some make extensive use of nominalizing particles, case marking is found in some, case copying throughout the noun phrase in a subset of these, agreement patterns can be found on certain modifiers, linkers exist in some languages for NP-internal constituents, a construct state attaches NP to a sister category in others, and so on. The positioning of items within the NP also exhibits variation.

We can understand all this variation better if we look at grammars from an efficiency and processing point of view. Predictions can be made for the existence of different structural devices, for their possible omission and obligatory presence, and for their positioning within the NP, on the basis of general principles supported by experimental and corpus findings from language performance.

Noun phrases pose two challenges for grammatical description and for parsing. First, NPs do not always contain nouns (Ns), i.e., the head category that projects to a mother NP and that makes it recognizable and different from other phrases (cf. Jackendoff 1977; Pollard and Sag 1994; Dryer 2004). An NP must therefore be constructible from a variety of other terminal categories that are dominated by NP, the precise nature of which can vary across languages. Second, it must be made clear in grammars and in performance which terminal categories are to be attached to a given NP, as opposed to some other NP or to other phrases.

I begin with a brief listing of some of the major syntactic and morphosyntactic devices that are found in NPs across languages and that are relevant to any discussion of construction and attachment. Several categories can be argued to construct NP, summarized in (6.1):

(6.1) Nouns (i.e., lexical items specialized for the category N) like *student* and *professor* in English.
Pronouns (personal, demonstrative, interrogative, etc.): *he*/*she*, *this*/*that*, *who*, and their counterparts in other languages; cf. Bhat (2004).
Various determiners including the definite article (in theories in which Determiner Phrase and NP are not distinguished; cf. Hawkins 1993, 1994; Payne 1993).

Nominalizing particles like Lahu *ve* (Matisoff 1972), Mandarin *de* (Li and Thompson 1981), and Cantonese *ge* (Matthews and Yip 1994: 113) can combine with a non-noun or pronoun to construct a mother NP, as in the examples of (i); see Lehmann (1984: 61–6).

(i) a. [$_{NP}$ chu ve] (Lahu)
fat NOMINALIZER
'one that/who is fat'

b. [$_{NP}$ [$_{VP}$ chī hūn] de] (Mandarin)
eat meat NOMINALIZER
'one who eats meat'

c. [$_{NP}$ [$_{VP}$ heui hōi-wúi] ge] (Cantonese)
go have-meeting NOMINALIZER
'those who are going to the meeting'

Classifiers in many languages perform syntactic functions that include the construction of NP (Aikhenvald 2003: 87–90), resulting in omission of nouns from NP and pronoun-like uses for classifiers, as in the following example from Jacaltec (Craig 1977: 149).

(ii) xal naj pel chubil chuluj naj hecal (Jacaltec)
Said CL Peter that will-come CL/he tomorrow
'Peter said that he will come tomorrow'

Case particles or suffixes construct a case-labeled mother or grandmother NP respectively; cf. Hawkins (1994: ch. 6) for detailed discussion, e.g., in Japanese, German, Russian.

(iii) a. [$_{NPACC}$ tegami o] (Japanese)
letter ACC

b. [$_{NPACC}$ den Tisch] (German)
the-ACC-SG-MASC table

c. [$_{NPACC}$ lip-u] (Russian)
lime tree-ACC-SG-II

Note that in the kind of Simpler Syntax model assumed here (Culicover and Jackendoff 2005) the dominating mother phrase in all of these examples will be an NP. For theories in which more complex phrasal projections from functional heads are assumed—e.g., a Determiner Phrase over an NP—the principles proposed here can be readily extended and can be viewed as providing a processing perspective on both NP and DP structure (I am grateful to Sten Vikner for discussion and confirmation of this point). A number of details will differ from the account proposed here, regarding

which of these maximal projections is actually constructed by particular daughters and regarding the attachments to each, but the same processing logic can in principle carry over to structural analyses incorporating DPs.

Various (morpho)syntactic devices signal the attachment of sister categories to a given NP, as summarized in (6.2):

(6.2) Adjective agreement as in Latin, in which adjectives agree in case, number, and gender features with some [$_{NP}$ N] (see Vincent 1987) permitting separation of noun phrase constituents, as in (ib).

(i) a. [$_{NP}$ illarum bonarum feminarum] (Latin)
that-GEN-PL-FEM good-GEN-PL-FEM woman-GEN-PL-FEM
'of those good women'

b. [$_{PP}$ [$_{NP_i}$ magno] cum[$_{NP_i}$ periculo]]
great-ABL-SG-NEUT with danger-ABL-SG-NEUT
'with great danger'

Case copying in 'word-marking' Australian languages like Kalkatungu (Blake 1987; Plank 1995) also signals attachment (to a similarly case-marked [$_{NP}$ N]), permitting separation of NP constituents as in (iib) from Blake (1987: 87):

(ii) a. [$_{NP_i}$ tyipa-yi thuku-yu yaun-tu] [$_{NP_j}$ yanyi]
this-ERG dog-ERG big-ERG white-man
itya-mi (Kalkatungu)
bite-FUT
'This big dog will bite the white man'

b. [$_{NP_i}$ tyipa-yi thuku-yu][$_{NP_j}$ yanyi] itya-mi [$_{NP_i}$ yaun-tu]
this-ERG dog-ERG white-man bite-FUT big-ERG

These case suffixes also construct a case-marked mother or grandmother NP, as in (6.1iii)—i.e., case markers can serve both to construct the dominating (case-labeled) NP and to attach the respective daughters with the same case to it.

Mandarin *de* similarly performs an attachment function in addition to the construction function illustrated in (6.1ic) above, attaching NP-dominated constituents together that stand to its left and right within the NP that it constructs. See the discussion of *de* in Lehmann (1984: 63–6) from which the following examples with their tonal diacritics are taken:

(iii) a. [shuìjiòu de rén $_{NP}$] (Mandarin)
Sleep NOMLZ/ATTACH person
'sleeping person'

b. [[bù hǎo] de lái-wǎng $_{NP}$]
not good NOMLZ/ATTACH come-go
'undesirable contact'

c. [[$_{S}$ wǒ lái] de dìfáng $_{NP}$]
I come NOMLZ/ATTACH place
'place from which I am coming'

d. [[$_{S}$ wǒ [$_{VP}$ jiǎn zh]] de jiǎndao $_{NP}$]
I cut paper NOMLZ/ATTACH scissors
'scissors with which I cut paper'

Classifiers also attach NP sisters to the NP that they construct, as in the following examples from Cantonese in which the classifier attaches a possessor to its head noun (iva) and a (preposed) relative clause to its head noun (ivb); see Matthews and Yip (1994: 107–12).

(iv) a. lóuhbáan ga chē (Cantonese)
boss CL car
'the boss's car'

b. ngóhdeih hái Faatgwok sihk dī yéh
we in France eat CL food
'the food we ate in France'

The repeated classifier *-ma* in the following example from Tariana functions like agreement in Latin (6.2i) and case-copying in Kalkatungu (6.2ii) to signal co-constituency between adjective and noun within NP (Aikhenvald 2003: 94–5): *nu-kapi-da-ma hanu-ma* (1SG-hand-CL:ROUND-CL:SIDE big-CL:SIDE), 'the big side of my finger.' Linkers such as *na* in Tagalog attach *ulól* ('foolish') and *unggó* ('monkey') into a single NP in *ulól na unggó* ('foolish monkey'); see Hengeveld et al. (2004: 553).

The construct state in Berber signals co-constituency between nouns or NPs in the construct state and a preceding noun (vb), quantity word (vc), preposition (vd), intransitive verb (ve), and transitive verb (vf) (Keenan 1988: 119–25):

(v) a. *Free form*: aryaz 'man' arba 'boy' tarbatt 'girl' (Berber)
Construct form: uryaz urba terbatt

b. [$_{NP}$ axam [$_{NP}$ uryaz]]
tent manCONSTRUCT
'tent of the man/the man's tent'

c. [$_{NP}$ yun uryaz]
One manCONSTRUCT
'one man'

d. [$_{PP}$ tama (n) [$_{NP}$ uryaz]]
near manCONSTRUCT
'near the man'

e. [$_{S}$ lla [$_{VP}$ t-alla [$_{NP}$ terbatt]]]
IMPF she-cry girlCONSTRUCT
'The girl is crying'

f. [$_{S}$ [$_{VP}$ i-annay [$_{NP}$ urba] [$_{NP}$tarbatt]]]
he-saw boyCONSTRUCT girl
'The boy saw the girl'

The construct state signals attachment of these immediate constituents but does not unambiguously construct any particular mother or grandmother phrase. The mother most immediately dominating [$_{NP}$ N] in the construct state can be NP, PP, or VP, etc.

A possessive (genitive) *-s* in English (and similar forms in other languages) signals the attachment of PossP to the head N, and also the construction of a grandmother (or mother) NP (NP*i* in (vi)):

(vi) [$_{NP}$*i* [$_{POSSP}$ [$_{NP}$*j* the king of England]-s] daughter]

6.2 Constructibility hypotheses

The Axiom of Constructibility was defined in Hawkins (1994: 379) as follows:

(6.3) *The Axiom of Constructibility*
For each phrasal node P there will be at least one word of category C dominated by P that can construct P on each occasion of use.

It appears that there is always some category C that enables the parser to recognize that C is dominated by a phrase of a particular type, NP, PP, or VP, etc., generally as a daughter or as a granddaughter. Building hierarchical phrase structure trees in syntactic representations on the basis of terminal elements is a key part of grammatical processing. If a given P cannot be properly constructed in parsing, its integration into the syntactic tree and its semantic interpretation are at risk of failing. More generally, I have argued (Hawkins 1993, 1994: ch. 6) that (6.3) motivates a lot of the grammatical

properties that formal syntacticians propose for heads of phrases, both lexical and functional, and that it provides a processing explanation for these properties.

There are numerous differences between different formal models of grammar with respect to the precise set of heads they define, and numerous disagreements exist with respect to particular categories; cf. Dryer (1992) and Corbett et al. (1993) for detailed summaries and discussion. I argued in Hawkins (1993, 1994: ch. 6) that the disputed categories generally have a construction function in parsing (whence the plausability of considering them heads at all), and that it is this that ultimately motivates the whole notion of head of phrase and its correlating properties.

6.2.1 NP construction

The Axiom of Constructibility leads to a prediction for the structure of NPs:

(6.4) *Prediction 1: NP Construction*
Any phrase that is of type NP must contain either (i) a lexical head N or pronoun (personal or demonstrative, etc.) or proper name, or (ii) some other functional category that can construct NP on each occasion of use in the absence of N or Pro or Name.

I expect NPs to contain either some lexical and inherent head category like a noun or pronoun or name, on the basis of which NP can always be recognized; alternatively (6.4) predicts that categories will be found in the NP that project uniquely to it and these categories will be especially productive, and indeed obligatory, in the absence of nouns, pronouns, and names. Examples of these functional categories unique to NP are given in (6.5):

(6.5) a. Lahu, Mandarin, and Cantonese nominalizers, as in (6.1ii).
b. Jacaltec classifiers, as in (6.1iii).
c. Certain non-nominal categories including numerals and adjectives may unambiguously construct NP in certain languages (Dryer 2004).
d. Spanish permits omission of nouns with certain restrictive adjectives plus the definite article as a constructor of NP (*lo difícil* 'the difficult thing') and has also expanded this option to other categories such as infinitival VPs in *el hacer esto fue fácil* (DEF to-do this was easy) 'doing this was easy' (Lyons 1999: 60; Dryer 2004).
e. Malagasy has expanded it to locative adverbs, as in *ny eto* (DEF here), meaning 'the one(s) who is/are here' (Anderson and Keenan 1985: 294).

f. Case marking on adjectives in, e.g., Latin and German permits them to function as referential NPs, Latin *bonī* (good-Nom-Masc-Pl) 'the good ones,' German *Gutes* (good-Nom-Neut-Sg) 'good stuff.'
g. In numerous languages the definite article signals a nominalization of some kind, Lakhota *ktepi kį wąyake* (kill DEF he-saw) 'he saw the killing' (Lyons 1999: 60), or the construction of a subordinate clause in noun phrase position, e.g., as subject or object, in Huixtan Tzotzil and Quileute (Lyons 1999: 60–1).
h. Head-internal relatives are structurally clauses that function as NPs and they are regularly marked as such by definiteness markers and/or case particles and adpositions, as in Diegueño (Gorbet 1976; Basilico 1996).
i. Free relatives can also consist of a clause functioning as an NP that is constructed by a nominalizing particle, e.g., in Cantonese *léih mh ngoi ge* (you not want Nominalizer) 'what you don't want' (Matthews and Yip 1994: 113).

The values of category C constructing NP can vary in these $_{NP}${C, X} structures, as can the values of X. There are language-particular conventions for the precise set of constructing categories (nominalizing particles, classifiers, definite articles, etc.) and for the different values of X (adjective, adverb, infinitival VP, S, etc.) that can combine with the relevant C to yield a noun phrase. But the very possibility and cross-linguistic productivity of omitting the noun/pronoun/name and of still having the phrase recognized as NP, in so-called 'nominalizations' and in the other structures illustrated here containing functional categories unique to NP that enable it to be constructed in language use, follows from the Axiom of Constructibility (6.3).

6.2.2 Lexical differentiation for parts of speech

A further prediction made by (6.4) is relevant for those languages whose lexical items are highly ambiguous with respect to syntactic category, even for major parts of speech like noun and verb. The Polynesian languages are often discussed in this context (see, e.g., Broschart 1997; Hengeveld et al. 2004). English has a large number of words that are ambiguous between noun and verb and there are many minimal pairs such as *they want to run/they want the run* and *to play is fun/the play is fun*. The article constructs NP and disambiguates between N and V here.

Languages without a unique class of nouns do not have lexical categories that can unambiguously construct NP on each occasion of use. If lexical predicates are vague as to syntactic category, then projection to NP is not

guaranteed by lexical entries and the Axiom of Constructibility is not satisfied. We accordingly expect the following:

(6.6) *Prediction 2: Lexical Differentiation*
Languages in which nouns are differentiated in the lexicon from other categories (verbs, adjectives, or adverbs) can construct NP from nouns alone. Languages without a unique class of nouns in the lexicon will make use of constructing particles in order to construct NP and disambiguate the head noun from other categories; such particles are not required (though they are not ruled out) in languages with lexically differentiated nouns.

Relevant data come from the Polynesian languages, which make extensive and obligatory use of NP-constructing particles such as 'definite' articles, extending their meanings into the arena of indefiniteness, see Lyons (1999: 57–60). Samoan *le*, Maori *te*, and Tongan *e* appear to be best analyzed as general NP constructors: they convert vague or ambiguous predicates into nouns within the NP constructed. Other (tense and aspect) particles construct a clause (IP) or VP and convert ambiguous lexical predicates into verbs (Broschart 1997). We have here a plausible motivation for the expanded grammaticalization of definite articles and of other particles in these languages (see §4.2 and Hawkins 2004: 82–92).

One way to test this link between NP-constructing particles and lexical differentiation would be to compare languages with and without lexically unique nouns by selecting various subsets of lexical predicates, quantifying numbers of category-ambiguous items (i.e., predicates like *run* and *play* in English, as opposed to *student* and *professor*, which are uniquely nouns), numbers of syntactic environments that require the definite article or other NP constructor, and corpus frequencies for these constructors. Hengeveld et al. (2004) provide a useful typology for lexical differentiation across languages and a language sample.

6.2.3 Head-initial and head-final asymmetries

Head-initial languages with VO in the verb phrase have predominantly head-initial orders in other phrases as well that permit early construction of these phrases in parsing, by projection from the respective heads (V projects to VP, N to NP, P to PP, etc.). OV languages have predominantly head-final phrases that favor late construction. Consistent head ordering minimizes processing domains for phrase structure recognition by shortening the distances between heads and this provides an explanation for the productivity of the two major language types, head-initial and head-final (see §5.1 and §7.1). There is,

however, an interesting asymmetry between them that can be seen in so-called non-lexical or functional head categories. The NP reveals this asymmetry clearly. See §5.8.3 for additional discussion.

A non-lexical category C within an NP that can construct it, in addition to an N, can be efficient in VO languages. Either $[_{NP}\ N \dots]$ or $[_{NP}\ C \dots]$ orders can construct NP immediately on its left periphery and can provide minimal phrasal combination domains (5.12) and lexical domains (5.15) linking, e.g., V and NP within a VP:

(6.7) $[_{VP}\ V\ [_{NP}\ N \dots]$
$[_{VP}\ V\ [_{NP}\ C \dots N \dots]$
|-------|

These structures are efficient by Minimize Domains (2.1), therefore, and I expect additional constructing categories C to be productive in VO and head-initial languages and to be especially favored when N itself is not initial in NP, e.g., in $[_{NP}$ C AdjP N]. The determiner and adjective positions of English exemplify this, with left-peripheral articles constructing NP in advance of N. Additional constructing categories in OV and head-final languages, on the other hand, do not have comparable benefits. They lengthen phrasal combination domains and other processing domains linking NP to V when NP precedes, whether the additional constructor precedes or follows N:

(6.8) $[[\dots N \dots C\ _{NP}]\ V\ _{VP}]$
$[[\dots C \dots N\ _{NP}]\ V\ _{VP}]$
|------------|

Additional constructors of NP can be inefficient in OV orders, therefore, and at variance with MiD and are predicted to be significantly less productive than their head-initial counterparts as a consequence.

(6.9) *Prediction 3: VO versus OV asymmetries*
Constructors of NP other than N, Pro, and Name, such as articles, are efficient for NP construction in VO languages and should occur frequently; they are not efficient for this purpose in OV languages and should occur less frequently.

There are many interesting tests that can be made of this prediction. One was discussed briefly in §5.8.3 and involved definite article data from *WALS* (see Dryer's 2005a,c chapters). *WALS* gives data on languages that have definite articles as a separate category from demonstrative determiners (from which definite articles have generally evolved historically; see Himmelmann 1997; C. Lyons 1999). If, as argued in Hawkins (2004: 82–93), it is processing

efficiency that drives the grammaticalization of definite articles out of demonstratives, then we expect to see a skewing in the distribution of definite articles in favor of head-initial languages. The figures in (6.10) show that VO languages do indeed have significantly more definite articles than OV languages. We also expect that non-rigid OVX languages should have more definite articles than OV languages with rigid verb-final order, since OVX languages have more head-initial phrases in their grammars, including head-initial NPs (Hawkins 1983), in which early construction of NP can be an advantage. This prediction is also borne out. The figures in parentheses refer to Dryer's 'genera' as do the percentages:

(6.10)

	Def word distinct from Dem	*No definite article*
Rigid OV	19% (6)	81% (26)
VO	58% (62)	42% (44)
Non-rigid OVX	54% (7)	46% (6)

6.3 Attachability hypotheses

Corresponding to the Axiom of Constructibility (6.3) I propose (6.11):

(6.11) *The Axiom of Attachability*
For each phrasal node P, all daughter categories {A, B, C, . . . } must be attachable to P on each occasion of use.

Moreover, by the principle of Minimize Forms (2.8) I expect that the amount of syntactic, morphosyntactic, or lexical encoding that is needed to signal attachability will be in proportion to the processing difficulty of making the attachment.

In other words, all daughters must be attachable (see Hawkins 1994: 61–4), and the more difficult the attachment is, the more grammatical or lexical information is required to bring it about. The use of explicit attachment devices under conditions of difficulty, and their possible omission when processing is easy, is efficient: activation of processing resources and greater effort are reserved for conditions under which they are most useful. This is supported by a wide range of grammatical and performance data that motivate Minimize Forms (2.8): form minimizations apply in proportion to the ease with which a given property P can be assigned in processing to a given form F. Rohdenburg's (1996, 1999) complexity principle provides further supporting data for this from English corpora: "In the case of more or less explicit grammatical options, the more explicit one(s) will be preferred in cognitively more complex environments" (Rohdenburg 1999: 101).

For attachments to NP, (6.11) and Minimize Forms (2.8) lead to the hypothesis in (6.12):

(6.12) *NP Attachment Hypothesis*
Any daughters {A, B, C, . . . } of NP must be attachable to it on each occasion of use through syntactic, morphosyntactic, or lexical encoding on one or more daughters, whose explicitness and differentiation are in proportion to the processing difficulty of making the attachment.

6.3.1 Separation of NP sisters

One clear factor that increases the difficulty of attaching constituents together as sisters is separation from one another.

(6.13) *Prediction 4: Separation of Sisters*
Morphosyntactic encoding of NP attachment will be in proportion to the degree of separation between sisters: the more distance, the more encoding.

Consider first some performance data from English involving relative clauses with explicit relativizers (*who*, *whom*, *which*, and *that*) versus zero. The relativizers construct a relative clause (recall §2.2.2). Their presence can also help to attach the relative to the head, especially when there is animacy agreement between relativizer and head noun (*the professor who* . . . , etc.), but also in the absence of such agreement (since relatives are known to attach to head nouns by phrase structure rules). Empirically, it turns out that the presence of the relativizer and the avoidance of zero is proportional to the distance between the relative clause and the head noun (among other factors relevant to this distinction, see the discussion of the predictability or otherwise of the relative clause in relation to relativizer omission in §2.2.2). The figures in (6.14) are taken from Quirk's (1957) corpus of spoken British English. They show that the use of explicit relativizers increases significantly, from 60 percent to 94 percent, when there is any separation between nominal head and relative.

(6.14) a. *Restrictive (non-subject) relatives adjacent to the head noun*
explicit relativizer = 60% (327) zero = 40% (222)

b. *Restrictive (non-subject) relatives separated from the head noun*
explicit relativizer = 94% (58) zero = 6% (4)

The figures in (6.15) measure the impact of larger versus smaller separations on relativizer retention and are taken from the Brown corpus (by Barbara Lohse, now Barbara Jansing, in Lohse 2000).

(6.15) a. *Separated relatives in NP-internal position*
which/that = 72% (142) zero = 28% (54)

b. *Separated relatives in NP-external position (i.e., extraposed)*
which/that = 94% (17) zero = 6% (1)

The relatives in (6.15b) have been completely extracted out of NP, in structures corresponding to *buildings will never fall down which we have constructed.* In (6.15a) they remain NP-internal but still separated, e.g., by an intervening PP as in *buildings in New York which we have constructed.* The corresponding relatives with zero relativizers are *buildings in New York 0 we have constructed* (6.15a) and *buildings will never fall down 0 we have constructed* (6.15b). There is a significant increase from 72 percent to 94 percent in relativizer retention when the separated relatives are extraposed, and a corresponding decrease in zero relativizers. These data support prediction 4 (6.13).

More generally, the avoidance of zero-marked relatives in these separated relative clauses is reminiscent of the distributional pattern we have now seen for these and other empty elements in §§2.2.2, 2.2.3, and 2.5.3. The absence of the relativizer, which is motivated by Minimize Forms (2.8), sets up an additional processing dependency between the reduced clause and the head noun. An explicitly marked relative clause can be recognized as such on the basis of the relativizer. Removal of the relativizer requires access to the head noun so that the reduced clause can be processed as a relative clause, first of all, and be attached to this noun within its NP. When the reduced clause and the head noun are adjacent, both MiF and MiD preferences are satisfied. The form of the relative clause is minimal and processing the additional dependencies between relative and head takes place in a minimal domain. But as the reduced relatives are increasingly separated from their nominal heads, the processing of the additional dependencies requires increasingly large and more complex domains, at variance with Minimize Domains (2.1). The cooperation between MiF and MiD under adjacency becomes an increasing competition under separation.

Notice that the reason we can now give for the adjacency of zero-marked relatives to their heads is intimately connected to the reason I gave for the preferred adjacency of lexically dependent elements to their heads in the V-PP-PP data of §5.2. There are additional dependencies in both, going beyond phrasal construction and parsing, that result in an adjacency preference by Minimize Domains (2.1): semantic dependencies in the one case involving the need for the verb to access the preposition and/or vice versa in order for a correct meaning to be assigned (e.g., *count on someone*); syntactic dependencies in the other whereby the type of clause in question and its adjunct status are recognized by reference to the head noun. The explanation for both types

of adjacency is, I believe, fundamentally the same and is predicted by Minimize Domains (2.1); there are more grammatical or lexical relations to be processed in zero dependencies and semantic dependencies, and more processing domains, making adjacency or proximity between the interdependent elements preferred in both performance and (when conventionalized) in grammars too. See Hawkins (2003) and (2004: 148–65) for further discussion and more examples.

Consider now some data from grammars involving case marking, of relevance to prediction 4 (6.13) and also to the interaction between MiF and MiD that we have just discussed. In languages that employ case copying as an attachment strategy we predict a possible asymmetry whereby explicit case marking can be retained on separated, but not on adjacent, sisters. Warlpiri exemplifies this. Contrast the Warlpiri pair (6.16) from Hale (1973) with the Kalkatungu examples (6.2ii) above from Blake (1987: 87), repeated here as (6.17):

(6.16) a. [$_{NP}$ tyarntu wiri-ngki]+tyu yarlki-rnu (Warlpiri)
dog big-ERG+me bite-PAST

b. [$_{NP_i}$ tyarntu-ngku]+tyu yarlku-rnu [$_{NP_i}$wiri-ngki]
dog-ERG+me bite-PAST big-ERG
'The big dog bit me.'

(6.17) a. [$_{NP_i}$ tyipa-yi thuku-yu yaun-tu] [$_{NP_i}$ yanyi] itya-mi
this-ERG dog-ERG big-ERG white-man bite-FUT
(Kalkatungu)

b. [$_{NP_i}$ tyipa-yi thuku-yu] [$_{NP_j}$ yanyi] itya-mi [$_{NP_i}$ yaun-tu]
this-ERG dog-ERG white-man bite-FUT big-ERG
'This big dog will bite the white man.'

Case copying in Kalkatungu occurs on every word of the NP, whether adjacent or not. Warlpiri case copying occurs only when NP sisters are separated (6.16b). When NP constituents are adjacent (6.16a), the ergative case marking occurs just once in the NP and is not copied. This pair of Australian languages illustrates the asymmetry underlying Moravcsik's (1995: 471) agreement universal:

(6.18) *Moravcsik's Agreement Universal*
If agreement through case copying applies to NP constituents that are adjacent, it applies to those that are non-adjacent.

Agreement can be absent under adjacency at the same time that it occurs in non-adjacent environments. What is ruled out is the opposite asymmetry: agreement when adjacent and not when non-adjacent. Since agreement is a type of attachment marking we see correspondingly that the explicit encoding

of attachment in performance and grammars is found under both adjacency ((6.14a) and (6.17a)) and non-adjacency ((6.14b), (6.16b), and (6.17b)). Zero coding is preferred only when there is adjacency and is increasingly dispreferred when there is not (compare (6.14a) with (6.14b) in performance and (6.16a) with (6.16b) in grammars). What is not found is the opposite of the English relativizer pattern and of Warlpiri case coding: explicit attachment coding under adjacency and zero coding for separated items.

An example of case copying in a nominative–accusative language comes from Hualaga Quechua (see Plank 1995: 43 and Koptjevskaja-Tamm 2003: 645). When a possessor phrase is separated from its possessed head, as in (6.19), the accusative case marker *-ta* appropriate for the whole NP is added to genitive case-marked *Hwan-pa*.

(6.19) Hipash-nin-ta kuya-: Hwan-pa-ta (Hualaga Quechua)
daughter-3POSS-ACC love-1 Juan-GEN-ACC
'I love Juan's daughter.'

6.3.2 Minimize NP attachment encoding

A further prediction that can be made on the basis of the NP Attachment Hypothesis (6.12) is:

(6.20) *Prediction 5: Minimize NP Attachment Encoding*
The explicit encoding of attachment to NP will be in inverse proportion to the availability of other (morphosyntactic, syntactic, and semantic–pragmatic) cues to attachment: the more such cues, the less encoding.

In other words, we predict less explicit attachment marking when there are other cues to attachment. Consider in this regard Haspelmath's (1999b: 235) universal regarding the omissibility of definite articles in NPs with possessors depending on the type of possession.

(6.21) *Haspelmath's Article-Possessor Universal*
If the definite article occurs with a noun that is inherently related to an accompanying possessor, such as a kinship term, then it occurs with nouns that are not so inherently related.

I suggest that this universal can be seen as a consequence of the attachment function of the definite article, linking a possessor to a head noun. Kinship involves necessary and inalienable relations between referents, which makes explicit signaling of the attachment less necessary with nouns of this subtype. The definite article can attach a possessor to a head noun in Bulgarian, Nkore-Kiga, and Italian (6.22a), but not when the head noun + possessor describes a

kinship relation like 'my mother' (6.22b); see Haspelmath (1999b: 236) and Koptevskaja-Tamm (2003):

(6.22) a. Bulgarian *kola-ta mi* Nkore-Kiga *e-kitabo kyangye*
car-DEF my DEF-book my
Italian *la mia casa*
DEF my house

b. Bulgarian *majka(*-ta) mi* Nkore-Kiga *(*o-)mukuru wangye*
mother(-DEF) my (DEF-)sister my
Italian *(*la) mia madre*
(DEF) my mother

Support for this attachment explanation comes from the fact that other attachment devices (in (6.2)) show a parallel sensitivity to inalienable possession, suggesting that omissibility is not a consequence of the semantics and pragmatics of definiteness as such in combination with inalienable possession. The Cantonese nominalizer/attachment marker *ge* can be omitted as an explicit signal of attachment for possessor + noun when there is an inalienable bond between them, like kinship, and especially when the possessor is a pronoun. Contrast *ngóh sailóu* (I younger-brother, i.e., 'my younger brother') with *gaausauh ge baahngūngsāt* (professor NOMLZ/ATTACH office, i.e., 'the professor's office'); see Matthews and Yip (1994: 107).

A particularly subtle test of the basic idea behind prediction 5 (6.20) has been made on Zoogocho Zapotec data by Sonnenschein (2005: 98–110). There are different formal means for marking possession in this language, by simple adjacency of nouns (6.23a), by a possessive prefix (6.23b), and by a postnominal possessor phrase headed by *che* (of) (6.23c):

(6.23) a. tao lalo (Zoogocho Zapotec)
grandmother Lalo, i.e., 'Lalo's grandmother'

b. x-kuzh-a'
POSS-pig-1SG, i.e., 'my pig'

c. tigr che-be'
tiger of-3INFORMAL, i.e., 'her tiger'

Sonnenschein tests the idea that there is a continuum from inalienable possession at the one end ('my head,' etc.) through frequently possessed items (like 'her pig') to not very frequently possessed items (like 'her tiger'). He shows on the basis of a corpus study that the amount of formal marking for possession correlates inversely with the frequency with which the relevant head nouns are in a semantic possession relation. Possession signaled by simple adjacency (6.23a) is used for head nouns that are always possessed

(like kinship terms and body parts). Possession signaled syntactically by a postnominal possessor phrase (6.23c) is used with head nouns that are generally unpossessed. And NPs that show either morphological (6.23b) or syntactic encoding (6.23c) are more variably possessed. This intermediate group also shows a preference for the morphological variant when the possession is more inherent and for the syntactic variant when the possession is less inherent, for example when a possessed house is under construction and the owners are not yet living in it.

Sonnenschein's quantification of the degree and frequency of possession correlating inversely with both the presence versus absence of possession marking and with its amount and complexity supports the role of additional semantic–pragmatic cues in signaling the attachment of possessor to possessed, resulting in form minimization.

6.3.3 Lexical differentiation and word order

Languages with lexical specialization for parts of speech such as adjectives can provide clear attachments to NP in many grammatical environments when phrase structure rules plus the lexicon are accessed in parsing—i.e., lexical specialization plus a grammar can often guarantee unambiguous attachment to NP. Languages without such specialization must rely on either morphosyntactic particles (see (6.2)) or on fixed word order. In (6.24) I formulate a prediction that can be made for fixed word orders on the basis of the NP Attachment Hypothesis in (6.12).

(6.24) *Prediction 6: Lexical Differentiation and Word Order*
Lexical specialization for the category Adjective will permit the relevant languages to order the Adjective before or after the Noun, attachment to NP being guaranteed by this part of speech plus phrase structure rules; lack of such specialization should lead to more fixed word orders (among other structural consequences), with consistent attachments to a leftward nominal head (VO languages) or to a rightward nominal head (OV languages).

Hengeveld et al. (2004) have indeed shown that a lack of lexical differentiation for basic parts of speech correlates with more fixed and typologically consistent word orders across phrasal categories. The NP Attachment Hypothesis (6.12) provides a reason why. They distinguish, inter alia, between languages that have a separate category of adjectives, which can unambiguously attach to NP, versus those that do not. The former (exemplified by English and Wambon, types 4–5 on the Hengeveld et al. Part of Speech hierarchy) all have separate nouns, verbs, and adjectives. The latter (exemplified by Samoan,

Warao, and Ngiti, types 1–3 for Hengeveld et al.) have no separate adjectives. They may or may not have separate nouns and verbs either, depending on their position on the Hengeveld et al. hierarchy.

Consider Basque, an OV language which has a separate class of adjectives ordered inconsistently after the noun, while adverbial modifiers of the verb precede the verb (Saltarelli 1988). In the absence of lexical (or morphosyntactic) differentiation between Adj and Adv, a non-head modifier that followed N and preceded V would be regularly ambiguous as to its attachment site, to the left or to the right:

(6.25) N Adj/Adv V [Head-Modifier in NP within SOV]

Using English morphemes, a grammatical distinction between *loud* (adjective) and *loudly* (adverb) can make clear the intended attachments in *music loud played* (attach *loud* to *music* within NP) versus *music loudly played* (attach *loudly* to *played* within VP). Alternatively, a consistently ordered adjective in an SOV language would avoid this attachment ambiguity:

(6.25′) Adj N Adv V [Modifier-Head in NP within SOV]

Conversely, a VO language with Modifier Head ordering in the NP but with postverbal adverbial modification would invite similar attachment ambiguities in the absence of lexical differentiation between *play loud music* and *play loudly music*:

(6.26) V Adv/Adj N [Modifier-Head in NP within VO]

A consistently ordered adjective would avoid them:

(6.26′) V Adv N Adj [Head-Modifier in NP within VO]

In VO (head-initial) languages, non-head categories can be consistently attached to heads on their left, therefore, and in OV (head-final) languages they can be consistently attached to the right. Languages with lexically differentiated categories, on the other hand, can tolerate ordering inconsistency when the lexicon supplies a clear class of adjectives that are known to attach to NP in relevant environments.

The precise predictions that we can make on the basis of (6.24) are complicated by the fact that category differentiation in the lexicon plus word order is only one grammatical means for solving attachment problems, morphosyntactic linking of different kinds being another (see (6.2)), and also (in spoken language) considerations of prosody and intonation (see, e.g., Kiaer 2007). Nonetheless, we expect any one attachment signaling device to be more productive in certain language types and structures than in others, namely in those for which the absence of any attachment device at all

would be problematic. For lexical differentiation we can reasonably make the following predictions, therefore:

(6.27) a. Lexically differentiated languages with a productive Adj category (Hengeveld et al.'s types 4–5) and with basic SOV will permit inconsistent HM order in NP, either as a basic order or as a frequent variant.

b. Such languages with basic VO (VSO or SOV) will permit inconsistent MH order in NP, either as a basic order or as a frequent variant.

(6.28) a. Lexically undifferentiated languages without a productive Adj category (Hengeveld et al.'s types 1–3) and with basic SOV will favor consistent MH order in NP.

b. Such languages with basic VO (VSO or SVO) will favor consistent HM order in NP.

The relevant language quantities, taken from Hengeveld et al.'s Table 4 (p. 549), are as follows:

(6.29) a. 10/13 SOV languages described in (6.27a) permit inconsistent HM orders in NP (e.g., Basque).

b. 4/8 VO languages described in (6.27b) permit inconsistent MH orders in NP (e.g., Arapesh).

(6.30) a. 5/6 SOV languages described in (6.28a) have consistent MH order in NP (e.g., Mundari).

b. 3/4 VO languages described in (6.28b) have consistent HM order in NP (e.g., Samoan).

The number of relevant SOV and VO languages available for testing in this sample is sometimes small. Nonetheless, a significant proportion of those predicted to permit word order inconsistency do so (6.29), while the languages predicted to be consistent in their orderings generally are so (6.30).

The ten SOV languages in Hengeveld et al.'s sample with inconsistent HM orders in NP referred to in (6.29a) are: Abkhaz, Alamblak, Bambara, Basque, Hittite, Koasati, Nasioi, Sumerian, Oromo, and Wambon. Of these, four have maximum inconsistency with basic and fixed HM (Bambara, Koasati, Nasioi, and Sumerian), four have basic HM with frequent departures in favor of the typologically consistent MH (Abkhaz, Basque, Nasioi, and Sumerian), while two have no basic order in NP (Hittite and Alamblak). The three SOV languages described in (6.29a) with consistent MH orders in NP are: Burushaski, Japanese, and Nama. MH is both basic and fixed in the latter two, while MH is basic and free in Burushaski.

The four VO languages described in (6.29b) with inconsistent MH orders in NP are: Arapesh, Berbice Dutch, Pipil, and Polish. Berbice Dutch and Pipil have basic MH (with fixed and free NP word orders respectively). Arapesh and Polish are classified as having no basic NP order (MHM).

The five SOV languages with basic and fixed MH order in NP referred to in (6.30a) are: Mundari, Hurrian, Quechua, Turkish, and Ket (with Warao being the sixth SOV language in this set and having basic and fixed HM). The three VO languages with basic and fixed HM in NP described in (6.30b) are: Samoan, Miao, and Tidore (with Tagalog the fourth in this set and having no basic order, i.e., MHM).

I have argued in this chapter that a number of cross-linguistic patterns and regularities in noun phrase syntax and morphosyntax can be captured and better understood when viewed from the processing perspective of the PGCH (1.1). Two hypotheses have been proposed, Constructibility (6.3) and Attachability (6.11), from which six predictions have been derived in §6.2 and §6.3 and tested on illustrative and quantified data. Grammars appear to have conventionalized the preferences of performance that are evident in languages with structural choices between, e.g., the presence or absence of a relative pronoun, of case marking, of an article or classifier, and between one ordering versus another. A processing efficiency approach can help us clarify why and how grammars make use of these various devices summarized in (6.1) and (6.2) and why different languages exhibit the cross-linguistic variants that they do in this area.

7

Ten differences between VO and OV languages

I have argued (§2.1 and §5.1) that we can give a simple and motivated explanation in efficiency terms for the productivity of two major language types, VO and OV, and more generally for head-initial and head-final languages. Within each of these, phrases that are consistently or harmonically ordered are, in turn, far more productive than their inconsistent or disharmonic counterparts, as shown throughout Chapter 5. These two types can be productive because there are two, and only two, logically possible ways for a grammar to be optimally efficient for phrase structure construction and recognition: namely when heads consistently precede or consistently follow their non-heads within the hierarchically arranged phrases of a tree structure. The productivity of two mirror-image types is accordingly predicted by Minimize Domains (2.1), in conjunction with certain basic assumptions about phrase structure and syntactic heads which are now shared across many grammatical models despite differences in detail (cf. Corbett et al. 1993; Brown and Miller 1996). MiD makes a further prediction for the less than optimal grammars and their distribution: the more efficient they are, the more exemplifying languages there will be; the less efficient, the fewer, in proportion to the degrees of efficiency that MiD defines (see §§5.4–5.5 and Hawkins 1990, 1994, 2004 for more quantified data).

The predicted productivity of head-initial and head-final grammars is supported by processing operations that go beyond the recognition and production of phrase structure. All the other syntactic and semantic relations that link heads of phrases can be processed more efficiently when heads are adjacent and the domains for these additional processing operations are minimal. So, head–complement relations between a verb and its argument NPs or PPs are more efficiently processed when V is adjacent to N and P respectively, whether in the order [V [N . . .]] and [V [P . . .]] or in the reverse [[. . . N] V] and [[. . . P] N]; see §§5.2–5.3 and Hawkins (2004). MiD predicts that the more distinct processing operations there are that link two heads, whether of phrasal construction and attachment (§5.1),

head–complement processing (§§5.2–5.3), or involving other types of dependencies as well (see §9.2), the tighter will be their adjacency in performance and grammars. This can be seen in the quantities of relevant performance instances in languages with choices, and in quantities of grammars in language samples (recall §1.3).

For VO and OV grammars Maximize Online Processing (2.16) provides a further motivation for the productivity of these symmetrical orderings. Verbs and direct objects exhibit symmetrical relations of dependency, when dependency is defined in processing terms as it is here in (2.1) (§2.1): Two categories A and B are in a relation of dependency iff the parsing of B requires access to A for the assignment of syntactic or semantic properties to B with respect to which B is zero-specified or ambiguously or polysemously specified. This definition is argued in Hawkins (2004: 18–25) to capture the intuition behind the dependency concept in a more empirically adequate and testable way than in definitions given in purely grammatical terms (Tesnière 1959; Hays 1964). Direct objects are dependent on verbs, by this definition, since they receive a thematic role from the verb. For the selection and processing of this role the parser has to access the verb. By MaOP this leads to a preference for V before NP ordering, so that the verb is already accessible when the NP is parsed and an unassignment of its thematic role is avoided online. However, a universal [V NP] preference is offset by the fact that verbs are also dependent on NPs for semantic disambiguation (recall Keenan's 1979 discussion of *run the race/run the water/run the advertisement/run his hand through his hair*; see §2.3.3). For syntactic property assignments to a verb the parser must also select between competing subcategorization frames based on co-occurring NPs and PPs, etc. This all favors [NP V] ordering by MaOP, so as to position these items before the verb and avoid the unassignment of syntactic and semantic properties of the verb at the verb. The result is symmetrical dependencies between verb and object which are reflected in clear symmetrical orderings across languages, with no single order being fully optimal (recall §4.4), and with adjustments being made in the respective grammars to compensate for processing delays and unassignments through, e.g., enriched case marking and 'argument differentiation' (for [NP V]) and enriched 'predicate frame differentiation' (for [V NP]); see §7.2 for details. When dependencies are asymmetrical, by contrast (see §2.3 and Chapter 8), the orderings are generally asymmetrical too with X before Y when Y is asymmetrically dependent on X. Symmetry and asymmetry in dependency, defined in parsing terms, appear to correlate with symmetry and asymmetry in ordering, therefore (see Hawkins 2002, 2004: ch. 8 for details).

7.1 Mirror-image weight effects and head orderings

The first consequence of the VO–OV typology, therefore, stemming as I have argued from a shared preference for minimal domains (§2.1), is a major difference in surface structure: syntactic heads are ordered in a mirror-image way across the two types ([V [P [N . . .]]] versus [[[. . . N] P] V], etc.). Weight effects also show a mirror image and follow a short-before-long preference for VO (§5.2) and a long-before-short preference for OV (§5.3) languages, both in performance and in grammars (§5.4). Numerous details supporting these mirror images have been summarized in the relevant sections of this book and in previous publications cited here and do not need to be repeated.

What does need to be put in proper perspective in this chapter is, first of all, a discussion of when the mirror image breaks down and the two language types show similarities. And second we need to examine certain adjustments that each type makes to its basic head ordering, resulting in further contrasts between them. The advantage of an efficiency approach to typological variation is that we can give some principled reasons and make predictions for the limits on mirror images, in exactly the cases in which non-mirror images and similarities seem to be found, and we can also give reasons and make predictions for the respective adjustments to the basic typology that result in further contrasts. And of course this basic typology itself involving head-initial and head-final orders has been argued to follow from the equal efficiency of numerous processing operations when heads consistently precede or follow their non-heads.

Efficiency gives us a principled way of capturing mirror images, therefore, plus the similarities, plus various contrasts that can be linked to the mirror image. The cross-linguistic patterns supporting this partitioning seem to be increasingly clear and fall out readily from several language samples. Purely grammatical approaches, on the other hand, can help us with detailed points of analysis for individual languages, but they have no independent apparatus beyond stipulation for predicting and partitioning the cross-linguistic variation in the ways we shall see in this chapter.

The areas of similarity between head-initial and head-final languages seem to involve, interestingly, phrases that are at the two extremes of weight and complexity: very short items, like single-word adjectives modifying nouns; and very long ones, like relative clauses modifying nouns or embedded finite clauses as complements of verbs. I have already commented on the adjectives in §4.4 and referred to Dryer's (1992) finding that there are systematic exceptions to Greenberg's mirror-image correlations when the category that modifies a head is a single-word item. In effect, both orders of (single-word)

adjective and noun (AdjN as in English, NAdj as in Basque) are productive in both VO and OV languages and there is no longer a mirror-image preference (cf. also §6.3.3). This actually follows from the definition of MiD given in §5.1. A short adjective will not add significantly to phrase structure processing domains linking NP to a preceding or following head (V, P, or N etc.), whether the adjective precedes or follows the head noun it modifies (cf., e.g., English *he* [*studied* [*easy books*]]), but a longer intervening adjective phrase would significantly lengthen these processing domains; see §5.5.

For the heavier relative clauses and finite complements the preferred structure is, as we shall see, similar across VO and OV types, but the distribution of variants is now different. For {N, Adj} and {V, O} all four logically possible combinations of word orders are productive. For {N, Rel} both VO and OV languages prefer head-initial NRel, although the OV languages also exhibit head-final RelN in some cases, which is not found in VO. Just three of the four combinations are now productive (see §2.5.2 and §7.3).

7.2 Predicate frame and argument differentiation in VO and OV languages

A second major consequence of the VO/OV typology that is of considerable psycholinguistic significance involves the crucial role of the verb in online processing. Frazier (1985) discusses a number of temporary ambiguities and garden paths that arise in English as a consequence of its verb position and online parsing routines, e.g., (7.1)–(7.4). The regions of temporary ambiguity are italicized.

(7.1) *Bill knew the answer* was correct.

(7.2) *While Mary was reading the book* fell down.

(7.3) *The candidate expected to win* lost.

(7.4) *Mary helped the boy and the girl* ran away.

The verb *know* in (7.1) can take both an NP direct object (like *the answer*) or a *that*-complement (like *that the answer was correct*), the verb *read* in (7.2) is ambiguous between transitive and intransitive frames, and *expected* in (7.3) is ambiguously active and passive. When sequences like *Bill knew*, *While Mary was reading*, and *The candidate expected* are parsed, these verbs are compatible with a plurality of co-occurrence possibilities, and it is this plurality that results in the temporary ambiguity and garden path.

More generally, temporary ambiguities like these arise because alternative verb co-occurrence frames or predicate frames are activated when the verb is encountered. The garden path results when the simplest and most expected frame selected from the alternatives in the italicized portion of the sentence turns out to be the wrong one subsequently when later material is encountered, i.e., there is a structural misassignment online that causes reanalysis and backtracking (see Hawkins 2004: 51–8).

The assumption that the verb does indeed activate alternative co-occurrence frames in online processing has been supported experimentally by Tanenhaus and Carlson (1989); see also Melinger, Pechmann, and Pappert (2009) for a summary of studies documenting the central role of the verb in online production and incremental processing. Noun phrases and their thematic roles, Agent, Patient, Instrument, do not in general select for and predict particular predicates because they are each compatible with too many, unless there is rich contextual information. But a verb can activate a finite set of co-occurrence frames online, and if the verb occurs first before its arguments, then these latter will gradually select the intended frame from the set of possibilities that have been activated. If the arguments occur first, before the verb, then alternative co-occurrence frames will not generally be activated early and gradually reduced, but instead the intended frame will need to be selected immediately following its activation at the verb in a way that takes account of preceding arguments.

The position of the lexical verb is highly significant for language performance, therefore, and we can expect, by the Performance–Grammar Correspondence Hypothesis (1.1), that there will be consequences of this for the grammars and lexicons of languages with different basic verb positions. The crucial point is: does the verb precede its co-occurring arguments (as in VSO and SVO languages), or do these latter precede the verb (i.e., SOV)? In the former case there is early activation of the verb's co-occurrence possibilities with temporary ambiguities and possible garden paths as the intended frame is gradually selected. In the latter, co-occurrence frame activation and selection based on preceding material will be almost instantaneous.

In Hawkins (1995) I formulated some predictions for different language types based on this difference. Languages with verb-final structures (like Japanese and Korean and the Dravidian languages, and also non-rigid SOV languages like German) should exhibit what I called greater 'predicate frame differentiation' and greater 'argument differentiation.' Predicate frame differentiation refers to the degree to which a verb is distinctive from others by virtue of its unique selectional restrictions or syntactic co-occurrence possibilities (so-called 'subcategorization'). A verb that is uniquely transitive is more differentiated than one that is ambiguously transitive or intransitive.

A verb that selects restricted direct objects for 'putting on clothing' according to the body part and the type of clothing in question (hats on the head, a coat over the rest of the body, etc.) is more differentiated than one (like English 'put on') that is compatible with many different types of body parts and clothing activities (see Hawkins 1995 and 1986). And a verb that lists only V-NP-NP for its ditransitive frame is more differentiated than one that lists both this and V-NP-PP. Argument differentiation refers to the degree to which arguments are, e.g., differentially case-marked versus ambiguous as to case, and the degree to which they are assigned a narrow set of thematic roles like Agent and Patient rather than a broader set as in English (see §2.4 and the examples below).

When the verb is the last constituent in the clause and the very next item to be parsed belongs in an altogether different clause, the parser must succeed instantly in selecting the correct frame and its arguments. The grammar and lexicon must, I hypothesize, help the parser of a verb-final language by ensuring that predicate frame selection is immediately successful. More differentiated predicate frames and arguments can accomplish this and in a number of observable ways.

First, subcategorization restrictions can be made tighter in SOV languages (e.g., avoid transitive/intransitive ambiguities for verbs like *read* in (7.2)) with the result that arguments can be paired with their predicates more uniquely and more easily. Second, additional selectional restrictions can be imposed so that certain verb–NP pairings are more constrained, frequently co-occurring and easily recognizable. This is exemplified in detail for German versus English selectional restrictions in Hawkins (1986) following Leisi (1975) and Plank (1984). Verbs like *put on* and *break* allow a broader range of direct objects in English (e.g., breaking both brittle and non-brittle objects like tendons and ropes, putting on a wider range of clothing objects) compared to German. Third, subjects and objects can be made less semantically diverse in SOV languages so that there are fewer co-occurrence possibilities to choose from and more constrained assignments of thematic roles to NPs (transitive subjects are agents, transitive objects are patients, etc.). E.g., we expect fewer assignments of thematic roles to a transitive subject such as Instrument (*A penny would buy two to three pins*), Location (*This tent sleeps four*), and Theme (*The book sold 10,000 copies*), of the kind we find in English; see §2.4. Direct object NPs should regularly be patients (*I kicked the book*) rather than themes (*I liked the book*). Fourth, surface coding devices can be grammaticalized for arguments that permit immediate thematic role recognition. This is what case marking generally does. It constrains the thematic roles that can be assigned to surface NPs, making them less semantically diverse compared with caseless languages. This has the advantage of both making thematic role

information available early online, prior to the verb, and of facilitating argument–predicate assignments when the verb is encountered. And indeed it has long been known that there is a strong correlation between SOV and 'rich case marking' (see Hawkins 2004: 246–50 for figures, further references, and discussion). Fifth, certain rule types like raisings and long Wh-movement can be avoided in SOV languages that reposition arguments away from the immediate predicates with which they belong semantically.

Constraining the verb's co-occurrences in this way means that at the time the verb is processed its intended predicate frame can be immediately recognized and selected in parsing, in accordance with Maximize Online Processing (2.26). There will be no temporary ambiguities or garden paths (i.e., unassignments or misassignments in verb processing), of the kind that arise in verb-before object-structures like (7.1) and (7.2) above. These could be avoided in English (as Frazier 1985 points out) in the event that this language had OV word order (contrast the order in (7.2), in which *read* would be unambiguously intransitive in an OV English, with *While Mary the book reading was*..., which would be unambiguously transitive). The final position of the verb has the further advantage that the considerable semantic variability in the interpretation of verbs (and of other 'function' categories; see Keenan 1979) can be resolved immediately at the verb by reference to its preceding arguments, in verb phrases corresponding to *run the race*/*run the water*/*run the advertisement*/*run the campaign*, etc., in English. My general prediction here is that such variability should be less extensive in OV languages, since their selectional restrictions are tighter. But still some generality in verbal semantics is unavoidable, with fine-tuning to the meaning of a DO argument, and by positioning verbs finally SOV languages succeed in immediately allowing the precise interpretation of the verb to be assigned at the verb itself, without the need for look ahead and without unassignments and misassignments.

This need for, and the benefits of, immediately assigning syntactic and semantic properties to the verb at the verb itself mean that typologically we expect to see verb-final languages with a more constrained set of verb co-occurrence possibilities. Specifically we expect to see more predicate frame differentiation and more argument differentiation, but less 'argument trespassing,' which I define to mean less movement of NP arguments into clauses in which they contract no semantic relations with their most immediate predicates (see Hawkins 1986, 1995). For verb-early (SVO, VSO, VOS) languages, however, no such constraints are predicted. These languages do not need to conventionalize devices that permit immediate and correct predicate frame recognition at the verb, because the parser still has the remainder of the clause in which to complete its predicate frame selection, and because the

intended predicate frame cannot be identified at the verb anyway and there needs to be a 'look ahead' to the verb's arguments. As a consequence, the need for immediate and correct decision-making at the verb, and the resulting need for clear predicate-frame differentiation, argument differentiation, and for local argument–predicate matching will impose much weaker requirements on the grammars and lexicons of such languages. Indonesian (SVO) reveals many similarities with English, for example, whereas Russian (SVO) does not (Müller-Gotama 1991, 1994). Within Germanic, English exemplifies these typological possibilities for an SVO language (as a result of historical changes in Middle and Early Modern English; see Hawkins 2012), whereas Icelandic is more conservative. But verb-final languages should be more constrained in these respects, whereas verb-early languages are predicted to be more variable. Some of these predictions are summarized in (7.5):

(7.5)
i. if SOV, then case system (or fixed SOV)
ii. if SOV, then narrow semantic range of basic grammatical relations (e.g., Acc = Patient)
iii. if SOV, then more consistent assignment of OBLIQUE thematic roles (e.g., Instruments, Locatives) to grammatically OBLIQUE arguments (rather than to SU and DO)
iv. if SOV, then no or limited raising
v. if SOV, then no or limited Wh-extraction out of clauses

These predictions were tested in Table 1 in Hawkins (1995: 142), which is reproduced here as Table 7.1, using data taken from Müller-Gotama (1991, 1994).

The predictions are supported, in this sample of languages at least.

The label that I have given to this contrast between OV and English-type VO languages is 'tight fit' versus 'loose fit' (Hawkins 1986: 121–7). What is at issue here is the mapping between forms and meanings. Tight-fit languages have richer, more complex, and more unique surface forms that map onto less ambiguous and more constrained meanings, i.e., there is more of a one-to-one correspondence between form and meaning. This simplifies the mapping between them, but at the expense of processing more forms. Loose-fit languages involve more complex form–meaning mappings (the selection between many different syntactic and semantic possibilities for one and the same grammatical construction and verb), but simpler processing of the forms themselves. This trade-off can be made more precise in terms of quantifiable degrees of predicate frame differentiation, of argument differentiation, and of argument trespassing (see Hawkins 1995).

The diachronic 'drift' of English (Sapir 1921) has been one of movement from a tighter-fit language like Modern German to a looser-fit language (see

TABLE 7.1 Grammatical correlations with verb position (from Müller-Gotama 1991, 1994)

Language	Basic V-position	Morphological Case System	Semantic Range of Basic GRs	Raising	Wh-extraction
Korean	SOV	yes	narrow	no[a]	no
Japanese	SOV	yes	narrow	no[b]	no
Malayalam	SOV	yes	narrow	no	no
Turkish	SOV	yes	narrow[c]	(no)[d]	no
German	SOV	yes	narrow	no[a]	limited
Russian	SVO	yes	narrow	no[a,b]	limited
Chinese	SVO	no	broader	no[a]	no
Sawu	V-initial	no[e]	narrow	no	no
Hebrew	SVO	no[f]	narrow	(no)[g]	yes
Indonesian	SVO	no	broader	limited[h]	yes
Jacaltec	VSO	no	narrow	yes	yes
English	SVO	no	broadest	yes	yes

Notes

[a] In Korean, German, and Russian there are structures that appear on the surface to have undergone Tough Movement. For Korean, Müller-Gotama (1991: 97–100) argues that this is not the correct analysis; Comrie and Matthews (1990) argue against Tough Movement in German and Russian; for Chinese, Müller-Gotama (1991: 277–8) also argues against a Tough Movement analysis, following Shi (1990).

[b] Japanese and Russian have apparent S-O Raising structures in which the embedded verb must be *be*. Müller-Gotama (1991: 211–12) argues against a raising analysis in these cases, citing relevant unpublished work on Russian by Comrie.

[c] Müller-Gotama (1991) provides semantic range data only for DOs in Turkish.

[d] There may be some limited S-O Raising in certain dialects of Turkish; cf. Müller-Gotama (1991: 219).

[e] Sawu does not have case morphology, but it does have a set of case prepositions, including one for ergative NPs.

[f] Modern Hebrew lacks morphological case markers. A case particle *et* serves as a marker of definite direct objects, however, and this leads Müller-Gotama (1991: 295) to assign 'yes' to Hebrew for case. However, by this criterion, many other languages can also be assigned 'yes' which lack morphological case and have free-standing case particles or prepositions. I have accordingly converted his 'yes' to 'no' here.

[g] Gundel et al. (1988) have argued against S-S and S-O Raising in Hebrew. Müller-Gotama (1991: 287–8) argues that there may be limited evidence for S-S Raising, though not for S-O Raising and Tough Movement, but he considers its existence "doubtful."

[h] S-O Raising exists in Indonesian; see Chung (1976) and Müller-Gotama (1991: 137–8). There is also a limited process of S-S Raising, involving obligatory passivization in the dependent clause and referred to by Chung (1976) as Derived Subject Raising.

Hawkins 2012). Once the new basic SVO word order was set, with new processing routines, the stage was set for further possible diachronic changes that involved the processor looking ahead in the parse string, which results in more temporary and full ambiguity, and more instances of dependent processing whereby less explicit and less formally marked elements receive their property assignments by accessing some item elsewhere in the clause on which

their interpretation depends. These are all processing correlates of the loose-fit type. So, as explicit case marking was lost on non-pronominals in English, NPs became dependent on their sister verbs for the assignment of properties that were often independently processable before. In earlier English, accusative and dative case could often be recognized on the basis of the NP itself, and the associated patient and recipient roles could be assigned without the parser having to access a verb later in the clause. But the zero-marked NPs of Middle English required access to the verb for (abstract) case and thematic role assignment. And once this dependency on the verb was conventionalized, the stage was set for broader thematic role assignment possibilities, sensitive to the verb's semantic properties, within the new NP-V-NP structure. Modern English transitive clauses with locative and instrumental subjects emerged (*this tent sleeps four*, *my lute broke a string*), perhaps by conventionalizing in the grammatical subject the range of pragmatic functions and thematic roles that occurred in the first constituent of the verb-second structures of earlier English (see Los and Dreschler 2012). Locative and instrumental thematic roles are normally assigned to PPs (*four can sleep in this tent*) but they could now be assigned to a transitive structure in which delayed predicate frame selection and look-ahead to the last NP had become a normal part of processing routines. Similarly raising structures could emerge like *Mary happened to win the prize* in which the adjacent NP-V... pair does not permit the assignment of a clear thematic role to the NP *Mary*, just as it cannot with *this tent sleeps*..., and instead subsequent material in the parse string needs to be accessed for thematic role and predicate frame selection.

Relative pronouns and complementizers are also deletable in Modern English (cf. *the man I know* for *the man who(m)/that I know*, *Bill knew the answer was correct* for *Bill knew that the answer was correct*). These deletions were extremely rare in earlier English (see Traugott 1992; Hawkins 2012). Immediate processability has again given way to temporary ambiguities (see (7.1)) and to dependent processing whereby the subordinate status of clauses is not explicitly and unambiguously signaled but is derived by accessing later items in the string like the finite verb (recall the discussion of *that*-trace parsing in §4.3). Likewise many verbs and nouns have become category-ambiguous in English (*run*, *play*, *show*, etc.) which had been lexically or morphologically unambiguous earlier and which now require co-occurring articles or infinitival *to* (*I want the play/to play*). More generally, Modern English has significantly more ambiguity and vagueness in its syntactic, lexical, and morphological categories than it once did, and more temporary ambiguity, and it relies to a much greater extent on co-occurring categories for the assignment of the relevant (disambiguating) properties. This is all characteristic of a 'loose-fit' language, involving multiple assignments of

meanings to forms, and a greater role for sentence-internal and also wider contextual disambiguation.

7.3 Relative clause ordering asymmetries in VO and OV languages

The postnominal relative clause type of English, exemplified in (7.6), is good for domain minimization:

(7.6) I [$_{VP1}$ saw [$_{NP}$ professors*i* [$_{S}$ that [$_{VP2}$ wrote*i* books]]]]

When a head noun like *professors* immediately follows the verb *saw*, the phrasal combination domain (PCD) for recognizing the two ICs of the matrix VP1 (namely V and NP) is an optimal two words (*saw professors*), giving a 2/2 = 100% IC-to-word ratio (cf. §2.1). The domains for processing the various lexical relations of co-occurrence with the verb are also minimal. If the relative clause were to precede *professors*, as in (7.7), these domains would be larger.

(7.7) I [$_{VP1}$ saw [{{books, wrote*i* $_{VP}$}, that $_{S}$} professors*i* $_{NP}$]]

Regardless of the internal ordering of relative clause constituents in (7.7), a four-word viewing window linking *saw* to *professors* would be required for the PCD of VP1 and for lexical processing.

(7.6) is also optimal for MaOP. It avoids misassignments of relative clause constituents to VP1, or their unassignment in the event that the co-occurrence requirements of the matrix verb are not met. For example, if *books* is the first relative clause constituent to be parsed in (7.7) it would be misassigned to VP1 as a direct object of *saw*; if the complementizer *that* occurred first, a clausal complement of V could be misassigned to VP, instead of assigning and attaching a relative clause to NP (see Hawkins 1994: 327). If the matrix verb were *greeted* rather than *saw* (i.e., *I* [*greeted* [[*books wrotei that*] *professorsi*]] and *I* [*greeted* [[*that wrotei books*] *professorsi*]]) then *books* and *that* would be unassigned to their dominating nodes online rather than misassigned since it is not normal to ?*greet books*, and *greet* does not take a sentential complement. Either way, misassignments or unassignments would be extensive for RelN orderings that are contained within head-initial phrases.

NRel is therefore optimal both for MiD and for MaOP in verb-early languages, and it is to this that I attribute its almost exceptionless co-occurrence with VO in typological samples (Lehmann 1984; Dryer 1992).

Languages with containing head-final phrases, on the other hand, cannot satisfy both MiD and MaOP simultaneously, since the two preferences conflict in this language type. Consider a misassignment structure in Japanese that has

been shown by Clancy et al. (1986) to result in processing difficulty (see also Antinucci et al. 1979):

(7.8) Zoo-ga [[[[kirin- taoshi-ta*(i)* $_{VP1}$] $_{S}$] shika-o*i* $_{NP}$] nade-ta $_{VP2}$]
elephant-NOM giraffe-ACC knocked down deer-ACC patted
'The elephant patted the deer that knocked down the giraffe.'

The online parse first misrecognizes *Zoo-ga* as the nominative subject of *taoshi-ta* within a simple main clause analysis of *Zoo-ga kirin-o taoshi-ta* (with the meaning 'the elephant knocked down the giraffe') and then reassigns words and phrases in accordance with the tree structure shown in brackets in (7.8). The online misassignments are quite severe in this example, as shown by the quantified complexity criteria for garden paths in Hawkins (2004: 51–5, 207).

In the event that the matrix subject NP (*zoo-ga*) does not match the co-occurrence requirements of the (lower) verb (*taoshi-ta*), then the words, phrases, and relations that are misassigned in (7.8) will be temporarily unassigned instead (see Hawkins 2004). Either way misassignments or unassignments are extensive for RelN in this structural position. But (7.8) is good for MiD. The PCD for the matrix VP2 permits processing of its two ICs (the NP headed by *shika-o* and the verb *nade-ta*) in the same two-word viewing window that renders the mirror-image head-initial VP1 optimal in (7.6) (IC-to-word ratio = 2/2, i.e., 100%). The lexical relations between the matrix verb and the NP can be recognized in this same optimal viewing window.

By contrast, if Japanese were to have an NRel construction co-occurring with OV (as in German relative clauses) and as shown in (7.9), the PCD for VP2 would be as inefficient as RelN is with VO in (7.7): *shika-o kirin-o taoshi-ta nade-ta* (2/4 = 50%). The domains for processing the lexical relations between the matrix verb (*nade-ta*) and NP would be larger as well since the head of NP (*shika-o*) is significantly further from the verb than in (7.8).

(7.9) Zoo-ga [[$_{NP}$ shika-o*i* [[kirin-o taoshi-ta*i* $_{VP1}$] $_{S}$]] nade-ta $_{VP2}$]

On the other hand, the misassignments and unassignments of the RelN structure in (7.9) would be avoided by the early positioning of the head noun in this NRel structure. The early construction of the appropriate NP and S to which *kirin-o* and *taoshi-ta* are to be attached means that these words can be correctly attached as they are processed, and this improves their Online Property to Ultimate Property ratios in accordance with MaOP (see §2.3 and Hawkins 2004: 208), making their online processing more efficient.

Hence whereas MiD and MaOP define the same preferences for head-initial (VO) languages, resulting in almost exceptionless NRel, head-final languages

must choose (for their basic word orders at least) either minimal domains at the expense of mis-/unassignments (i.e., no maximal online processing), or mis-/unassignment avoidance (through adherence to MaOP) at the expense of minimal domains. This is shown in (7.10).

(7.10)

	MiD	*MaOP*
VO & NRel	+	+
VO & RelN	−	−
OV & RelN	+	−
OV & NRel	−	+

A processing approach to this distribution can motivate the absence of the VO & RelN combination, therefore, and also the co-occurrence of either RelN or NRel with OV. Neither is fully optimal in OV languages. Support for this competing motivation explanation comes from more fine-tuned predictions that we can make regarding when the NRel variant will be selected and when RelN, in OV languages. In what follows we will see further support for the two contributing principles, MiD and MaOP, in the different kinds of subtypes of OV languages that choose NRel versus RelN respectively.

First, notice that the more rigidly head-final a language is (cf. Greenberg 1963), the more containing phrases there will be that can dominate NP and that are also head-final. Consequently there will be more containing phrases whose IC-to-word ratios are improved by having the NP head-final as well, which means that RelN should be favored. RelN is therefore strongly motivated for rigid OV languages like Japanese, by MiD, but is less motivated for non-rigid OV languages like German, in which the verb often precedes NP and so prefers a head-initial NRel. This prediction appears to be borne out, as seen in the following data from *WALS* (Dryer 2005d,e; Dryer with Gensler 2005):

(7.11)

WALS	*Rel–Noun*	*Noun–Rel or Mixed/Correlative/Other*
Rigid sov {o, x} v	50% (17)	50% (17)
Non-rigid sov ovx	0% (0)	100% (17)

The rigid OV languages are evenly split between a prenominal relative clause ordering on the one hand, and on the other hand either exclusively postnominal relatives, mixed pre- and postnominal, a correlative as found in Hindi, or some other 'head-internal' relative clause type to which I return in §7.4. Non-rigid OV languages (like German and Persian) are exclusively of the NRel/mixed/correlative/other type, with the great majority being just NRel. This correlation between rigid OV and RelN, and non-rigid OV and

NRel, supports the generalization that NPs are head-final in proportion to the head finality of their containing phrases, in accordance with MiD.

The data of (7.11) reveal a further pattern, however. In the competition between MiD and MaOP, the latter is the stronger principle. None of the non-rigid OV languages have RelN—i.e., MaOP wins out every time in these languages despite the fact that at least some phrases containing NP in non-rigid OV languages will be head-final, which should result in RelN by MiD. In the rigid OV languages a full half are also pulled away from MiD's preferred RelN. Overall only one third of OV languages (17/51) in *WALS* have the MiD-preferred RelN, all of them rigid OV, while two thirds of OV languages (34/51) have the NRel structure favored by MaOP, either as the only strategy or in combination with some other. I shall return in §9.3 to a discussion of why MaOP is plausibly the stronger principle in this competition.

Second, we can now predict that whichever of the two non-optimal variants in OV languages (NRel or RelN) is conventionalized as a basic order, then non-basic orders, morphosyntactic devices, or other structural responses should also be conventionalized that can realize the suppressed preference to some extent at least, and in proportion to the degree of preference that this latter defines. In other words, MiD and MaOP are universal processing preferences, and even though a given basic word order cannot incorporate both simultaneously in combination with OV, and only one can be selected, the other cannot be ignored altogether. The basic NRel & OV orders of German, motivated by MaOP, tolerate a lengthy VP PCD that is at variance with MiD only to a limited degree and beyond this degree MiD asserts itself through extrapositions from NP (see §3.3) in order to minimize the processing domains for VP. Some OV languages have conventionalized a basic 'adjoined relative clause' corresponding to these extrapositions of German and they routinely separate the head noun from its relative, e.g., Diyari (Dryer 2005d). These structures, exemplified by (7.12) from Austin (1981), can be regularly advantageous for MiD, as long as distances between the nominal head and the relative are not too great, and at the same time the initial positioning of the head noun is good for MaOP.

(7.12) ŋani wila-ni yata-la ŋana-yi [yinda-nini] (Diyari)
1SG.SUBJ woman-LOC speak-FUT AUX-PRES cry-REL.DS
'I'll talk to the woman who is crying.'

Similarly RelN in rigid OV languages, motivated by MiD, tolerates misassignments and unassignments only to a certain extent, allowing MaOP to assert itself in numerous otherwise unexplained typological properties. Christian Lehmann (1984: 168–73) points out that the syntactic form and content of prenominal relatives is much more restricted than that of their postnominal

counterparts. Prenominal relatives are more strongly nominalized, with more participial and non-finite verb forms (as in Dravidian, Lehmann 1984: 50–2, cf. Hawkins 1994: 387–92), with more deletion of arguments, and more obligatory removal of modal, tense, and aspect markers, and so on. They generally admit only restrictive and not appositive interpretations. All of this makes them shorter, and reduces the temporal duration and number of online unassignments and also misassignments by the criteria defined in Hawkins (2004: 52–4), i.e., it reduces the negative consequences for MaOP within a structure favored by MiD.

Third, we make predictions for performance selections of relative clause options in the different language types. The worse the relevant structures are by the quantitative metrics associated with MiD and MaOP in Hawkins (2004), the less frequently they should be selected. See §3.3 for Extraposition from NP data from German corpora. Cantonese preposes a RelN from a direct object position after V to a clause-initial position favored by MiD in proportion to the complexity of the relative clause and the strain on the PCD for a postverbal VP in this typologically mixed structure (see Matthews and Yeung 2001 and (5.3) in §5.1). Degrees of misassignment and unassignment in OV languages should also result in fewer selections for the relevant structures in OV languages, just as they do in VO (Hawkins 2004: 55–61).

7.4 Related relative clause asymmetries

There are a number of relative clause types that are generally referred to as 'head-internal' and that are strongly characteristic of OV and not of VO languages; cf. Downing (1978), Keenan (1985), Cole (1987) for early discussions of this feature of OV languages. The head noun does not stand on the left or right periphery of a relative clause sister and is not external to it in these languages, but remains as a constituent within the 'relative' clause itself syntactically, while receiving the semantic interpretation of an extracted English-type head plus relative—i.e., the subordinate clause serves to delimit the reference of the semantic head by making a truth-conditional claim within this clause that is predicated of the head. An example cited in Dryer (2005d) from Couro and Langdon (1975) is the following from Mesa Grande Diegueño:

(7.13) ['ehatt gaat kw-akewii]=ve=ch nye-chuukuw (Mesa Grande Diegueño)
dog cat REL.SUBJ-chase=DEF=SUBJ 1OBJ-bite

'The dog that chased the cat bit me.'

The subject relative prefix on the verb (*kw-*) signals that *'ehatt* ('dog') is the semantic head, as shown in the gloss. The absence of this prefix would signal that *gaat* ('cat') is the head, i.e., with subject *'ehatt* and object *gaat* occurring in the same clause-internal SOV order but with the meaning 'the cat that the dog chased bit me.'

A head-internal variant of (7.13), in which the semantic head is actually fronted syntactically to the left of the subordinate clause, but still internal to it and not external by various syntactic tests, is found in Yavapai. The following example is from Lehmann (1984: 120) who draws on Kendall (1974):

(7.14)	ʔna-č	[ʔkwa-m	čan	qwaqata	čičkyat-i]	ʔ-wal-kəm
	I-NOM	knife-INSTR	John	meat	cut-PRET	1-seek-IMPF

(Yavapai)

'I seek the knife with which John cut the meat.'

For detailed formal analyses of the syntax of head-internal relatives such as these, see especially Cole (1987) and Basilico (1996).

Psycholinguistically the fronting of the semantic head *ʔkwa-m* in (7.14) is interesting because it supports the preference for head noun fillers (and for other fillers, like Wh-words) to precede their gaps or subcategorizers, i.e., it supports the Fillers First principle of Hawkins (2004: 203–5), which goes back to Janet Fodor (1983). This principle is normally realized by both fronting and extracting the filler from the clause in which its gap or subcategorizer occurs, e.g., into a clause-external complementizer position within generative analyses. But we see in Yavapai that the same psycholinguistically motivated movement (cf. the arguments in Hawkins 2004) can be realized by just fronting the filler, without extracting it.

Yet another variant of the head-internal type is the correlative of Hindi (Keenan 1985), which I referred to in the last section. Dryer (2005d) points out that in a correlative the (semantic) head occurs inside its (relative) clause, as in the Mesa Grande Diegueño example, but this clause is now outside the main clause that predicates something of the semantic head, and there is an anaphoric NP in the main clause that refers to the semantic head and relative. He gives the following example from Bambara (taken from Bird and Kante 1976):

(7.15)	[muso	min	taara],	o	ye	fini	san	(Bambara)
	woman	REL	leave	3SG	PAST	cloth	buy	

'the woman who left bought the cloth'

Keenan's (1985) similar example from Hindi is given in (7.16):

(7.16) Jis a:dmi ka kutta bema:r hai, us a:dmi ko
CORREL man GEN dog sick is that man ACC
mai ne dekha (Hindi)
I ERG saw
'I saw the man whose dog is sick.' (lit. 'which man's dog is sick, that man I saw')

Dryer (2005e) confirms the crucial typological point in this context, that head-internal relatives are strongly characteristic of SOV languages. In fact, he states that he knows of only one solid exception to it, the verb-initial language Kutenai which has head-internal relatives (data from his own fieldwork). He also points out that in Tukang Besi head-internal relatives are a minority variant alongside NRel (cf. Donohue 1999).

This strong association between the different subtypes of head-internal relatives and OV calls for an explanation. The natural hypothesis to pursue is that neither RelN nor NRel is fully optimal in OV languages; recall (7.10) in the last section. Head-internal relatives provide a compromise that affords advantages of both domain minimization and un-/misassignment avoidance, without having to choose exclusively in favor of one or the other. So, the verbal head of the subordinate clause in Yavapai (7.14), *čičkyat-i* ('cut-PRET') is adjacent to the matrix verb *ʔ-wal-kəm* ('1-seek-IMPF'), which permits the subordinate clause to be constructed and immediately attached to the matrix VP, which is good for MiD. The fronting of the semantic head *ʔkwa-m* ('knife') is motivated, as we have mentioned, by Fillers First, which in turn is motivated by MaOP. Similarly, in Mesa Grande Diegueño (7.13) the verbal head of the subordinate clause *akewii* ('chase') immediately precedes the functional heads *ve* and *ch*, which can be argued to project to and construct NP (cf. §6.2.1 and Basilico 1986), which in turn immediately precede the matrix verb *nye-chuukuw* ('1OBJ-bite'), all of which is optimal for MiD. Meanwhile there are no gaps in this structure and no misassignments or unassignments resulting from their processing, which is good for MaOP. The exact processing cost of converting a syntactic clause-internal head into a semantic representation in which the head is semantically clause-external and modified by the remainder of its clause has still to be determined.

There is still more to say about the precise parsing of these fascinating structures, therefore. What is significant in this context is that their typological distribution provides a further contrast between OV and VO languages that can be linked, I would argue, to the striking split between RelN and NRel options summarized in (7.10) and (7.11) above. Neither RelN nor NRel is optimal for both MiD and MaOP in OV languages. The result is considerable variation between alternative relative clause structures that prioritize one or

the other efficiency principle in these languages, including head-internal relatives that are not technically syntactic head-modifier structures and that appear to offer a compromise between both principles.

7.5 Complementizer ordering asymmetries in VO and OV languages

The explanation proposed here for the asymmetry in head noun and relative clause orderings across languages can be supported by a very similar asymmetry in the positioning of complementizers and their sister clauses. VO languages are exceptionlessly complementizer-initial, as in English (*I believe* [*that the boy knows the answer*]), i.e., there appear to be no VO languages with SComp (Hawkins 1990). OV languages are either complementizer-initial (Persian) or complementizer-final (Japanese), and the same general explanation can be given for this as I have given for relative clauses. I give the summary chart in (7.17):

(7.17)

	MiD	*MaOP*
VO & CompS	+	+
VO & SComp	−	−
OV & SComp	+	−
OV & CompS	−	+

An OV & SComp combination is illustrated for Japanese in (7.18). The OV & CompS combination is found in Persian and was illustrated in §5.8.1. Generally this structure involves extraposition of CompS to the right of V.

(7.18) a. Mary ga [[kinoo John ga kekkonsi-ta to $_S$]
Mary NOM yesterday John NOM married that
it-ta $_{VP}$] (Japanese)
said
'Mary said that John got married yesterday.'

b. [kinoo John ga kekkonsi-ta to $_S$] Mary ga [it-ta $_{VP}$]

Final complementizers are completely absent from head-initial languages, it seems. The use of an initial complementizer, as in (7.17a), is motivated both by MiD and MaOP.

(7.19) a. I [$_{VP}$ believe [$_S$ that the boy knows the answer]]

The two ICs of VP, V and S, can be recognized in a two-word viewing window (*believe that*), giving an IC-to-word ratio of 2/2 = 100%. Recognition of the sentence complement co-occurrence for the verb *believe* is also accomplished

in this minimal domain. (7.19a) is also optimal for MaOP since it provides immediate information about the subordinate clause boundary and avoids the misassignment or unassignment of NPs within it.

A final complementizer, as in (7.19b), would produce a much longer CRD for VP. If this complementizer is required in order to construct the subordinate clause sister to V, then the CRD for VP proceeds from *believe* to *that* and will have a low IC-to-word ratio of 2/7 = 29%. If this subordinate clause can be recognized prior to *that*, the ratio for VP will still be significantly lower than the 100% of (7.19a), and the domain for processing the lexical relation between the verb and its complement will be longer as well.

(7.19) b. I [$_{VP}$ believe [the boy knows the answer that $_{S}$]]

The structure of (7.19b) also results in the misassignment of *the boy* as a direct object of *believe*, or its unassignment in the event that *the boy* does not match the co-occurrence requirements of a matrix verb such as *realize*. Neither MaOP nor MiD motivate final complementizers, therefore, and both motivate the initial complementizer of (7.19a).

In head-final languages a final complementizer as in (7.20a) produces minimal CRDs for VP and minimal domains for lexical processing, but also potential misassignments or unassignments online prior to recognition of the complementizer.

(7.20) a. I [[the boy the answer knows that $_{S}$] believe $_{VP}$]

If any case marking on *the boy* is compatible with matrix clause attachment and combination with *I*, there will be a misassignment. If it is not and if any case marking does not unambiguously construct a clause and a clause boundary (by Grandmother Node Construction; see §5.8.2 and Hawkins 1994: 361) there will be an unassignment delay in subordinate clause recognition and in attachment to the matrix VP. If the subordinate clause is preposed, as in (7.20b), in order to provide good IC-to-word ratios for the matrix clause, the resulting bottom-up parse will delay recognition of S as a subordinate clause rather than a matrix one, leaving this decision unassigned until the final complementizer (cf. Frazier and Rayner 1988; Hawkins 1994).

(7.20) b. [the boy the answer knows that $_{S}$] I [believe $_{VP}$]

Either way there will be misassignments or unassignments in (7.20a) and unassignments in (7.20b). An initial complementizer as in (7.20c) avoids these mis- or unassignments, but at the expense of a long CRD for VP and a long domain for processing the complement relation between S and *believe*.

(7.20) c. I [[$_{S}$ that the boy the answer knows] believe $_{VP}$]

Once again, MiD and MaOP cannot both be satisfied simultaneously in SOV languages. (7.20a) and (7.20b) are good for MiD and bad for MaOP; (7.20c) is bad for MiD and good for MaOP.

As in the case of relative clause typology, more fine-tuned predictions can be made by this approach. First, the more containing head-final phrases there are, the greater will be the MiD motivation for SComp. Hence SComp should be strongly favored in rigid OV languages like Japanese, less so in non-rigid OV languages like Persian and German. This seems to be the case, though systematic testing is still required.

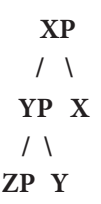
Second, non-basic word orders, morphosyntactic and other responses are predicted for languages and structures in which one or the other principle is suppressed. Extraposition of a direct object CompS to the right of V is obligatory in German and in many other OV and CompS languages (Hawkins 1990, 1994: 302–8); see §5.8.1. The initial complementizer is motivated by MaOP at the expense of MiD, and the latter then reasserts itself in extrapositions, which apply more productively to complex finite embeddings than to shorter infinitival embeddings (Hawkins 1990, 1994, and also Hawkins 1986 for detailed discussion of German). Rich case marking is productive in combination with OV and can avoid many misassignments and unassignments in the SComp structures that are motivated by MiD at the expense of MaOP (cf. again §5.8.1).

Moreover, not all CompS structures are good for MiD in all VO languages and environments. In VOS languages such as Malagasy, a direct object CompS would create a long VP prior to the subject and an inefficient CRD for the containing S and it is accordingly obligatorily postposed to the right (Hawkins 1994: 298–9). Clausal subjects in SVO languages also create lengthy clausal domains and are predicted to undergo regular extraposition to the right, in proportion to the improvements in their IC-to-word ratios (cf. Erdmann 1988, Hawkins 1994: 190–6, and §2.5.1 for data from English that support this).

7.6 Fewer phrasal constructors and more affixes in OV languages

I pointed out in §5.8.3 that there are some typological patterns in SOV languages that have not yet been explained grammatically, and which suggest that the small processing delay between the construction of YP and its attachment to a mother XP in head-final structures has grammatical consequences. Recall (5.2) from §5.1 repeated here as (7.21):

(7.21)



Head-final

The parser has to wait one word after YP has been constructed at Y, until X constructs the XP to which YP can be attached—i.e., there is a one-word processing delay between the construction of YP and its attachment to XP.

Fewer free-standing X words follow Y in OV languages. When a preposition precedes NP as in English, the preposition is typically a free-standing word. But when P follows NP in an OV language there are fewer free-standing postpositions, even when the language has some, and what corresponds to a preposition in VO will often surface as a suffix on nouns; see Tsunoda et al. (1995) and §5.8.3 for further details. In general there are many more X affixes on Y in OV languages—e.g., derivational and inflectional suffixes on nouns and on verbs—and these X affixes can then construct YP and XP simultaneously at Y, the former through Mother Node Construction (cf. (5.8) in §5.1), the latter through Grandmother Node Construction (Hawkins 1994: 361–6). This is an efficient solution to the one-word processing delay since construction and attachment of YP can now take place at one and the same word, Y, just as YP can be immediately attached to an already constructed XP in the mirror-image head-initial structure (cf. (5.1) in §5.1).

Consider the precise form of sentence complementizers and their distribution. In head-initial languages these are typically free-standing words that construct subordinate clauses. In head-final languages such words correspond more commonly to participial and other subordinate clause indicators affixed to verbs and they are much less productive as independent words. Dryer (2009) gives relevant figures from his typological sample. Of the languages that have free-standing complementizers, 74 percent (140) occur initially in CP within VO languages. Just 14 percent (27) occur finally in CP within OV languages, i.e., in structure (7.21). The remaining 12 percent (22) also occur initially, in a CP within an OV basic order. The morphemes that correspond to free-standing complementizers in OV languages are frequently affixes, therefore, and these affixes indicate subordinate clause status. Adding affixes to verbs means that both S and its subordinate status are constructed simultaneously on the last word of the subordinate clause. See Hawkins (1994: 387–93) for detailed exemplification of several different types of subordinating affixes in OV grammars and their corresponding parsing routines. These figures from Dryer (2009) are summarized in (7.22):

(7.22) *Complementizer Words* (% of total languages with complementizers)

VO	Initial	74%	(140)
	Final	0%	(0)
OV	Initial	12%	(22)
	Final	14%	(27)

I also pointed out more generally in §5.8.3 that head-final languages typically lack single-word categories that can construct mother nodes as an addition or alternative to the syntactic heads with which they combine. Corresponding to the prepositions of VO languages there are regularly suffixes on nouns in OV. Corresponding to the initial complementizers of VO languages there are regularly suffixes on verbs in OV. Corresponding to the productive definite articles of VO languages, this category is simply not typically grammaticalized in OV languages.

I argued in §5.8.3 that either N or a freestanding Art can construct NP immediately on its left periphery and provide efficient and minimal phrasal combination domains (PCDs) in VO languages (§2.1). Art-initial should be especially favored when N is not initial in the NP, e.g., when adjectives can precede N as in English; see (5.35) repeated here as (7.23):

(7.23) $[_{VP}$ V $[_{NP}$ N . . . Art . . .]
$[_{VP}$ V $[_{NP}$ Art . . . N . . .]
|------|

In OV languages any additional constructor of NP will lengthen these processing domains, whether it follows or precedes N, by constructing the NP early and extending the processing time from the construction of NP to the processing of V. Additional constructors of NP are therefore inefficient in OV orders, as was shown in (5.36), given here as (7.24):

(7.24) [[. . . N . . . Art $_{NP}$] V $_{VP}$]
[[. . . Art . . . N $_{NP}$] V $_{VP}$]
|----------|

Definite articles are found predominantly in VO rather than OV languages, therefore. This supports the processing efficiency point I am making here. The data of (5.37) given here as (7.25) are taken from *WALS* (Dryer 2005c) and compare frequencies for VO and (rigid) OV languages in which there are distinct words for definite articles and for demonstrative determiners (see also §6.2.3):

(7.25)	*Def word distinct from Dem*	*No definite article*	[*WALS* data]
Rigid OV	19% (6)	81% (26)	
VO	58% (62)	42% (44)	

7.7 Complexity of accompanying arguments in OV vs VO languages

Related to the patterns in §7.6 is an observation that has been made regarding the grammar and the processing of verb phrases in SOV languages, by Nichols, Peterson, and Barnes (2004) and Ueno and Polinsky (2009) respectively. OV languages have an 'intransitive bias' compared with VO languages. They prefer to have fewer preceding left-branching arguments before the verb and to use more intransitive one-place predicates, compared to transitive two-place predicates, than VO languages. Ueno and Polinsky (2009) established this by examining corpora and usage statistics for adults and children speaking Japanese and Turkish (OV) versus English and Spanish (VO). One-place predicates (i.e., SV structures) were significantly preferred over two-place predicates (i.e., SOV) in OV compared to VO languages. At a grammatical level Nichols et al. (2004) note that OV languages will regularly have an intransitive bias morphologically and derive transitive verbs from intransitives, rather than deriving intransitives from basic transitives, as in many other languages.

The consequence of this contrast is that verb phrases with a left-branching direct object as in (7.21) above—i.e., with YP preceding X within XP—are less common than their mirror-image counterparts in VO languages with YP following X ((5.1) in §5.1). This accords well with the observations made in §7.3 and §5.8.3 about left-branching structures generally and their structural and semantic limitations, and it supports the point made in §7.2 about the importance of the verb in the online processing of clauses. Ueno and Polinsky see their data as supporting the fact that it is beneficial to encounter the verb as early as possible. The strength of this preference is clearly not sufficient to pull the verb leftwards in all OV languages, and override the (Minimal Domain) benefits of having all or most heads to the right, matching the comparable benefits of heads to the left in VO languages (see §5.1 and §7.1). But it is a contrary preference nonetheless, and provides evidence for a tension within OV languages that is interesting to think about in a diachronic context and as a contributing motivation for change (see §4.4). There are competing efficiencies within all languages that motivate the existence of variants, each of which possesses some but not all processing advantages (cf. §9.3). For OV languages rich case marking combined with narrow thematic role assignments permits some aspects of argument structure and argument interpretation to be accessed early in advance of the verb (see §7.2), but the greater frequency of intransitives suggests that earlier access to the verb is still a real and competing preference in these languages. The reader is referred to Nichols et al. (2004) and Ueno and Polinsky (2009) for full grammatical and corpus details.

7.8 The ordering of obliques in OV vs VO languages

There is an interesting asymmetry between OV and VO languages with respect to the positioning of so-called 'oblique phrases.' This asymmetry has not received the attention it deserves in the grammatical literature, but it is addressed in Dryer with Gensler (2005) and it can be given an explanation in processing-efficiency terms. An oblique phrase (following Dryer with Gensler 2005) is a syntactic NP, DP, or PP that functions most frequently as an adjunct of a verb, such as *with the key* in (7.26), or less frequently as a non-subject, non-direct object, and (in languages that make the distinction) non-indirect object phrase that is nonetheless a complement or argument of a verb, like *on the desk* in (7.27):

(7.26) John opened [the cupboard] [with the key]

(7.27) John put [the book] [on the desk]
V O X

The different ordering possibilities for X in VO and OV languages and their respective quantities in Dryer with Gensler's chapter are given in Table 7.2. VO languages show considerable consistency. Of the three logically possible orderings for V, O, and X, a full 98 percent of those with a basic order for these categories have X in rightmost position (VOX), none have X intervening between V and O (VXO), and just 2 percent (three entries, all of them dialects of Chinese) have initial X (XVO).

The VOX pattern is found in English. For XVO Dryer with Gensler give the following example from Mandarin:

(7.28) Mandarin Chinese (Li and Thompson 1981: 390)

tāmen	[zài	fángzi-hòumian]	xiūli	diànshìjī
they	at	house-behind	repair	television
		X	V	O

'They repair televisions behind their house.'

For twenty-two languages with VO (10 percent of the total) more than one ordering of X is productive and none could be considered basic or dominant. See Dryer (2005f) for the criteria by which a decision on basicness and dominance is reached in the typological samples of *WALS*. Essentially this decision is based on frequency, given the current state of information available for most languages. The dominant order is either the exclusive order in the event that there are no competitors, or it is at least twice as frequent as the next most frequent competitors when one or more alternative orders exist.

OV languages reveal a very different picture. All three logically possible basic orders of V, O, and X are now productive, occurring with the following

TABLE 7.2 {V, O, X} orders in the *WALS* Map 84 (Dryer with Gensler 2005: 342–5)

VO	1.	VOX	189 lgs	(98% of basic orders)
	2.	VXO	0	
	3.	XVO	3	(2%)
	4.	More than one of 1–3, none dominant	22	(10% of VO)
OV	5.	XOV	45	(43% of basic orders)
	6.	OXV	23	(22%)
	7.	OVX	37	(35%)
	8.	More than one of 5–7, none dominant	39	(27% of OV)

relative frequencies: XOV (43%) > OVX (35%) > OXV (22%). And the number of OV languages without a single basic order is significantly higher than the 10% figure for VO, at 27 percent, as shown in Table 7.2.

Examples of the XOV, OXV, and OVX types given in Dryer with Gensler (2005) are respectively from Slave, Nagatman, and Kairiru:

(7.29) Slave (Athapaskan, Rice 1989: 997)

t'eere	[deno	gha]	?erákee?ee	wihsį
girl	REFL.mother	for	parka	3.made
	X		O	V

'The girl made a parka for her mother.'

(7.30) Nagatman (isolate, Papua New Guinea, Campbell and Campbell 1987: 8)

[mo	me]	[ke	na]	hohui-në-taya
fish	OBJ	these	with	look.for-1.SUBJ-3PL.OBJ
O		X		V

'We look for fish with these.'

(7.31) Kairiru (Oceanic, Papua New Guinea, Wivell 1981: 151)

ei	porritamiok	a-pik	[gege-ɪ	nat	nai]
3SG	axe	3SG-take	from-3SG	child	that
	O	V	X		

'He/She took the axe from that child.'

7.8.1 The patterns

The data of Table 7.2 reveal a number of patterns, summarized in Table 7.3 using Roman numerals in square brackets to identify each one. Pattern [I] has already been commented on: the consistency of VO vs OV variability.

Pattern [II] involves a strong preference for the adjacency of verb and direct object ('verb–object bonding' in the terminology of Tomlin 1986). 100 percent of VO languages have the verb and object adjacent (i.e., there are no VXO orders); 78 percent of OV languages also show this preference (all but OXV). Overall 274 out of 297 languages with basic orders have V and O adjacent to one another, or 92 percent.

A further pattern, found in 87 percent of all basic orders, is object and X on the same side of V, i.e., Pattern [III]. 98 percent of VO languages exemplify this (all but XVO), as do 65 percent of OV (all but OVX).

A final pattern, [IV], found in 84 percent of all basic orders, is O before X, i.e., the direct object precedes the oblique phrase. 98 percent of VO languages exemplify this, as do 57 percent of OV.

These patterns are summarized in Table 7.3.

TABLE 7.3 Basic patterns in Table 7.2

[I] *VO consistency vs OV variability*
VO lgs: 98% of basic orders are VOX; 10% have no dominant order
OV lgs: all three basic orders productive; 27% have no dominant order

[II] *V&O adjacency*
VO lgs: 100% have adjacent V&O (no VXO)
OV lgs: 78% have adjacent V&O (all but OXV)
Across all lgs, adjacent basic orders of V&O = 92% (274/297)

[III] *O and X on same side of V*
VO lgs: 98% have O and X on same side (all but XVO)
OV lgs: 65% have O and X on same side (all but OVX)
Across all lgs, basic orders of O and X on same side = 87% (257/297)

[IV] *O before X*
VO lgs: 98% have OX, 2% XO (Chinese dialects only)
OV lgs: 57% have OX (60/105), 43% XO (45/105)
Across all lgs, basic orders of O before X = 84% (249/297)

We can now compare these conventionalized ordering patterns for V, O, and X across languages with relevant performance data within languages involving these and similar categories and apply our efficiency principles of Chapter 2 in order to try and explain the grammatical patterns.

7.8.2 Verb and direct object adjacency [Pattern II]

Consider some corpus data involving the VOX/VXO alternation in English, i.e., with structures containing a postverbal direct object and a PP as in (7.26) and (7.27) above. It has been observed traditionally (by, e.g., Ross 1967) that the VXO departure from the basic VOX is characteristic of direct objects that are particularly 'heavy,' giving rise to the label 'Heavy NP Shift,' e.g., *John opened* [*with the key*] [*the cupboard that had been closed for centuries*]. Minimize Domains (2.1) predicts that it is the relative weight of the NP versus the PP, rather than the absolute size or complexity of NP, that should condition the shift. This has since been confirmed in production experiments conducted by Stallings (1998) and Stallings and MacDonald (2011), in corpus data presented in Hawkins (1994), and in both corpus and experimental data given in Wasow (2002).

There are just 22/480 instances of VXO shifting (5 percent) in Hawkins's (1994: 183) data, repeated here in (7.32), the incidence of which increases in the bottom line of the left-hand columns as the weight difference of O > X increases. English direct objects are abbreviated here as oNP, and (prepositional) PPs as mPP (with the prepositional head constructing the mother PP at the left edge). The [V oNP mPP] basic order of English can be seen as a conventionalization of domain minimization preferences for the processing of phrasal combination domains (PCDs) (5.12) and lexical domains (LDs) (5.15); cf. §4.3 and §5.2. Direct object NPs are significantly shorter in general than postverbal PPs; see (7.36) below. The same short-before-long preference evident in the relative ordering of two PPs (see §5.2) will therefore favor oNP before mPP. In addition, oNPs are always complements, whereas PPs are more often *not* in a lexical relation with V, i.e., there are more VO lexical processing domains than VX domains. Direct object adjacency to V is preferred for both PCD and LD processing, therefore.

(7.32)

	oNP>mPP by			oNP=mPP	mPP>oNP by		
n = 480	5+	3–4	1–2		1–2	3–7	8+
[V oNP mPP]	0% (0)	43% (6)	95% (38)	100% (68)	99% (209)	99% (103)	100% (34)
[V mPP oNP]	100% (9)	57% (8)	5% (2)	0% (0)	1% (2)	1% (1)	0% (0)

The preferred adjacency of V and oNP in the top line can be seen overwhelmingly when mPP>oNP in the right-hand columns, when weight differences are equal, and even when oNP>mPP by 1–2 words. Heavy NP Shift to [V mPP oNP] in the second line becomes significant only when oNP>mPP by larger (3–4 and 5+) weight differences.

Wasow (1997, 2002) gives further corpus data showing how the different kinds of V-PP relations impact shifting ratios to [V PP NP] in English, in addition to weight. He distinguishes:

– 'opaque collocations' between V and PP, defined as lexical combinations whose meanings are non-compositional and require a processing domain that includes both V and PP for their appropriate meaning assignment, e.g., *take into account NP*;
– 'transparent collocations,' i.e., those that should be comprehensible to a speaker who knows the literal meaning of each of the words in it, but has never before encountered them in this combination, e.g., *bring to someone's attention NP*; and
– 'non-collocations,' i.e., compositional combinations that are not lexically listed, e.g., *take to the library NP.*

Shifting rates for sample instances of these three types in his data are shown in (7.33):

(7.33) opaque collocations = 60%
transparent collocations = 47%
non-collocations = 15%

Wasow's 'collocations' correspond to what we are calling a 'lexical combination' here; see §5.2. The transparency or opacity of the collocation seems to reflect whether there is an additional dependency between V and PP, as this term was defined in §2.1 (see Hawkins 2004: 18–25). His data indicate that V-PP adjacency is preferred in proportion to the number of lexical combination and lexical dependency relations holding between V and PP. When one adds this finding to the clear weight effect in (7.32) the prediction made by MiD (3.10) in §2.1 is supported: the shifting preference to [V PP NP] is proportional to the number of lexical and syntactic relations whose processing domains can be minimized and to the extent of the minimization preference for a particular order within different processing domains (see further §9.2).

Consider now the corpus data from Japanese involving the mirror-image alternation of XOV/OXV, i.e., sentence pairs such as (7.34a) and (b) which we considered in §5.3:

(7.34) a. (Tanaka ga) [[Hanako kara $_{PP}$] [sono hon o $_{NP}$] katta $_{VP}$]
Tanaka NOM Hanako from that book ACC bought
X O V
(Japanese)
'Tanako bought that book from Hanako.'

b. (Tanaka ga) [[sono hon o $_{NP}$] [Hanako kara $_{PP}$] katta $_{VP}$]
O X V

In a head-final VP like this, long-before-short phrases will generally provide minimal phrasal combination domains, since the categories that construct mother phrases (V, P, Comp, case particles, etc.) are on the right. If the direct object is a long complement clause headed by the complementizer *to*, for example, as in (7.18) above repeated here, and if this complement intervenes between the subject *Mary ga* and the verb *it-ta* as it does in (7.18a), then the matrix PCD will proceed from the subject to the verb and will be very long. In (7.18b), on the other hand, the distance from *Mary ga* to the verb is very short, and the added distance from the complementizer *to* to the verb compared with (7.18a) is not considerable, making (7.18b) a much more efficient structure overall.

(7.18) a. Mary ga [[kinoo John ga kekkonsi-ta to $_S$] it-ta $_{VP}$]
Mary NOM yesterday John NOM married that said
(Japanese)
'Mary said that John got married yesterday.'

b. [kinoo John ga kekkonsi-ta to $_S$] Mary ga [it-ta $_{VP}$]

By similar reasoning, a preference for (7.34a) or (7.34b) is predicted in proportion to the relative weight difference between the object phrase (NPo) and the head-final PP (PPm). The longer NPo or PPm is predicted to precede the shorter phrase.

The following Japanese data were reported in Hawkins (1994: 152) and were collected by Kaoru Horie (see (5.22) in §5.3):

(7.35)

	NPo>PPm by			NPo=PPm	PPm>NPo by		
n = 244	5+	3–4	1–2		1–2	3–8	9+
[PPm NPo V]	21% (3)	50% (5)	62% (18)	66% (60)	80% (48)	84% (26)	100% (9)
[NPo PPm V]	79% (11)	50% (5)	38% (11)	34% (31)	20% (12)	16% (5)	0% (0)

There is much more variability in Japanese than in the English (7.32), but a clear mirror-image pattern is still evident between them. There is a strong preference for the adjacency of V and NPo in the top line of the right-hand columns when PPm>NPo, and a similar adjacency preference approaching 2-to-1 when weights are equal, and also when NPo>PPm by 1–2 words. The equal or majority preposing of NPo is seen in the left-hand columns only of the second line when NPo exceeds PPm by larger (3–4 and 5+) weight differences. This mirror-image shifting preference has been corroborated in the experimental and corpus studies of Yamashita and Chang (2001) and

Yamashita (2002), and it underscores the prediction of MiD: the directionality of weight effects depends on the language type. Heavy phrases shift to the right in English-type (head-initial) structures, and to the left in head-final Japanese.

The resistance to NPo preposing when weight differences are small in (7.35) matches the same resistance to postposing in English (7.32) under these conditions and supports the preferred adjacency of V and NP for reasons of lexical processing: all direct objects are in a lexical relation with V, whereas most PPs will not be, in both languages.

These performance data enable us to make sense of the conventions of basic ordering, in accordance with the Performance–Grammar Correspondence Hypothesis (1.1); see §1.1. Basic orders follow the preferences of performance within the respective language types. The basic [V oNP mPP] order of English does this by providing minimal domains for phrase structure and lexical processing. Postverbal PPs are significantly longer than object NPs on average, in the corpus data of Hawkins (1994), as shown in (7.36):

(7.36) *English* (Hawkins 1994: 183)

mPP > oNP	73%	(349/480)
oNP > mPP	13%	(63/480)
oNP = mPP	14%	(68/480)

PPs are accordingly preferred to the right of NPs in performance, and this can result in a conventionalized grammatical rule of ordering that positions oNP before mPP (cf. §4.3). There may also be a convention that positions arguments of the verb adjacent to the verb and before adjuncts. Both rules can be argued to have been conventionalized in English. Alternatively one might argue that these are simply the preferred orders one would expect in English anyway, even without a grammatical convention, given the weight differences between mPP and oNP, and the fact that direct objects are always arguments of the verb whereas most PPs are not. Some evidence against this 'pure performance' account, and in favor of a grammatical convention, comes from the data in the left-hand columns of (7.32). The putatively basic and conventionalized order is regularly maintained in many structures in which oNP is heavier than mPP, except when weight differences between them are large (5+ words) and lead to Heavy NP Shift. In comparable postverbal data from other languages (such as Hungarian) in which the relative orderings are undoubtedly grammatically free (see Kiss 1987, 2002), one finds an immediate sensitivity to weight effects and clear ordering preferences between phrases with only one- and two-word weight differences (Hawkins 1994: 133). The absence of such effects in English provides an interesting (performance) argument for the existence of a grammatical

ordering convention, one that is retained in a subset of data that would have shifted in response to weak performance preferences in languages without the basic ordering convention.

Japanese direct objects (NPo) are also shorter than (postpositional) PPs on average, but less consistently than in English, based on the data of Hawkins (1994: 152) at least:

(7.37) *Japanese* (Hawkins 1994: 152)

PPm > NPo	41%	(100/244)
NPo > PPm	22%	(53/244)
NPo = PPm	37%	(91/244)

The relative weights are now more evenly divided, though PPm > NPo exceeds NPo > PPm in frequency by almost 2-to-1. An XOV convention for Japanese is supported by the processing advantages for both phrasal combination and lexical domains, therefore, and the resistance to the OXV conversion (i.e., [NPo PPm V]) in the left-hand columns of (7.35) provides some further support for this, albeit weaker support than for the corresponding VOX convention in English.

7.8.3 Object and X on same side of verb [Pattern III]

So far we have considered alternations of VOX vs VXO and of XOV vs OXV. In both cases O and X occur on the same side of V, in conformity with Pattern [III] of Table 7.3. There is a third OV type, OVX, with O and X on opposite sides. Object and X on the same side is significantly preferred overall, however, and I shall argue that the principle of MiD can account for this as well. MiD also makes predictions for the internal structure of X in the minority languages with OVX which can be tested on the *WALS* database. Consider first some important general properties of, and differences between, O and X in VO and OV languages.

The more equal weight distribution between O and X in Japanese (7.37) compared with English (7.36) matches an important grammatical difference between these two languages: there is greater structural differentiation between O and X in English. Direct objects are always NPs in this language, oblique phrases are generally PPs, and English like many or most VO languages has a productive class of freestanding prepositions as discussed in §7.5, making PP and NP distinct phrases, with PP at least one word longer in its minimal word content since it contains P as well as NP. Japanese, by contrast, employs postposition-like case particles for its nominative, accusative, and dative NPs (*ga*, *o*, and *ni* respectively), whose surface syntax and weight make

them very similar to oblique PPs like *Hanako kara* in (7.34), requiring often subtle tests to distinguish NP from PP (Kuno 1973).

This relative lack of structural differentiation between O and X is characteristic of OV languages in general. Even when single-word postpositions occur (§7.5, §5.8.3), they are less productive as a category in OV languages than prepositions are in VO. Some OV languages have just one or two, and some have none at all. As a result oblique phrases are syntactically often NPs, just like direct objects, and suffixes are very common in OV languages for the syntactic and semantic relations expressed by prepositions in VO languages; see (7.43) below. Again, oblique phrases are (case-marked) NPs. In addition, there are many OV languages like Japanese whose case markers are postpositional particles or clitics, which again makes O and X similar in surface syntax.

The result of all this is that there is less weight differentiation between O and X in OV languages and more processing motivation for the occurrence of alternative orders (e.g., XOV and OXV) that maximize efficiency in PCDs, depending on whatever weights O and X happen to have in individual sentences. The productivity of both XOV and OXV in the grammatical ordering conventions of OV languages matches this greater performance variability, therefore.

There is also variability in the positioning of the head within the PP or NP that constitutes the X phrase in OV languages, as I pointed out when discussing the distinction between rigid and non-rigid OV languages in §5.8.1. MiD predicts that when X regularly precedes V, as in Japanese, we should see head-final XPs (Xm), with P to the right of PP, and N to the right of NP. In this way the PCDs linking, e.g., P and a clause-final V will be minimal, as in (7.34b) above. But in those (non-rigid) OV languages in which X follows V, we should see more head-initial XPs (mX) since the corresponding processing domains linking V and the head of XP will then be shorter.

This prediction was tested briefly in §5.8.1 on certain word order correlations. Let us lay out relevant data in more detail here. The figures given below refer to 'genera' in the *WALS* database (given in parentheses), not to individual languages (see Dryer 1989 for discussion of genera and Dryer 2011 for a list of the genera in *WALS*).

For correlations involving postpositions vs prepositions within a PP as XP, there is a clear tendency in the predicted direction: one third of (non-rigid) OVX languages have either prepositions or no dominant order within PP and are transitional between the overwhelmingly postpositional XOV and OXV and the predominantly prepositional VO, as shown in (7.38) (Dryer 2005b; Dryer with Gensler 2005):

(7.38)

	Postpositions	*Prepositions or No dominant order*
XOV	97% (32)	3% (1)
OXV	94% (15)	6% (1)
OVX	67% (14)	33% (7)
VO	14% (22)	86% (134)

There is a similar tendency in the direction of a head-initial noun-before-genitive order within NPs in OVX languages, as seen in (7.39) (Dryer 2005g; Dryer with Gensler 2005):

(7.39)

	Genitive–Noun	*Noun–Genitive or No dominant order*
XOV	97% (30)	3% (1)
OXV	89% (16)	11% (2)
OVX	69% (18)	31% (8)
VO	27% (45)	73% (124)

On the other hand, the (two-thirds) majority of P and N orders remain head-final for these XPs in OVX languages, a point to which I return below.

For NPs consisting of a head noun plus an adjunct there is striking support for head-initial ordering in OVX languages. Prenominal relative clauses, for example, are completely absent in (non-rigid) OVX languages, as I showed in (7.11) in §7.3, in contrast to rigid OV languages, making OVX almost identical in its correlations to VO. The figures in (7.40) give the *WALS* counts (Dryer 2005d,e; Dryer with Gensler 2005) separately for XOV and OXV in addition to OVX, and add the corresponding quantities for VO languages (the three exceptional VO & RelN languages are all Chinese dialects):

(7.40)

	Rel–Noun	*Noun–Rel or Mixed/Correlative/Adjoined*
XOV	57% (13)	43% (10)
OXV	36% (4)	64% (7)
OVX	0% (0)	100% (17)
VO	3% (3)	97% (116)

Prenominal adjectives are entirely absent from OVX languages as well, making OVX even more head-initial than VO languages, as shown in (7.41) (Dryer 2005h; Dryer with Gensler 2005):

(7.41)

	Adj–Noun	*Noun–Adj or No dominant order*
XOV	44% (16)	56% (20)
OXV	33% (6)	67% (12)
OVX	0% (0)	100% (23)
VO	29% (51)	71% (122)

The figures for a definite article evolving out of a demonstrative determiner were given in (7.25) for rigid OV and VO languages only (§7.6). These are now broken down also into separate figures for XOV, OXV, OVX, and VO languages in (7.42). It is significant that OVX languages have almost the same proportion of separate definite articles as VO, in contrast to other OV (Dryer 2005c; Dryer with Gensler 2005):

(7.42)

	No definite article	*Def word distinct from Dem*
XOV	83% (20)	17% (4)
OXV	75% (6)	25% (2)
OVX	46% (6)	54% (7)
VO	42% (44)	58% (62)

The existence of case suffixes or postpositional clitics is strongly correlated with OV, whereas VO languages prefer no case affixes and either prepositional clitics or occasionally case prefixes, with some case suffixes/postpositional clitics as well. The correlations for OVX are again quite different from other OV languages and closer to those of VO, as shown in (7.43) (Dryer 2005i; Dryer with Gensler 2005):

(7.43)

	Case suffixes or Postpos clitics	*Case prefixes or Prep clitics*	*No case affixes*
XOV	73% (27)	3% (1)	24% (9)
OXV	88% (15)	0% (0)	12% (2)
OVX	48% (10)	5% (1)	48% (10)
VO	27% (42)	16% (25)	58% (91)

Since these case affixes play a significant role in the online construction of mother and grandmother nodes (see §5.8.3, §7.6, and Hawkins 1994), the greater frequency of prefixes and of unaffixed heads is expected in VO languages, and is also characteristic of OVX, whereas suffixes predominate in XOV and OXV.

There are clearly many more head-initial phrasal orderings in OVX languages, as predicted by MiD. Dryer with Gensler (2005) point to a further one: the fronting of a finite Aux to clause-initial position is characteristic of many OVX languages throughout Africa, e.g., Supyire (Gur), Tunen (Bantu), Grebo (Kru), Ma'di (Central Sudanic). I return to this in §7.10.

In both OV and VO languages we have seen in Table 7.3 that O and X are preferentially positioned on the same side of V. This pattern makes sense given the preference for head-ordering consistency and can be argued to be motivated by it. Processing domains for a plurality of phrases in combination with V can be shorter when heads are ordered consistently, to the left or to the right; cf. §§5.2–5.3. Consistency in turn is motivated by MiD.

Positioning O and X on different sides of V would require inconsistent head orderings within NP and XP in order to make processing domains optimally efficient. But in the languages in which O and X occur on opposite sides with any frequency at all (OV languages) we have seen that XP often is an NP. Hence a head ordering good for the postverbal NP will not be good for the preverbal one in these languages, and vice versa, lessening its desirability.

Positioning O and X on the same side of V with a head ordering consistent with that of V is also compatible with O and X occurring efficiently in any relative ordering, before V (OV languages) or after V (VO). Positioning O and X on different sides limits their efficient occurrences to one side of V only and to one relative ordering.

Inconsistent head ordering and opposite positioning for O and X also has negative consequences for the formulation of grammatical conventions. It limits the generality of the principles that define phrase structure, resulting in less general X-bar rules and more linearization exceptions and stipulations, as I argued in Hawkins (1983: 183–9) following Jackendoff (1977).

For all these interrelated reasons head ordering is generally consistent and O and X are positioned on the same side of V in the majority of languages, while there are correlations between opposite positioning and head inconsistency in the minority of languages with OVX.

7.8.4 Direct object before oblique phrases [Pattern IV]

There is also some head variability in VO languages. Some, a minority, have Postpositions (7.38), Genitive-N (7.39), Rel-N (7.40), and Adjective-N (7.41), compatible with preverbal positioning for these phrases by MiD. Yet only in the most extreme head-inconsistent languages (such as Chinese) do we find basic XVO orders mirroring the productive OVX. V and O are still adjacent in Chinese [Pattern I], so this dispreference is in part the result of O and X not being on the same side (contrary to Pattern II). But why don't we get the mirror image XVO more productively in VO languages? I suggest that there is a second general principle interacting with MiD here, which pushes X to the right of (subject and) object complements, even in VO languages, namely (7.44):

(7.44) *Argument Precedence* (AP)
Arguments precede X

This is a linear precedence principle of language production that may have parallels in comprehension. It is supported empirically by the productivity of (S)OVX versus the rarity of the mirror-image (S)XVO, as we have seen. Theoretically it may be motivated by the greater frequency and accessibility of the argument NPs for a given verb versus the diversity and lesser frequency

of each XP type that can co-occur with that verb: every clause has at least a subject argument for the relevant verb, and object arguments are more frequent for each transitive verb than any single type of XP. Arguments are also more definite and more salient and foregrounded than adjunct XPs. These kinds of frequencies and accessibility differences are associated with earlier production in a number of online production models (Bock 1982; Bock and Levelt 1994; Kempen and Hoenkamp 1987; MacDonald et al. 1994; Stallings and MacDonald 2011; Levelt 1989). Earlier access to arguments before adjuncts, and their earlier linearization, may be motivated by their greater frequency and accessibility, therefore.

7.8.5 VO consistency vs OV variability [Pattern I]

I have argued that V&O adjacency [Pattern II], and O and X on the same side [Pattern III] are ultimately motivated by the principle of Minimize Domains. A second general principle of production, Argument Precedence (7.44), appears to underlie O before X [Pattern IV]. It is interesting now to observe how these three patterns [II]–[IV] all converge and reinforce each other in VO languages, resulting in consistent VOX. But in OV languages they are partially opposed, and it is this, I suggest, that leads to variability: more variation in performance in response to the different ordering preferences of individual sentences; more variable basic orderings in OV grammars; and also fewer basic orders in OV grammars than in VO grammars.

This interaction is shown in (7.45):

(7.45)

	V & O Adjacency	*O & X on Same Side*	*O before X*
VOX	+	+	+
XVO	+	–	–
VXO	–	+	–
XOV	+	+	–
OXV	–	+	+
OVX	+	–	+

VOX languages conform to all three patterns, the VO competitors (XVO and VXO) to at most one. The degree of preference for VOX is considerable and our approach predicts that it will be highly favored both in performance and in grammars. The OV language types, on the other hand, each conform to two patterns, as shown in (7.46). I have added the number of exemplifying languages from Table 7.1 under each word order type.

(7.46)

	VOX	>	XOV/OXV/OVX	>	XVO/VXO
	3		2		1
Lgs:	189		45/23/37		3/0

The efficiency advantages for each OV type are more equal, and selections in both performance and grammars should be more equal as well. Preferences will depend on the degree of weight difference between X and O in individual sentences, and on the (argument or adjunct) status of X. There may also be a difference in the relative strength of our two general principles, Minimize Domains and Argument Precedence (7.44), which can impact the selections in cases of competition.

The clear quantitative pattern evident in (7.46) is that there are more exemplifying languages, the more efficiency principles there are that a given type can exemplify. This provides suggestive evidence for the point made in §4.4 that the most common language types are those that can incorporate the most efficiency principles. In the structures and datasets considered here involving obliques the preferences all conspire and cooperate in languages that have basic VO order. This is not always the case for this language type as a whole, however. Cross-linguistic frequencies suggest that there are a number of common types, SVO, non-rigid SOV, rigid SOV, etc., that are overall as efficient as they can be given the competitions and inherent tensions between principles and the different ways of realizing them in grammars (e.g., through head-initial or head-final orders); see §2.5 and §9.3 for further elaboration on this.

In summary, four empirical patterns have been proposed here using data from *WALS*:

(7.47) Pattern [I] VO consistency vs OV variability
Pattern [II] Verb & Object adjacency
Pattern [III] O and X on same side of V
Pattern [IV] O before X

and two general principles: Minimize Domains and Argument Precedence (7.44). MiD motivates Patterns [II] and [III] and AP motivates [IV]. The interaction between them results in the general pattern [I] and in the quantitative distribution of exemplifying languages.

7.9 Wh-fronting in VO and OV languages

There are languages like English that regularly front Wh-words in direct questions, thereby forming a filler-gap dependency as in (7.48).

(7.48) Who*i* do you think [that John saw 0*i*]?

Alternative strategies for forming Wh-questions include keeping the Wh-word 'in situ,' i.e., in its original position, or else some form of adjacency and attachment to the verb within what is often referred to as a 'focus'

position (see Hawkins 1994: 373–9 and Hawkins 2004: 136). The fronting option is highly correlated with basic verb position and is especially favored in V-initial languages (see (7.49) and (7.50) below). The adjacency/attachment and focus option is especially favored in SOV languages (Kim 1988, Hawkins 1994, and §7.10), for example in Kannada (Bhat 1991) and Turkish (Erguvanli 1984). There are interesting correlations between the structure of Wh-questions across languages and verb position, therefore, some of which will be explored here. In particular I will be concerned with the fronting option as in (7.48).

Some 29 percent of the world's languages have this option, according to Dryer's (2005j) *WALS* figures given in (7.51) (rather more 'genera' have it, 43 percent, in Dryer's 1991 figures shown in (7.49)). The fronting structure raises two questions in this context. First, why is the displacement of Wh almost always to the left and not to the right? I have addressed this in Hawkins (2004: 171–7 and 203–5) and subsumed it under a 'Fillers First' generalization, motivated ultimately by MaOP (see §2.3.1 for a summary). Second, why is Wh-fronting correlated with basic verb position?

The correlations can be seen in different language samples (see Hawkins 1999 for a summary of data from Greenberg's 1963 and Ultan's 1978 earlier samples). Dryer (1991) gives the following figures from his genetically and areally controlled sample, in terms of genera:

(7.49) *Wh-fronting and basic verb position* Dryer (1991)

V-initial:	23/29 genera = 79%
SVO:	21/52 genera = 40%
SOV:	26/82 genera = 32%
Total:	70/163 genera = 43%

The more recent figures from *WALS* counting languages rather than genera are given in (7.50) and (7.51), with the latter showing quantities of Wh-fronting for just VO versus OV (Dryer 2005j; Dryer with Gensler 2005):

(7.50) *Wh-fronting and basic verb position* *WALS*

VSO:	42/61 languages = 69%
VOS:	10/15 languages = 67%
SVO:	65/262 languages = 25%
SOV:	52/280 languages = 19%

(7.51) *Wh-fronting and VO versus OV* *WALS*

VO:	143/377 languages	=	38%
OV:	69/353 languages	=	19.5%
Total:	212/730 languages	=	29%

These different language samples show a strong preference for Wh-fronting in V-initial genera and languages (79 percent in (7.49), and 69 percent and 67 percent in (7.50) for VSO and VOS respectively). They also show that this option is not preferred in SOV languages and genera (68 percent do not have it in (7.49), 81 percent do not in (7.50)). The figures in (7.51) indicate that a VO language is roughly twice as likely to have Wh-fronting as an OV language. (7.49) and (7.50) show that there is therefore a relative frequency ranking of V-initial > SVO > SOV for Wh-fronting.

Why should this verb position ranking be correlated with filler-gap structures in Wh-questions? The definition of a Filler-Gap Domain given in Hawkins (2004: 175) suggests an answer and is reproduced here as (7.52):

(7.52) *Filler-Gap Domain* (FGD)
An FGD consists of the smallest set of terminal and non-terminal nodes dominated by the mother of a filler and on a connected path that must be accessed for gap identification and processing; for subcategorized gaps the path connects the filler to a coindexed subcategorizer and includes, or is extended to include, any additional arguments of the subcategorizer on which the gap depends for its processing; for non-subcategorized gaps the path connects the filler to the head category that constructs the mother node containing the coindexed gap; all constituency relations and co-occurrence requirements holding between these nodes belong in the description of the FGD.

The verb is the subcategorizer for all Wh-words that are verbal arguments and that must be linked to it in the relevant domains for lexical processing (see (5.15) in §5.2). The verb is also the head category that constructs the VP and/or S nodes containing the gap sites for most adjunct Wh-words. As a result FGDs and lexical domains linking Wh to the verb will be most minimal in verb-early languages and will become larger as the verb occurs further to the right. Domains for the syntactic processing of adjunct relations will become larger as well. A leftward Wh-word (predicted independently by MaOP (3.26)) accordingly sets up an MiD prediction that co-occurring verb positions will be progressively dispreferred as the verb stands further to the right and FGDs and other domains linking the filler to the verb become larger. The filler-gap structure is increasingly not selected, therefore, and in situ or verb attachment structures are used instead. This explains why these non-fronting strategies are especially favored in SOV languages (see Kim 1988, and also Hawkins 1994: 373–9 for a parsing account of why attachment to a finite verb can achieve the same scope effects for Wh-words as leftward movement and adjacency to the clause over which Wh has scope; see also §7.10). In a similar way I have

argued that gaps give way to resumptive pronouns in relative clauses when there are increasing FGD sizes (Hawkins 1999, 2004: 186–90).

The distinction between rigid and non-rigid SOV languages is also relevant here. Wh-fronting should be more productive with non-rigid than with rigid SOV, since the distance between Wh*i* and V*i* in Wh*i*[SOV*i*X] is smaller than in it is in Wh*i*[SXOV*i*]; see also Aldai (2011). Conversely, in situ and V-attached Wh-phrases should be more frequent in rigid than in non-rigid SOV languages. *WALS* (Dryer 2005j; Dryer with Gensler 2005) enables us to test this. In (7.53) I give the correlations for Wh-fronting with verb, direct object, and oblique phrase ordering.

(7.53) *Wh-fronting and {v, o, x} orders WALS*

vo (i.e. vox & xvo):	61/167 languages	=	40%
ovx	12/35 languages	=	34%
{x, o} v	9/68 languages	=	13%
ov total	21/103 languages	=	20%

As we move down (7.53) from VO languages to non-rigid OVX to rigid {X, O} V the frequency of Wh-fronting declines. The figures for all VO and OV languages (40 percent versus 20 percent) are close to those given in (7.51) (38 percent to 19.5 percent), i.e., without regard for the position of obliques. The interesting contrast here is between non-rigid OVX and its rigid counterpart with both O and X preceding V (i.e., with basic order XOV or OXV; see §7.8). A full 34 percent of non-rigid OV languages have Wh-fronting compared with just 13 percent for those with rigid OV. This is in the direction predicted and it enables us to amplify our predicted frequency ranking as shown in (7.54):

(7.54) *Wh-fronting frequency ranking*
V-initial > svo > non-rigid sov > rigid sov

A major difference now remains to be explained between relative clauses and wh-questions. Wh-fillers to the right of their gaps or subcategorizers are largely unattested. Head noun fillers to the right, i.e., RelN, are productive, but limited to (primarily rigid) OV languages; see (7.11) in §7.3. Why this difference?

I offer the following. An NP is contained in numerous mother phrases (PP, VP, NP, etc.) that will favor head finality for this NP in OV languages by MiD (2.1). Wh-phrases, by contrast, typically occur in the highest, matrix clause and there is generally no dominating phrase, head-final or otherwise, to exert a preference for a consistent positioning of the Wh-word either to the right or left of its sister clause. The MaOP preference for Fillers First (§2.3.1) is therefore unopposed by MiD, whereas the corresponding preference for NRel is systematically opposed in all head-final phrases containing

NP. Indirect questions (*I wonder* [*who John saw*]) do embed the questioned clause within a matrix clause, but their frequency appears to be limited compared with direct questions and their impact on a grammaticalized ordering should be correspondingly small. The variation in relative clause filler-gap ordering is therefore the result of a competition between MaOP and MiD, whereas for Wh-filler-gaps the preferences of MaOP are generally unopposed and favor Fillers First.

7.10 Verb movement in VO and OV languages

A number of VO languages have interesting verb movement rules whose conditions of application suggest that their ultimate motivation is to signal the onset of a clause and to construct a node such as S, by Mother Node Construction (see (5.8) in §5.1). The Celtic languages, like Welsh, front a verb to the left of S, but only when that verb is finite, whether lexical or auxiliary, as shown in (7.55) from Sadler (1988).

(7.55) a. Gwelodd Mair y ddamwain. (Welsh)
See-PAST-3SG Mary the accident
'Mary saw the accident.'

b. Gwnaeth Siôn ennill.
Do-PAST-3SG John win
'John won.'

The non-finite verb *ennill* in (7.55b) remains in situ within the VP. More generally, all non-finite verbs in Welsh remain in situ and to the right of the subject. The finite verbs *gwelodd* and *gwnaeth*, on the other hand, are fronted out of the VP and to the left of the subject. Their finiteness means that they are dominated by a full clause, which consists of both tense and an NP subject in addition to VP, and these verbs carry both tense and subject agreement information and hence they can serve as constructors of the S in parsing. If this is indeed the motivation for fronting the finite verb, then we have a simple explanation for why non-finite verbs do not front: such verbs project only to VP and do not bear the additional morphological properties of sentencehood. Hence they cannot construct S and so this motivation for moving them is lacking (see Hawkins 1994: 381–5 for detailed discussion of these Welsh parsing routines).

Similar verb-fronting rules can be found in Germanic languages like Danish, resulting in verb-first structures, and in verb-second after an initial constituent, as shown in (7.56) (examples from Bredsdorff 1962):

(7.56) a. I går så jeg ham på gaden. (Danish)
Yesterday saw I him on street-the
'Yesterday I saw him on the street.'

b. Har De vaeret I Danmark?
Have-PRES you been to Denmark'
'Have you been to Denmark?'

The Germanic verb-fronting rules apply only when the verb is finite and can construct a clause and signal its onset, whether in absolute initial position (verb-first) or after a preposed constituent (verb-second). In contrast to the Celtic languages, however, in which verb fronting occurs in both main and subordinate clauses, verb fronting in Germanic applies in main clauses only, while subordinate clauses are constructed by other categories such as complementizers which are generally in complementary distribution with verb fronting, making this latter a main clause phenomenon only (see further Hawkins 1994).

Verb-fronting rules in head-initial languages like these make the parsing of S fundamentally similar to, and consistent with, the parsing of all other phrases in these languages. For example, in Welsh prepositions occur initially in PP, adjectives precede their complements within AdjP, the NP is initiated by Det or N, a non-finite V is leftmost within its VP, and so on. The left-peripheral positioning of finite verbs can be motivated by the fact that all constructing categories are left-peripheral, therefore, and that finite verbs can construct an S. Similar considerations apply to Danish, and even to the mixed typology language German, which despite a number of verb- and auxiliary-final orders is argued in Hawkins (1994: 393–400) to be a fundamentally left-peripheral constructing language (a head-initial wolf in head-final sheep's clothing!). In all of these head-initial languages, therefore, a fronted finite V signals the onset of S.

But just as Wh-movement applies exclusively to the left and there is no corresponding movement to the right (§7.9), so too finite verb movement is to the left and there appears to be no comparable finite verb postposing, certainly not in VO languages and not even in head-final languages with OV. In the event that the OV language is rigidly verb-final, then the verb will already be final, matching the head orders in all other phrases. In non-rigid OV languages, however, such rightward movements could in principle exist, but it is interesting that even here they seem not to occur. Just as Wh-movement applies to the left and is found in a subset of non-rigid SOV languages (see §7.9), so too verb movement is to the left and occurs in a subset of non-rigid SOV languages as well. The verbs that front in OV languages appear to be finite, either auxiliary or lexical, and they may or may not be

accompanied by additional non-finite verbs with which they combine in the clause.

Dryer with Gensler (2005) discuss several examples of SAuxOVX structures in the non-rigid OV languages of Africa, i.e., with the auxiliary verb in second position, much as in German main clauses (see Hawkins 1986: ch. 11). The following example (taken from Heath 1999) is from Koyraboro Senni (Songhai, spoken in Mali):

(7.57) ay ga nooru wiri [ay baaba ga] (Koyraboro Senni)
1SG FUT money seek 1SG father on
Aux O V X
'I will seek money from my father.'

Tunen, a Bantu language of the Cameroon with non-rigid SOVX word order, exemplifies it as well (example from Dugast 1971):

(7.58) mɛ ná mìɔkɔ bokòn [ò nuìy] (Tunen)
1SG PAST stones raised LOC river
Aux O V X
'I raised the stones from the river.'

In Grebo (Kru, spoken in Liberia) this SAuxOV alternates with SVO in clauses that lack auxiliaries (example from Innes 1966):

(7.59) a. ne yi bla du-I (Grebo)
I IMPF rice pound-IMPF
Aux O V
'I am pounding rice.'

b. ne du-da bla ne
I pound-DIST.PAST rice PART
V O
'I pounded rice.'

Aldai (2011) gives a detailed discussion of verb fronting in non-rigid OV languages with particular reference to Basque and links the movement of the verb explicitly to the fronting of a Wh-element. The verbal complex in modern Basque, consisting of finite and non-finite elements, almost always fronts to a position immediately adjacent to the fronted Wh-word, as shown in (7.60a,b), and Aldai argues that this is motivated by the preference for a more minimal domain linking Wh-words and the verbs that either subcategorize for them or that construct the phrase that contains the gap site for the Wh-filler. In other words, Aldai builds on the correlation between Wh-fronting and verb position shown in §7.9 and sees the fronting of the verb as a consequence of this fronting of Wh (examples from Aldai 2011: 1095):

(7.60) a. Zer idatzi du Martin-ek? (Basque)
What.ABS written has Martin-ERG?
'What has Martin written?'

b. Nor-k idatzi du liburua?
Who-ERG written has book.ABS?
'Who has written the book?'

In earlier stages of recorded Basque, especially in the sixteenth century, verb fronting (i.e., #Wh-Verb-X) alternated with non-fronted verbs (#Wh-X-Verb) and Aldai tests a number of predictions deriving from Hawkins (2004) on earlier Basque corpora. For example an adjacent Wh-V is predicted to be preferred and more frequent when Wh is an argument rather than an adjunct of V (since there are both lexical and syntactic domains for processing; see §5.2 and §9.2), and when a potentially intervening X is long and complex rather than short (§5.2), and he finds these predictions to be supported. Aldai also points to other non-rigid OV languages that have a similar verb-fronting rule to Basque, e.g., Ingush (Nakh-Daghestanian, North Caucasus), as seen in (7.61) which contrasts with the normal OV order of (7.62) (examples from Nichols 1994):

(7.61) a. Miča vuod iz? (Ingush)
Where go he?
'Where is he going?'

b. Fi diež va iz?
What doing is he?
'What is he doing?'

c. Hanaa dennad da:s sowĝat?
Who.DAT gave father.ERG present?
'Who did Father give a present to?'

(7.62) da:s woєaa kita:b sielxan delar. (Ingush)
Father.ERG son.DAT book.NOM yesterday gave
'Father gave son a book yesterday.'

For fronted auxiliaries in a very different language, the Australian Warlpiri, and their parsing, see, e.g., Kashket (1988).

Aldai's (2011) discussion linking verb fronting to Wh-movement in non-rigid OV languages is interesting and his historical Basque data are impressive. His proposed typology of SOV languages in terms of degrees of rigidity (Aldai 2011: 1093) is also extremely useful. But there are additional considerations that are relevant here, I believe, that will be more or less applicable to

different subtypes of SOVX and also to relevant VO languages like Celtic and Germanic that have verb fronting (see (7.55) and (7.56)).

First, we need to account for the asymmetry in the directionality of movement: finite verbs front, they are not postposed, in both VO and OV languages. A possible reason for this can be found in Ueno and Polinsky's (2009) argument, based on both usage data and grammars, that it is preferable to encounter the verb of a sentence as soon as possible (recall §7.7). One-place predicates were significantly preferred over two-place predicates in their OV language usage data, thus reducing the amount of material to be processed prior to V. OV grammars have an intransitive bias morphologically (see Nichols et al. 2004). Left-branching subordinate clauses such as relatives with OV are significantly reduced in expressive power compared with their postnominal counterparts in VO languages (see Lehmann 1984: 168–73 and §5.8.3), which again reduces the amount of material to be processed prior to a clause-final V. In §7.2 I drew attention to the importance of the verb in online processing as an integrator and predictor of clause structure. All of these considerations motivate early placement of the verb, in order to make clear what the structure of the clause is, and how the NPs and PPs and other phrases within it are linked together. They are not sufficient to prevent verb-final positioning altogether, even rigid verb finality, when that is consistent with all the other phrases in the language and gives minimal domains for phrase structure processing (see §7.1). But they do provide a certain inner tension in favor of verb fronting (see §4.4 and §9.3), which results in the verb movement rules of non-rigid SOV languages that we have seen, and in verb frontings in VO languages such as Welsh and Danish.

Second, it is also relevant that non-rigid SOV languages are typologically mixed throughout the grammar and contain numerous head-initial structures in addition to head-final ones; see §7.8.3. The fronting of a finite verb provides one more head-initial structure in addition to those discussed in §7.8.3.

Third, I believe that the key point about finite verb fronting is its construction potential, as I mentioned at the beginning of this section. That is why the fronted verb must be finite and bear distinctive morphology or be syntactically an auxiliary or some other category that is capable of signaling the onset of S, not just of VP. In fact, if the fronted verb has been extracted out of VP in the relevant language, then it will construct S directly as its mother, rather than as a grandmother over a mother VP (see Hawkins 1994: 381–402 for detailed parsing routines for S and VP nodes in different languages).

Fourth, the precise conditions of application for verb fronting in different languages can vary and can conventionalize different preferences. For Basque the reduction in Filler Gap Domains after Wh-movement has clearly been a

major motivation whose diachronic origin can be seen in Aldai's (2011) corpus data. In other languages the precise conditions of application point to additional motives. The Germanic V-2 rule brings some lexical verbs forward and adjacent to a fronted Wh, as in (7.63a) in German, but not when there is a finite auxiliary as in (7.63b):

(7.63) a. Wen kennst Du? (German)
Who.ACC know.2SG you.2SG?
'Who do you know?'

b. Wen hast Du gesehen?
Who.Acc have.2SG you.2SG seen?
'Who have you seen?'

Instead, some have argued for a more pragmatic motivation for the Germanic V-2 structure. The initial constituent of German provides a topical grounding, and frequently temporal and spatial bounding, for the event described in the clause constructed by the finite verb. See, e.g., Los and Dreschler (2012) for a succinct summary of this view and literature review. Their proposal is reminiscent of Jacobs' (2001) "frame-setting" generalization for topics and their predications: "In (X, Y), X is the *frame* for Y if X specifies a domain of (possible) reality to which the proposition expressed by Y is restricted" (Jacobs 2001: 656). Reinhart (1982) talks simply about "aboutness" in this context.

The point I would add to this here is that having a finite verb to construct S means that S becomes a sister to the initial constituent, within the structure [XP [$_S$ Vf...]]. XP and S can now engage in the kinds of semantic and pragmatic relations that sisters can engage in, depending on what the precise phrases are. In (7.63) the XP provides an argument for the lexical verb in S. On other occasions it may be a more pragmatic and frame-setting relationship. A clear semantic relation involving quantifier and operator scope can be seen in some of the remains of the Germanic V-2 rule in Modern English, specifically in Subject–Auxiliary Inversion and its conditions of application.

Consider the following minimal pairs discussed in Hawkins (1986: 229–30) and taken from Klima (1964), Lieberman (1974):

(7.64) a. At no time was John prepared.
b. In no time John was prepared.

(7.65) a. With no job would John be happy.
b. With no job John would be happy.

(7.66) a. Not even a year ago did he manage to make a profit.
b. Not even a year ago he managed to make a profit.

In (7.64a), with finite verb fronting, the scope of negation extends to the whole S constructed by *was*, i.e., the sentence means 'John was not prepared at any time.' (7.64b) without fronting does not have this meaning. What is negated is exclusively the material within the preposed constituent. Hence, John *was* prepared, and in a period of time that was not considerable. (7.65a) is similarly paraphrasable by 'John would not be happy with any job,' with negation extending to the whole S constructed by *would*. (7.65b) means 'John would be happy not to have any job,' with negation limited to the preposed constituent. (7.66a) is paraphrasable by 'he did not manage to make a profit even a year ago,' whereas (7.66b) means that he did manage to make a profit, and not even a year ago.

In all of these examples constructing an S in the (a) sentences on the basis of the fronted finite verb creates a clausal sister of the preposed negative phrase, and this syntactic sisterhood relation makes the S available, we can hypothesize, to contract semantic relations with the preposed phrase, specifically to be in its scope. In the (b) examples the preposed phrase occurs inside a single clause (i.e., [$_S$ PP NP VP]) and there is no sisterhood relation between the fronted XP and a second constituent, [$_S$ Vf...]. It is the sisterhood between the two constituents, I suggest, that licenses the scope relations that we see in the (a) sentences.

Finally, the verb movements we have seen in this last section of the chapter can be found in non-rigid SOV languages and in VO languages but not, it seems, in rigid SOV. They provide revealing evidence of the inherent tension that exists in SOV languages between the advantages and disadvantages of positioning the verb late. Having the verb in absolute final position is characteristic of strongly head-final languages, in which all the other phrases are head-final too and in which there is pressure for 'Cross-Category Harmony' (Hawkins 1983), arguably as a result of Minimize Domains (2.1). The overall syntax of these languages appears to be the primary motive of rigid verb finality, therefore. A further advantage of verb-finality is that all relevant properties of the verb, in terms of its syntactic and semantic co-occurrences, can be gathered up at the verb, and the precise interpretation of the verb can be assigned immediately by reference to its preceding arguments, in accordance with Maximize Online Processing (2.26). I have argued (see §7.2) that these co-occurrences are of necessity more constrained in an OV language, when the verb follows its co-occurring arguments, than in a VO language when it precedes and when there needs to be look-ahead anyway to subsequent phrases that follow V. VO languages may, but do not have to, develop their look-ahead routines further, resulting in a greater tolerance for unassignments and misassignments online and in more 'dependent' processing

generally. I have argued that this is exactly what English has done in the course of its history (see Hawkins 2012 and §7.2).

The disadvantages of verb-late are that VP and/or S nodes are constructed late and preceding phrases are left unattached to a mother node (see §5.8.1 and §5.8.3), resulting in online unassignments in parsing. This creates a pressure to postpose phrases to the right of V and to create non-rigid SOV, in conjunction with mixed head ordering (see §5.8.1 and §7.8). It results in less complex arguments preceding V, a preference and a bias for intransitives over transitives, and a reduction in the syntax and expressive power of certain subordinate clause types (§5.8.3 and §7.7). It also results in more affixal node construction (§7.6). These pressures structure various hierarchies of phrasal postposing in SOV languages and result in a typology of non-rigid SOV (see Hawkins 1986: ch. 10 for a detailed illustration from German, and Aldai 2011: 1093). For example, heavier phrases will postpose to the right of V before lighter phrases do, adjuncts before complements, and so on.

Verb-early languages have complementary advantages and disadvantages. Their mother nodes are created early and permit immediate online attachments (§5.8), but many aspects of verbal processing are delayed, resulting in unassignments and misassignments online (§7.2). Both VO and OV language types can be fully and equally efficient, I have argued, for Minimize Domains, which explains their roughly equal distribution typologically (see §7.1). But their respective advantages and disadvantages in terms of MaOP appear to be complementary, and a large number of the differences between them that we have seen in this chapter seem to reflect this complementarity and the reverse positioning of the verb.

8

Asymmetries between arguments of the verb

In the last chapter I considered asymmetries between languages depending on the basic position of the verb. Many cross-linguistic generalizations also show asymmetries between the arguments of verbs, regardless of the verb's position, i.e., between what are loosely called 'subjects,' 'direct objects,' and 'indirect objects.' See Primus (1999) for discussion of why these are loose terms and for the descriptive tools that can make them more precise. The asymmetries are visible in patterns of co-occurrence between the arguments of verbs, in rule applicability patterns, formal marking patterns, and in linear ordering. They have been described in a number of frameworks and approaches in terms of hierarchies of grammatical relations, hierarchies of morphological cases and verb agreement, hierarchies of thematic roles, and in terms of linear precedence preferences derived from these hierarchies. This chapter raises the question: why should there be such asymmetries and hierarchies among the arguments of a verb, with these correlating patterns? Some answers are proposed that draw on the efficiency principles of this book, Minimize Domains (2.1), Minimize Forms (2.8), and Maximize Online Processing (2.26).

8.1 Illustrating the asymmetries

One example of such an asymmetry involves argument co-occurrence: if a verb is subcategorized to take a direct object NP (DO), it is also subcategorized to take a subject (SU), but the converse fails, cf. Blake (1990). A more general formulation is given in Primus's (1995, 1999) Subcategorization Principle as follows:

(8.1) The assignment of a lower-ranking case by a predicate P implies asymmetrically the assignment of a higher-ranking case by P, in the unmarked case.

Illustrative support for this comes from case assignment quantities for nominative-, accusative-, and dative-taking (non-compound) verbs in German. Verbs that assign accusative also assign nominative, those that assign dative generally assign both nominative and accusative. These implications are not reversible, and the numbers of verbs assigning each case declines as shown in (8.2), from Primus (1999) using data from Mater (1971):

(8.2) Non-compound verbs in German: Nominative = 17,500
Accusative = 9,700 (all select Nom)
Dative = 5,450 (5,100 select Nom & Acc)

Patterns in the applicability of rules reveal a similar asymmetry among the grammatical relations SU > DO > IO. This ranking can be found in the top three positions of Keenan and Comrie's (1977) Accessibility Hierarchy (SU > DO > IO > OBL > GEN) for syntactic rules such as Relative Clause Formation. I summarized this hierarchy in §2.2.3 and illustrated its predictions for gap and resumptive pronoun strategies. Here I summarize their 'cut-off' data. Relativization can cut off (i.e., cease to apply) in an asymmetrical manner with languages permitting relativization on all positions above, but not below, the relevant points on the hierarchy, as shown in (8.3):

(8.3) *Rules of relative clause formation and their cut-offs within the clause* (Keenan and Comrie 1977)
SU only: Malagasy, Maori
SU & DO only: Kinyarwanda, Indonesian
SU & DO & IO only: Basque
SU & DO & IO & OBL only: North Frisian, Catalan
SU & DO & IO & OBL & GEN: English, Hausa

A similar asymmetry between the top two positions of this hierarchy, SU > DO, can be seen in verb agreement across languages. Verb agreement may apply to, and be marked on, the SU only, but if it applies to DO, it applies to SU as well. Exemplifying languages, distinguished according to the nominative-accusative or ergative-absolutive basis for their verb-marking pattern (cf. Comrie 1978; Blake 2001; Dixon 1994; Tallerman 1998; Primus 1999) are given in (8.4):

(8.4) *Verb agreement rules across lgs*
SU only: (Nom-Acc) French, Finnish, Tamil
(Abs-Erg) Avar, Khinalug
SU & DO only: (Nom-Acc) Hungarian, Mordva
(Abs-Erg) Eskimo, Coos

For a succinct and accessible summary of verb agreement patterns and for illustrative data from many of these languages, the reader is referred to Tallerman (1998: 158–62).

Morphological marking on the arguments themselves also reveals asymmetries. When several arguments are case-marked, the higher-ranked ones may receive a less complex formal marking (morphologically and phonologically) than lower ones, often zero. This was illustrated in §2.2.1 for various hierarchies proposed by Greenberg (1966). Primus (1995, 1999) follows Greenberg in reformulating the pattern for case marking as what she calls an 'Allomorphism Principle.' I reproduce it here with a different name as the Case Form Principle, following a suggestion made by Martin Haspelmath (p.c.):

(8.5) *Case Form Principle*
The higher the rank of a case on the Case Hierarchy of a language . . . , the less complex is, as a preference, its morpho-phonological realization. *Corollary*: where there is a case system in which just one case has zero allomorphs that case will be the one with the highest rank on the Case Hierarchy.

Relevant data from Avar (Comrie 1978: 338), in which a higher-ranked absolutive noun (*jas*) is zero-marked while a noun in the ergative case has explicit case marking (*vas-as*:), were discussed in a different context in §3.5 and are repeated here in (8.6):

(8.6) Vas-as: jas-0 j-ec:ula (Avar)
boy-ERG girl-ABS SG.FEM.ABS-praise
'The boy praises the girl.'

With regard to linear ordering, subjects precede objects in the vast majority of the world's languages—for example in 96 percent of Tomlin's (1986) sample (i.e., in SOV, SVO, and VSO languages); cf. (8.7):

(8.7) *Relative frequencies of basic word orders in Tomlin's (1986) sample* (402 lgs)

SOV	(168)	VSO	(37)	VOS	(12)	OVS	(5)
SVO	(180)					OSV	(0)
87%		9%		3%		1%	

More generally, there are, according to Primus (1999), hierarchy-based linear ordering preferences among arguments, both across grammars and within languages, as summarized in the next section.

These asymmetries pose challenges of both a descriptive and an explanatory nature for grammatical theory and typology. What are the best descriptive mechanisms for capturing the generalizations that underlie these cross-linguistic patterns? A recurring mechanism has involved the hierarchical

arrangement of syntactic, morphological, and semantic properties that are assigned to different arguments. But what then explains the rankings of these properties on the hierarchies and why do rules, formal marking omissions, and linear precedence favor the higher positions on hierarchies? Efficiency provides many of the answers.

8.2 Hierarchies

The hierarchy idea has a long history. In the area of morphosyntax a key publication was Greenberg (1966), in which asymmetries between the ranked categories of numerous grammatical areas formed the basis for markedness or feature hierarchies such as Singular > Plural > Dual > Trial/Paucal; see §2.2.1. Morphological inventories across grammars, declining morphological distinctions, and increased formal marking provided the primary evidence for these hierarchies, while performance data involving frequency of use offered further support. Croft (1990/2003) gives an extensive summary and discussion of Greenberg's hierarchies, incorporating most of the more recent hierarchies that have been proposed in the typological literature.

The most extensive application of hierarchies to argument structure can be found in the work of Beatrice Primus (Primus 1993, 1995, 1999). She has proposed hierarchies for all the major properties that impact the arguments of a verb, e.g., for case morphology (8.8), (configurational) syntax (8.9), and thematic roles such as (8.10) and she formulates the kinds of linguistic generalizations illustrated in the last section involving co-occurrence, rule applicability, formal marking, and linear ordering in terms of these hierarchies:

(8.8) Case Morphology: Nominative > Accusative > Dative > Other
Absolutive > Ergative > Dative > Other

(8.9) Syntax: higher structural position (c-command*ing*) > lower position (c-command*ed*)

(8.10) Semantics (theta-roles): Agent > Recipient > Patient
Experiencer > Stimulus
Possessor > Possessed

There are many descriptive advantages to Primus's 'generalized hierarchy' approach. First, by decomposing the grammatical relations subject, direct object, and indirect object into these more basic component properties she can avoid problems that afflict models and generalizations formulated in terms of the composite notions, subject and direct object, etc. For example, NPs defined as 'nominative' and 'accusative' are strongly aligned with other syntactic and semantic properties and often permit a more abstract

generalization in terms of SU and DO, but NPs defined as 'absolutive' in ergative–absolutive systems correspond to both the subject (of intransitive clauses) and the direct object (of transitives) in nominative–accusative languages, destroying the validity of these more general grammatical relations. By contrast, valid cross-linguistic generalizations can be readily captured in terms of the component atoms of these hierarchies, along the lines of (8.8)–(8.10): zero case marking on NP arguments can be described as applying to the highest case on the relevant case hierarchy, verb agreement may be with the top two positions on the case hierarchy for the relevant language, or with an NP in the highest c-commanding position, and so on.

Second, this approach makes testable predictions for permitted variation across languages, as in the examples just given. A rule like Passive can now promote an NP with a certain case, or with a certain theta-role or in a certain syntactic position. The result will be a set of partially different rule outputs across languages, depending on which property has been conventionalized in the conditions of application for Passive in the relevant language.

Third, areas of mismatch between the hierarchies make a novel and principled set of predictions for cross-linguistic and intralinguistic variation in linear ordering. Consider the interaction between the case hierarchies of (8.8) and the first theta-role hierarchy of (8.10) shown in (8.11) (for nominative–accusative case systems) and (8.12) (for ergative–absolutive ones):

(8.11) Case: Nom [Ag] > Acc [Pat] > Dat [Recip]
Theta: Ag [Nom] > Recip [Dat] > Pat [Acc]

(8.12) Case: Abs [Pat] > Erg [Ag] > Dat [Recip]
Theta: Ag [Erg] > Recip [Dat] > Pat [Abs]

If one assumes, as Primus does, that higher-ranked positions on each hierarchy are preferably linearized to the left of their respective lower-ranked positions, then (8.11) indicates that nominative agents in a case-marked language like German or Russian should precede other arguments in ditransitive clauses. Accusative patients and dative recipients, on the other hand, should exhibit variation: the accusative is higher than the dative on the case hierarchy, whereas the recipient is higher than the patient on the theta-role hierarchy. Hence both orders should be found and Primus (1998, 1999) shows (for German) that they are. More generally she argues that hierarchy mismatches result in productive variation, whereas correspondences between the hierarchy rankings generally result in consistent linear precedences.

The hierarchy mismatches for ergative–absolutive languages in (8.12) are extensive and Primus uses them to explain why the majority of languages

classified hitherto as object before subject (OS) are those with ergative–absolutive morphology, including OSV languages (Dyirbal, Hurrian, Siuslaw, Kabardian, Fasu) and OVS languages (Apalai, Arecuna, Bacairi, Macushi, Hianacoto, Hishkaryana, Panare, Wayana, Asurini, Oiampi, Teribe, Pari, Jur Luo, Mangarayi). The absolutive is the highest case on the case hierarchy, motivating O before S in the traditional classification, while the agent is the highest theta-role, motivating S before O. Ergative–absolutive languages with the SOV classification favor theta-based positioning, those with OVS or OSV favor positioning based on case. This predicts the existence of both types, though not, as far as I can tell, the much greater frequency of SOV, and additional considerations will need to be invoked to explain this (see §8.5 and §9.3 below).

The specific questions raised by these hierarchies and their correlating properties are the following.

Why do syntactic and morphological rules favor the higher positions, e.g., verb agreement with a nominative case-marked NP rather than with an accusative when there is agreement with only one argument, or relative clause formation being favored with whatever clustering of properties defines a subject in preference to a direct object, etc.?

Why does zero marking favor higher positions? Why isn't dative preferably zero-marked over nominative and accusative or over absolutive and ergative?

Why should linear ordering have anything to do with hierarchy positions? And to the extent that it does, why are higher positions favored to the left of a clause, rather than to the right? In terms of processing, why should the parser or formulator want to receive or produce higher positions first? And why should the case hierarchy which prefers absolutives first be so much weaker than the theta-role hierarchy which prefers agents first in languages with ergative–absolutive morphology?

In the remaining sections of this chapter I consider these questions in relation to the efficiency principles of Chapter 2. Notice first that any explanation we offer for asymmetric rule applications, formal marking, and linear ordering must have *generality*. It would be ad hoc to propose a theory of zero allomorphy, for example, that worked only for case assignment and not for other inflectional morphemes, or a theory of linear ordering that worked only for the linear ordering of arguments and not for other linear orderings as well. The efficiency principles of Chapter 2 do have such generality.

8.3 Minimize Domains

This principle can account for rule applicability asymmetries such as relativization in (8.3). It was defined in (2.1) and is repeated here:

(2.1) *Minimize Domains* (MiD) cf. Hawkins (2004: 31)
The human processor prefers to minimize the connected sequences of linguistic forms and their conventionally associated syntactic and semantic properties in which relations of combination and/or dependency are processed. The degree of this preference is proportional to the number of relations whose domains can be minimized in competing sequences or structures, and to the extent of the minimization difference in each domain.

Relative clause formation involves a dependency between the head of the relative and the position relativized on (i.e., the gap, subcategorizer, or resumptive pronoun coindexed with the head). Relatives on lower hierarchy positions have larger and more complex relativization domains: e.g., a relative on a DO necessarily contains a co-occurring SU (and more phrasal nodes), a relative on a SU need not contain (and regularly does not contain) a DO; a relativized IO contains SU and DO, etc.

It is these co-occurrence asymmetries between arguments plus the added phrasal complexity associated with multiple argument clauses and with OBL and GEN positions that, I believe, underlies the Keenan and Comrie (1977) cut-off points in grammars (8.3). In Hawkins 2004: 177–90, I gave a quantification of this structural complexity that explicitly linked the Accessibility Hierarchy to the Minimize Domains principle of (2.1). This also accounts for corresponding data from performance in languages like English whose grammars permit a range of relativizations from simple to more complex (see Hawkins 1999, 2004 for detailed illustration and literature review). Keenan and S. Hawkins (1987) presented the following results from a controlled repetition experiment conducted on adults and (11-year-old) children. Accuracy levels declined in accordance with the Accessibility Hierarchy supporting Keenan and S. Hawkins's and Keenan and Comrie's claim that there was added complexity and more processing difficulty down the positions of the hierarchy.

(8.13) *Accuracy percentages for English relativizations in a controlled repetition experiment*

	SU	DO	IO	OBL	GEN-SU	GEN-DO
Adults	64%	62.5%	57%	52%	31%	36%
Children	63%	51%	50%	35%	21%	18%

For more recent performance data and relevant discussion, see Diessel (2004) and Diessel and Tomasello (2006); see also §2.2.3.

Further support for this type of explanation comes from the distribution of gaps to resumptive pronouns in languages like Hebrew that have both, i.e., in structures like (8.14). This distribution, across grammars and within languages like Hebrew (cf. Ariel 1999 for detailed exemplification), was discussed in §2.2.3.

(8.14) ha-isha$_i$ [she-Yon natan (la$_i$) et ha-sefer] (Hebrew)
the woman that John Gave (to-her) ACC the book
'the woman John gave the book to'

Gaps are harder to process and prefer the smaller and easier relativization domains associated with higher positions of various complexity hierarchies; resumptive pronouns prefer the lower and harder-to-process environments. Hawkins (1999, 2004) summarized the grammatical patterns as follows:

(8.15) *Gaps*
If a (relative clause) gap is grammatical in position P on a complexity hierarchy H, then gaps will be grammatical on all higher positions of H.

(8.16) *Resumptive Pronouns*
If a resumptive pronoun is grammatical in position P on a complexity hierarchy H, then resumptive pronouns will be grammatical in all lower and more complex positions that can be relativized at all.

The result is an asymmetrical distributional skewing across languages. The figures from Table 2.1 in §2.2.3 showing the distribution of gaps to pronouns in relativizations on the different positions of the Accessibility Hierarchy are summarized in (8.17):

(8.17)

	SU	DO	IO/OBL	GEN
Gaps =	24 [100%]	17 [65%]	6 [25%]	1 [4%]
Pros =	0 [0%]	9 [35%]	18 [75%]	24 [96%]

The reverse directionality of these implications (gaps from low to high, pronouns from high to low) is explained by the structural complexity correlate of the Accessibility Hierarchy, by the added processing load of relativizing on lower positions, and by the fact that resumptive pronouns make processing easier than gaps. The pronouns have the advantage that they provide an explicit argument for predicates within the relative clause, which minimizes various domains for clause-internal processing. They also serve to more clearly identify and flag the position relativized on (see Hawkins 1999).

Many language-particular rule formulations support this pattern of gaps in easier relativization environments and pronouns in more complex ones; see §4.3. The Cantonese data with gaps in smaller and pronouns in larger relativization domains ((2.19) and (2.20) above) are repeated here as (8.18) and (8.19) respectively:

(8.18) a. [Ngo5 ceng2 0_i] go2 di1 pang4jau5$_i$ (Cantonese)
I invite those CL friend
'friends that I invite'

b. *[Ngo5 ceng2 keoi5dei6$_i$)] go2 di1 pang4jau5$_i$
I invite them those CL friend

(8.19) [Ngo5 ceng2 (keoi5dei6$_i$) sik6-faan6] go2 di1 pang4jau5$_i$
I invite (them) eat-rice those CL friend
'friends that I invite to have dinner'

8.4 Minimize Forms

This principle is relevant for rule applicability asymmetries such as verb agreement (8.4) and for the formal marking asymmetries of Primus's (1995, 1999) Case Form Principle (8.5). MiF was defined in (2.8) and is repeated here:

(2.8) *Minimize Forms* (MiF)
The human processor prefers to minimize the formal complexity of each linguistic form F (its phoneme, morpheme, word, or phrasal units) and the number of forms with unique conventionalized property assignments, thereby assigning more properties to fewer forms. These minimizations apply in proportion to the ease with which a given property P can be assigned in processing to a given F.

As discussed in §2.2, the processing of linguistic forms and of conventionalized property assignments requires effort (for the speaker articulatory as well as processing effort). Minimizing forms and form–property pairings reduces that effort and it does so by fine-tuning the minimization of forms to information that is already active in processing, through accessibility, frequency, and inferencing. MiF states that minimization is accomplished both by reducing the set of formal units in a form or structure and by reducing the number of forms in a language with unique property assignments. And it makes general predictions from which the verb agreement patterns and formal marking asymmetries between arguments can be shown to follow.

There is considerable support for the idea that a reduction in form processing is an advantage; see §2.2. MiF adds a second factor to this reduction in

forms alone, involving form–property pairings in the (grammatical and lexical) conventions of a language. It is not efficient to have a distinct form F in a language for every possible property P that one might wish to express. To do so would greatly increase the number of form–property pairs needed and the length and complexity of each proposition. More frequently selected properties are conventionalized in single lexemes or unique categories, phrases, and constructions. Less frequently used properties must then be expressed through word and phrase combinations and through meaning enrichments in context (cf., e.g., Sperber and Wilson 1995; Levinson 2000). The predictions of MiF were summarized in (2.9):

(2.9) *Form Minimization Predictions*
   a. The formal complexity of each F is reduced in proportion to the frequency of that F and/or the processing ease of assigning a given P to a reduced F (e.g., to zero).
   b. The number of unique F:P1 pairings in a language is reduced by grammaticalizing or lexicalizing a given F:P1 in proportion to the frequency and preferred expressiveness of that P1 in performance.

These predictions are supported by the performance frequency data that Greenberg (1966) gave for his feature hierarchies like Sing > Plur > Dual > Trial/Paucal (recall §2.2.1). The usage figures he cited for Sanskrit noun inflections with different number morphemes were given in (2.11):

(2.11) Singular = 70.3%; Plural = 25.1%; Dual = 4.6% (Sanskrit)

I.e., these hierarchies are **performance frequency rankings** defined on entities within common grammatical and/or semantic domains, and these rankings are reflected in cross-linguistic patterns that conventionalize morphosyntax and allomorphy in accordance with (2.9a,b). The ultimate causes of the frequencies can be quite diverse, of course, and include real-world frequencies of occurrence, communicative biases in favor of animates rather than inanimates, and syntactic and semantic complexity. What is significant is that grammars have conventionalized form–meaning pairings and allomorphy distinctions in accordance with performance frequencies, whatever their causes. The greater the frequency of each of these categories, the more accessible, and the easier each is to assign in performance, as a default or as an expectation. As a result less formal marking and fewer morphemes and words are needed in grammars to signal the relevant property values; see §2.2.

Let us now relate these considerations to the verb agreement (8.4) and case morphology patterns (8.8). (2.9a) leads to a general prediction for the amount

of formal marking down all performance frequency hierarchies, defined in (2.12) repeated here from §2.2.1:

(2.12) *Quantitative Formal Marking Prediction*
For each hierarchy H the amount of formal marking (i.e., phonological and morphological complexity) will be greater or equal down each hierarchy position.

Corresponding to the relative frequency data of Sanskrit noun inflections, we find that the amount of formal marking increases (or at least does not decrease) down the number hierarchy across languages. In Manam the 3rd singular suffix on nouns is zero, the 3rd plural is *-di*, the 3rd dual is *-di-a-ru*, and the 3rd paucal is *-di-a-to* (Lichtenberk 1983). Formal marking increases from singular to plural, and from plural to dual, and is equal from dual to paucal.

(2.12) can subsume Primus's Case Form Principle for case marking (repeated here for convenience) under a more general prediction.

(8.5) *Case Form Principle*
The higher the rank of a case on the Case Hierarchy of a language . . . , the less complex is, as a preference, its morpho-phonological realization. *Corollary*: where there is a case system in which just one case has zero allomorphs that case will be the one with the highest rank on the Case Hierarchy.

The existence of languages like Avar with a zero absolutive and explicit ergative, see (8.6), (and of languages with zero nominatives and explicit accusatives) is now predicted by (2.12) and (8.5) in conjunction with her case hierarchies in (8.8). The reverse of these patterns (with, e.g., zero ergatives and explicit absolutives) is predicted not to occur.

This same logic can be extended to account for some more subtle differences in case marking in particular languages, for example in 'differential object marking' languages like Hebrew, Turkish, and Spanish (see Aissen 2003). These languages differentiate between NPs based on their definiteness or animacy and a long-standing explanatory intuition has been that the more subject-like (animate or definite) a direct object is, the more likely it is to be explicitly case-marked, thereby disambiguating subject from object. But Newmeyer (2005: 158) points out that differential object marking is also productive in environments where its absence could not possibly lead to ambiguity and he proposes instead a more general frequency-based explanation along the lines discussed here: there is simply more explicit coding for the less frequent types of direct object (animates, etc.), in differential object-marking languages. This is exactly what (2.9a) asserts: less formal marking in proportion to the frequency of occurrence in a given grammatical domain.

(2.9b) leads to prediction (2.13) for morphological inventories:

(2.13) *Morphological Inventory Prediction*
For each hierarchy H (A > B > C) if a language assigns at least one morpheme uniquely to C, then it assigns at least one uniquely to B; if it assigns at least one uniquely to B, it does so to A.

A distinct dual implies a distinct plural and singular in the grammar of Sanskrit, and a distinct dative implies a distinct accusative and nominative in the case grammar of Latin and German (or a distinct ergative and absolutive in Basque). In effect, the existence of a certain morphological number or case low in the hierarchy implies the existence of unique and differentiated numbers and cases in all higher positions.

(2.13) accounts for the verb agreement pattern of (8.4), repeated here:

(8.4) *Verb agreement rules across lgs*
SU only: (Nom–Acc) French, Finnish, Tamil
(Abs–Erg) Avar, Khinalug
SU & DO only: (Nom–Acc) Hungarian, Mordva
(Abs–Erg) Eskimo, Coos

If a language has a verb agreement rule with a single argument and assigns an affix that agrees with it in some way, this affix will be linked to the most frequently occurring argument, i.e., the subject. More precisely it will be linked to the highest position on the relevant hierarchy that has been conventionalized for agreement in the language in question. This could be the nominative-marked argument, or the absolutive one, or the highest c-commanding NP, and so on (Primus 1999). If a language has verb agreement with a second argument, the affix signaling this agreement will be linked to the next most frequently occurring argument on the relevant hierarchy, direct object, or accusative, ergative, and so on.

The dual predictions of (2.12) and (2.13) (derived from (2.9ab) and ultimately from MiF (2.8)) can now clarify an apparent contradiction that exists between verb agreement and case marking. The existence of overt agreement is preferred with subjects across languages, whereas overt case marking is least preferred on subjects! This contradiction is apparent only and reflects the existence (2.9b) versus the formal marking (2.9a) of morphological distinctions. The existence of a grammaticalized verb agreement rule is *directly* proportional to the relative frequencies of the arguments that the verb agrees with, with subject agreement more common than direct object agreement. More precisely, verb agreement with a nominative case-marked NP is more common than agreement with an accusative, agreement with an absolutive is more common than with an ergative, etc. depending on which specific

hierarchy the agreement rule is linked to in Primus's theory; cf. (8.8)–(8.10). The amount of formal marking for cases, on the other hand, is *inversely* proportional to the frequencies of these arguments (see (8.5)). More frequent arguments are more often reflected in morphological distinctions and in rules of morphosyntax that refer to them, therefore, but they are at the same time less explicitly and formally coded, as in all markedness hierarchies.

There is a further corollary and a prediction made by these hierarchies for the formal complexity of the verb agreement affixes themselves. Like the case affixes on nouns, the amount of formal marking on verbs should decline (or be equal) for agreement affixes linked to the different positions on the relevant hierarchies, with, e.g., nominative and absolutive agreement affixes having more zero allomorphy than accusative and ergative respectively. This deserves systematic investigation and cross-linguistic quantification using Primus's (1999) generalized hierarchy theory, but has not been attempted here.

8.5 Maximize Online Processing

This principle is relevant for the linear precedence correlates of the argument hierarchies in (8.8)–(8.12) and is repeated here from §2.3.

(2.26) *Maximize Online Processing* (MaOP), cf. Hawkins (2004: 51)
The human processor prefers to maximize the set of properties that are assignable to each item X as X is processed, thereby increasing O(nline) P(roperty) to U(ltimate) P(roperty) ratios. The maximization difference between competing orders and structures will be a function of the number of properties that are unassigned or misassigned to X in a structure/sequence S, compared with the number in an alternative.

MaOP motivates a number of linear ordering preferences across languages, that are summarized in (2.27) in §2.3, including a displaced filler Wh-phrase being preposed to the left of its (gap-containing) clause (§2.3.1 and §7.9), a topic to the left of its dependent predication in topic-prominent languages (§2.3.2), and so on. The reverse orderings are argued to be less efficient when there are asymmetric dependencies like this. If B depends on A for certain property assignments, such as gap-filling, coindexation or semantic enrichments of various kinds, then ordering B after A means that these properties can be assigned immediately to B as it is processed by reference to an already accessible A. But if B precedes A, there will be processing delays through 'unassignments' or 'misassignments' (Hawkins 2004: ch. 8). MaOP can also account for the fact that rich case marking is preferred in verb-final languages, thereby providing early access to case marking and theta-role assignments to arguments prior to the verb, while rich verb agreement is preferred in

verb-initial languages, providing early access to argument structure at the verb (see §7.2 and Hawkins 2004: 246–50).

This efficiency logic can explain why configurationally higher (c-commanding) nodes preferably precede lower (c-commanded) ones in structures such as s[NP vp{NP, V}] vs s[vp{NP, V} NP] (see Primus's structural hierarchy (8.9)). In other words, (external) subjects precede verbs and their direct objects within VP regardless of VO or OV ordering (recall the figures in (8.7)). The asymmetrically c-commanded positions are syntactically dependent on c-commanding positions for numerous syntactic relations, such as anaphora (Reinhart 1983), i.e., they involve a dependency of B on A again. Hence positioning B first would delay access to A and to the properties that are assigned to B by reference to it.

Let us now consider whether this efficiency logic can be extended to Primus's (1999) other linear ordering preferences defined on thematic roles and cases; see (8.10) and (8.8) repeated here as (8.20) and (8.21) respectively:

(8.20) Agent precedes Recipient precedes Patient
Experiencer precedes Stimulus
Possessor precedes Possessed

(8.21) Case 1 (Nom/Abs) precedes Case 2 (Acc/Erg) precedes Case 3 (Dative)

Primus (2002) explains the preferred theta-role orderings of (8.20) in terms of the 'thematic dependence' of lower theta-roles on higher theta-roles, and she argues that accessing the higher theta-role first is more efficient for processing. Thus, a patient requires a co-occurring agent and is thematically dependent on it; for something to be a stimulus for an experiencer verb there has to be an actual experiencer; if something is to be possessed, there must be a possessor; and so on. Having the NP with the independent theta-role precede the dependent one means that this independent theta-role is accessible to the processor at the time that the dependent theta-role is processed, permitting an immediate and full online interpretation of the thematically dependent NP, in accordance with MaOP (2.26).

The intuition that underlies her proposal is appealing. A patient is only a patient by virtue of the action of some agent, an item is only possessed if there is a possessor, and so on. Understanding that something is a patient, or possessed, is only complete when one has access to the agent and the possessor respectively, and from this perspective positioning the latter first makes sense. The dependent NP follows the NP on which it is dependent for its full interpretation.

A possible objection to this explanation comes from verb-final languages in which the thematic role interpretation of the second NP can depend even

more strongly on the following verb than on the preceding NP. Compare a structure like *Mary the boy kicked* with *Mary the boy pleased*. The thematic role assigned to *the boy* in these two sentences, patient versus experiencer respectively, is more dependent on the verb than it is on the first NP, *Mary*, and the relative sequencing of one thematic role before the other does not help here. Or at least, if thematic dependence is invoked to explain relative NP orders, one might expect verbs to precede their non-subject arguments in transitive clauses in all languages.

A solution to this problem can be found in the fact that verbs are also dependent semantically on their VP-internal arguments, and hence the dependencies go in both directions, from V to NP and from NP to V. Recall the different interpretations of *run* in *run the race*/*run the water*/*run the advertisement*/*run the campaign*, etc. as discussed in Keenan's (1979) paper on function versus argument categories. The V-NP relationship is therefore a symmetrical one in terms of these semantic dependencies and interpretations and in online processing, and this is what I have argued in §7.2 and in Hawkins (2002, 2004) ultimately underlies the symmetrical ordering possibilities that result in productive VO and OV languages. Between agents and patients, on the other hand, and between possessors and possessed, there is a genuine asymmetry and so Primus's explanation for the asymmetrical linear precedence patterns of (8.20) can go through.

A similar explanation for the preferred ordering of cases in (8.21), however, may be more problematic. It is true that there are asymmetries between the cases: an accusative is generally assigned as a second case only in the presence of a nominative. But there is a difference here between this asymmetry in grammar and a dependency in processing. When parsing an accusative, the parser does not generally need prior access to a nominative in order to recognize the accusative as such. On the contrary, accusatives are generally distinctively case-marked, while nominatives may be zero. Similarly the parsing of ergatives does not require prior access to absolutives in order to be recognizably ergative since ergatives are generally distinctively case-marked (Comrie 1978; Dixon 1994). Hence the hierarchic asymmetry between these cases in grammars does not seem to translate into a linear ordering advantage for positioning the higher case first so that the value of the lower case can be recognized and immediately assigned by reference to the already parsed higher case, in accordance with MaOP.

This may explain why in hierarchy mismatches such as (8.11) and (8.12) the theta-role hierarchies are so much more powerful than the case hierarchies and result in plentiful SO classifications but infrequent OS in ergative–absolutive languages (recall §8.2 above). In effect, prior positioning of the agent generally outranks prior positioning of the morphologically highest

case, here the absolutive. The linear ordering preference for cases, (8.21), is evidently weaker than that for thematic roles, (8.20), which suggests that its causality might be different.

Gibson (1998: 59) offers a processing explanation for the [nominative] subject before [accusative] object ordering preference, in languages like Finnish and German, and we might invoke his explanatory intuition in this context (recall §3.5), but in modified form. He argues that ordering object before subject in these languages, in effect accusative before nominative, poses greater working memory demands and is more complex in his working memory model, since the initial accusative requires and predicts a co-occurring nominative, whereas a nominative does not require and predict an accusative. Hence there is more online memory cost associated with accusative before nominative and greater working memory load in processing.

I have argued in §3.5 against the precise form of this explanation (see also Hawkins 2002, 2004, 2007) on the grounds that there are numerous linear ordering patterns across languages that go against it. It is often not the case that XY is preferred over YX when Y predicts X and not vice versa. Topic-marked phrases (e.g., Japanese *wa*) predict a following predication and are generally initial. Fronted Wh-words predict a following gap, and Wh is almost never moved to the right (Hawkins 2002; Polinsky 2002). Complementizers are cross-linguistically preferred *before* the subordinate clauses that they construct (CPs) and generally predict (Hawkins 1990). For ergative case marking (almost all of them SOV and VSO languages) the ergative is the predicting case which requires an accompanying absolutive and most of these languages position the ergative first.

But let us now reconsider Gibson's idea within the context of a broader multi-factor processing approach to grammars. The linear orderings that are exceptions to his working memory load explanation, the fronted Wh-words and topics, etc., can be motivated by their competing MaOP advantages (2.26). The dependent categories B prefer to have prior access to the As on which they are dependent in processing. In other words, there are independent reasons for these linear orderings that override working memory load increases on these occasions and that could be argued to make the added working memory load tolerable.

Similarly if we accept Primus's (1999) explanation for the thematic role orderings in (8.20), we have a ready explanation for the ergative-before-absolutive preference in ergative languages. What we are then missing is a convincing explanation for the competing absolutive-before-ergative positioning in a minority of these, since we have suggested that MaOP does not provide a good processing explanation for them. Gibson's working memory load account gives us an alternative that we can appeal to instead. MaOP is

generally the stronger principle that overrides this working memory motive in Wh and topic positioning, and there are also independent factors that favor initial complementizers before a predicted complement clause as well (cf. §7.5). In ergative languages the strength of hierarchy (8.21) positioning absolutives first is weaker than that of hierarchy (8.20) positioning agents first. This would follow if working memory load is a weaker factor in this construction and MaOP a stronger one. I return in the next chapter (§9.3) to a consideration of the relative strength between these two ordering principles, within the context of a broader discussion of competing motivations. For nominative-accusative languages, notice that there is no conflict: both MaOP and Gibson's working memory account go together to predict consistent nominative-before-accusative orders.

I began this chapter by examining patterns of asymmetry between the arguments of multi-argument clauses, namely co-occurrence asymmetries, rule applicability asymmetries, formal marking asymmetries, and linear ordering asymmetries (§8.1). These asymmetries have been described in terms of hierarchies and various predictions have been formulated using them for constraints on variation, both in performance and across grammars (§8.2). In the remaining sections I argued that these hierarchies are explainable in terms of processing and performance, and specifically in terms of the efficiency principles defined in Chapter 2, Minimize Domains (§8.3), Minimize Forms (§8.4), and Maximize Online Processing (this section). I have also suggested that a prediction-based principle proposed by Ted Gibson in terms of working memory load has a role to play in its interaction with these principles in explaining some linear orderings derived from Primus's case hierarchy (8.21).

9

Multiple factors in performance and grammars and their interaction

In the course of this book I have presented evidence for three general principles of efficiency (MiD, MiF, and MaOP, §§2.1–3.3) and for more specific principles derived from them such as the head adjacency generalization of §5.1 and Fillers First (§2.3.1). I have also appealed to and adopted some additional ideas from the recent psycholinguistic literature on ease of processing such as predictability (see §2.2.2 and §3.1). The result is a model of efficiency principles of different types and of different levels of generality that makes predictions for performance data and for cross-linguistic variation in grammars. An important part of any such model involves the manner in which principles interact. Sometimes they cooperate and reinforce each other; sometimes they compete and the competitions may result in variation, one variant following the one principle, another the other. I have made frequent reference to such interactions when discussing various datasets in previous chapters. But I have not yet pulled things together and given the matter the attention it deserves. For example, one principle may be stronger than another, in general or in a particular grammatical construction. Why should this be, and can we predict which will be the stronger one? Each of my general principles is also gradient in the sense that it may apply in a stronger or weaker way to a given structure, depending on, e.g., the weight difference between phrases to be linearized. This gradience will have consequences for the ability of one principle to withstand competition from another. Principles can also reinforce one another and a structure that they predict collectively will be more favored than an alternative motivated by fewer principles.

In this chapter I focus specifically on the interaction of multiple principles in an attempt to better understand how principles work together. This is, I submit, not something that is well understood in the current research literature. There are many interesting observations regarding gradience, cooperation, and relative strength among competing principles, but these are often just described or stipulated, and there is no general understanding of why principles should interact and exhibit the relative strengths that they do.

I believe that many patterns of interaction, even in grammars, can be explained when viewed from an efficiency and ease of processing perspective. Just as efficiency and ease of processing can be argued to structure the basic principles of performance and grammar themselves—ultimately MiD, MaOP, and MiF—so too the manner of their cooperation and their relative strength in competition seem to be determined by these same general forces of efficiency and ease of processing. In §§9.1–9.3 I return to some relevant data from performance and grammars introduced in earlier chapters, and I add new data, in order to illustrate this. Three general patterns will be presented that can be seen when principles interact, involving their degree of preference (§9.1), cooperation (§9.2), and competition (§9.3). §9.4 gives a brief summary of the central point of this chapter.

9.1 Pattern One: Degree of preference

(9.1) **Pattern One**: each principle P applies to predict a set of outputs {P}, as opposed to a competing set {P'} possibly empty, in proportion to the degree of preference defined by P for {P} over {P'} within a theory of processing ease and efficiency.

We have seen several instances of efficiency principles applying in proportion to the degree of preference they define for competing structures of one and the same type. The relative weight of postverbal constituents in English determines the strength with which Minimize Domains applies to effect a short-before-long linearization in permutable postverbal PPs such as (9.2), discussed in §2.1 and §5.2:

(9.2) a. The man [$_{VP}$ looked [$_{PP_1}$ for his son] [$_{PP_2}$ in the dark and quite derelict building]]
1 2 3 4 5

b. The man [$_{VP}$ looked [$_{PP_2}$ in the dark and quite derelict building] [$_{PP_1}$ for his son]]
1 2 3 4 5 6 7 8 9

The bigger the weight difference, the higher the percentage of short-before-long PPs (PPS = shorter PP, PPL = longer PP) as shown in (9.3):

(9.3)

n = 323	PPL > PPS by 1 word	by 2–4	by 5–6	by 7+
[V PPS PPL]	60% (58)	86% (108)	94% (31)	99% (68)
[V PPL PPS]	40% (38)	14% (17)	6% (2)	1% (1)

The gradient effects of MiD can also be seen in the conventionalized constituent orders of grammars in cross-linguistic samples. The Greenbergian correlations (§5.4) show that lexical heads and other phrasal constructing categories are preferably positioned so as to minimize phrasal combination domains (see (5.12)) throughout a sentence, in both head-initial and head-final structures. The data in (9.4a) and (b) reveal a strong correlation between verb-initial order in VP and prepositions before NP, i.e., with V and P adjacent, and between verb-final order and postpositions after NP, again with V and P adjacent (in phrases corresponding to [*drove* [*to the cinema*]] and [[*the cinema to*] *drove*]] respectively). (9.4c) and (9.4d) with non-adjacent V and P (i.e., [*drove* [*the cinema to*]] and [[*to the cinema*] *drove*] respectively) are both infrequent. (9.4a) and (b) involve the most minimal domains for phrase structure processing: the adjacent V and P permit construction of VP and attachment of two immediate constituents to it, V and PP, resulting in optimal IC-to-word ratios. (9.4c) and (9.4d) involve longer domains for VP construction and attachment of these ICs to it, and lower ratios, since V and P are separated by NP (see (5.6) in §5.1):

(9.4) a. $[_{VP}$ V $[_{PP}$ P NP]] = 161 (41%) IC-to-word: 2/2 = 100%

b. $[[_{NP}$ P $_{PP}]$ V $_{VP}]$ = 204 (52%) IC-to-word: 2/2 = 100%

c. $[_{VP}$ V $[_{NP}$ P $_{PP}]]$ = 18 (5%) IC-to-word: 2/4 = 50% (NP=2)

d. $[[_{PP}$ P NP] V $_{VP}]$ = 6 (2%) IC-to-word: 2/4 = 50%

MiD-preferred (9.4a)+(b) = 365/389 (94%)

Departures from the optimal head-adjacent orderings in grammars show the gradient nature of this MiD preference clearly. These departures, like (9.4c) and (d), are much less common than the orderings that are optimal. Dryer (1992) has shown that when heads are not adjacent, there is a cross-linguistic preference for single-word items to intervene between them before whole phrases do. Single-word adjectives can intervene between a verb and a noun in English (***saw** yellow **books***), but adjective phrases are postposed to the right of N (***saw books** yellow with age*). We have also seen evidence in §5.5 for a center-embedding hierarchy in which constituents can intervene between P and N in proportion to their respective weights. The single-word adjective intervenes in English and many other languages ([***under*** [*yellow* ***books***]]) whereas the adjective phrase cannot (*[***under*** [*yellow with age* ***books***]]). More generally, the more structurally complex the center-embedded constituent is (Rel > Possp > Adj) and the longer the PCD for its containing phrase (here PP), the fewer languages there are. For the environment $[_{PP}$ P [__ N $_{NP}]]$ we therefore have the following center-embedding hierarchy, with language frequencies for the combination in question shown in percentages and taken from the sample of Hawkins (1983):

(9.5)	Prep lgs:	AdjN	32%	NAdj	68%
		PosspN	12%	NPossp	88%
		RelN	1%	NRel	99%

In §2.2 we saw evidence for Minimize Forms applying in a gradient manner to reduce morphemes in phonological size and assign priority to their grammaticalization and lexicalization in proportion to relative frequency of occurrence in performance. This explains the structuring of the Greenbergian markedness hierarchies and other morphological patterns in §2.2.1 and §8.4. One such was the Quantitative Formal Marking Prediction, defined in (2.12):

(2.12) *Quantitative Formal Marking Prediction*
For each hierarchy H the amount of formal marking (i.e., phonological and morphological complexity) will be greater or equal down each hierarchy position.

Corresponding to the relative frequency data for Sanskrit noun inflections shown in (9.6), the amount of formal marking increases (or at least does not decrease) down the number hierarchy Sing > Plur > Dual > Trial/Paucal across grammars (cf. (2.10) in §2.2.1). In Manam the 3rd singular suffix on nouns is zero, the 3rd plural is *-di*, the 3rd dual is *-di-a-ru*, and the 3rd paucal is *-di-a-to* (Lichtenberk 1983).

(9.6) Singular = 70.3%; Plural = 25.1%; Dual = 4.6% (Sanskrit)

Formal marking increases from singular to plural in Manam, and from plural to dual, and is equal from dual to paucal, matching the relative frequencies for the relevant categories in languages such as Sanskrit. This illustrates and supports the gradient nature of MiF in performance and grammars.

9.2 Pattern Two: Cooperation

(9.7) **Pattern Two:** the more principles there are that define a collective preference for a common set of outputs {P}, as opposed to a proper subset or complement set {P'} motivated by fewer principles, the greater will be the preference for and size of {P} compared with {P'}.

9.2.1 Performance

Performance data illustrating Pattern Two come from the stronger preference in English for postverbal prepositional phrases and particles to be adjacent to a verb when that adjacency is supported both by syntactic weight and by lexical-semantic dependencies with the verb, rather than by just one of these principles alone.

Recall that a phrasal combination domain (2.3) is a domain for the processing of a syntactic relation of phrasal combination or sisterhood between constituents (§2.1). Some of these sisters may contract additional relations of a semantic or lexical nature, whose processing requires a lexical domain (5.15) sufficient to recognize the lexical combination in question and assign the appropriate syntactic and semantic properties to it; see §5.2 and Hawkins (2004: 111–17). Entailment tests were used in Hawkins (2004) and in Lohse, Hawkins, and Wasow (2004) to provide evidence for this lexical listing of verb–preposition and verb–particle combinations.

For example, the preposition *on* in *John counted on his father* cannot be processed independently of *counted*, nor can *counted* be processed independently of *on*. This sentence does not entail 'John counted,' removing the PP, nor does it entail 'John did something on his father,' removing the verb and replacing it with a general Pro-Verb (Hawkins 2000, 2004). *John played on the playground* is different: it entails 'John played' and 'John did something on the playground,' i.e., the meanings of *played* and *on* can be processed independently of one another.

V-PP combinations such as *count on his father* show a stronger adjacency to V than those classified as independent by the entailment tests. There is cooperation and reinforcement among the principles favoring adjacency between V and the dependent PP when two prepositional phrases follow a verb (§5.2). Specifically in the data of Hawkins (2000): 82 percent (265/323) had the syntactically preferred short PP adjacent to V preceding a longer one, i.e., their PCDs were minimal; 73 percent (151/206) had a lexically interdependent PP adjacent to V, i.e., their lexical processing domains were minimal. For PPs that were *both* shorter *and* lexically dependent, the adjacency rate to V was 96 percent, which was (statistically) significantly higher than for each factor alone. This supports Pattern Two (see §9.3.1 for further data and discussion of these postverbal PPs in this context).

Similarly, *John lifted the child up* with a verb–particle construction entails both 'John lifted the child' and 'the child WENT up.' *John washed the dishes up* does not entail 'the dishes WENT up,' but it does entail 'John washed the dishes' (Lohse et al. 2004). These tests provide evidence for lexical domains of processing in addition to syntactic domains such as PCDs.

The dependency of a particle (as in *wash X up*) has a significant and consistent effect on its preferred adjacency to the verb compared with independent particles (*lift X up*): NPs that intervene and split verbs and particles creating larger PCDs are systematically less tolerated when the particles are also dependent and when lexical-semantic processing prefers a smaller distance between them as well, again supporting Pattern Two. This is shown in Figure 9.1 taken from Lohse et al. (2004) and Hawkins (2011). The split particle

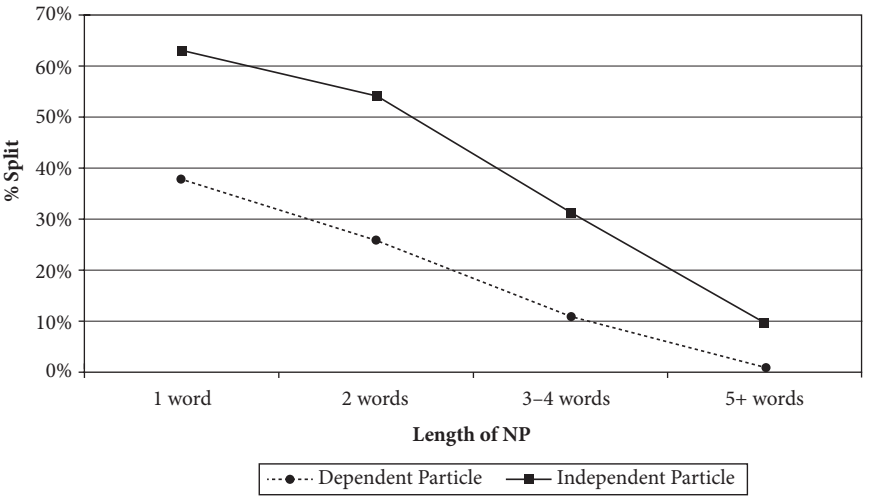


FIGURE 9.1 Split vs joined particles by NP length and particle type

ratios (V NP Part) for dependent particles are systematically lower than those for independent particles. More processing domains linking the verb and the particle result in tighter adjacency.

Similarly, when relative clauses have zero relativizers, the reduced relative clause becomes dependent on the head noun for recognition as an actual relative clause and for attachment to N within the NP. This added dependency results in more adjacent head noun–relative clause combinations than we find with explicit relativizers. Once again, more processing domains linking the noun and the relative clause result in tighter adjacency, supporting Pattern Two. In §6.3.1 I gave the following figures, from Quirk (1957) and Lohse (2000) respectively.

(6.14) a. *Restrictive (non-subject) relatives adjacent to the head noun*
explicit relativizer = 60% (327) zero = 40% (222)
b. *Restrictive (non-subject) relatives separated from the head noun*
explicit relativizer = 94% (58) zero = 6% (4)

(6.15) a. *Separated relatives in NP-internal position*
which/that = 72% (142) zero = 28% (54)
b. *Separated relatives in NP-external position (i.e. extraposed)*
which/that = 94% (17) zero = 6% (1)

The ratio of zero to explicit relativizers is as shown here for adjacent (6.14a) versus non-adjacent (6.14b), and for separated but still NP-internal (6.15a) versus separated and NP-external relatives (6.15b). Zero clearly prefers

adjacency over non-adjacency and NP-internal separation over NP-external. If we compare the quantities vertically, we see that as many as 98 percent (222/226) of zero relatives stand adjacent to their heads in (6.14), while 98 percent (54/55) prefer NP-internal position if they are separated at all. The corresponding adjacency figures for relatives with explicit relativizers are less, at 85 percent (327/385) in (6.14) and 89 percent (142/159) in (6.15).

This strong correlation between zero relatives and their adjacency to the nominal head can be seen, I argued in §6.3.1, as a reflection of the Minimize Forms preference and of Minimize Domains working together. Removing the relativizer makes the form of the relative clause more minimal. This sets up additional dependencies on the head noun, for recognition of the clause as a relative clause and attachment to the head. These dependencies prefer minimization. When head noun and relative are adjacent, both general principles cooperate, but when there is separation, there is increasing competition and we see more and more explicit relativizers.

The explanation for this adjacency of zero-marked relatives to their heads is essentially the same, I argued in §6.3.1, as that which I have given for the preferred adjacency of lexically dependent elements to their heads in the V-PP-PP data of §5.2 and in the verb particles of Figure 9.1. In both cases there are additional dependencies, going beyond phrasal construction and parsing, that result in an adjacency preference by Minimize Domains (2.1): semantic dependencies in the one case involving the need for the verb to access the preposition and particle and/or vice versa, in order for a correct meaning to be assigned; syntactic dependencies in the other whereby the type of clause in question and its adjunct status are recognized by reference to the head noun. The explanation for both types of adjacency is predicted by Minimize Domains (2.1); there are more grammatical or lexical relations to be processed in zero dependencies and semantic dependencies, and hence more processing domains in both cases, making adjacency or proximity between the interdependent elements preferred, in performance and (when conventionalized) in grammars too. See Hawkins (2003) and (2004: 148–65) for further elaboration on this explanation for the adjacency of zero-marked dependencies, and further examples of the descriptive generalization that underlies it involving tighter adjacency between more interdependent elements.

9.2.2 Grammars

Grammatical support for Pattern Two comes from certain basic word order types that are supported by three versus two versus just one preference principle, with correlating quantities of grammars. For example, in §7.8.5 I examined the ordering of oblique (X) phrases in relation to a verb and

(direct) object in the data of Dryer with Gensler (2005), e.g., *(Mary) opened the door with the key*. Three preference patterns were visible in the language quantities. I derived the first two (verb and object adjacency and object and X on the same side) from Minimize Domains, the third (object before X) from an independent Argument Precedence principle (§7.8.4). These three preferences are shown in the three columns of (9.8). A plus sign in the row alongside each basic word order, VOX, XVO, etc., indicates that the preference in question is satisfied in that basic order, a minus sign indicates that it is not.

(9.8)

	V & O Adjacency	*O & X on Same Side*	*O before X*
VOX	+	+	+
XVO	+	−	−
VXO	−	+	−
XOV	+	+	−
OXV	−	+	+
OVX	+	−	+

VOX languages conform to all three patterns. The OV language types, on the other hand, each conform to only two, while the other VO competitors (XVO and VXO) conform to at most one. Correlating with this are the quantities of grammars shown in (9.9). The more principles a grammatical type exemplifies, the more such grammars there are, which supports Pattern Two.

(9.9)

	VOX	>	XOV/OXV/OVX	>	XVO/VXO
	3		2		1
Lgs:	189		45/23/37		3/0

Notice that it is not being claimed here that these three principles are equally strong, merely that they have a mutually reinforcing and additive effect. Some indication of what their relative strengths are can be seen in the quantitative data of Dryer and Gensler (2005) and in the numbers of grammars that conform to each, as summarized in §7.8. In that section I also gave a detailed discussion of what the plausible advantages might be for each of these preferences, in terms of processing ease and efficiency, following Hawkins (2008).

Further grammatical support for Pattern Two comes from the fact that complements prefer adjacency to their heads over adjuncts in the basic order conventions of numerous phrases in English and other languages and are generated adjacent to the head in the phrase structure grammars of Jackendoff (1977) and Pollard and Sag (1987); see Hawkins (2004: 131–6) for details. Tomlin's (1986) Verb–Object Bonding principle provides further cross-linguistic support for this since he points to languages in which it is impossible or dispreferred for adjuncts to intervene between a verbal head and its subcategorized direct object complement.

Why should complements be closer to the head? The basic reason I offered is that complements also prefer head adjacency over adjuncts in performance, in the double PP data of §5.2. The explanation for this in turn is that there are more combinatorial and/or dependency relations that link complements to their heads than link adjuncts to their heads. Hence there are more processing domains that can be minimized when complement and head are adjacent, in accordance with Pattern Two.

First, recall that complements are listed in a lexical co-occurrence frame that is defined by, and activated in online processing by, a specific head such as a verb. Adjuncts are not so listed and they occur in a wide variety of phrases with which they are semantically compatible (see Pollard and Sag 1994). Processing the lexical co-occurrence frame of a verb within a particular sentence therefore favors a minimal Lexical Domain (5.15) linking that verb to the complements or arguments selected. The reality of this preference was seen in the lexical dependency effects between verbs and PPs in the English double PP data.

Second, there are more productive relations of semantic and syntactic interdependency between heads and complements than between heads and adjuncts—i.e., there are more cases in which the meaning or grammar of one of these categories requires access to the other for its assignment. Processing these interdependencies favors minimal domains.

A direct object, for example, receives its thematic from the transitive verb, typically a subtype of Dowty's (1991) Proto-Patient role, depending on the particular verb. Adjuncts, by contrast, do not receive thematic roles (see Grimshaw 1990). A direct object is also syntactically required by a transitive verb for grammaticality, whereas adjuncts are not syntactically required sisters. Assigning the appropriate thematic role to the required direct object favors a minimal lexical domain.

Conversely transitive verbs regularly undergo Keenan's (1979) 'function category range reduction' by the direct object NP. Recall the different senses of *run* in *run the race*, *run the water*, *run the advertisement*, and so on. The verb is lexically dependent on the direct object for its precise interpretation, but it is not dependent on an accompanying adjunct such as *in the afternoon* (in, e.g., *run the race in the afternoon*). Similarly intransitive verbs are frequently dependent on PP complements for their interpretation, as we saw with *count on X*.

Grammaticalization of empty categories and of zero agreement markers, etc., within small processing domains provides yet more support for Pattern Two (9.7), matching the performance support for adjacency of zero forms summarized in the last section. Thus, there may be zero agreement among noun phrase constituents that are adjacent to one another within an NP, but explicit agreement when they are separated, as in Warlpiri; see §6.3.1 and

Moravcsik's (1995: 471) agreement universal cited in that section and repeated here:

(6.18) *Moravcsik's Agreement Universal*
If agreement through case copying applies to NP constituents that are adjacent, it applies to those that are non-adjacent.

Grammars have conventionalized the same preference for zero marking that we saw in the English relative clause data summarized in §9.2.1. Both Minimize Forms (2.8) and Minimize Domains (2.1) are fulfilled under adjacency, but as the distance between them expands, cooperation between these principles increasingly becomes competition; see §9.3.2.

A similar conventionalized adjacency effect is found in the optional case-marking deletion rule of Kannada exemplified in (2.53) in §2.5.3 and repeated here (from Bhat 1991: 35):

(2.53) a. Avanu ondu pustaka-vannu bareda. (Kannada)
he (NOM) one book-ACC wrote
'He wrote a book.'

b. Avanu ondu pustaka bareda.
he (NOM) one book wrote

c. A: pustaka-vannu avanu bareda.
That book-ACC he (NOM) wrote
'That book he wrote.'

The accusative *-vannu* can be removed when the direct object (*pustaka* 'book') is adjacent to the final verb in (2.53b), but not when it is moved away as in (2.53c). The minimal direct object form is only grammatical when the verb on which the direct object has become dependent for assignment of the accusative case and patient thematic role is within the most minimal processing domain, i.e., when both Minimize Forms (2.8) and Minimize Domains (2.1) are complied with.

Finally, gaps in relative clause positions high on the Keenan–Comrie Accessibility Hierarchy are motivated by Minimize Forms and they occur in minimal domains for relative clause processing, in accordance with Minimize Domains; see Table 2.1 in §2.2.3. As these domains expand, cooperation between them again becomes competition; see §9.3.2.

9.3 Pattern Three: A Competition Hypothesis

(9.10) **Pattern Three:** when there is competition between two principles A and B, where each predicts a (partially) different set of outputs in

> the competing structures to which both apply, {A} versus {B}, then each continues to apply (a) in proportion to its intrinsic degree of preference, as in Pattern One; (b) each may be reinforced by supporting principles, as in Pattern Two; but (c) the relative strength of A over B will be in proportion to the relative degree of preference defined for {A} > {B} within a theory of processing ease and efficiency.

According to Pattern Three the relative strength of principles under competition, even in grammars, reflects the overall processing ease and efficiency advantages for the set {A} over {B}, taking into account the intrinsic strength of each and reinforcement from supporting principles if any. But what is it precisely about the recorded examples in the literature of one principle being stronger than another that actually makes them so? The general hypothesis I propose is that, just as the principles themselves (Minimize Domains, etc.) are ultimately explained by ease of processing and efficiency, so too is their **relative** strength when they are opposed to one another. In other words, the outputs {A} defined by A are preferred over outputs {B} defined by B, because the {A} structures are easier or more efficient to process when they are in conflict.

9.3.1 Performance

Notice first that even when there is competition, the degrees of preference defined by single principles (Pattern One) and cooperation and reinforcement among principles (Pattern Two) continue to assert themselves where relevant, as specified in (a) and (b) respectively of (9.10). Consider (9.10a), for example. In the postverbal English data of (9.2) comprising two PPs, if one of them is a PP that is lexically dependent on V (Pd) (e.g., *count* [*on his father*] [*in his youth*]), i.e., V-Pd-Pi, and prefers adjacency to V for lexical processing, it will nonetheless be postposed to the right of a lexically independent PP (Pi), in proportion to its heaviness, e.g., *count* [*in his youth*] [*on his much-adored and generous father*]. In other words, syntactic weight continues to assert itself gradiently, even when opposed by a lexical adjacency principle for a V-Pd pair (see §5.2). This is shown in the data of (9.11) in which a heavy dependent Pd like *on his . . . father* is postposed to the right and separated from V in proportion to its increasing weight relative to an independent Pi like *in his youth*.

(9.11)

	Pd = Pi	Pd > Pi by 1word	by 2–4	by 5+
[V Pd Pi]	83% (24)	74% (17)	33% (6)	7% (2)
[V Pi Pd]	17% (5)	26% (6)	67% (12)	93% (28)

Wiechmann and Lohmann (2013) have investigated these postverbal PPs of English further, testing the patterns of Hawkins (2000) in a larger and more controlled database (over three times more data points), and this leads them to revise some of my earlier conclusions. Their findings are directly relevant to Pattern Three (9.10). First of all, they show that although syntactic weight applies to more structures to define an ordering preference between PPs based on short before long (many double PP sequences do not include one that is dependent, leaving weight alone to decide the ordering preference), when the two actually compete it is the lexical-semantic dependencies that are stronger and that exert a greater preference for adjacency with V. The lexical-semantic factor is 1.7 times stronger than the syntactic one in these conflicts in their data. They arise whenever a semantically dependent phrase is longer than an independent one, e.g., *dwelling* [*for a few moments*] [*on what the police have done or have not done*] (V-Pi-Pd), to take an example from their data. An arrangement minimizing the domain for lexical-semantic processing was favored over one minimizing the domain for syntactic processing significantly more often in these cases of conflict. This result also replicates the experimental findings conducted on the use of these structures by older and younger adults in Marblestone (2007). Again there was a preference for short lexical-semantic domains of processing over short syntactic ones in cases of conflict in her data.

The explanation that suggests itself for this is that the processing disadvantages of not assigning meanings online to words like *count* and *on*, or *dwell* and *on*, are more severe than delays in syntactic phrasal processing through weight—i.e., in this structure at least, the disadvantages for lexical-semantic processing are stronger and priority is given to their more rapid resolution. Wiechmann and Lohmann (2013) formulate this idea in terms of a model of language production like Levelt's (1989) whereby conceptual-semantic planning is prior to constituent assembly. According to this view the combination *count on his father* is part of an early planning phase verbalizing the central proposition. Independent PPs like the time adjunct *in his youth* may be planned either later than this or simultaneously with it, resulting in less adjacency overall and in later production.

Whether this is a viable explanation or not will need to await further analysis of head-final languages, in which adjunct phrases often precede arguments and semantically dependent phrases (see §7.8). In the interim I suggest instead that the greater priority for positioning semantically interdependent words together follows simply from efficiency since meanings can be assigned sooner to words in online processing this way. The general principle I would invoke here is Maximize Online Processing (2.26) in §2.3. Within the set of linguistic properties that are assigned to forms in online

processing, semantic properties are the most important ones for communication. Hence, it is important to maximize the assignment of these properties to each form as it is encountered. Whatever the best formulation and modeling of this insight turns out to be, I would argue that the relative strength of conflicting principles in this case follows from general principles of processing ease and efficiency in language use.

Wiechmann and Lohmann (2013) provide further quantitative and statistical data of relevance to (9.10). In addition to measuring lexical-semantic dependency domains and phrasal constituency domains for postverbal PP structures, they examine two further factors which have been argued in the literature to affect postverbal orderings and which are tested in Hawkins (2000). The first is the Manner > Place > Time (MPT) generalization (cf. Quirk, Greenbaum, Leech, and Svartvik 1985), and the second is information structure, specifically the Given before New principle (see below). Both of these are shown by Wiechmann and Lohmann to be weak predictors, compared to lexical-semantic dependencies and syntax, but they are not insignificant, as I had argued. MPT raises the accuracy of ordering predictions (compared to the combined lexical semantic and syntactic preferences), specifically it raises the 'classification accuracy,' from 74.6 percent to 76.8 percent. Adding information status raises it further from 76.8 percent to 78.7 percent. What is interesting about these small increases is that they show the effect of reinforcing principles amidst competition, i.e., they show the effect of (b) in (9.10). When additional factors such as MPT and Given before New are at play, they also contribute to the overall ordering prediction for PPs and they reinforce one other, therefore, and help to offset a competing principle.

Why should information status be among the weakest predictors? Specifically, why should pragmatic principles such as Given before New be so much less powerful than both phrasal syntax and lexical-semantic dependencies in these unconventionalized orderings of PPs? Not only Wiechmann and Lohmann, but also Wasow (2002), Kizach (2010), and Hawkins (1994) have all argued that syntactic processing preferences are stronger in these cases and trump pragmatic ones most of the time in datasets that systematically control for both.

A possible answer is this. Notice first that pragmatic meanings differ from lexical-semantic dependencies in that one word does not need access to another in order for the processor to actually assign a basic meaning to it. Rather, these pragmatic meanings involve the larger organization of the discourse and are more 'stylistic.' Second, the set of structures that involve different weights between categories and on which MiD defines a preference for shorter PCDs far exceeds the set which are differentiated pragmatically

(either by Given before New or by New before Given) and for which a pragmatic theory defines a preference. Every syntactic category (V, Adj, etc.) in every phrase and on every occasion of use is subject to a MiD ordering preference. Pragmatic differences in information structure, on the other hand, affect a much smaller number of clauses in a language, namely those containing two NPs that differ in the pragmatic values in question. Hence, even in the competing structures containing two NPs, the syntactic processing advantages continue to apply both to the NPs and their ordering and to all the other categories and phrases in the clauses that contain them, making the overall advantages of following MiD more extensive than the advantages afforded by information structure ordering. Third, there is also a certain indeterminacy in the pragmatics literature over whether Given before New or New before Given is the preferred ordering for discourse processing (see Hawkins 1994 for detailed discussion and critique), which makes it less clear what the overall benefits of different pragmatic orders are for processing ease and efficiency. Given items have the advantage of recent activation, but they delay new and newsworthy information. There are also issues of cross-linguistic generality to be considered here, and especially the behavior of head-final languages with respect to pragmatic ordering (see Hawkins 1994 for discussion of Japanese in this regard).

The descriptive and explanatory details of these pragmatic principles of ordering require further investigation. What is emerging from controlled datasets, however, is that pragmatic principles are much weaker than lexical-semantic and syntactic ones. The general hypothesis I propose for this is that they have much less of an effect on overall processing ease and efficiency, for the kinds of reasons enumerated here, and hence they win fewer of the conflicts in the competition sets {A} and {B} in (9.10), though they do have a small reinforcing effect when they support stronger principles (see (9.7)).

As a further example of competition between principles in performance, consider the decreasing numbers of English zero relativizers in relative clauses that are increasingly separated from their heads (§6.3.1 and §9.2.1), alternating with explicit relativizers, and the decreasing numbers of gaps in Hebrew relatives, alternating with resumptive pronouns (Ariel 1999 and §2.2.3). These distributions are in accordance with (c) in Pattern Three (9.10) which states that the relative strength of A over B will be in proportion to the relative degree of preference defined for {A} > {B} within a theory of processing ease and efficiency. As the domains for processing the relative clause and its link to the head noun expand in English, the Minimize Domains pressure to lessen the dependencies of the reduced clause on the head increases, and explicit relativizers become more common, trumping the Minimize Forms preference for zero. Similarly, expanding the filler-gap domains in Hebrew increases the

pressure for more local processing of grammatical and lexical relations within the relative clause, on the basis of an explicit resumptive pronoun. Once again MiD increasingly trumps MiF in more complex relatives.

9.3.2 Grammars

In the grammars of relative clauses across languages we have seen competition between a filler before gap (or filler before subcategorizer) processing preference (Fodor 1978; Hawkins 1999, 2004) and MiD's word order preferences. The head noun is the filler in a relative clause construction (*the book*i *that John read 0*i) and NRel order is preferred in all languages by the filler-before-gap preference, ed from Maximize Online Processing (2.26), whereas MiD (2.1) prefers NRel in VO languages and RelN in OV. This is shown in (9.12).

(9.12)

	Minimize Domains	*Fillers before Gaps*
VO & NRel	+	+
VO & RelN	−	−
OV & RelN	+	−
OV & NRel	−	+

Empirically, Fillers before Gaps is the stronger principle in this competition within OV languages in the *WALS* data of Dryer (2005d) and Dryer with Gensler (2005), as shown in (9.13) (which was given as (7.11) in §7.3):

(9.13)

WALS	*Rel–Noun*	*Noun–Rel or Mixed/Correlative/Other*
Rigid SOV {O, X} V	50% (17)	50% (17)
Non-rigid SOV OVX	0% (0)	100% (17)

Only one third of OV languages (17/51) in *WALS* have the MiD-preferred RelN, all of them rigid OV languages whose containing head-final phrases (VP, etc.) define the strongest overall preference for head finality in NPs; see §7.3. Two thirds of OV languages (34/51) have NRel with the head noun filler before its gap, therefore, either as the only strategy or in combination with some other one.

Why should this be? I hypothesize here that the processing disadvantages of gaps before fillers are quantifiably very severe (through garden paths/misassignments and unassignments online; see Hawkins 2004: 205–10, Fodor 1978, and §7.3), and they outweigh the MiD disadvantages of head-ordering inconsistencies resulting from NRel in combination with SOV (e.g., Persian and German). Not only are regular garden paths and unassignments extremely inefficient for the RelN structure, but the disadvantages of the reverse NRel in an OV language can be, and are, readily relieved by postposing transformations

such as Extraposition from NP (see §3.3 and Hawkins 2004: 142–6 for detailed discussion of this option in German with supporting data taken from Uszkoreit et al. 1998). I suggest that the greater relative strength of Fillers before Gaps (motivated by MaOP) over MiD in the competition sets here follows from the greater overall efficiency of giving priority to the former—i.e., relative strength does not need to be stipulated. It follows from the overall processing ease and efficiency benefits stemming from one or the other principle in these areas where they compete.

Matching the performance preference for more resumptive pronouns rather than gaps in larger relativization domains, we have grammatical hierarchies such as the Keenan–Comrie Accessibility Hierarchy (2.17) in which gaps are increasingly replaced by obligatory resumptive pronouns down the hierarchy, i.e., in what I have argued to be more complex and less minimal environments; see Table 2.1 in §2.2.3 and also Hawkins (1999, 2004: ch.7) for this and other complexity hierarchies. The competition between Minimize Forms and Minimize Domains is resolved in a very principled way in these conventionalized data, with MiD increasingly winning the competition over the MiF-motivated gap as relativization domains expand and the pressure for local processing of verb–argument relations within the relative clause increases accordingly. The choice between {A} and {B} structures here (gaps versus resumptive pronouns) is in proportion to the overall preference for one over the other, therefore, with MiD asserting itself increasingly through resumptive pronouns in relatives on DO (35 percent), on IO/OBL (75 percent), and on GEN (96 percent).

In §7.9 I explained the decreasing numbers of Wh-fronting grammars in terms of a similar competition, between Maximize Online Processing and Minimize Domains. The frequency ranking was given as (7.54) based on figures taken from Dryer (1991, 2005j) and Dryer with Gensler (2005):

(7.54) *Wh-fronting frequency ranking*
V-initial > svo > non-rigid sov > rigid sov

Increasing distance between Wh and the verb, as subcategorizer and/or constructor of the phrasal projection containing an adjunct Wh, creates increasingly long domains for processing the relationship between them. The fronting of Wh and creation of a filler before gap/subcategorizer structure is motivated by MaOP. The increasing potential distances to the verb are gradually disfavored by MiD and result in more local strategies for Wh-questions, especially in non-rigid and rigid SOV languages, namely Wh-in-situ and Wh-attachment to the verb (see Kim 1988 and Hawkins 1994: 373–9). The increasing pressure exerted by MiD in these {A} versus {B} competition sets provides further support for (c) in Pattern Three (9.10).

Competition between MiF and both MaOP and MiD was discussed in §2.5.3 and §7.2 involving the grammaticalization of rich case marking across languages. In languages in which verbs occur late it was argued to be advantageous to have NPs clearly case-marked so that cases and thematic roles can be assigned early online, prior to the verb, in accordance with MaOP. Supporting data for this were given in (2.52), repeated here:

(2.52)	V-final with R-case:	89% (Nichols 1986), 62–72% (Dryer 2002; Hawkins 2004: 249)
	V-initial with R-case:	38%, 41–47% respectively
	SVO with R-case:	14–20% (Dryer 2002)

The Kannada data (2.53) in §2.5.3 and §9.2.2 show that zero case marking in this verb-final language is only tolerated when the NP stands immediately adjacent to the verb on which it now depends for case and thematic role assignment—i.e., MiF and MiD work together here when zero forms are adjacent to their interdependent elements (the verb), and there is only a short delay in case and thematic role assignment, at variance with MaOP. But both MaOP and MiD trump MiF and insist on explicit case marking when the NP bearing the case is not adjacent to the verb. Again we can make sense of the competition here between {A} (case-marked) and {B} (non-case-marked) NPs in terms of our efficiency principles and we can understand their relative strength by examining the degree to which they apply in different structural variants.

As a final example of competing principles, and involving case marking again, recall Primus's (1999) hierarchies for case and theta-roles discussed in §8.5:

(9.14)	Theta:	Ag [Erg]	>	Pat [Abs]
	Case:	Abs [Pat]	>	Erg [Ag]

Primus shows that these and other hierarchies underlying NP ordering are preferably linearized from left to right, e.g., agent before patient, which in ergative languages creates competition between the theta-role hierarchy and her case hierarchy which prefers absolutive [patient] before ergative [agent]. The great majority of ergative languages favor the theta-role hierarchy, but a minority favor Abs before Erg basic order and these have often been misanalyzed, Primus argues, as OSV languages (Dyirbal, Hurrian, Siuslaw, Kabardian, Fasu) or OVS (Apalai, Arecuna, Bacairi, Macushi, Hianacoto, Hishkaryana, Panare, Wayana, Asurini, Oiampi, Teribe, Pari, Jur Luo, Mangarayi). Agent and patient are semantic notions and their preferred relative ordering is argued by Primus to have online processing advantages that trump those of the purely formal case morphology hierarchy, i.e., absolutive before

ergative. A patient is thematically and conceptually dependent on an agent for bringing about the action described in an event, and it is therefore preferable for event construal to have the agent presented first. Similarly it is preferable to position an experiencer before a stimulus and a possessor before the possessed: the latter depend on the former for their status and construal in each case. The conflicting preference, I suggested in §8.5, may be motivated by a variant of the memory cost explanation that Gibson (1998) proposed for nominative before accusative in languages with these cases, namely: an accusative predicts a following nominative, whereas an initial nominative does not necessarily predict an accusative (since nominatives are most often the subjects of intransitive verbs). Similarly, an ergative predicts a following absolutive, adding to online working memory load. The reverse absolutive before ergative reduces this. The fact that the great majority of ergative languages position the ergative first shows that the ordering preference based on thematic roles is the stronger principle.

I suggested, following Primus's discussion, that agent and patient and the other theta-roles of her hierarchies are semantic notions and that their preferred relative ordering has both online processing advantages as well as online advantages for the communication of meanings (in the form of efficient transitive event descriptions) that trump the memory cost advantages deriving from a purely formal case morphology hierarchy. This is a further instance of (c) in (9.10).

9.4 Summary

I have defined three general patterns in this chapter that can be seen when multiple principles cooperate and compete in data from performance and grammars: (9.1), (9.7), and (9.10). My basic hypothesis, derived from these patterns, is that the strength of principles, and especially their relative strength in competition, reflects the processing ease and efficiency of the competing datasets {A} versus {B} predicted by different principles. One principle A will be stronger than another B in proportion to the greater processing ease and efficiency of {A} over {B}, according to some independently motivated theory of ease and efficiency of the kind I have been defining and illustrating in this book. The patterned preferences in the data come from single principles applying in a gradient manner (§9.1), from the cumulative preferences of mutually reinforcing principles (§9.2), and from the relative strengths of both cooperating and competing principles (§9.3). By examining the interaction of principles from the same explanatory perspective that I have proposed for the principles themselves (in terms of the Performance–Grammar

Correspondence Hypothesis in §1.1), the patterns of interaction begin to make sense. In this perspective processing ease and efficiency are paramount.

There are many details in the performance and grammatical patterns I have discussed that require further investigation, and there are many details of their interaction that await further specification. But I believe that this general way of looking at cooperation and competition may help us in better understanding other examples in the literature: it suggests further areas to examine, and it sheds some light on why principles interact in the ways they do.

In a nutshell, think of the interaction between principles in terms of ease and efficiency of use. If you believe, as I have argued throughout this book, that ease and efficiency shape all of performance and grammar and their respective principles, then it makes sense that the way principles interact with one another, and their relative strengths in conflict, should be determined by ease and efficiency as well. That is the basic idea I propose.

10

Conclusions

I began this book by formulating the PGCH in §1.1:

(1.1) *Performance–Grammar Correspondence Hypothesis* (PGCH)
Grammars have conventionalized syntactic structures in proportion to their degree of preference in performance, as evidenced by patterns of selection in corpora and by ease of processing in psycholinguistic experiments.

The hypothesis predicts that there should be a correspondence between preferences in performance, in languages with choices and variants, and preferences in the grammatical conventions of languages with fewer choices and variants. These predictions have been tested throughout this book on comparative data from performance and grammars, involving several types of structures. They reveal striking parallels. In §§2.1–2.3 I formulated three general principles of efficiency that appear to underlie the shared patterns of performance and grammars, Minimize Domains, Minimize Forms, and Maximize Online Processing. They are repeated here:

(2.1) *Minimize Domains* (MiD)
The human processor prefers to minimize the connected sequences of linguistic forms and their conventionally associated syntactic and semantic properties in which relations of combination and/or dependency are processed. The degree of this preference is proportional to the number of relations whose domains can be minimized in competing sequences or structures, and to the extent of the minimization difference in each domain.

Combination = Two categories A and B are in a relation of combination iff they occur within the same syntactic mother phrase or maximal projection (phrasal combination), or if they occur within the same lexical co-occurrence frame (lexical combination).

Dependency = A category B is dependent on a category A iff the parsing of B requires access to A for the assignment of syntactic or semantic properties to B with respect to which B is zero-specified or ambiguously or polysemously specified.

(2.8) *Minimize Forms* (MiF)
The human processor prefers to minimize the formal complexity of each linguistic form F (its phoneme, morpheme, word, or phrasal units) and the number of forms with unique conventionalized property assignments, thereby assigning more properties to fewer forms. These minimizations apply in proportion to the ease with which a given property P can be assigned in processing to a given F.

(2.26) *Maximize Online Processing* (MaOP)
The human processor prefers to maximize the set of properties that are assignable to each item X as X is processed, thereby increasing O(nline) P(roperty) to U(ltimate) P(roperty) ratios. The maximization difference between competing orders and structures will be a function of the number of properties that are unassigned or misassigned to X in a structure/sequence S, compared with the number in an alternative.

The data of this book are not predicted by theories in which grammars are autonomous from processing, i.e., by theories in which processing gives nothing back to grammars, even though grammars may be constantly accessed in language use. If the PGCH is on the right track, therefore, there are some clear consequences for grammatical theory, for typology, for historical linguistics, and also for psycholinguistics, that we need to make explicit.

I shall first summarize some of the data supporting the PGCH (§10.1). I then discuss some consequences for grammatical theory and language typology (§10.2). Finally, I draw attention to bigger issues in linguistics and psycholinguistics that are raised by this research program that can benefit from further research and clarification.

10.1 Support for the PGCH

The precise predictions made by the PGCH for grammars were formulated in §1.2 as follows:

(1.2) *Grammatical predictions of the PGCH*
(a) If a structure A is preferred over an A' of the same structural type in performance, then A will be more productively grammaticalized, in proportion to its degree of preference; if A and A' are more equally favored, then A and A' will both be productive in grammars.
(b) If there is a preference ranking A>B>C>D among structures of a common type in performance, then there will be a corresponding

hierarchy of grammatical conventions (with cut-off points and declining frequencies of languages).

(c) If two preferences P and P' are in (partial) opposition, then there will be variation in performance and grammars, with both P and P' being realized, each in proportion to its degree of motivation in a given language structure.

Consider (1.2a). In §2.1, §5.2, and §5.3 I summarized performance data from English and Japanese showing clear preferences for minimal domains in phrase structure production and recognition (EIC effects in my earlier work, MiD effects here). Word orders in which one structure had a smaller domain than another were preferred (short constituent before long constituent, or long before short), and the preferences were highly correlated with degree of minimization in the respective order. Orderings that were equally minimal were all productive. These same preferences can be seen in the Greenbergian head-ordering correlations (§5.1 and §5.4), i.e., in the grammatical preference for minimal phrasal combination domains with adjacent heads. V-initial order in the VP co-occurring with a P-initial PP is significantly more frequent than V-initial with P-final. A V-final VP co-occurring with a P-final PP is just as minimal as V-initial with P-initial, and both head orderings are highly productive in grammars (see §7.1). Domain minimizations can also be seen in the different directionalities for rearrangement rules, to the right or left as a function of the position of the constructing or head category within the rearranged constituent, and in numerous disharmonic word orders (see §5.8.1, §7.5, and §7.8.3, and Hawkins 2004: 138–42).

Performance data from English also showed a clear preference for complement PPs to be closer to their verbal heads than adjunct PPs (§5.2), a preference that was similarly derived from MiD as formulated in (2.1). Complements involve the processing of more relations than adjuncts (both phrasal combination and lexical co-occurrence relations of different kinds), each of which prefers a minimal processing domain with the verb, resulting in stronger adjacency. This same preference for complement adjacency can be seen in grammatical models and in cross-linguistic ordering preferences (see §5.4 and §9.3).

Data from English involving minimal formal marking in adjunct clauses revealed another performance pattern of relevance to (1.2a) (see §6.3.1). Under conditions of adjacency to the head, both minimal and explicit forms (relative clauses with zero and explicit relativizers) are productive, whereas under non-adjacency formal marking is preferred and zero is dispreferred in proportion to the distance from the head. This is in accordance with MiF, since increasing the distance between adjunct and head makes their processing more difficult,

which increases the preference for explicit surface coding of the adjunct relation. Similarly in grammars that have conventionalized the presence or absence of morphosyntactic attachment marking on noun phrase constituents according to their adjacency to one another—for example, in the form of case copying—it is the zero option that is systematically disallowed under non-adjacency (§6.3.1). Under adjacency either formal marking or zero or both can be conventionalized. Dispreferred performance options are the first to be eliminated in grammars (recall the discussion of mechanisms of change in §4.4), while productive options in performance are productive across grammars.

In §3.5 I referred to data from Japanese, German, and Finnish supporting the asymmetric preference for subjects before objects in performance (SOV over OSV, or SVO over OVS). Correspondingly a clear asymmetric preference for subjects before objects can be found in the basic ordering conventions of grammars, again in accordance with (1.2a). (For more fine-tuned ordering predictions for arguments of the verb, in terms of their case marking and thematic roles, and for discussion of these in terms of MaOP, see §8.5.)

With respect to (1.2b) I have argued that many hierarchies structuring grammatical conventions are correlated with preference patterns and ease of processing in performance. Greenberg's (1966) markedness hierarchies were supported both by performance data and by grammars (see §2.2.1). Declining performance frequencies are matched by increases in formal marking in morphological paradigms (e.g., by non-zero versus zero forms). This is in accordance with MiF, since less frequent categories require more activation and processing effort in order to assign the relevant properties, whence more explicit coding down the hierarchies (see (2.9a) in §2.2). Declining performance frequencies are also matched by decreases in the very existence of uniquely distinctive morphemes (duals and datives, etc.) and by declining feature combinations in paradigms (plural *and* gender distinctions, etc.). This is also in accordance with MiF's predictions (see (2.9b) in §2.2).

In §2.2.3 I discussed Keenan and Comrie's Accessibility Hierarchy from the perspective of (1.2b), building on their intuition that this grammatical hierarchy is ultimately structured by declining ease of processing through increasing complexity and declining frequencies of use for the different hierarchy positions. Performance data from languages like Hebrew support the declining preference for gaps in more complex and less frequent filler-gap domains, and the increasing complementary preference for resumptive pronouns. Since resumptive pronouns make identification of the position relativized on easier, MiF enabled me to predict that there would be reverse hierarchies for gaps and resumptive pronouns across grammars too, with gaps declining in increasingly

large domains, and pronouns in smaller ones. The Keenan–Comrie language sample supported this prediction exceptionlessly (see Table 2.1 in §2.2.3).

Other complexity hierarchies for filler-gap domains with Wh-movement were discussed in §7.9 and involved the increasing distance between a fronted Wh and the verb as subcategorizer. The frequency of the Wh-fronting strategy declines, the further the verb occurs to the right, in accordance with MiD. In Hawkins (2004: 192–203) I also discussed some additional hierarchies for gaps linked to Wh-words and to heads of relative clauses involving more complex phrasal embeddings for gaps than those in the Keenan–Comrie hierarchy. Performance data supporting these hierarchies were again given from different languages with numerous options (including English, Akan, and Scandinavian), while the grammatical cut-offs and conversions from gaps to resumptive pronouns were supported by grammatical data from representative languages.

There are also hierarchies of permitted center-embedded constituents in grammatical conventions, and I argued (in §5.5) that these hierarchies are structured by degrees of minimization in phrasal combination domains (in accordance with MiD): a more complex center-embedded constituent implies the grammaticality of less complex constituents (whose presence within the relevant configuration is independently motivated to occur there).

Prediction (1.2c) links the variation that results from competing preferences in performance to variation across grammars. For example, in the English double PP data of §2.1, §5.2, and §9.3.1 there is sometimes a competition between domain minimization for lexical dependency processing and minimization for phrasal combination processing, namely when the longer PP is in a lexical dependency relation with the verb. The short-before-long preference conflicts here with the preferred adjacency of lexical dependents. Numerous minimization competitions of this sort between different processing domains were illustrated in §2.5 and §9.3, with predicted consequences for structural variants in performance and grammars. For example, Extraposition from NP in German often benefits one phrasal combination domain (for VP) at the expense of another (for NP), and some quite precise predictions can be made for when it will apply and when not, predictions that have been supported by corpus data (see §2.5.1).

In grammars, the typology of relative clause ordering exhibits variation in OV languages (productive RelN and NRel, plus different strategies altogether; see §§7.3–7.4) compared with unity in VO languages (almost all are NRel). This was attributed to the conflicting demands of MiD and MaOP in OV languages and to their mutual reinforcement in VO (§7.3 and §9.3.2). More generally, I argued in Hawkins (2004: 244–6) that asymmetric ordering preferences in performance and grammars arise when different preferences

converge on a single order, whereas symmetries (VO and OV) have many complementary preferences and dispreferences in each order. Whichever order is selected in a grammar will conventionalize one preference at the expense of another, and compensating syntactic and morphosyntactic properties can then be predicted as a consequence. For example, relative clause extrapositions of NRel are found in (non-rigid) OV languages: NRel is motivated by MaOP, permitting early construction of a clausal S through a relativizer or complementizer, while the extrapositions are often good for MiD as well; see §7.3 and §9.3.2. Languages with RelN and OV have consistent head finality, which is good for MiD (§7.1), but the reduced expressiveness within the relative clause compared with NRel is also good for MaOP since it limits the complexity of left-branching preverbal constituents and the delays in attaching them to a mother node within a bottom-up parse (see §5.8.3, §7.5, and §7.6). Similarly, more case morphology in verb-late languages is good for online argument processing and argument–predicate matching in a structure that optimizes online verb processing, all of which is in accordance with MaOP, but at the expense of more form processing, contrary to MiF; see §2.5.3 and §7.2.

Both performance data (from corpora and processing experiments) and grammars provide support for the efficiency principles MiD, MiF, and MaOP, therefore. In many cases I could point to closely matched data from usage and grammar, as in several of the examples just summarized, and these provide the strongest support for the PGCH. In other cases I have given data from performance only or from grammars only supporting the general principles.

10.2 The performance basis of grammatical conventions and cross-linguistic patterns

At several points I have contrasted the grammatical generalizations and universals proposed in this book with those formulated in purely grammatical models. For example, in §5.4 I referred to the head-ordering parameter of generative grammar in the context of Greenberg's correlations, and I discussed Dryer's (1992) finding that there are systematic exceptions to these when the non-head category is a non-branching category like a single-word adjective. The single-word/multi-word distinction involves what is most plausibly a difference in terminal elements among items of the same category type (e.g., adjective phrases). To the extent that single-word and multi-word phrases have different positionings, with one being regularly subsumed under head ordering and the other not, their common category type precludes a straightforward grammatical generalization distinguishing them. And even if some difference in grammatical (sub)type could be justified, it would still

need to be shown why one (sub)type regularly follows consistent head ordering, whereas the other may or may not.

Performance data, and the principle of MiD, provide us with an explanation (see §5.5). When one phrasal combination domain competes with another in performance and involves a difference of only a single word, the preference for the more minimal domain is very small, since there is little gain and both orders are productive. But when multiple words are added to a domain, there is a much stronger preference for minimization (see, e.g., the English postverbal PP ordering data of §5.2). So too in grammars. The difference between [$_{VP}$ V [$_{NP}$ N Adj]] and [$_{VP}$ V [Adj N $_{NP}$]] (with a single-word adjective) is small, and cross-linguistic variation can be expected, destroying the generality of a head-ordering parameter that favors V-initial order in VP and N-initial in NP. The difference between [$_{VP}$ V [$_{NP}$ N AdjP]] and [$_{VP}$ V [AdjP N] $_{NP}$] (containing a multi-word adjective phrase) is much greater, however, so consistent head ordering is predicted (contrast English *yellow book* with *book yellow with age*). More generally, the head-ordering parameter is a quantitative preference, not an absolute universal, and the degree of consistency for different pairs of categories is predicted by MiD to be proportional to the (average) complexity of the potentially intervening non-heads. This results in the prediction that single-word adjectives can precede nouns more readily than adjective phrases in head-initial languages (and follow nouns more readily in head-final languages). It results in the center-embedding hierarchy for NP constituents ((5.24) in §5.5)) whereby single-word adjectives can also precede nouns more readily than (the typically longer) possessive phrases in head-initial languages, while these latter can precede nouns more readily than relative clauses (which are typically longer still). The important point here is that there is a principled performance basis for consistent head ordering and for the different levels of consistency that can be expected with different pairs of categories, as a function of their terminal content. Grammars do not predict this. They can stipulate a head-ordering parameter, but the stipulation is partly right, partly wrong, and of course unexplanatory. Performance preferences and principles such as MiD can predict the degrees of adherence to consistent head ordering, and they provide us with an independent motivation and explanation for this fundamental property of grammars.

I have also given an explanation for adjacency effects in the internal structuring of phrases in phrase structure grammars (see §5.4 and §9.2, and also Hawkins 2004: 131–6). Why should complements be closer to their heads than adjuncts? Performance data reveal adjacency preferences in proportion to the number of processing relations between items. The more relations there are, the tighter the adjacency. Complements (being lexically listed

co-occurrences with frequent dependencies in processing) exhibit many more such relations and dependencies with a head than adjuncts. Hence when a basic order is conventionalized in a grammar, the convention fixes preferences that can be seen in performance in the absence of the convention (as in the English postverbal PP orderings of §5.2), and complements stand closer to their heads.

In §§5.6–5.8 I compared the predictions made for infrequent and non-occurring word orders across languages by a purely grammatical principle, FOFC (the Final-Over-Final Constraint), with the efficiency principles developed here. FOFC rules out various combinations of a head-initial daughter phrase within a head-final mother, i.e., [[$_{YP}$ Y ZP] X $_{XP}$]. I argued that there are, first of all, many other word order 'disharmonies' that are just as infrequent or non-occurring as these, and that FOFC does not account for them whereas efficiency does, e.g., those with the mirror-image [$_{XP}$ X [ZP Y $_{YP}$]]. FOFC is not general enough, therefore. Second, efficiency appears to be able to subsume FOFC under a larger and more inclusive generalization deriving from a performance theory. And third, it gives us an explanation for the disharmonies, whereas FOFC is stipulated. Infrequent and non-occurring word order conventions are those that involve non-minimal processing domains (at variance with MiD) and delayed phrasal construction (at variance with MaOP); see §5.7. I argued (in §5.8.4) that grammatical studies testing FOFC have nonetheless been descriptively invaluable and have given a level of precision to word order studies going beyond that of typology. But a functional typology incorporating a theory of processing ease and efficiency gives a better explanation.

We have seen numerous examples in this book of zero marking in morphosyntax and in morphological paradigms alternating with non-zero and explicit forms. The distribution of zero forms on Greenberg's markedness hierarchies is systematically skewed to the higher positions, i.e., to those that are most frequent in performance and easiest to access and process (see §2.2.1 and §8.4). In §6.3.1 it was shown that agreement marking on noun phrase constituents could be zero under adjacency, alternating with explicit marking when there were discontinuities (as in Warlpiri case copying). This mirrored the strong preference for explicit relativizers in English under non-adjacency. Relativizer omission has also been shown to be correlated with the predictability of actually having a relative clause following various noun phrase elements (different determiners, superlative adjectives, etc.; see §2.2.2).

The clear generalization that emerges from these facts is that zero is associated with greater processing ease, i.e., with greater frequency and accessibility and predictability, permitting the assignment of default values like nominative (rather than accusative and dative) and singularity (rather than

plurality and duality). Zero is associated with smaller domains linking the zero-marked constituents to others, and also with the level of expectancy for actually having a following relative clause. Hence there are declining zeros down the various hierarchies. These associations between zero and ease of processing enrichment all follow from MiF. But the greater efficiency of more minimal forms then sets up additional dependencies on accessible items for the processing of syntactic and semantic relations, and these additional dependencies favor more minimal domains (for gap-filling, phrasal attachment, thematic role assignment, etc.), by MiD. Once again, there appears to be no purely grammatical explanation for these hierarchies and for the respective environments favoring zero and non-zero forms. Grammars have conventionalized the preferences of performance, and to explain grammars you have to explain performance and provide a theory of ease of processing.

The Keenan–Comrie Accessibility Hierarchy has been one of the most discussed empirical generalizations about relative clause formation across languages, since it was first proposed in the 1970s. It has received very little attention in the formal grammatical literature, however. Hierarchies of relativizable positions within a clause do not fit well with a parameterized theory of Universal Grammar when they do not involve the kinds of bounding nodes (IP, NP, etc.) that generative grammarians normally invoke for filler-gap dependencies. Yet there is a large body of psycholinguistic evidence to support Keenan and Comrie's original intuition of increasing processing complexity down the hierarchy (see Hawkins 1999 and 2004: 171–7 for summaries). Once again, gaps (i.e., zero marking of the position relativized on) favor the easier-to-process and higher positions on the hierarchy (§2.2.3). The gaps are efficient by MiF since a nominal head is already available and accessible in this structure. As the size of filler-gap domains increases, this benefit is outweighed by the increasing non-minimality of both coindexing and lexical processing domains for the subcategorizer, and gaps are replaced by resumptive pronouns. The pronouns provide local domains for lexical co-occurrence processing and for adjunct processing in the event that the filler is an adjunct, and only co-indexing with the filler needs a larger processing domain (see Hawkins 2004: 183–6 for illustrative quantification of these different domains in relative clause processing). The result is systematic and reverse implicational statements in these relative clause and Wh-movement universals, with gaps going from low to high positions, and resumptive pronouns from high to low on these hierarchies.

This mirror-image structuring is not predicted by any grammatical theory that I know of, and for good reason. It is not a consequence of any deep grammatical principle. It reflects the processing architecture of language users and the interaction between efficiency benefits deriving from less form

processing on the one hand, and from the increasing strain on those benefits in non-minimal filler-gap and lexical domains. This interaction is reflected in performance preferences and in the variation patterns that have become 'frozen' in grammars (see §4.4).

In Chapter 6 a range of grammatical details and variation patterns in noun phrase syntax were discussed that can be described, but not readily explained, in purely grammatical terms. A lot of this variation, I argued, could be better understood in terms of efficiency, and I proposed an Axiom of Constructibility (6.3) and an Axiom of Attachability (6.11). Predictions were made for the existence of alternative constructors of NP, for their possible omission and obligatory presence, for their positioning (see §6.2), and for the existence and nature of different attachment markers, e.g., for attaching possessive phrases to their head nouns (see §6.3.2).

Chapter 7 pulled together ten major typological differences between VO and OV languages, involving their co-occurring properties such as subcategorization and selectional restrictions, different relative clause types, complementizers and their ordering, articles, the ordering of obliques, Wh-fronting, and verb movement. To my knowledge this is the first time that such a typology of verb position correlations has been attempted. Once again, as in Chapters 5 and 6, the variation patterns can be usefully described in formal grammatical terms, but the explanation for the variation points to principles of efficiency and processing ease. In particular, the very existence and productivity of both types can be seen, I argued, as a consequence of their equal and optimal efficiency for phrasal construction by MiD. When heads are consistently initial or consistently final, processing domains for phrase structure recognition and production are optimal. The details of the departures from consistency, the resulting disharmonies and correlations with non-rigid SOV, etc., can also be better understood in efficiency terms, especially in relation to the interaction between MiD and MaOP.

Chapter 8 linked my efficiency principles to various asymmetries between the arguments of verbs, involving their co-occurrence, the rules they undergo, their formal marking, and their linear ordering.

Finally in Chapter 9 I turned to the general question of how multiple principles actually work together to make predictions for performance data and grammars. Consistent with the methodology employed throughout this book, my theorizing about the interaction of multiple factors has been driven by clear empirical patterns. Three types of general patterns have been observed. The first set illustrate a Degree of Preference generalization, whereby data are distributed gradiently, even grammatical data, in accordance with preference principles that derive ultimately from ease and efficiency in language processing (§9.1). The second set illustrate Cooperation, whereby the

more principles there are that define a collective preference for a common set of outputs {P}, the greater is the preference for and size of that {P} (§9.2). The third set of data involve Competition and they suggest that when there is competition between two principles A and B, each may continue to follow Degree of Preference and Cooperation, as before, but crucially the relative strength of their outputs {A} defined by A and {B} defined by B will be in proportion to the overall ease and efficiency of one compared with the other. Just as the basic principles of performance and grammar are shaped by processing ease and efficiency (MiD, MaOP, etc.), so too is the manner of their interaction. This way of looking at cooperation and competition can help us, it was argued, to better understand how principles work together and why they have the relative strengths they do, instead of just observing that they do.

The main point of this section is that principles derived from performance data give us an explanation for many fundamental properties of grammars for which there are either no current principles, or else stipulations with varying degrees of adequacy. To the extent that the PGCH is supported, a theory that accounts only for grammars misses significant generalizations. It also fails to incorporate the very principles that can lead to more adequate grammars, given that there is a correspondence between performance and grammars (§1.1). Formal models that stand the greatest chance of being descriptively adequate, according to this view, are those that can accommodate functional principles as grammatical conventions, such as *Simpler Syntax* (Culicover and Jackendoff 2005; see, e.g., p. 41) and those variants of Optimality Theory that seek some functional grounding and explanation for their constraints; see, e.g., Bresnan, Dingare, and Manning (2001), Bresnan and Aissen (2002), and Haspelmath (1999a). Formalization can make explicit how functional generalizations have been conventionalized exactly and it can define all and only the predicted outputs of the grammar with precision.

10.3 Some bigger issues

The first big issue that needs to be raised in a book about efficiency concerns efficiency itself. What is it? What makes some linguistic structures more efficient than others? I have defined three principles that are supposedly instances of it, on the basis of which numerous predictions have been made and tested. But what do these principles follow from?

In §2.4 I proposed that efficiency relates to the most basic function of language which is, as I see it, to communicate information from the speaker (S) to the hearer (H), and I gave the following definition:

(2.38) Communication is efficient when the message intended by S is delivered to H in rapid time and with the most minimal processing effort that can achieve this communicative goal.

This definition introduced two key criteria that concern the manner in which efficient information is conveyed, namely speed, and degree of processing effort, specifically ease rather than complexity in processing. Timing is relatively straightforward to quantify, but processing ease has a number of possible causes, and the precise interaction between them is not, I believe, very clear or well understood in the psycholinguistic literature. I shall comment on this presently. Notice first that the degree of processing effort that is required for communication to be efficient must be defined in a relative way, namely relative to the communicative goal of successfully delivering the intended message, and this degree can vary, depending on the speaker's assessment of the hearer's state of mind, their mutual knowledge, mutual manifestness, and so on (see Sperber and Wilson 1995; Levinson 2000; Hawkins 1991). Some messages in some speech acts require very little encoding, and very little processing effort, for efficient communication, others require more, and unless the required minimal encoding is given the communicative goal is not met.

The efficiency principle MaOP (2.26) reflects the preference for speed in processing since it sets a premium on introducing grammatical and semantic properties and assigning them to their relevant linguistic forms as quickly as possible (avoiding unassignments and misassignments online). Obviously speech is a linear string of sounds and words, and not all the properties of the message can be presented at the outset. MaOP states that property assignments to forms should not generally be delayed as these forms are processed, resulting in partial assignments and look-ahead to subsequent items. Meanwhile MiD (2.1) and MiF (2.8) capture one major aspect of minimal processing effort: the preferred reduction in word sequences and grammatical structures that must be processed for recognition (and production) of these structures and their meanings (i.e., MiD), and the preferred reduction in the meaningful units themselves, i.e., morphemes and words (MiF).

There are other research programs dedicated to efficiency, which have really taken off in the years since my 2004 book, some of which I referred to in §3.1 (see Jaeger and Tily 2011 for overview). They add exciting new dimensions and predictions to those considered here, especially with respect to online measurements of processing ease and efficiency. Some of the bigger issues that I will raise in this last section with respect to my research program will also

arise in these others, however, and so the points I make here may have more general applicability.

In addition to the definition of efficiency itself, the second big issue that arises in this context is: what exactly makes processing easier? In §3.1 I summarized three big ideas in the processing literature (see that section for precise references, which will be omitted here for the sake of succinctness).

The first involves the size of processing domains in working memory, e.g., domains for filler-gap processing, phrasal combination domains, and so on. The idea of a 'constrained capacity' on working memory is not much help here, I suggested, since all the interesting effects of larger domains (weight effects in ordering, etc.) take place well within whatever capacity constraints there are supposed to be, and since different languages seem to have different capacities. The conclusion I derive from much of the working memory literature, and from comparisons of different domain sizes within and across languages, is simply that the more items there are to process, of relevance to some parsing task, the harder it is—i.e., size is a gradient notion and processing difficulty increases when there are more forms and their properties to process and to hold in working memory, phonological, morphological, syntactic and semantic, in a given domain. And the more grammatical and semantic properties that need to be assigned to a particular construction, the more such processing domains there are, and the greater will be the collective preference for minimizing processing effort, by reducing domain sizes. This is, in essence, what the MiD principle captures.

The second big idea involves interference, and this may also be reflected in MiD's preferences, though less directly. Small domains for, e.g., filler-gap processing will contain fewer alternative possible gap sites (see Hawkins 2004: 198 for discussion of grammaticality and acceptability distinctions that reflect the presence or absence of such alternatives), small domains for argument–predicate matching will contain fewer alternative arguments and predicates that need to be considered, and so on.

The third big idea involves frequency, predictability, expectedness, and surprisal. There are different terms here, and differences in detail, but the basic intuition seems to be much the same: the more expected something is, the easier it is to process. We have seen the linguistic consequences of this in Greenberg's markedness hierarchies (§2.2.1), and in English relativizer omissions (§2.2.2), and in other structures as well, and I have suggested that at least some of these effects can be subsumed under MiF (2.8): the more expected categories are easier to process and have more zero forms, for example. There are many other empirical consequences of predictability and expectedness, however, including for online processing and the distribution of linguistic properties and information (see Jaeger and Tily 2011).

What is currently in need of clarification and of more research is how these big ideas, involving the different sources of ease and complexity in processing, fit together and impact one another. In §§9.2–9.3 I discussed some quite precise patterns of cooperation and competition in both performance data and grammars between the efficiency principles that I have defined here. To the extent that my principles are relevant to and instantiate these big ideas in processing, these patterns can suggest larger principles of interaction. The distribution of gaps to resumptive pronouns in relative clause hierarchies points to what is at first cooperation between MiD and MiF, which increasingly becomes competition in structurally larger relativization domains (§2.2.3, §9.3.1, and §9.3.2). The distribution of relative clause orders within VO and OV languages points to a similar competition between MaOP (favoring fillers before gaps) and MiD (§7.3 and §9.3.2), as does the distribution of Wh-fronting and verb position, as filler-gap domains become larger (§7.9 and §9.3.2). MaOP and MiF compete with respect to the presence or absence of case marking in SOV languages. The former is generally stronger, resulting in explicit marking in these NP-early languages (§2.5.3 and §7.2). If the case marking is removed, the additional dependencies of NP on the verb for case and thematic role assignment create more domains for processing which, by MiD, favor adjacency between that NP and V (§9.2.2 and §9.3.2).

In many of these interactions I was able to see a clear reason for why one principle was stronger than another in the competing datasets ({A} versus {B}; see (9.10) in §9.3), or for why a single principle like MiD made a stronger prediction for one structural variant over another, depending on the overall benefit for the sentence as a whole from, e.g., applying Extraposition from NP or not in German (§2.5.1). To the extent that my MiD principle captures the basic insight of the working memory literature involving domain sizes, and to the extent that MiF captures the predictability and expectedness intuition, these patterns can help us to understand how these larger forces work together, and how they combine or compete to make one structure easier to process than another. My goal here is not actually to address psycholinguistic theories directly. It is a more linguistic goal, to understand structural variation better, and especially cross-linguistic variation. I have pointed to certain revealing patterns in performance and grammars, I have defined efficiency principles that are supported by them, I have proposed some ideas for why they have the relative strengths they do when they cooperate or compete, and I hope that both linguists and psychologists will take note of these patterns and draw the appropriate conclusions from them within their respective theories.

There is one processing puzzle in particular that merits attention with respect to the interaction between these big ideas and which has not been

given the attention it deserves. Predictability clearly aids processing (see the relativizer omission data in §2.2.2 and the references summarized in §3.1). But why does it not also have the reverse effect, of greatly increasing the number of items held in working memory within our various processing domains? If a verb in a verb-early language can activate numerous possible predicate frames before the following argument(s) select(s) from among them (as I suggested in §7.2 based on experimental evidence about temporary ambiguities and garden paths in languages like English), why aren't the processing domains for verb–argument co-occurrences so clogged up with alternative possibilities that they make this a very complex domain? Perhaps to some extent they are, and this is one of the benefits of having the reverse order of arguments before verbs in an OV language. But then why is predictability such an advantage and how does it offset this putative disadvantage? Presumably it benefits online processing when predicted arguments are actually encountered, i.e., the precise benefit may lie in Maximize Online Processing. But is this benefit then actually opposed by Minimize Domains? Is the predictability of argument structure (good for MaOP) opposed in verb-early languages by larger domains of processing (disfavored by MiD), that are replete with activated co-occurrences that will be eliminated once the intended argument has been found? How would we calculate the precise domains in this case (since there are huge numbers of activated co-occurrences for verbs, with the strength of activation for each varying enormously)? And how do we reconcile the MiD disadvantages calculated this way for VO languages with the roughly equal distribution of VO and OV found across languages? This distribution is expected when domains are calculated simply as we have done here, by looking at actually occurring items in the parse string and the distances between their heads, and not including alternative and eventually dismissed activations (see §7.1). Putting this another way, this distribution is expected when we look at integration domains in the sense of Gibson (1998, 2000), but not at prediction domains. I don't see clarity on this issue in the psycholinguistic literature. What I do see clearly is that cross-linguistic data support the simpler definition of syntactic domains and their contents given here, based on actual surface structure items and the grammatical and semantic relations that hold between them.

More generally, I believe that grammatical data (i.e., data involving the rules and conventions of different languages, correlations between properties within grammars, and their distribution throughout the globe) can be extremely useful for psychologists to examine, in addition to their performance data from experiments and corpora. My proposal for weight effects in the performance of different languages, and specifically for mirror-image weight effects reflecting the internal structure of phrases

(see §§5.2–5.3), came not from an examination of the theoretical literature in psycholinguistics, but from Greenberg's word order correlations (§5.4), i.e., from cross-linguistic grammatical typology. The theoretical literature from psycholinguistics that I reviewed suggested that all languages should behave like English and rearrange heavy items to the right (§3.5). But since many grammars do not do this in their rearrangement rules and basic orders, it had to be the case, I reasoned, that heavy items in free word order structures (i.e., in those that have not been fixed by convention; see §4.3) will be *pre*posed to the left in rigid SOV languages, and *post*posed to the right in SVO languages like English, reflecting the general pattern that we see in the word order rules themselves, long before short in the former, short before long in the latter. Increasingly this is being shown to be the case in psycholinguistic experiments on these languages and in corpora (see §3.2 and §3.5).

This inference, from grammatical conventions to performance preferences, is possible precisely because grammars have conventionalized the preferences of performance, in accordance with the PGCH (§1.1). Hence psychologists can benefit from looking, not just at cross-linguistic variation (Jaeger and Norcliffe 2009), but at grammars as well, and they need to help linguists in understanding these conventions better. In addition to weight effects, all the grammatical patterns summarized in the last section (§10.2) provide evidence for performance preferences and their respective strengths. The preferred adjacency of some items to others, hierarchies of different limits on phrasal combination and filler-gap domains, gaps alternating with resumptive pronouns, other alternations of explicit and zero forms, asymmetric orderings for some pairs of elements, symmetrical for others—all of these patterns, I have argued, are shaped by different processing preferences, the nature of which can be inferred from the fixed conventions. These can then be further tested in experiments (both psycholinguistic and increasingly neurolinguistic experiments) and in corpora from relevant languages and their users, and they can contribute to psycholinguistic (and neurolinguistic) theories of online processing.

In addition to providing patterns of relevance for the big issues in defining processing ease and complexity (§3.1), grammatical data can help with the resolution of more specific problems, involving the relationship between production and comprehension (§3.2), differences between different types of performance data (§3.3), and locality versus antilocality effects (§3.4). For example, the grammatical evidence on locality is highly relevant to the current debate in psycholinguistics on locality versus antilocality preferences in performance. Certain experimental results have shown a facilitation at the processing of the verb in verb-final languages when items preceding the verb

are longer (see §§3.3–3.4). The processing of the verb itself seems to be easier when it is more expected, and the reasons proposed include the delay in encountering the verb beyond some average and expected length for clauses, as well as other causes of expectedness. But these findings do not mean that the sentence as a whole, or certain processing domains within it, such as the VP phrasal combination domain, are simpler. Far from it. The evidence from grammars is crystal-clear and highly suggestive. There is not a single word order universal that I am aware of in which non-local domains are preferred over local ones, unless there is some independently motivated competing principle that motivates limited forms of non-locality (§9.3). In other words, grammars tell us that locality, and MiD (2.1) as defined here, are almost always supported. The greater expectedness of the verb in larger domains evidently provides only a weak facilitation, at that one item, within structures that are harder to process overall, on account of their greater size and complexity. Once again, grammatical data provide suggestive evidence for further testing, as well as a valuable check on current theoretical ideas and on the interpretation of certain experimental findings.

In §3.5 and §8.5 I gave a further example of the usefulness of cross-linguistic grammatical data in relation to processing ideas developed in psycholinguistics. Gibson (1998) proposed a prediction-based explanation for the preferred order of subjects before objects in numerous languages. The accusative case in these languages predicts a co-occurring nominative, adding to working memory load, whereas the reverse does not, whence the subject (nominative) before object (accusative) preference. But this doesn't work for languages with different case-marking systems: subjects with ergative case in languages like Avar predict a co-occurring absolutive object, and yet the preferred linearization in these languages is ergative before absolutive. The explanation doesn't work either for numerous other items that predict co-occurring items to the right (fronted Wh-words, initial topics, etc.). It may, however, help us to explain a subset of ergative languages in which the non-predicting absolutive precedes the predicting subject, i.e., with object before subject, but the relative infrequency of this grammaticalization compared with ergative (and agent) subjects first, now provides suggestive evidence about the weakness of this working memory explanation compared with what I have argued to be the competing motivation in terms of MaOP.

In certain cases grammatical data from different language types can provide important evidence for teasing apart properties and processing issues that are often confounded in the more familiar languages. Languages with typologically mixed head-initial/head-final structures (for example German) provide relative clauses that can disentangle structural complexity and frequency effects. German has many postnominal relatives in combination with OV

(NRel & OV). In these structures it does not have larger relativization domains for relatives on SU, DO, and IO with ditransitive verbs (see Hawkins 2004: 171–80). Rather these will be identical in size and proceed from the nominal head external to Rel to the final verb within Rel in NRel. Yet Diessel and Tomasello (2006) found evidence for increasingly harder processing and delayed acquisition down SU > DO > IO, in accordance with the Keenan–Comrie Accessibility Hierarchy. This could be explained in terms of declining frequencies for positions that are structurally identical (see Diessel 2004 for further cases of frequency as a better predictor for order of acquisition than structural complexity).

The (mirror-image) typological inconsistency of Chinese, with RelN & SVO, is also relevant here. In Hawkins (2004: 177–80) I argued that relativizations on SU and DO in this language were identical in domain size, according to the calculation of domains proposed there. Experimental findings for Mandarin have, interestingly, been mixed, with Hsiao and Gibson (2003) showing DO easier to process, and Lin and Bever (2006) showing a SU preference in accordance with the Accessibility Hierarchy. Corpus frequency data supporting the AH ranking is given in Wu (2009). The important point in all these examples is that grammars provide important structural details that can inform psycholinguistic hypotheses and testing and that can, in certain well-chosen languages, help to tease apart factors that are confounded in other languages.

I have raised questions (in §3.3) about the need for greater awareness of how different methods of data collection can lead to different findings about processing ease and complexity. The online, offline, and corpus data summarized there for Extraposition from NP in German (see Konieczny 2000) revealed significant differences, which ultimately led to the locality versus antilocality debate. We need greater clarity on what these different paradigms are each good for. Online measurements at the verb may be telling us nothing about the overall complexity of the clause in which the verb occurs, while corpora may be revealing the assessments of overall complexity made by language producers, irrespective of the benefits or otherwise at individual words. This may explain why corpora from different languages are generally very much in line with the global predictions of my Minimize Domains, which is not an online measure (in Hawkins 2004, though an online variant of it was defined in Hawkins 1994). There needs to be further discussion of this. A further methodological development that I find most exciting involves the use of artificial language learning experiments, and iterated artificial language learning experiments, of the kind described in Tily and Jaeger (2011). These web-based experiments are already producing results whereby constructed languages reveal learning biases and advantages that are in accordance with

typological universals, involving, e.g., word order, case marking, and other phenomena. Perhaps some of the tricky issues involving competition between preferences (§9.3) can be elucidated in this way, as can the basic principles of processing ease themselves.

The final issue that I wish to raise in this last section concerns the notion of dependency. A number of linguists and psychologists have chosen to define structural relations in grammars using dependency as a primitive, essentially following early proposals by Tesnière (1959) and Hays (1964). In Hawkins (2004: 18–25, 97–101) I argued instead for the separation of structural relations like combination, extended combination, and c-command, from the notion of dependency, which I defined in parsing terms, and I suggested that this was the best way to capture what I take to be the original intuition behind it: category B is dependent on category A when B receives some property or properties by reference to A, which are not intrinsic to B and which B would not receive on its own without the language user having access to A. So, in some languages like Latin (Vincent 1987) and Kannada (Bhat 1991), case is explicitly marked on an NP and that case can often be recognized by looking at the NP alone and a linked thematic role can be assigned without having to access the governing verb. In other languages, like English, most NPs do not bear explicit case, and the verb does have to be accessed for case- and thematic-role assignment (see §7.2), and so it does too in Kannada in the event that the case marking is deleted (see §2.5.3) or in the event of ambiguity through case syncretisms in the declensional classes of Latin. It is useful, in a cross-linguistic context therefore, to be able to invoke the notion of dependency when the verb must be accessed for these property assignments, in contrast to the explicit case-marking examples where it need not be.

The definition which I gave in (2.1) was accordingly the following, which I repeat here as (10.1):

(10.1) *Dependency*
A category B is dependent on a category A iff the parsing of B requires access to A for the assignment of syntactic or semantic properties to B with respect to which B is zero-specified or ambiguously or polysemously specified.

In Hawkins (2004: 23) I also defined a dependency assignment as follows:

(10.2) *Dependency Assignment*
If the parsing of a word or phrase B requires access to another A for the assignment of property P to B, B being zero-specified or ambiguously or polysemously specified with respect to P, then P is a *dependency assignment* and B is dependent on A with respect to P.

Coindexation (P) is a dependency assignment to a pronoun or anaphor (B) from its antecedent (A). So is thematic role assignment (P) to an object NP (B) from a transitive verb (A). The appropriate lexical-semantic content (P) is a dependency assignment to transitive *run* (B) when the parser accesses *the water* (A) as opposed to *the race* (A) as direct object; see §7.2. The selection of the syntactic category noun (P) in Tongan is a dependency assignment to an ambiguous or polysemous predicate (B) when the parser accesses an NP-constructing particle or function category (A); see §6.2.2. These dependency assignments are all made within relevant processing domains of terminal elements in surface structure in the usual way.

Not only does this parsing definition capture the original intuition behind dependency, with real benefits for language comparison, it can also solve some problems that arise in grammatical definitions, and it can contribute to a further performance explanation for grammatical phenomena. Very briefly, one issue that is not clear in purely grammatical discussions (and in psycholinguistic models that make use of them, e.g., Gibson 1998, 2000; Temperley 2007) is whether the dependency between B and A is symmetrical or asymmetric—i.e., does A also depend on B, when B depends on A, or not? Tesnière (1959) insisted that all dependencies are asymmetrical. But this misses a generalization. Sometimes the dependencies go in both directions: NPs can depend on verbs for their case and thematic role, as we have seen, but the verb also depends on NP(s) for the precise meaning (in *run the race*/*the water*/*the ad*, etc.) and predicate frame selection. On other occasions the dependencies are asymmetrical and only B depends on A (an anaphor on its antecedent, for example). The empirical consequences of this are significant: symmetries in dependency are reflected in symmetrical orderings (OV and VO), dependency asymmetries in asymmetrical orderings (antecedent before anaphor); see §2.3 and Hawkins (2004: ch.8). A ready explanation can be given for these ordering correlations with symmetry and asymmetry in terms of MaOP, therefore, but it requires a different definition for dependency, ultimately a processing one, as given here.

It is my hope that many others, psychologists, linguists, and computer scientists will find the data and the patterns and the theoretical discussion of this book useful and will want to join me in resolving some of the many issues of this sort that remain.

References

Abney, S. P. (1987). 'The English Noun Phrase in its Sentential Aspect.' Ph.D. dissertation, MIT, Cambridge, Mass.

Aikhenvald, A. Y. (2003). *Classifiers: A Typology of Noun Categorization Devices*. Oxford: Oxford University Press.

Aissen, J. (1999). 'Markedness and Subject Choice in Optimality Theory,' *Natural Language & Linguistic Theory*, 17: 673–711.

Aldai, G. (2003). 'The Prenominal [-Case] Relativization Strategy of Basque: Conventionalization, Processing and Frame Semantics.' MS, Departments of Linguistics, USC and UCLA.

Aldai, G. (2011). '*Wh*-questions and SOV Languages in Hawkins' (2004) Theory: Evidence from Basque,' *Linguistics*, 49(5): 1079–1135.

Anderson, J. R. (1983). *The Architecture of Cognition*. Cambridge, Mass.: Harvard University Press.

Anderson, S. R. (1976). 'On the Notion of Subject in Ergative Languages,' in C. N. Li (ed.), *Subject and Topic*. New York: Academic Press, 1–23.

Anderson, S. R. and Keenan, E. L. (1985). 'Deixis,' in T. Shopen (ed.), *Language Typology and Syntactic Description*, vol. 3: *Grammatical Categories and the Lexicon*. Cambridge: Cambridge University Press, 259–308.

Antinucci, F., Duranti, A., and Gebert, L. (1979). 'Relative Clause Structure, Relative Clause Perception and the Change from SOV to SVO,' *Cognition*, 7: 145–76.

Ariel, M. (1990). *Accessing Noun-Phrase Antecedents*. London: Routledge.

Ariel, M. (1999). 'Cognitive Universals and Linguistic Conventions: The Case of Resumptive Pronouns,' *Studies in Language*, 23: 217–69.

Arnold, J. E., Wasow, T., Losongco, A., and Ginstrom, R. (2000). 'Heaviness vs. Newness: The Effects of Structural Complexity and Discourse Status on Constituent Ordering,' *Language*, 76: 28–55.

Austin, P. (1981). *A Grammar of Diyari, South Australia*. Cambridge: Cambridge University Press.

Bach, E. and Chao, W. (2009). 'On Semantic Universals and Typology,' in M. H. Christiansen, C. Collins, and S. Edelman (eds.), *Language Universals*. Oxford: Oxford University Press, 152–73.

Bard, E. G., Robertson, D., and Sorace, A. (1996). 'Magnitude Estimation of Linguistic Acceptability,' *Language*, 72: 32–68.

Basilico, D. (1996). 'Head Position and Internally Headed Relative Clauses,' *Language*, 72: 498–532.

Bauer, W. (1993). *Maori*. London: Routledge.

Berman, R. A. (1984). 'Cross-linguistic First Language Perspectives on Second Language Acquisition Research,' in R. W. Andersen (ed.), *Second Languages: A Cross-linguistic Perspective*. Rowley, Mass.: Newbury House, 13–36.

Berwick, R. C. and Chomsky, N. (2011). 'The Biolinguistic Program: The Current State of its Evolution and Development,' in A. M. di Sciullo and C. Boeckx (eds), *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty*. Oxford: Oxford University Press, 19–41.

Bever, T. G. (1970). 'The Cognitive Basis for Linguistic Structures,' in J. R. Hayes (ed.), *Cognition and the Development of Language*. New York: Wiley, 279–362.

Bhat, D. N. S. (1991). *Grammatical Relations: The Evidence against their Necessity and Universality*. London: Routledge.

Bhat, D. N. S. (2004). *Pronouns*. Oxford: Oxford University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., and Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow, Essex: Pearson Education Ltd.

Biberauer, T., Holmberg, A., and Roberts, I. (2007). 'Disharmonic Word-Order Systems and the Final-over-Final Constraint (FOFC),' in A. Bisetto and F. Barbieri (eds), *Proceedings of XXXIII Incontro di Grammatica Generativa*. Bologna: University of Bologna, 86–105.

Biberauer, T., Holmberg, A., and Roberts, I. (2008). 'Structure and Linearization in Disharmonic Word Orders,' in C. C. Chang and H. J. Haynie (eds), *Proceedings of the 26th West Coast Conference on Formal Linguistics*. Somerville, Mass.: Cascadilla Proceedings Project, 96–104.

Bird, C. and Kante, M. (1976). *An Kan Bamanakan Kalan: Intermediate Bambara*. Bloomington: Indiana University Linguistics Club.

Blake, B. (1987). *Australian Aboriginal Grammar*. London: Croom Helm.

Blake, B. (1990). *Relational Grammar*. London: Routledge.

Blake, B. (2001). *Case*, 2nd edn. Cambridge: Cambridge University Press.

Bock, K. (1982). 'Towards a Cognitive Psychology of Syntax: Information Processing Contributions to Sentence Formulation,' *Psychological Review*, 89: 1–47.

Bock, K. and Levelt, W. J. M. (1994). 'Language Production: Grammatical Encoding,' in M. A. Gernsbacher (ed.), *Handbook of Psycholinguistics*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bornkessel, I. (2002). 'The Argument Dependency Model: A Neurocognitive Approach to Incremental Interpretation,' Ph.D. dissertation, University of Potsdam.

Bowerman, M. (1988). 'The "No Negative Evidence" Problem: How Do Children Avoid Constructing an Overly General Grammar?' in J. A. Hawkins (ed.), *Explaining Language Universals*. Oxford: Blackwell, 73–101.

Braine, M. D. S. (1971). 'On Two Types of Models of the Internalization of Grammars,' in D. I. Slobin (ed.), *The Ontogenesis of Grammar*. New York: Academic Press, 153–86.

Bredsdorff, E. (1962). *Danish: An Elementary Grammar and Reader*. Cambridge: Cambridge University Press.

Bresnan, J. and Aissen, J. (2002). 'Optimality and Functionality: Objections and Refutations,' *Natural Language and Linguistic Theory*, 20: 81–95.

Bresnan, J., Dingare, S., and Manning, C. D. (2001). 'Soft Constraints Mirror Hard Constraints: Voice and Person in English and Lummi,' in M. Butt and T. H. King (eds), *Proceedings of the LFG 01 Conference*. Stanford: CSLI Publications, 13–32.

Broschart, J. (1997). 'Why Tongan Does it Differently: Categorial Distinctions in a Language without Nouns and Verbs,' *Linguistic Typology*, 1: 123–65.

Brown, K. and Miller, J. (1996). *Concise Encyclopedia of Syntactic Theories*. Oxford: Elsevier Science.

Bybee, J. (1988). 'The Diachronic Dimension in Explanation,' in J. A. Hawkins (ed.), *Explaining Language Universals*. Oxford: Blackwell, 350–79.

Bybee, J. (2001). 'Mechanisms of Change in Grammaticalization: The Role of Frequency,' in R. D. Janda and B. D. Joseph (eds), *Handbook of Historical Linguistics*. Oxford: Blackwell, 602–23.

Bybee, J. and Hopper, P. (eds) (2001). *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins.

Bybee, J., Pagliuca, W., and Perkins, R. D. (1990). 'On the Asymmetries in the Affixation of Grammatical Material,' in W. Croft et al. (eds), *Studies in Typology and Diachrony*. Amsterdam: Benjamins, 1–42.

Campbell, C. and Campbell, J. (1987). *Yadë Grammar Essentials*. Unpublished Manuscript, Summer Institute of Linguistics.

Campbell, L. (2004). *Historical Linguistics: An Introduction*, 2nd edn. Cambridge, Mass.: MIT Press.

Choi, H.-W. (2007). 'Length and Order: A Corpus Study of (the) Korean Dative-Accusative Construction,' *Discourse and Cognition*, 14: 207–27.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT Press.

Chomsky, N. (1973). 'Conditions on Transformations,' in S. R. Anderson and P. Kiparsky (eds), *Festschrift for Morris Halle*. New York: Holt, Rinehart and Winston, 232–86.

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, Mass.: MIT Press.

Chomsky, N. (2000). 'Minimalist Inquiries: The Framework,' in R. Martin, D. Michaels, and J. Uriagereka (eds), *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*. Cambridge, Mass.: MIT Press, 89–156.

Chomsky, N. (2004). 'Beyond Explanatory Adequacy,' in A. Belletti (ed.), *The Cartography of Syntactic Structures*, vol. 4: *Structures and Beyond*. Oxford: Oxford University Press, 104–31.

Chomsky, N. (2005). 'Three Factors in Language Design,' *Linguistic Inquiry*, 36: 1–22.

Chung, S. (1976). 'On the Subject of Two Passives in Indonesian,' in C. N. Li (ed.), *Subject and Topic*. New York: Academic Press, 57–98.

Cinque, G. (2005). 'Deriving Greenberg's Universal 20 and its Exceptions,' *Linguistic Inquiry*, 36: 315–32.

Clancy, P. M., Lee, H., and Zoh, M. (1986). 'Processing Strategies in the Acquisition of Relative Clauses,' *Cognition*, 14: 225–62.

Cole, P. (1976). 'An Apparent Asymmetry in the Formation of Relative Clauses in Modern Hebrew,' in P. Cole (ed.), *Studies in Modern Hebrew Syntax and Semantics*. Amsterdam: North Holland, 231–47.

Cole, P. (1987). 'The Structure of Internally Headed Relative Clauses,' *Natural Language and Linguistic Theory*, 5: 277–302.

Comrie, B. (1978). 'Ergativity,' in W. P. Lehmann (ed.), *Syntactic Typology: Studies in the Phenomenology of Language*. Austin: University of Texas Press.

Comrie, B. (1989). *Language Universals and Linguistic Typology*, 2nd edn. Chicago: University of Chicago Press.

Comrie, B. and Matthews, S. (1990). 'Prolegomena to a Typology of Tough Movement,' in W. Croft, K. Denning, and S. Kemmer (eds), *Studies in Typology and Diachrony: Papers Presented to Joseph H. Greenberg on his 75th Birthday*. Amsterdam: John Benjamins, 43–58.

Corbett, G. G., Fraser, N. M. and McGlashan, S. (eds) (1993). *Heads in Grammatical Theory*. Cambridge: Cambridge University Press.

Couro, T. and Langdon, M. (1975). *Let's Talk 'Iipay Aa: An Introduction to the Mesa Grande Diegueño Language*. Ramona, Calif.: Ballena Press.

Craig, C. G. (1977). *The Structure of Jacaltec*. Austin: University of Texas Press.

Croft, W. (1990). *Typology and Universals*. Cambridge: Cambridge University Press.

Croft, W. (1996). 'Linguistic Selection: An Utterance-based Evolutionary Theory of Language Change,' *Nordic Journal of Linguistics*, 19: 99–139.

Croft, W. (2000). *Explaining Language Change: An Evolutionary Approach*. Harlow: Pearson Education.

Croft, W. (2003). *Typology and Universals*, 2nd edn. Cambridge: Cambridge University Press.

Culicover, P. W. (1993). 'Evidence against ECP Accounts of the That-t Effect,' *Linguistic Inquiry*, 26: 249–75.

Culicover, P. W. (1999). *Syntactic Nuts: Hard Cases, Syntactic Theory and Language Acquisition*. Oxford: Oxford University Press.

Culicover, P. W. and Jackendoff, R. (2005). *Simpler Syntax*. Oxford: Oxford University Press.

Dahl, Ö. (2004). *The Growth and Maintenance of Linguistic Complexity*. Amsterdam: John Benjamins.

De Smedt, K. J. M. J. (1994). 'Parallelism in Incremental Sentence Generation,' in G. Adriaens and U. Hahn (eds), *Parallelism in Natural Language Processing*. Norwood, NJ: Ablex.

Diessel, H. (1999a). 'The Morphosyntax of Demonstratives in Synchrony and Diachrony,' *Linguistic Typology*, 3: 1–49.

Diessel, H. (1999b). *Demonstratives: Form, Function and Grammaticalization*. Amsterdam: John Benjamins.

Diessel, H. (2004). *The Acquisition of Complex Sentences*. Cambridge: Cambridge University Press.

Diessel, H. and Tomasello, M. (2006). 'A New Look at the Acquisition of Relative Clauses,' *Language*, 81: 882–906.

Dixon, R. M. W. (1994). *Ergativity*. Cambridge: Cambridge University Press.

Donohue, M. (1999). *A Grammar of Tukang Besi*. Berlin and New York: Mouton de Gruyter.

Downing, B. T. (1978). 'Some Universals of Relative Clause Structure,' in J. Greenberg (ed.), *Universals of Human Language*, vol. 4: *Syntax*. Stanford: University of Stanford, 375–419.

Dowty, D. R. (1991). 'Thematic Proto-Roles and Argument Selection,' *Language*, 67: 547–619.

Dryer, M. S. (1980). 'The Positional Tendencies of Sentential Noun Phrases in Universal Grammar,' *Canadian Journal of Linguistics*, 25: 123–95.

Dryer, M. S. (1989). 'Large Linguistic Areas and Language Sampling,' *Studies in Language*, 13: 257–92.

Dryer, M. S. (1991). 'SVO languages and the OV:VO Typology,' *Journal of Linguistics*, 27: 443–82.

Dryer, M. S. (1992). 'The Greenbergian Word Order Correlations,' *Language*, 68: 81–138.

Dryer, M. S. (2002). 'Case Distinctions, Rich Verb Agreement, and Word Order Type,' *Theoretical Linguistics*, 28(2): 151–7.

Dryer, M. S. (2004). 'Noun Phrases without Nouns,' *Functions of Language*, 11(1): 43–76.

Dryer, M. S. (2005a). 'Order of Object and Verb,' in M. Haspelmath et al. (eds), *The World Atlas of Language Structures*. Oxford: Oxford University Press, 338–41.

Dryer, M. S. (2005b). 'Order of Adposition and Noun Phrase,' in M. Haspelmath et al. (eds), *The World Atlas of Language Structures*. Oxford: Oxford University Press, 346–9.

Dryer, M. S. (2005c). 'Definite Articles,' in M. Haspelmath et al. (eds), *The World Atlas of Language Structures*. Oxford: Oxford University Press, 154–7.

Dryer, M. S. (2005d). 'Order of Relative Clause and Noun,' in M. Haspelmath et al. (eds), *The World Atlas of Language Structures*. Oxford: Oxford University Press, 366–9.

Dryer, M. S. (2005e). 'Relationship between the Order of Object and Verb and the Order of Relative Clause and Noun,' in M. Haspelmath et al. (eds), *The World Atlas of Language Structures*. Oxford: Oxford University Press, 390–3.

Dryer, M. S. (2005f). 'Determining Dominant Word Order,' in M. Haspelmath et al. (eds), *The World Atlas of Language Structures*. Oxford: Oxford University Press, 371.

Dryer, M. S. (2005g). 'Order of Genitive and Noun,' in M. Haspelmath et al. (eds), *The World Atlas of Language Structures*. Oxford: Oxford University Press, 350–3.

Dryer, M. S. (2005h). 'Order of Adjective and Noun,' in M. Haspelmath et al. (eds), *The World Atlas of Language Structures*. Oxford: Oxford University Press, 354–7.

Dryer, M. S. (2005i). 'Position of Case Affixes,' in M. Haspelmath et al. (eds), *The World Atlas of Language Structures*. Oxford: Oxford University Press, 210–13.

Dryer, M. S. (2005j). 'Position of Interrogative Phrases in Content Questions,' in M. Haspelmath et al. (eds), *The World Atlas of Language Structures*. Oxford: Oxford University Press, 378–81.

Dryer, M. S. (2009). 'The Branching Direction Theory of Word Order Correlations Revisited,' in S. Scalise, E. Magni, and A. Bisetto (eds), *Universals of Language Today*. Berlin: Springer.

Dryer, M. S. (2011). 'Genealogical Language List,' available online at <http://wals.info/languoid/genealogy>.

Dryer, M. S. with Gensler, O. D. (2005). 'Order of Object, Oblique, and Verb,' in M. Haspelmath et al. (eds), *The World Atlas of Language Structures*. Oxford: Oxford University Press, 342–5.

Dugast, I. (1971). *Grammaire du Tunen*. Paris: Klincksieck.

Elman, J., Karmiloff-Smith, A., Bates, E., Johnson, M., Parisi, D., and Plunkett, K. (eds) (1996). *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, Mass.: MIT Press.

Erdmann, P. (1988). 'On the Principle of "Weight" in English,' in C. Duncan-Rose and T. Vennemann (eds), *On Language, Rhetorica, Phonologica, Syntactica: A Festschrift for Robert P. Stockwell from his Friends and Colleagues*. London: Routledge, 325–39.

Erguvanli, E. E. (1984). *The Function of Word Order in Turkish Grammar*. Berkeley: University of California Press.

Erteschik-Shir, N. (1992). 'Resumptive Pronouns in Islands,' in H. Goodluck and M. Rochemont (eds), *Island Constraints: Theory, Acquisition and Processing*. Dordrecht: Kluwer, 89–108.

Erteschik-Shir, N. and Lappin, S. (1979). 'Dominance and the Functional Explanation of Island Phenomena,' *Theoretical Linguistics*, 6: 41–86.

Fodor, Janet D. (1978). 'Parsing Strategies and Constraints on Transformations,' *Linguistic Inquiry*, 8: 425–504.

Fodor, Janet D. (1983). 'Phrase Structure Parsing and the Island Constraints,' *Linguistics and Philosophy*, 6: 163–223.

Fodor, Janet D. (2001). 'Setting Syntactic Parameters,' in M. Baltin and C. Collins (eds), *The Handbook of Contemporary Syntactic Theory*. Oxford: Blackwell, 730–8.

Fox, B. A. and Thompson, S. A. (2007). 'Relative Clauses in English Conversation: Relativizers, Frequency and the Notion of Construction,' *Studies in Language*, 31(2): 293–326.

Francis, E. J. (2010). 'Grammatical Weight and Relative Clause Extraposition in English,' *Cognitive Linguistics*, 21(1): 35–74.

Francis, W. N. and Kučera, H. (1982). *Frequency Analysis of English Usage: Lexicon and Grammar*. Boston: Houghton Mifflin.

Frazier, L. (1979). *On Comprehending Sentences: Syntactic Parsing Strategies*. Bloomington: Indiana University Linguistics Club.

Frazier, L. (1985). 'Syntactic Complexity,' in D. Dowty, L. Karttunen, and A. Zwicky (eds), *Natural Language Parsing*. Cambridge: Cambridge University Press, 129–89.

Frazier, L. and Fodor, Janet D. (1978). 'The Sausage Machine: A New Two-Stage Parsing Model,' *Cognition*, 6: 291–326.

Frazier, L. and Rayner, K. (1988). 'Parameterizing the Language Processing System: Left- vs. Right-Branching Within and Across Languages,' in J. A. Hawkins (ed.), *Explaining Language Universals*. Oxford: Blackwell, 247–79.

Gast, V. (2007). 'From Phylogenetic Diversity to Structural Homogeneity: On Right-branching Constituent Order in Mesoamerica,' *Sky Journal of Linguistics*, 20: 171–202.

Gell-Mann, M. (1992). 'Complexity and Complex Adaptive Systems,' in J. A. Hawkins and M. Gell-Mann (eds), *The Evolution of Human Languages*. Redwood City, Calif.: Addison-Wesley, 3–18.

Gibson, E. (1998). 'Linguistic Complexity: Locality of Syntactic Dependencies,' *Cognition*, 68: 1–76.

Gibson, E. (2000). 'The Dependency Locality Theory: A Distance-based Theory of Linguistic Complexity,' in Y. Miyashita, P. Marantz, and W. O'Neil (eds), *Image, Language, Brain*. Cambridge, Mass.: MIT Press, 95–112.

Givón, T. (1979). *On Understanding Grammar*. New York: Academic Press.

Gorbet, L. (1976). *A Grammar of Diegueño Nominals*. New York: Garland.

Greenberg, J. H. (1963). 'Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements,' in J. H. Greenberg (ed.), *Universals of Language*. Cambridge, Mass.: MIT Press, 73–113.

Greenberg, J. H. (1966). *Language Universals, with Special Reference to Feature Hierarchies*. The Hague: Mouton.

Greenberg, J. H. (1978). 'How Does a Language Acquire Gender Markers?' in J. H. Greenberg, C. A. Ferguson, and E. A. Moravcsik (eds), *Universals of Human Language*, Vol. 3: *Word Structure*. Stanford: Stanford University Press, 47–82.

Greenberg, J. H. (1995). 'The Diachronic Typological Approach to Language,' in M. Shibatani and T. Bynon (eds), *Approaches to Language Typology*. Oxford: Clarendon Press, 143–66.

Grice, H. P. (1975). 'Logic and Conversation,' in P. Cole and J. Morgan (eds), *Speech Acts*. New York: Academic Press, 41–58.

Grimshaw, J. (1990). *Argument Structure*. Cambridge, Mass.: MIT Press.

Gundel, J. K. (1988). 'Universals of Topic-Comment Structure,' in M. Hammond, E. Moravcsik, and J. R. Wirth (eds), *Studies in Syntactic Typology*. Amsterdam: John Benjamins, 209–39.

Gundel, J. K., Houlihan, K., and Sanders, G. (1988). 'On the Function of Marked and Unmarked Terms,' in M. T. Hammond, E. A. Moravcsik, and J. R. Wirth (eds), *Studies in Syntactic Typology*. Amsterdam: John Benjamins, 285–301.

Haiman, J. (1983). 'Iconic and Economic Motivation,' *Language*, 59: 781–819.

Hale, J. (2003). 'The Information Conveyed by Words in Sentences,' *Journal of Psycholinguistic Research*, 32: 101–23.

Hale, K. (1973). 'Person Marking in Walbiri,' in S. Anderson and P. Kiparsky (eds), *Festschrift for Morris Halle*. New York: Holt, Rinehart and Winston, 308–44.

Hall, C. J. (1992). *Morphology and Mind*. London: Routledge.

Haspelmath, M. (1999a). 'Optimality and Diachronic Adaptation,' *Zeitschrift für Sprachwissenschaft*, 18: 180–205.

Haspelmath, M. (1999b). 'Explaining Article-Possessor Complementarity: Economic Motivation in Noun Phrase Syntax,' *Language*, 75: 227–43.

Haspelmath, M. (2002). *Morphology*. London: Arnold.

Haspelmath, M. (2008a). 'Creating Economical Morphosyntactic Patterns in Language Change,' in J. Good (ed.), *Language Universals and Language Change*. Oxford: Oxford University Press, 185–214.

Haspelmath, M. (2008b). 'Parametric Versus Functional Explanations of Syntactic Universals,' in T. Biberauer (ed.), *The Limits of Syntactic Variation*. Amsterdam: John Benjamins, 75–107.

Haspelmath, M., Dryer, M. S., Gil, D., and Comrie, B. (eds) (2005). *The World Atlas of Language Structures*. Oxford: Oxford University Press.

Hauser, M., Chomsky, N., and Fitch, W. T. (2002). 'The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?' *Science*, 298: 1569–79.

Hawkins, J. A. (1983). *Word Order Universals*. New York: Academic Press.

Hawkins, J. A. (1985). 'Complementary Methods in Universal Grammar: A Reply to Coopmans,' *Language*, 61: 569–87.

Hawkins, J. A. (1986). *A Comparative Typology of English and German: Unifying the Contrasts*. Austin: University of Texas Press.

Hawkins, J. A. (ed.) (1988). *Explaining Language Universals*. Oxford: Blackwell.

Hawkins, J. A. (1990). 'A Parsing Theory of Word Order Universals,' *Linguistic Inquiry*, 21: 223–61.

Hawkins, J. A. (1991). 'On (In)definite Articles: Implicatures and (Un)grammaticality Prediction,' *Journal of Linguistics*, 27: 405–42.

Hawkins, J. A. (1993). 'Heads, Parsing, and Word Order Universals,' in G. G. Corbett, N. M. Fraser, and S. McGlashan (eds), *Heads in Grammatical Theory*. Cambridge: Cambridge University Press, 231–65.

Hawkins, J. A. (1994). *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.

Hawkins, J. A. (1995). 'Argument-Predicate Structure in Grammar and Performance: A Comparison of English and German,' in I. Rauch and G. F. Carr (eds), *Insights in Germanic Linguistics*, 1. Berlin: Mouton de Gruyter, 127–44.

Hawkins, J. A. (1998). 'Some Issues in a Performance Theory of Word Order,' in A. Siewierska (ed.), *Constituent Order in the Languages of Europe*. Berlin: Mouton de Gruyter, 729–81.

Hawkins, J. A. (1999). 'Processing Complexity and Filler-Gap Dependencies,' *Language*, 75: 244–85.

Hawkins, J. A. (2000). 'The Relative Order of Prepositional Phrases in English: Going Beyond Manner-Place-Time,' *Language Variation and Change*, 11: 231–66.

Hawkins, J. A. (2001). 'Why are Categories Adjacent?' *Journal of Linguistics*, 37: 1–34.

Hawkins, J. A. (2002). 'Symmetries and Asymmetries: Their Grammar, Typology and Parsing,' *Theoretical Linguistics*, 28(2): 95–149.

Hawkins, J. A. (2003). 'Why are Zero-Marked Phrases Close to their Heads?', in G. Rohdenburg and B. Mondorf (eds), *Determinants of Grammatical Variation in English*. Berlin: Mouton de Gruyter, 175–204.

Hawkins, J. A. (2004). *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.

Hawkins, J. A. (2007). 'Processing Typology and Why Psychologists Need to Know About It,' *New Ideas in Psychology*, 25: 87–107.

Hawkins, J. A. (2008). 'An Asymmetry between VO and OV Languages: The Ordering of Obliques,' in G. Corbett and M. Noonan (eds), *Case and Grammatical Relations: Essays in Honour of Bernard Comrie*. Amsterdam: John Benjamins, 167–90.

Hawkins, J. A. (2009). 'An Efficiency Theory of Complexity and Related Phenomena,' in G. Sampson, D. Gil, and P. Trudgill (eds), *Language Complexity as an Evolving Variable*. Oxford: Oxford University Press, 252–68.

Hawkins, J. A. (2011). 'Discontinuous Dependencies in Corpus Selections: Particle Verbs and their Relevance for Current Issues in Language Processing,' in J. Arnold and E. Bender (eds), *Readings in Cognitive Science*. Stanford, Calif.: CSLI Publications, 269–91.

Hawkins, J. A. (2012). 'The Drift of English Towards Invariable Word Order from a Typological and Germanic Perspective,' in T. Nevalainen and E. C. Traugott (eds), *The Oxford Handbook of the History of English*. Oxford: Oxford University Press, 622–32.

Hawkins, J. A. and Buttery, P. (2010). 'Criterial Features in Learner Corpora: Theory and Illustrations,' *English Profile Journal*, 1:e4.1–23.

Hawkins, J. A. and Filipović, L. (2012). *Criterial Features in L2 English: Specifying the Reference Levels of the Common European Framework*. Cambridge: Cambridge University Press.

Hays, D. G. (1964). 'Dependency Theory: A Formalism and Some Observations,' *Language*, 40: 511–25.

Heath, J. (1999). *A Grammar of Koyraboro (Koroboro) Senni*. Köln: Rüdiger Köppe Verlag.

Heine, B. (2008). 'Contact-induced Word Order Change without Word Order Change,' in P. Siemund and N. Kintana (eds), *Language Contact and Contact Languages*. Amsterdam and Philadelphia: John Benjamins, 33–60.

Heine, B. and Kuteva, T. (2002). *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press.

Heine, B. and Reh, M. (1984). *Grammaticalisation and Reanalysis in African Languages*. Hamburg: Buske.

Hengeveld, K., Rijkhoff, J., and Siewierska, A. (2004). 'Parts-of-Speech Systems and Word Order,' *Journal of Linguistics*, 40: 527–70.

Himmelmann, N. P. (1997). *Deiktikon, Artikel, Nominalphrase: Zur Emergenz syntaktischer Struktur*. Niemeyer: Tübingen.

Hodler, W. (1954). *Grundzüge einer germanischen Artikellehre*. Heidelberg: Carl Winter.

Hoekstra, T. and Kooij, J. G. (1988). 'The Innateness Hypothesis,' in J. A. Hawkins (ed.), *Explaining Language Universals*. Oxford: Blackwell, 31–55.

Holmberg, A. (2000). 'Deriving OV Order in Finnish,' in P. Svenonius (ed.), *The Derivation of VO and OV*. New York: Oxford University Press, 123–52.

Hopper, P. J. and Traugott, E. C. (1993). *Grammaticalization*. Cambridge: Cambridge University Press.

Hsiao, F. and Gibson, E. (2003). 'Processing Relative Clauses in Chinese,' *Cognition*, 90: 3–27.

Hudson, R. A. (1987). 'Zwicky on Heads,' *Journal of Linguistics*, 23: 109–32.

Innes, G. (1966). *An Introduction to Grebo*. London: SOAS, University of London.

Jackendoff, R. (1977). *X-bar Syntax: A Study of Phrase Structure*. Cambridge, Mass.: MIT Press.

Jacobs, J. (2001). 'The Dimensions of Topic-Comment,' *Linguistics*, 39: 641–81.

Jaeger, T. F. (2006). 'Redundancy and Syntactic Reduction in Spontaneous Speech.' Unpublished doctoral dissertation, Stanford University, Stanford, Calif.

Jaeger, T. F. and Norcliffe, E. (2009). 'The Cross-linguistic Study of Sentence Production: State of the Art and a Call for Action,' *Language and Linguistics Compass*, 3(4): 866–87.

Jaeger, T. F. and Wasow, T. (2008). 'Processing as a Source of Accessibility Effects on Variation,' in R. T. Cover and Y. Kim (eds), *Proceedings of the 31st Annual Meeting of the Berkeley Linguistic Society*. Ann Arbor: Sheridan Books, 169–80.

Jaeger, T. F. and Tily, H. (2011). 'On Language "Utility": Processing Complexity and Communicative Efficiency,' *WIREs: Cognitive Science*, 2(3): 323–35.

Jurafsky, D. (1996). 'A Probabilistic Model of Lexical and Syntactic Access and Disambiguation,' *Cognitive Science*, 20: 137–94.

Just, M. A. and Carpenter, P. A. (1992). 'A Capacity Theory of Comprehension: Individual Differences in Working Memory,' *Psychological Review*, 99(1): 122–49.

Kashket, M. B. (1988). 'Parsing Warlpiri, a Free Word Order Language,' in S. P. Abney (ed.), *The MIT Parsing Volume, 1987–88*, Parsing Project Working Papers No. 1. Cambridge, Mass.: MIT, Center for Cognitive Science, 104–27.

Kayne, R. S. (1994). *The Antisymmetry of Syntax*. Cambridge, Mass.: MIT Press.

Keenan, E. L. (1975). 'Variation in Universal Grammar,' in R. Fasold and R. Shuy (eds), *Analyzing Variation in English*. Washington, DC: Georgetown University Press, 136–48. Reprinted in Keenan (1987), 46–59.

Keenan, E. L. (1979). 'On Surface Form and Logical Form,' *Studies in the Linguistic Sciences*, 8(2): 163–203. Reprinted in Keenan (1987), 375–428.

Keenan, E. L. (1985). 'Relative Clauses', in T. Shopen (ed.), *Language Typology and Syntactic Description*, vol. II: *Complex Constructions*. Cambridge: Cambridge University Press, 141–70.

Keenan, E. L. (1987). *Universal Grammar: 15 Essays*. London: Croom Helm.

Keenan, E. L. (1988). 'On Semantics and the Binding Theory,' in J. A. Hawkins, *Explaining Language Universals*. Oxford: Blackwell, 105–44.

Keenan, E. L. and Comrie, B. (1977). 'Noun Phrase Accessibility and Universal Grammar,' *Linguistic Inquiry*, 8: 63–99. Reprinted in Keenan (1987), *Universal Grammar: 15 Essays*. London: Croom Helm, 3–45.

Keenan, E. L. and Hawkins, S. (1987). 'The Psychological Validity of the Accessibility Hierarchy,' in Keenan (1987), 60–85.

Kempen, G. (2000). 'Could Grammatical Encoding and Grammatical Decoding Be Subserved by the Same Processing Module?' *Behavioral and Brain Sciences*, 23: 38–9.

Kempen, G. (2003). 'Cognitive Architectures for Human Grammatical Encoding and Decoding.' MS, Max-Planck-Institute for Psycholinguistics, Nijmegen.

Kempen, G. and Hoenkamp, E. (1987). 'An Incremental Procedural Grammar for Sentence Formulation,' *Cognitive Science*, 11: 201–58.

Kempson, R. M., Meyer-Viol, W., and Gabbay, D. (2001). *Dynamic Syntax*. Oxford: Blackwell.

Kendall, M. B. (1974). 'Relative Clause Formation and Topicalization in Yavapai,' *IJAL*, 40: 89–101.

Kiaer, J. (2007). 'Processing and Interfaces in Syntactic Theory: The Case of Korean.' Ph.D. thesis, King's College London.

Kim, A. H. (1988). 'Preverbal Focusing and Type XXIII Languages,' in M. Hammond, E. Moravcsik, and J. R. Wirth (eds), *Studies in Syntactic Typology*. Amsterdam: John Benjamins, 147–69.

Kimball, J. (1973). 'Seven Principles of Surface Structure Parsing in Natural Language,' *Cognition*, 2: 15–47.

King, R. (1969). *Generative Grammar and Historical Linguistics*. Englewood Cliffs, NJ: Prentice Hall.

Kirby, S. (1999). *Function, Selection and Innateness*. Oxford: Oxford University Press.

Kiss, K. E. (1987). *Configurationality in Hungarian*. Dordrecht: Reidel.

Kiss, K. E. (2002). *The Syntax of Hungarian*. Cambridge: Cambridge University Press.

Kizach, J. (2010). 'The Function of Word Order in Russian Compared with Danish and English.' Ph.D. thesis, Department of English, University of Aarhus.

Kizach, J. (2012). 'Evidence for Weight Effects in Russian,' *Russian Linguistics*, 36(3): 251–70.

Klima, E. (1964). 'Negation in English,' in J. Fodor and J. Katz (eds), *The Structure of Language*. Englewood Cliffs, NJ: Prentice Hall.

Konieczny, L. (2000). 'Locality and Parsing Complexity,' *Journal of Psycholinguistic Research*, 29(6): 627–45.

König, E. and Gast, V. (2009). *Understanding English-German Contrasts*. Berlin: Erich Schmidt.

Koptevskaja-Tamm, M. (2003). 'Possessive Noun Phrases in the Languages of Europe,' in F. Plank (ed.), *Noun Phrase Structure in the Languages of Europe*. Berlin: Mouton de Gruyter, 621–722.

Kuno, S. (1973). *The Structure of the Japanese Language*. Cambridge, Mass.: MIT Press.

Kwon, N., Gordon, P. C., Lee, Y., Kluender, R., and Polinsky, M. (2010). 'Cognitive and Linguistic Factors Affecting Subject/Object Asymmetry: An Eye-Tracking Study of Prenominal Relative Clauses in Korean,' *Language*, 86: 546–82.

Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, W. (1994). *Principles of Linguistic Change*, vol. 1: *Internal Factors*. Oxford: Blackwell.

Labov, W. (2007). 'Transmission and Diffusion,' *Language*, 81: 344–87.

Lehmann, Christian (1984). *Der Relativsatz*. Tuebingen: Gunter Narr Verlag.

Lehmann, Christian (1995). *Thoughts on Grammaticalization*. Munich: Lincom Europa.

Leisi, E. (1975). *Der Wortinhalt*, 5th edn. Heidelberg: Quelle und Meyer.

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, Mass.: MIT Press.

Levinson, S. C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.

Levinson, S. C. (2000). *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge, Mass.: MIT Press.

Levy, R. (2008). 'Expectation-based Syntactic Comprehension,' *Cognition*, 106: 1126–77.

Lewis, R. and Vasishth, S. (2005). 'An Activation-based Model of Sentence Processing as Skilled Memory Retrieval,' *Cognitive Science*, 29: 375–419.

Li, C. N. and Thompson, S. A. (1981). *A Functional Reference Grammar of Mandarin Chinese*. Berkeley: University of California Press.

Lichtenberk, F. (1983). *A Grammar of Manam*. Honolulu: University of Hawaii Press.

Lieberman, M. (1974). 'On Conditioning the Rule of Subject-Auxiliary Inversion,' *NELS*, 5: 77–91.

Lightfoot, D. W. (1991). *How to Set Parameters*. Cambridge, Mass.: MIT Press.

Lin, C. C. and Bever, T. G. (2006). 'Subject Preference in the Processing of Relative Clauses in Chinese.' Poster presented at the 25th West Coast Conference on Formal Linguistics, Somerville, Mass.

Lindblom, B. and Maddieson, I. (1988). 'Phonetic Universals in Consonant Systems,' in L. M. Hyman and C. N. Li (eds), *Language, Speech and Mind: Studies in Honour of Victoria A. Fromkin*. London: Routledge, 62–78.

Lohse, B. (2000). 'Zero versus Explicit Marking in Relative Clauses.' MS, Department of Linguistics, USC.

Lohse, B., Hawkins, J. A. and Wasow, T. (2004). 'Domain Minimization in English Verb-Particle Constructions,' *Language*, 80: 238–61.

Los, B. and Dreschler, G. (2012). 'From a Multifunctional First Constituent to a Multifunctional Subject,' in T. Nevalainen and E. C. Traugott (eds), *The Oxford Handbook of the History of English*. Oxford: Oxford University Press.

Lupyan, G. and Dale, R. (2010). 'Language Structure is Partly Determined by Social Structure,' *PLoS One*, 5(1): e8559, doi:10.13l1/journal.pone.0008559.

Lyons, Christopher (1999). *Definiteness*. Cambridge: Cambridge University Press.

Lyons, John (1977). *Semantics*, 2 vols. Cambridge: Cambridge University Press.

MacDonald, M. C. (1999). 'Distributional Information in Language Comprehension, Production and Acquisition: Three Puzzles and a Moral,' in B. MacWhinney (ed.), *The Emergence of Language*. Mahwah, NJ: Lawrence Erlbaum Associates.

MacDonald, M. C., Pearlmutter, N. J., and Seidenberg, M. S. (1994). 'The Lexical Nature of Syntactic Ambiguity Resolution,' *Psychological Review*, 101(4): 676–703.

MacWhinney, B. (1982). 'Basic Syntactic Processes,' in S. Kuczaj (ed.), *Language Acquisition: Syntax and Semantics*. Mahwah, NJ: Lawrence Erlbaum Associates.

Maddieson, I. (2005). 'Consonant Inventories,' in M. Haspelmath et al. (eds), *The World Atlas of Language Structures*. Oxford: Oxford University Press, 10–13.

Manber, U. (1989). *Introduction to Algorithms*. Reading, Mass.: Addison-Wesley.

Manning, C. D. (2003). 'Probabilistic Syntax,' in R. Bod, J. Hay, and S. Jannedy (eds), *Probability Theory in Linguistics*. Cambridge, Mass.: MIT Press, 289–341.

Marblestone, K. L. (2007). 'Semantic and Syntactic Effects on Double Prepositional Phrase Ordering across the Lifespan.' Ph.D. dissertation, University of Southern California.

Marcus, M. (1980). *A Theory of Syntactic Recognition for Natural Language*. Cambridge, Mass.: MIT Press.

Mater, E. (1971). *Deutsche Verben*. Leipzig: VEB Bibliographisches Institut.

Matisoff, J. A. (1972). 'Lahu Nominalization, Relativization, and Genitivization,' in J. Kimball (ed.), *Syntax and Semantics*, vol. 1. New York and London: Academic Press, 237–57.

Matthews, S. and Yeung, L. Y. Y. (2001). 'Processing Motivations for Topicalization in Cantonese,' in K. Horie and S. Sato (eds), *Cognitive-Functional Linguistics in an East Asian Context*. Tokyo: Kurosio Publishers, 81–102.

Matthews, S. and Yip, V. (1994). *Cantonese: A Comprehensive Grammar*. London: Routledge.

Matthews, S. and Yip, V. (2003). 'Relative Clauses in Early Bilingual Development: Transfer and Universals,' in A. G. Ramat (ed.), *Typology and Second Language Acquisition*. Berlin: Mouton de Gruyter.

McMahon, A. M. S. (1994). *Understanding Language Change*. Cambridge: Cambridge University Press.

McWhorter, J. (2001). 'The World's Simplest Grammars are Creole Grammars,' *Linguistic Typology*, 5: 125–66.

Meillet, A. (1912). 'L'Évolution des Formes Grammaticales,' *Scientia*, 12–26 (Milan). Reprinted (1948) in *Linguistique Historique et Linguistique Générale*. Paris: Champion, 130–48.

Melinger, A., Pechmann, T., and Pappert, S. (2009). 'Case in Language Production,' in A. Malchukov and A. Spencer (eds), *Oxford Handbook of Case*. Oxford: Oxford University Press, 384–401.

Miller, G. A. and Chomsky, N. (1963). 'Finitary Models of Language Users,' in R. D. Luce, R. Bush, and E. Galanter (eds), *Handbook of Mathematical Psychology*, vol. 2. New York: Wiley, 419–92.

Miyamoto, E. T. (2006). 'Understanding Sentences in Japanese Bit by Bit,' *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society*, 13: 247–60.

Mobbs, I. (2008). '"Functionalism," the Design of the Language Faculty, and (Disharmonic) Typology.' MS, King's College Cambridge.

Moravcsik, E. (1995). 'Summing up Suffixaufnahme,' in F. Plank (ed.), *Double Case: Agreement by Suffixaufnahme*. Oxford: Oxford University Press, 451–84.

Müller-Gotama, F. (1991). 'A Typology of the Syntax-Semantics Interface.' Unpublished Ph.D. dissertation, University of Southern California.

Müller-Gotama, F. (1994). *Grammatical Relations: A Cross-linguistic Perspective on their Syntax and Semantics*. Berlin: Mouton de Gruyter.

Newmeyer, F. J. (1998). *Language Form and Language Function*. Cambridge, Mass.: MIT Press.

Newmeyer, F. J. (2005). *Possible and Probable Languages: A Generative Perspective on Linguistic Typology*. Oxford: Oxford University Press.

Nichols, J. (1986). 'Head-marking and Dependent-marking Grammar,' *Language*, 62: 56–119.

Nichols, J. (1994). 'Ingush,' in R. Smeets (ed.), *The Indigenous Languages of the Caucasus*, vol. 4: *North East Caucasian Languages*, part 2. Delmar, NY: Caravan Books, 79–146.

Nichols, J., Peterson, D. A., and Barnes, J. (2004). 'Transitivizing and Detransitivizing Languages,' *Linguistic Typology*, 8: 149–211.

O'Grady, W. (2005). *Syntactic Carpentry: An Emergentist Approach to Syntax*. Mahwah, NJ: Lawrence Erlbaum Associates.

Oirsouw, R. A. van (1987). *The Syntax of Coordination*. London: Croom Helm.

Payne, J. (1993). 'The Headedness of Noun Phrases: Slaying the Nominal Hydra,' in G. G. Corbett et al. (eds), *Heads in Grammatical Theory*. Cambridge: Cambridge University Press 114–39.

Phillips, C. (1996). 'Order and Structure.' Ph.D. dissertation, MIT, Cambridge, Mass.

Pinker, S. and Jackendoff, R. (2009). 'The Components of Language: What's Specific to Language and What's Specific to Humans,' in M. H. Christiansen, C. Collins, and S. Edelman (eds), *Language Universals*. Oxford: Oxford University Press, 126–51.

Plank, F. (1984). 'Verbs and Objects in Semantic Agreement: Minor Differences between English and German that Might Suggest a Major One,' *Journal of Semantics*, 3(4): 305–60.

Plank, F. (ed.) (1995). Double Case: Agreement by Suffixaufnahme. Oxford: Oxford University Press.

Plank, F. (ed.) (2003). *Noun Phrase Structure in the Languages of Europe*. Berlin: Mouton de Gruyter.

Polinsky, M. (2002). 'Efficiency Preferences: Refinements, Rankings, and Unresolved Questions,' *Theoretical Linguistics*, 28(2): 177–202.

Pollard, C. and Sag, I. A. (1987). *Information-based Syntax and Semantics*, vol. 1: *Fundamentals*, CSLI Lecture Notes No. 13. Stanford: CSLI.

Pollard, C. and Sag, I. A. (1994). *Head-driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Primus, B. (1993). 'Syntactic Relations,' in J. Jacobs, A. von Stechow, W. Sternefeld, and T. Vennemann (eds), *Syntax: An International Handbook of Contemporary Research*, vol. 1. Berlin: Mouton de Gruyter, 686–705.

Primus, B. (1995). 'Relational Typology,' in J. Jacobs, A. von Stechow, W. Sternefeld, and T. Vennemann (eds), *Syntax: An International Handbook of Contemporary Research*, vol. 2. Berlin: Mouton de Gruyter, 1076–1109.

Primus, B. (1999). *Cases and Thematic Roles: Ergative, Accusative and Active*. Tuebingen: Max Niemeyer Verlag.

Primus, B. (2002). 'How Good is Hawkins' Performance of Performance?' *Theoretical Linguistics*, 28(2): 203–9.

Prince, A. and Smolensky, P. (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*. Cambridge, Mass.: MIT Press.

Quirk, R. (1957). 'Relative Clauses in Educated Spoken English,' *English Studies*, 38: 97–109.

Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. Harlow: Longman.

Radford, A. (1997). *Syntactic Theory and the Structure of English*. Cambridge: Cambridge University Press.

Reali, F. and Christiansen, M. H. (2006a). 'Processing of Relative Clauses is Made Easier by Frequency of Occurrence,' *Journal of Memory and Language*, 57: 1–23.

Reali, F. and Christiansen, M. H. (2006b). 'Word Chunk Frequencies Affect the Processing of Pronominal Object-Relatives,' *Quarterly Journal of Experimental Psychology*, 60: 161–70.

Reinhart, T. (1982). 'Pragmatics and Linguistics: An Analysis of Sentence Topics,' *Philosophica*, 27: 53–94.

Rice, K. (1989). *A Grammar of Slave*. Berlin: Mouton de Gruyter.

Rijkhoff, J. (2002). *The Noun Phrase*. Oxford: Oxford University Press.

Rizzi, L. (1991). *Relativized Minimality*. Cambridge, Mass.: MIT Press.

Rohdenburg, G. (1974). *Sekundäre Subjektivierungen im Englischen und Deutschen*. Bielefeld: Cornelson-Velhagen und Klasing.

Rohdenburg, G. (1996). 'Cognitive Complexity and Grammatical Explicitness in English,' *Cognitive Linguistics*, 7: 149–82.

Rohdenburg, G. (1999). 'Clausal Complementation and Cognitive Complexity in English,' in F.-W. Neumann and S. Schülting (eds), *Anglistentag 1998: Erfurt*. Trier: Wissenschaftlicher Verlag, 101–12.

Ross, J. R. (1967). 'Constraints on Variables in Syntax,' Ph.D. dissertation, MIT, Cambridge, Mass.

Ross, M. D. (1996). 'Contact-Induced Change and the Comparative Method: Cases from Papua New Guinea,' in M. Durie and M. D. Ross (eds), *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*. Oxford: Oxford University Press, 180–217.

Ross, M. D. (2001). 'Contact-Induced Change in Oceanic Languages in North-West Melanesia,' in A. Y. Aikhenvald and R. M. Dixon (eds), *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*. Oxford: Oxford University Press, 134–66.

Rumelhart, D., McClelland, J. L., and the PDP Research Group (eds) (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge, Mass.: MIT Press.

Saah, K. K. and Goodluck, H. (1995). 'Island Effects in Parsing and Grammar: Evidence from Akan,' *Linguistic Review*, 12: 381–409.

Sadler, L. (1988). *Welsh Syntax: A Government-Binding Approach*. London: Croom Helm.

Sadock, J. and Zwicky, A. (1985). 'Speech Act Distinctions in Syntax,' in T. Shopen (ed.), *Language Typology and Syntactic Description*, vol. 1: *Clause Structure*. Cambridge: Cambridge University Press.

Saltarelli, M. (1988). *Basque*. London and New York: Routledge.

Sapir, E. (1921). *Language: An Introduction to the Study of Speech*. New York: Harcourt Brace.

Schuh, R. G. (1983). 'The Evolution of Determiners in Chadic,' in E. Wolff and H. Meyer-Bahlburg (eds), *Studies in Chadic and Afroasiatic Linguistics*. Hamburg: Helmut Buske Verlag, 157–210.

Schütze, C. T. and Gibson, E. (1999). 'Argumenthood and English Prepositional Phrase Attachment,' *Journal of Memory and Language*, 40: 409–31.

Shannon, C. (1948). 'A Mathematical Theory of Communications,' *Bell Systems Technical Journal*, 27: 623–56.

Shannon, T. F. (1992). 'Toward an Adequate Characterization of Relative Clause Extraposition in Modern German,' in I. Rauch, G. F. Carr, and R. L. Kyes (eds), *On Germanic Linguistics: Issues and Methods*. Berlin: Mouton de Gruyter.

Sheldon, A. (1974). 'On the Role of Parallel Function in the Acquisition of Relative Clauses in English,' *Journal of Verbal Learning and Verbal Behavior*, 13: 272–81.

Shi, D. (1990). 'Is There Object to Subject Raising in Chinese?' in K. Hall, J.-P. Koenig, M. Meacham, and S. Reinman (eds), *Proceedings of the 16th Annual Meeting of the Berkeley Linguistics Society*. Berkeley: Berkeley Linguistics Society, 305–14.

Shibatani, M. (1976). *The Grammar of Causative Constructions*. New York: Academic Press.

Slobin, D. I. (1973). 'Cognitive Prerequisites for the Development of Grammar,' in C. Ferguson and D. I. Slobin (eds), *Studies of Child Language Development*. New York: Holt, Rinehart and Winston, 175–208.

Sonnenschein, A. H. (2005). *A Descriptive Grammar of San Bartolomé Zoogocho Zapotec*. München: Lincom Europa.

Sperber, D. and Wilson, D. (1995). *Relevance: Communication and Cognition*, 2nd edn. Oxford: Basil Blackwell. 1st edn (1986).

Stallings, L. M. (1998). 'Evaluating Heaviness: Relative Weight in the Spoken Production of Heavy-NP Shift.' Ph.D. dissertation, USC.

Stallings, L. M. and MacDonald, M. C. (2011). 'It's Not Just the "Heavy NP": Relative Phrase Length Modulates the Production of Heavy-NP Shift,' *Journal of Psycholinguistic Research*, 40(3): 177–87.

Stefanowitsch, A. (2008). 'Negative Entrenchment: A Usage-based Approach to Negative Evidence,' *Cognitive Linguistics*, 19(3): 513–31.

Svenonius, P. (2000). 'Introduction,' in P. Svenonius (ed.), *The Derivation of VO and OV*. Amsterdam: John Benjamins, 1–26.

Szmrecsanyi, B. (2004). 'On Operationalizing Syntactic Complexity,' *JADT* (7es Journées internationales d'analyse statistique des données textuelles).

Tallerman, M. (1998). *Understanding Syntax*. London: Arnold.

Tanenhaus, M. K. and Carlson, G. N. (1989). 'Lexical Structure and Language Comprehension,' in W. Marslen-Wilson (ed.), *Lexical Representation and Process*. Cambridge, Mass.: MIT Press, 529–61.

Temperley, D. (2007). 'Minimization of Dependency Length in Written English,' *Cognition*, 105: 300–33.

Tesnière, L. (1959). *Éléments de Syntaxe Structurale*. Paris: Klincksieck.

Thomason, S. (2001). *Language Contact: An Introduction*. Edinburgh: Edinburgh University Press.

Thomason, S. and Kaufman, T. (1988). *Language Contact, Creolization and Genetic Linguistics*. Berkeley: University of California Press.

Tily, H. and Jaeger, T. F. (2011). 'Complementing Quantitative Typology with Behavioral Approaches: Evidence for Typological Universals,' *Linguistic Typology*, 15: 479–90.

Tomasello, M. (2003). *Constructing a Language*. Cambridge, Mass.: Harvard University Press.

Tomlin, R. S. (1986). *Basic Word Order: Functional Principles*. London: Croom Helm.

Traugott, E. C. (1992). 'Syntax,' in R. M. Hogg (ed.), *The Cambridge History of the English Language*, vol. 1: *The Beginnings to 1066*. Cambridge: Cambridge University Press, 168–289.

Traugott, E. C. and Heine, B. (eds) (1991). *Approaches to Grammaticalization*. Amsterdam: John Benjamins.

Traugott, E. C. and König, E. (1991). 'The Semantics-Pragmatics of Grammaticalization Revisited,' in E. C. Traugott and B. Heine (eds), *Approaches to Grammaticalization*. Amsterdam: John Benjamins, 189–218.

Trudgill, P. (2009). 'Sociolinguistic Typology and Complexification,' in G. Sampson, D. Gil, and P. Trudgill (eds), *Language Complexity as an Evolving Variable*. Oxford: Oxford University Press, 98–109.

Trudgill, P. (2011). *Sociolinguistic Typology*. Oxford: Oxford University Press.

Tsao, F.-F. (1978). *A Functional Study of Topic in Chinese*. Taipei: Student Book Company.

Tsunoda, T., Ueda, S., and Itoh, Y. (1995). 'Adpositions in Word-Order Typology,' *Linguistics*, 33: 741–61.

Ueno, M. and Polinsky, M. (2009). 'Does Headedness Affect Processing? A New Look at the VO-OV Contrast,' *Journal of Linguistics*, 45(3): 675–710.

Ultan, R. (1978). 'Some General Characteristics of Interrogative Systems,' in J. H. Greenberg (ed.), *Universals of Human Language*. Stanford: Stanford University Press, 211–48.

Uszkoreit, H., Brants, T., Duchier, D., Krenn, B., Konieczny, Oepen, S. and Skut, W. (1998). 'Studien zur performanzorientierten Linguistik: Aspekte der Relativsatzextraposition im Deutschen,' *Kognitionswissenschaft*, 7: 129–33.

Vasishth, S. and Lewis, R. (2006). 'Argument-head Distance and Processing Complexity: Explaining both Locality and Anti-locality Effects,' *Language*, 82: 767–94.

Vincent, N. B. (1987). 'Latin,' in M. B. Harris and N. B. Vincent (eds), *The Romance Languages*. Oxford: Oxford University Press, 26–78.

Wasow, T. (1997). 'Remarks on Grammatical Weight,' *Language Variation and Change*, 9: 81–105.

Wasow, T. (2002). *Postverbal Behavior*. Stanford: CSLI Publications.

Wasow, T., Jaeger, T. F., and Orr, D. M. (2011). 'Lexical Variation in Relativizer Frequency,' in H. Simon and H. Wiese (eds), *Expecting the Unexpected: Exceptions in Grammar*. Berlin: Mouton de Gruyter, 175–95.

Wiechmann, D. and Lohmann, A. (2013). 'Domain Minimization and Beyond: Modeling PP Ordering,' *Language Variation and Change*, 25(1): 65–88.

Wivell, R. (1981). 'Kairiru Grammar.' MA thesis, University of Auckland.

Wu, F. (2009). 'Factors Affecting Relative Clause Processing in Mandarin.' Unpublished doctoral dissertation, University of Southern California, Los Angeles.

Yamashita, H. (2002). 'Scrambled Sentences in Japanese: Linguistic Properties and Motivation for Production,' *Text*, 22: 597–633.

Yamashita, H. and Chang, F. (2001). 'Long Before Short Preference in the Production of a Head-Final Language,' *Cognition*, 81: B45–B55.

Yamashita, H. and Chang, F. (2006). 'Sentence Production in Japanese,' in M. Nakayama, R. Mazuka, and Y. Shirai (eds), *Handbook of East Asian Psycholinguistics*, vol. 2: *Japanese*. Cambridge: Cambridge University Press, 291–7.

Yngve, V. H. (1960). 'A Model and an Hypothesis for Language Structure,' *Proceedings of the American Philosophical Society*, 104: 444–66.

Zipf, G. (1949). *Human Behavior and the Principle of Least Effort*. New York: Hafner.

Zwicky, A. M. (1985). 'Heads,' *Journal of Linguistics*, 21: 1–29.

This page intentionally left blank

Index of Languages

Index of Authors

Index of Subjects