In this major new book, John A. Hawkins argues that processing can provide a simple, functional explanation for syntactic rules of ordering, and for many other syntactic phenomena across languages. Constituents are ordered the way they are so that the processing of syntactic structure can be rapid and efficient in language use. Insights from generative syntax, typological studies of language universals and psycholinguistic studies of language processing are combined to show that there is a profound correspondence between performance and grammar. The theory explains such phenomena as the Greenbergian word-order universals, many ordering parameters of current generative theory, and also numerous hitherto unexplained ordering regularities across languages and in the rule formulations of the better-studied languages. The major ordering principle proposed, Early Immediate Constituents, is tested on performance data from ten distinct languages. The result is a unified theory of free and fixed word orders, in which ordering is driven by efficient structure recognition in both cases and not by pragmatics in the one case and innate syntax in the other.

This is the first book to systematically investigate the processing basis of syntactic rules and of cross-linguistic regularities. Other areas of syntax that are discussed include universals of relative clause formation and the head of phrase generalization. The book will be welcomed by all those interested in syntax, language universals, psycholinguistics, and cognitive science generally.
A performance theory of order and constituency
In this series

52 Michael S. Rochemont and Peter W. Culicover: English focus constructions and the theory of grammar
53 Philip Carr: Linguistic realities: an autonomist metatheory for the generative enterprise
54 Eve Sweetser: From etymology to pragmatics: metaphorical and cultural aspects of semantic structure
55 Regina Blass: Relevance relations in discourse: a study with special reference to Sisala
56 Andrew Chesterman: On definiteness: a study with special reference to English and Finnish
57 Alessandra Giorgi and Giuseppe Longobardi: The syntax of noun phrases: configuration, parameters and empty categories
58 Monic Charette: Conditions on phonological government
59 M. H. Klaiman: Grammatical voice
60 Sarah M. B. Fagan: The syntax and semantics of middle constructions: A study with special reference to German
61 Anjum P. Saleemi: Universal Grammar and language learnability
62 Stephen R. Anderson: A-Morphous Morphology
63 Lesley Stirling: Switch reference and discourse representation
64 Henk J. Verkuyl: A theory of aspectuality: the interaction between temporal and atemporal structure
65 Eve V. Clark: The lexicon in acquisition
66 Anthony R. Warner: English auxiliaries: structure and history
67 P. H. Matthews: Grammatical theory in the United States from Bloomfield to Chomsky
68 Lilijana Progovac: Negative and positive polarity: a binding approach
69 R. M. W. Dixon: Ergativity
70 Yan Huang: The syntax and pragmatics of anaphora
71 Knud Lambrecht: Information structure and sentence form: Topic, focus, and the mental representations of discourse referents
72 Luigi Burzio: Principles of English stress
73 John A. Hawkins: A performance theory of order and constituency

Supplementary volumes

Michael O. Siadhail: Modern Irish: grammatical structure and dialectal variation
Annick De Houwer: The acquisition of two languages from birth: a case study
Liliane Haegeman: Theory and description in generative syntax: A case study in West Flemish
A. E. Backhouse: The lexical field of taste: A semantic study of Japanese taste terms

Earlier issues not listed are also available
This book is dedicated to my collaborators, with gratitude.
Contents

Preface xi
Acknowledgments xiv
List of abbreviations xvii

1 Introduction 1
1.1 A processing approach to syntax 1
1.2 Some proposed processing explanations for grammar 3
1.3 Some problems for processing explanations 7
1.4 The approach of the present book 10
1.5 Towards a theory of language universals 16

2 The grammaticalization of processing principles 19
2.1 How is a processing principle expressed in a grammar? 19
2.2 A theory of structural complexity 24
2.3 Complexity predictions for grammatical variation across languages 37
2.4 Structural Domains and their semantic correlates 50

3 Early immediate constituents 57
3.1 The basic idea 57
3.2 Some principles of parsing 58
3.3 Some preliminary evidence for EIC 64
3.4 Defining and measuring EIC 69
3.5 EIC predictions for performance 83
3.6 EIC predictions for the grammar 95
3.7 EIC's predictions for typologically different structures 102
3.8 The relationship of EIC to other principles 106
## Contents

4  Testing EIC's performance predictions 121
   4.1 Arrangements of free word orders in multiple branching structures 122
   4.2 Binary branching structures 171
   4.3 Rearrangements of basic word orders in performance 180
   4.4 Comparing EIC with pragmatic predictions 214
   4.5 Temporary ambiguity preferences 242

5  Testing EIC's grammatical predictions 250
   5.1 Binary branching structures 251
   5.2 Multiple branching structures 269
   5.3 A look at English 282
   5.4 Grammatical rearrangements 296
   5.5 Word-order hierarchies 315
   5.6 Left–right asymmetries 321

6  Grammaticalized node construction 342
   6.1 The head of phrase generalization 343
   6.2 Some additional construction principles 359
   6.3 Some predicted category and ordering alternations 380

7  Conclusions 409
   7.1 Summary of the book 411
   7.2 Major findings and further issues 423

   Envoi 440
   Notes 442
   References 470
   Index of names 483
   Index of languages 486
   Subject index 488
To what extent are the grammatical conventions of human languages shaped by performance? In particular, can rules and principles of the syntax be explained by considerations of processing, and by the temporal sequencing of speech? There has been some discussion of this question in the past, among linguists and psychologists, and the consensus seems to be: only to a limited extent. That is, processing may nibble at the edges of the grammar but does not appear to be a major determinant of syntactic conventions.

I believe that this consensus is mistaken. It comes from looking at insufficient data from too few languages, and from accepting too uncritically certain common assumptions in theories of grammar and in theories of performance. When we expand the data and the language sample, and pursue other theoretical assumptions, we reach a very different conclusion: namely, that grammars are profoundly shaped by processing. I believe that even highly abstract and fundamental properties of syntax can be derived from simple principles of processing that are needed anyway, in order to explain how language is used.

This conclusion has been reached after a close analysis of linear ordering in performance and in grammars. It is further motivated by the correspondences between on-line procedures for recognizing constituent structure and the grammatical devices that make such recognition possible. It is also supported by grammatical constraints on relativization, on movement, and on various other phenomena across languages. As a result, performance now joins the core of an explanatory model of syntax. The role of an innate Universal Grammar is correspondingly reduced, while that of (ultimately innate) processing mechanisms is increased. The processing mechanisms in question are those that make it possible for humans to recognize (and produce) grammatical structure in a rapid and efficient manner, and they can be seen as the syntactic counterparts of the highly efficient (and again innate) mechanisms for sound and word recognition that have been experimentally validated in numerous psycholinguistic studies.
The methodology that I have employed in reaching this conclusion has consisted in integrating and extending ideas from three research traditions: generative syntax; typological studies of language universals; and psycholinguistics, specifically language processing. The relation between performance and competence has been discussed hitherto on the basis of only a limited range of languages, primarily English. By systematically juxtaposing psycholinguistic and generative-syntactic insights with comparative data from many languages, a lot of the issues appear in a new light. Most of the book is devoted to ordering phenomena, and in this area my collaborators and I have gathered performance data from ten typologically distinct and genetically diverse languages, which are then juxtaposed with grammatical data from hundreds of languages taken from current samples. By taking a large-scale comparative perspective, the set of grammatical generalizations about human language can be extended from the absolute universals and parameters of the generative paradigm to the implicational and distributional universals of the typological approach. By matching the performance data with the implicational structuring of grammatical conventions, and with statistical preferences across grammars, a profound correspondence between performance and grammar becomes clearly visible.

Implicationally related properties such as “if SOV, then postpositions in PP” co-occur because they are optimal for processing and involve minimal complexity: fewer computations are required over smaller numbers of entities in order to reach a given parsing decision, involving in this case the recognition of constituent structure. Where there are implicational hierarchies, the higher positions in the hierarchy involve less complexity than the lower positions. Similarly, frequently occurring structures across languages involve less complexity than infrequently occurring structures, all things being equal. This form of explanation is proposed for Keenan and Comrie's (1977) Accessibility Hierarchy for relativization. A processing explanation is also given for certain absolute universals of constituent structure organization, such as the head of phrase generalization.

In addition to arguing the case for “Performance-Driven Grammar,” this book presents a number of more specific findings. It is argued, on the basis of a detailed textual analysis of several languages, that free word order is not driven by pragmatic considerations, as is generally believed, but by syntactic processing, specifically by the need to recognize syntactic structure rapidly on-line. The major ordering principle proposed, Early Immediate Constituents, explains both free ordering and the grammaticalized ordering conventions of fixed word orders. Ordering, in our account,
becomes the handmaiden of constituent structure recognition, and this unification of performance and grammar means that performance data can also contribute to syntactic argumentation in deciding about the best constituent structure analysis and about whether there is a grammaticalized order at all in a given structure and language.

The book also offers a new theory of syntactic complexity, defined as an additive function applied to grammatically or psycholinguistically significant subsets of nodes within a tree diagram. This complexity metric exhibits clear parallels with metrics that are being developed in other grammatical areas, and it makes possible a reduction of Frazier's (1979a, 1985) various parsing preferences (Minimal Attachment, etc.) to just a single principle, namely "choose the least complex structure when the parser has a choice and all else is equal."

In a nutshell, what I have attempted in this book is an integration and extension of what I consider to be mutually reinforcing insights from different subdisciplines of linguistics and psycholinguistics. This kind of integration is never straightforward, and its success or otherwise will ultimately depend on the details, and on whether rather subtle properties of syntax, whose explanation is not obviously one of performance, can nonetheless be convincingly derived from the proposed processing axioms and principles. I invite the reader to decide.
Acknowledgments

I am indebted to numerous individuals, institutions, and funding sources, for assisting me in the completion of this book. The people who must be mentioned first are my collaborators, who collected data for me and tested EIC’s predictions on them, and to whom this book is dedicated. They are, in alphabetical order: Kaoru Horie (Japanese); Charles Kim (Korean); Stephen Matthews (Hungarian); Christine Mooshammer (German); Beatrice Primus (German and Rumanian); Charlotte Reinholtz (Turkish); Anna Siewierska (Polish); Yakov Testelec (Caucasian languages); and Maria Vilkuna (Finnish). Other individuals whose published data on linear ordering in performance and grammar have been crucial include: Matthew Dryer; Peter Erdmann; Lyn Frazier; Katalin Kiss; Chryssoula Lascaratou; and Tom Shannon.

Some of these linguists (Kiss, Primus, Siewierska, Testelec, Vilkuna, and myself) are currently members of the Constituent Order Group of the Programme in Language Typology, funded by the European Science Foundation, and one of the research goals of this group has been the testing of EIC (cf. the forthcoming proceedings of the project). The financial assistance provided by the ESF for this purpose is hereby gratefully acknowledged. Other individuals (Horie, Kim, Matthews, Reinholtz, and myself) received funds from the University of Southern California Faculty Research and Innovation Fund (FRIF), which are also gratefully acknowledged, as is a summer stipend which I was awarded by the National Endowment for the Humanities (FT–34150–90) in 1990.

The basic thesis of this book, that grammar is in large measure performance-driven, was first developed while I was a senior research associate at the Max-Planck-Institut für Psycholinguistik in Nijmegen, directing a generously funded interdisciplinary project on Explanation in Universal Grammar (cf. e.g. my edited volume Explaining Language Universals, Basil Blackwell, Oxford, 1988). I am grateful to the Max-Planck-Gesellschaft and also to the Netherlands Institute for Advanced Study (NIAS) for supporting this
Acknowledgments

project, and for providing an environment in which ideas could be shared across traditional boundaries, and in which the seeds for the current book could be planted. This theme was also explored in a workshop on "The Evolution of Human Languages," organized by Murray Gell-Mann and myself at the Santa Fe Institute, New Mexico, and funded by that Institute (cf. J. A. Hawkins and M. Gell-Mann, editors, *The Evolution of Human Languages*, Addison Wesley, Reading, Mass., 1992). The SFI's support is gratefully acknowledged as well.

The first draft of this book was completed back in 1987, and has been revised and extended since then in response to two very helpful reviews from Cambridge University Press, received in 1988. I am grateful to the reviewers for their detailed comments. A second draft was completed during a whole year of sabbatical and study leave from the University of Southern California during 1990–1991, for which I am also most thankful. The third and final draft was completed in August 1993.

Portions of the material in this book have been presented at the following universities and research institutes (arranged by country), whose invitations and remunerations are gratefully acknowledged: University College London; the Free University of Berlin, the universities of Bochum, Hannover, München, Wuppertal and the Saarland; the Max-Planck-Institut für Psycholinguistik, Nijmegen; Bell Communications Research New Jersey, the University of Delaware, MIT, Stanford University, UCLA and USC. This material was also presented at: the International Round Table on Heads in Linguistic Theory (Talking Heads), University of Surrey, 1991 (*in absentia*); the annual meeting of the Deutsche Gesellschaft für Sprachwissenschaft, Universität Bremen, 1992; and the International Symposium on Japanese Syntactic Processing, Duke University, 1991. All of these institutions and meetings provided useful feedback, which I have tried to incorporate in the final text wherever possible.

I have also benefited enormously from discussing my data and findings with many individuals too numerous to mention. I would like to single out the following, however, whose comments and perspectives have been especially valuable (in alphabetical order): Steven Abney, Elaine Andersen, Joseph Aoun, Bernard Comrie, Anne Cutler, Matthew Dryer, Anne Dunlea, Peter Erdmann, Janet Fodor, Erica Garcia, Jim Gee, Kaoru Horie, Joachim Jacobs, Ed Keenan, Charles Kim, Ekkehard König, Manfred Krifka, Stephen Levinson, Hans-Heinrich Lieb, William Marslen-Wilson, Peter Matthews, Stephen Matthews, Beatrice Primus, Pieter Seuren, Anna Siewierska, Neil Smith, Lynne Stallings, Yakov Testelec, Lorraine
Acknowledgments

Tyler, and Jean-Roger Vergnaud. None of these individuals necessarily agrees with me or with the use that I have made of their input.

A penultimate word of gratitude must go to the University of Southern California in Los Angeles, which has sustained me all these years and provided a rich and fulfilling environment in which I could pursue my work. Special mention must be made of Dean Marshall Cohen, who has presided with such distinction over the Division of Humanities during the last ten years, and to whom the Linguistics Department is profoundly indebted.

Finally, a cryptic message for long-suffering Annie: “The damn thing’s done! Remember Rowley Manor?”
Abbreviations

Abs  Absolutive case
Acc  Accusative case
Adj(P)  Adjective (Phrase)
Adv(P)  Adverb (Phrase)
Agg  Aggregate
Agg  Aggregate
AgP  Agreement Projection
AH  Accessibility Hierarchy
Art  Article
AtC  Attachment Construction
AtP  Attachment Promotion
Aux(P)  Auxiliary verb (Phrase)
Aux  Finite auxiliary verb
AuxNeg  Negative auxiliary verb
AuxT/A  Auxiliary verb designating tense or aspect
Caus  Causative
CLS  Conservation of Logical Structure
Comp  Complement of a head (in X-bar theory)
Comp(P)  Complementizer (Phrase)
Cond  Conditional
Cont  Continuous
CP  Complementizer Phrase
CRD  Constituent Recognition Domain
Dat  Dative case
Decl  Declarative
Def  Definite
Dem  Demonstrative determiner
Dep  Verbal dependent
Det(P)  Determiner (Phrase)
DO  Direct Object
Abbreviations

E  Expression node
EIC  Early Immediate Constituents
Erg  Ergative case
ES  Expanded sample
Fem  Feminine
G  Grandmother node
Gen  Genitive case or Possessive Phrase
Ger  Gerund
GN  Given–New
GNC  Grandmother Node Construction
GNCC  Grandmother-node-constructing category
GPSG  Generalized phrase-structure grammar
I  Importance (scores for an IC in a text)
IC  Immediate constituent
IC-1  Initial immediate constituent
ICA  Immediate Constituent Attachment
IC_m  An IC constructed on its right periphery by an MNCC or other constructing category
IMD  Immediate Matrix Disambiguation
Imp  Imperative
Indef  Indefinite
IND-OBJ  Indirect Object
Inf  Infinitive
Infl(P)  Inflection (Phrase)
IO  Indirect Object
L-to-R  Left to right
LIPOC  Language-Independent Preferred Order of Constituents
Loc  Locative
m  Mother-node-constructing category or other constructing category
M  Mother node
Masc  Masculine
mIC  An IC constructed on its left periphery by an MNCC or other constructing category
MNC  Mother Node Construction
MNCC  Mother-node-constructing category
NG  New–Given
Nom  Nominative case
Noml  Nominalizing particle
N(P)  Noun (Phrase)
NP    Accusative non-pronominal NP
O     Direct Object
OBJ   Direct Object
OBL   Oblique argument
OCOMP Object of Comparison
OSV   Object–Subject–Verb
OVS   Object–Verb–Subject
P     Predictability (scores for an IC in a text)
Part  Particle
Perf  Perfect
Pl    Plural
PNCC  Phrasal-node-constructing category
Poss(P) Possessive (Phrase)
P(P)   Preposition or Postposition (Phrase)
Pred(P) Predication (Phrase)
Prep  Preposition
Pres  Present tense
PrNMH Prepositional Noun-Modifier Hierarchy
Pro   Pronoun
PS rules Phrase-Structure rules
QP    Quantifier Phrase
QU    Question particle
RD    Relativization Domain
Rel   Relative clause
S     Sentence or Subject
Ș     Subordinate clause
SD    Structural Domain
Sg    Singular
SNC   Sister Node Construction
SOV   Subject–Object–Verb
Spec  Specifier
SU    Subject
Subdn Subordination marker
SUBJ  Subject
Sr    Root clause
Ss    Subordinate clause
SVO   Subject–Verb–Object
θ-roles Thematic roles
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top(P)</td>
<td>Topic (Phrase)</td>
</tr>
<tr>
<td>TU</td>
<td>Task Urgency</td>
</tr>
<tr>
<td>UG</td>
<td>Universal Grammar</td>
</tr>
<tr>
<td>$V_f$</td>
<td>Finite verb</td>
</tr>
<tr>
<td>$V_{nf}$</td>
<td>Non-finite verb</td>
</tr>
<tr>
<td>VOS</td>
<td>Verb–Object–Subject</td>
</tr>
<tr>
<td>V(P)</td>
<td>Verb (Phrase)</td>
</tr>
<tr>
<td>VP</td>
<td>Subordinate Verb Phrase</td>
</tr>
<tr>
<td>VSO</td>
<td>Verb–Subject–Object</td>
</tr>
<tr>
<td>XP</td>
<td>A phrase of any type</td>
</tr>
</tbody>
</table>
1 Introduction

1.1 A processing approach to syntax

Since the publication of Chomsky (1957) we have seen the development of a rich tradition of generative research into syntax. This tradition began by incorporating structuralist insights (cf. Saussure 1916; Bloomfield 1933; Harris 1951) into the kinds of grammatical rules that were widely employed in traditional grammars such as Jespersen (1924), thereby making these latter more explicit and more predictive in relation to grammaticality data (cf. Chomsky 1956). It led to a hypothesis about the role of grammatical knowledge within a larger theory of performance (cf. Chomsky 1965). It has contributed fundamental new insights into basic properties of natural language syntax and also into semantics, e.g. "constituent-command" (c-command) (cf. Reinhart 1983). It has elevated the goals of grammatical analysis beyond the facts of individual languages towards the construction of a theory of Universal Grammar (or UG), with built-in parameters of variation (cf. Chomsky 1981, 1982). And it has led to the formulation of numerous alternative models based on competing foundational concepts, whose empirical consequences are currently being worked out (cf. e.g. Aoun 1985, 1986; Gazdar, et al. 1985; Chomsky 1986; Blake 1990; and the papers in Baltin and Kroch 1989).

The philosophy that has been proposed as an explanatory background to, and motivation for, these generative descriptions is the innateness hypothesis (cf. e.g. Chomsky 1968). The universals of grammar are claimed to be innately represented in the human species and to be part of our genetic endowment, and it is this that is argued to explain the remarkable rapidity of language acquisition by children, despite the impoverishment and insufficiency of the input data to which the child is exposed (cf. Hoekstra and Kooij 1988).

Contrasting with these generative theories are approaches to syntax that are generally referred to as "functional." The basic tenet of works such as
Introduction

Givon (1979), Hopper and Thompson (1980), and DuBois (1987) is that fundamental properties of syntax, for example the grammatical properties of transitivity or of ergativity, are shaped by their functions in discourse, i.e. syntax is the way it is because of the functions that it is used to perform in actual communication. Functional approaches may appeal also to a more cognitive explanation for grammatical structure, whereby properties of the real world, as interpreted by human cognitive mechanisms, are reflected in natural language syntax, e.g. in grammatical areas for which Haiman (1985) has proposed an iconicity explanation. Some have questioned whether it makes sense for discourse as such to be causally prior, and have recast Hopper and Thompson's (1980) discourse basis for transitivity in more cognitive and semantic terms, cf. DeLancey (1987) and Croft (1990).

These functional approaches are typically associated with the typological paradigm of Universal Grammar (cf. Hawkins 1988a for a summary). But in fact the equation of functionalism with typology fails in both directions: there can be, and are, functional approaches to generative syntax; and typology, as a theory and enumeration of different language types, is as compatible with a generative paradigm as it is with a discourse-based or cognitive one.

A functional approach to generative syntax is simply one that seeks to explain the properties of formal syntactic rules, principles, or constructions in terms of language use. Not all such properties, and not even all universals of language, will necessarily be explainable in innateness terms, as Newmeyer (1990a) points out, and hence the generative enterprise does not exclude functional explanations in principle. Moreover, there is a class of functional explanations that are particularly appropriate for generative syntactic descriptions, namely those that derive from the processing of syntax in real time, as exemplified in the work of Janet Fodor and Lyn Frazier (cf. section 1.2). And Newmeyer (1990b, 1991) even suggests that such processing considerations might explain why some grammatical properties have become biologized and innate in the evolution of our species, whereas others have not. As for the nature of typology, the parameters of generative theory and their correlated properties define possible and impossible language types, just as implicational universals do within the tradition of Greenberg (1966) (cf. further Hawkins 1988a).

As we survey the current syntactic scene, therefore, there is no shortage of syntactic models for describing the grammatical details of various languages in terms of proposed universal primitives. There is also a large body of functional proposals that seek to provide motivations for syntactic phenomena in discourse and cognitive-semantic terms. But what has not yet received
the systematic attention it deserves is the question: to what extent is syntax the way it is because of processing? Sentences need to be produced and understood in real time. As psycholinguists develop models of production and comprehension, and as more and more data become available especially about how language is recognized on-line, it becomes natural to ask what role such processing constraints impose on grammars. That is the purpose of the present book. I will seek to show that the explanatory role of this particular functional consideration is considerable, that it can potentially replace many principles of UG (i.e. there is a performance explanation for much of what is regarded as the innate and universal core of syntax), and moreover that it is often processing, and not discourse or pragmatics, that shapes syntax when there is a functional explanation.

This book explores the processing basis of syntax, therefore, with particular reference to the order of meaningful elements, and with reference to certain general properties of constituent structure. I believe that this approach can illuminate a number of facts and regularities that are currently mysterious, or apparently arbitrary, by placing them in the context of the mechanisms by which language is produced and understood. We will proceed from questions such as: how can listeners work out what the syntactic structure of the intended sentence is, given the phonetic speech stream provided by the speaker? and what structures do speakers and hearers find more or less complex to manipulate in real time? When we view the data from this perspective, explanations can often be given for what are currently pure stipulations in the grammar, and one descriptive approach to the facts can be motivated over another.

1.2 Some proposed processing explanations for grammar

There is a handful of proposed processing explanations for grammatical phenomena in the current research literature, and these need to be reviewed at the outset. The assumption that underlies them is that there are general psychological mechanisms by which language is produced (cf. e.g. Levelt 1989) and comprehended (cf. e.g. Fodor, et al 1974) in real time, and that explain how it is possible for language users to communicate information rapidly and efficiently using the language code. Since its inception, the field of psycholinguistics has provided a growing body of experimental findings that shed light on these mechanisms. A further widely held assumption is that there are innate neural structures for the processing of language of the same basic kind that regulate other cognitive and perceptual capacities, such as
vision. Achieving efficiency and rapidity in comprehension is argued by Jerry Fodor (1983) to require a reflex-like and neurally specialized response to the incoming speech signal and to the mental representations derived from it, in the manner of his modularity thesis. For example, speech sounds can be recognized at an extremely rapid rate, up to 25 phonemes per second, whereas non-speech sounds cannot be identified as rates exceeding 7–9 items per second (cf. Lieberman 1992). Hence it is reasonable to assume that we are biologically pre-programmed to perceive human speech. There are also evidently neural mechanisms for recognizing and accessing individual words in the incoming speech signal, and these mechanisms are what underlie the extremely rapid and efficient word-recognition processes for which psycholinguists such as Marslen-Wilson and Tyler (1980) have presented experimental evidence. What they have shown is that words are recognized either at or prior to their “uniqueness point,” i.e. that point at which a word becomes uniquely different from all the other words in the language that share the same initial sound segments in the incoming speech stream. Given the rapidity of the speech stream, the recognition of the intended word at or prior to its uniqueness point means that there must be an equally rapid and efficient process mapping speech sounds into a mental lexicon in which words are encoded and grouped according to their left-to-right phonological structure, as in a dictionary (cf. further Hawkins and Cutler 1988).

At a more abstract level, we are evidently equipped with mechanisms for recognizing on-line syntactic structure, such as the groupings of words into phrases and constituents. This higher sentence structure is derived crucially from the lexical information provided by word recognition, which in turn presupposes phonological identification. Some proposals for universal syntactic parsing mechanisms can be found in Frazier (1979a), Frazier and Fodor (1978), and Inoue (1991). Again, there are plausibly innate neural mechanisms mediating between word recognition and the construction of syntactic representations. There are also plausibly innate mechanisms for interrelating all the levels of linguistic representation that need to be postulated in an explicit model of grammar and that contribute to the real-time understanding of any utterance. Crucially important, of course, is the mapping from word recognition and syntactic representation onto some representation of meaning, which in turn serves as input to a pragmatically enriched discourse representation.

For all these processes, there is evidence for performance mechanisms that operate obligatorily and immediately in response to the sense data within their respective domains. Psycholinguists disagree, in fact, over the extent to which these domains are “encapsulated” or autonomous from the
"central processes" of natural language understanding, i.e. specialized for the task at hand (cf. Marslen-Wilson and Tyler's 1987 arguments against Jerry Fodor's 1983 modularity thesis). But they are agreed about their fundamental speed and efficiency, and also presumably about the fact that these processing mechanisms are biologically pre-programmed in some form.

Some attempts to explain grammars in terms of processing mechanisms are the following. In phonology there are quite striking regularities in the composition of vowel inventories across languages, as Crothers (1978) has shown. Languages exhibit vowel systems of varying size, comprising three vowels, four, five, six, and so on, and they expand these inventories in largely predictable ways, starting from a basic three-way distinction between [i], [u], and [a]. Lindblom et al. (1984) have provided a processing explanation for this structuring, arguing that these expanding inventories are those that achieve "sufficient perceptual differences at acceptable articulatory costs" within the available phonetic space, and they have corroborated their predictions using computer simulations of the relevant phonetic variables. For a similar explanation for expanding consonant inventories, cf. Lindblom and Maddieson (1988). We have here a production and perception explanation for phonological structuring across languages.

In morphology, Hawkins and Gilligan (1988) document a clear preference for suffixing versus prefixing and infixing across languages, following early work by Greenberg (1957) in this area; and Hawkins and Cutler (1988) offer an explanation for this preference in terms of principles of morphological processing. Cf. Hall (1991) for a revised and more detailed explanation.

In syntax, Kuno (1973a, 1974) and Dryer (1980) have shown that there is a correlation between the perceptual difficulty of center embeddings and their frequent ungrammaticality in English and many other languages, cf. example (1.1) from English and the cross-linguistic summary of data provided in Hawkins (1988b):

(1.1) a. *Did §[that John failed his exam] surprise Mary?  
b. Did it surprise Mary §[that John failed his exam]?

The even more systematic avoidance of self-embedded relative clauses across languages, as in (1.2), has been given a processing explanation in Frazier (1985) in terms of the excessive number of non-terminal syntactic nodes that need to be postulated within a local viewing window of three adjacent words:

(1.2) *The farmer §[that the cow §[that gave bad milk] kicked] died.
Frazier (1979b, 1985) has examined the word-order universals of Greenberg (1966) from the perspective of the two-stage parsing model developed in Frazier and Fodor (1978), in which the first stage has a limited viewing window of five to six adjacent words. She argues that certain syntactic sequences, such as (1.3) (a postnominal relative clause within a noun phrase co-occurring with a postposition within a postpositional phrase) would regularly exceed this limit, and she predicts that they will be non-occurring:

\[(1.3) \text{pp[NP[Head Noun \Rel. Clause] Postposition]}
\text{Y Complement X}
\text{of Y}\]

Janet Fodor (1978, 1984) has proposed a parsing explanation for the Nested Dependency Constraint, illustrated in (1.4):

\[(1.4) \text{What are boxes easy to store } \text{in } \text{?}\]

The nested reading \text{store boxes in what} is grammatical, but the intersecting \text{store what in boxes} is not, and Fodor argues that this constraint has been grammaticalized in response to pressure from the parser. The cross-linguistic generality of this constraint has yet to be established, and for at least one language (Dutch) there are systematic exceptions, cf. Steedman (1984).

Frazier (1985) discusses "garden-path" sentences, which are known to cause processing difficulty, and notes that they do nonetheless occur quite frequently in English. She argues, however, that structures that would garden-path the hearer on every occasion of use regardless of the particular lexical items chosen will be blocked by the grammar. An example would be (1.5b) which will always be misanalyzed as a main clause on-line, prior to reanalysis as a subordinate clause when the matrix verb is encountered (cf. Bever 1970):

\[(1.5) \text{a. That Mary is sick surprised Bill.}
\text{b. *Mary is sick surprised Bill.}\]

Frazier labels this the Impermissible Ambiguity Constraint.

A rather different tradition of research has also contributed to parsing explanations for grammatical principles, that of computational linguistics, exemplified in works such as Marcus (1980) and Berwick and Weinberg.
Some problems for processing explanations

(1984). One principle that has been much discussed is subjacency, the proposal that no rule can relate two constituents separated by more than one bounding node (cf. e.g. van Riemsdijk and Williams 1986). Berwick and Weinberg offer an explanation for this in terms of the bound on left context within their proposed parser.

1.3 Some problems for processing explanations

These explanations are very suggestive, but they raise more questions than they answer, in my opinion, and they do not yet enable us to conclude that processing has shaped grammar to any significant extent. Why not?

The first issue that has to be resolved involves the precise status of many of the sentences under discussion. Are they unacceptable in performance and also ungrammatical (i.e. not generated by the relevant grammar)? Or unacceptable, yet still grammatical? If the latter, then any processing difficulty associated with e.g. center embeddings (cf. (1.1a)) or self-embeddings (cf. (1.2)) will not be reflected in any grammatical rules or constraints, since grammars will freely generate such structures. If there are rules or constraints blocking these structures, then processing may be a possible explanation.

The Standard Theory of generative grammar (cf. Chomsky 1965) was founded on the premise that processing difficulty could determine only the acceptability status of sentences, and not their grammaticality. The rules of the grammar generate all and only the grammatical sentences of a given language, and these are then assigned different degrees of acceptability within a performance model that contains grammar as one of its components. I will call this the assumption of pure acceptability. Naturally, if we accept this assumption, then processing difficulty will not be reflected in grammars, and performance and competence will be quite orthogonal.

One possibility that is admitted by Chomsky involves processing explanations that hold at the level of the evolution of the species, cf. Chomsky and Lasnik (1977) and also Newmeyer (1990b, 1991). Considerations of performance might have shaped the human language faculty in its evolution, favoring some innate universals of grammar over others.²

The processing explanations summarized in section 1.2 are clearly predicated on a rejection of the assumption of pure acceptability. That is, they assume that there is some grammatical mechanism blocking sentences of the relevant type, and that processing explains why it operates the way it does. But some of these explanations raise a second set of problems: not all
languages appear to have responded to the processing difficulty in question; and not all sources of processing difficulty seem to result in a regular grammatical response.

For example, if processing difficulty is supposed to explain the absence of center embeddings in English and in many other languages, why is it that exactly the same type of center embedding does nonetheless show up in certain languages, such as Japanese? The structure of (1.6a) is grammatical in Japanese, even though many center-embedded Ss are regularly preposed within their containing clauses in Japanese performance, as in (1.6b) (especially longer ones, cf. ch. 4.1.5), thereby confirming the processing difficulty they cause. Example (1.6a) is nonetheless both grammatical and acceptable:

(1.6) a. Mary-ga ₃[kinoo John-ga kekkonsi-ta to] it-ta.
   “Mary yesterday John married that said,” i.e. Mary said that John got married yesterday.


The productivity of (1.6a) is an embarrassment for whatever explanation is proposed for the status of (1.1a) in English and many other languages. The logical problem here can be summarized as follows: if one is going to appeal to fundamental properties in the architecture of human processing mechanisms in order to explain why grammars block certain structures, then why are they not blocked in all grammars? The same problem arises for self-embedded relative clauses, which are systematically avoided in some languages, but tolerated in others (cf. example (1.9) below).

Conversely, some sources of processing difficulty seem to invite more of a grammatical response than others. Frazier (1985) points out that garden paths occur quite frequently in English (cf. the famous sentence *The horse raced past the barn fell* and many others), whereas center embeddings are regularly avoided. Both types of structures are known to cause processing difficulty, so why the differential response?

In order to avoid these kinds of insufficiencies, another set of processing explanations makes (at least implicitly) what I will call the assumption of absolute unprocessability. Some absolute limit on human processing resources is proposed within a given domain, and a grammatical response is then explained on the grounds that this limit would be exceeded by structures of a certain type. The word order combination of (1.3) above would regularly exceed the five-to-six-word viewing window of the first stage of Frazier and
Fodor's (1978) parser. Subjacency violations would exceed the bound on left context within Berwick and Weinberg's (1984) parser. Hence the relevant structures are necessarily non-occurring.

This assumption leads to a third set of problems, however, because these proposals construct strong claims on rather weak foundations. We just don't know, in the current state of the art, what the absolute limits on the human processing system are. Nor has there been enough cross-linguistic research to establish with confidence that certain types of grammatical phenomena are not attested. Even if we grant that the parser operates in two stages along the lines of Frazier and Fodor (1978), does it necessarily follow that immediate constituents whose recognition exceeds the five to six words of the first-stage viewing window are unprocessable? How good is the evidence for a five-to-six-word limit? Can't other parsing or memory systems interact in these cases and help out? I shall argue in ch. 5.1.3 that there are, in fact, languages whose constituent orders will regularly exceed Frazier's (1985) proposed limit. Similarly, how good is the evidence for Berwick and Weinberg's (1984) parser? And is subjacency a genuine universal? Many surface structures of English are prima facie counterexamples to both subjacency and the Berwick and Weinberg parser, since the WH-word can be separated by two or more bounding nodes from its trace, as in (1.7), and they are real counter-examples to the extent that successive Comp-to-Comp movements (preserving subjacency) are either linguistically or psycholinguistically unmotivated:

(1.7) Who did you say that you imagined Fred would consider asking Susan to bring \( \emptyset \_i \) to the party?

A further problem that will plague all proposals based on the assumption of absolute unprocessability is the directionality-of-explanation issue. Does the grammar respond to the parser, or vice versa? Smith (1986) draws attention to this in his review of Berwick and Weinberg's parsing explanation for subjacency: "one could equally well argue that the (arbitrary) existence of subjacency as a principle of grammar explained the bound on left-context in the parser, as opposed to vice versa" (p. 225).

To summarize, the first set of problems for any processing explanation for grammars involves disentangling the unacceptability-versus-ungrammaticality status of sentences. Secondly, grammars do not respond uniformly to processing difficulty: not all grammars block a structure that is difficult to process; and not all sources of difficulty seem to result in a regular grammatical response. And thirdly, assumptions about the absolute unprocessability of certain structures are largely premature at this point.
1.4 The approach of the present book

I believe that there are solutions to these problems. I believe, moreover, that the impact of processing mechanisms on grammars has been absolutely fundamental, and that by solving these problems we will be able to see the extent of this impact more clearly.

First, I reject the assumption of pure acceptability. Despite the problems alluded to above, there is simply too much anecdotal evidence suggesting that grammars can respond, and have responded, to processing, by incorporating rules, constraints or principles that owe their raison d'être to the need for language users to produce and comprehend language in a rapid and efficient manner in real time. Once we admit that processing can determine grammatical properties in any one area or for just one structural type, we have of course disproved the assumption, and it becomes important to establish to what extent performance is a significant determinant of grammars, and how performance interacts with those aspects of grammar that are independent of it and perhaps derivable from an innate UG. Among the processing explanations enumerated in section 1.2, at least some of them will, it seems to me, be able to survive critical scrutiny in some form, and thereby disprove Chomsky's (1965) assumption of the relationship between grammar and performance.

With the benefit of hindsight, I would suggest that the assumption of pure acceptability was not well argued to begin with. It seems reasonable enough for verb-particle constructions in English such as I called my long lost uncle from America who had been trying to reach me for several weeks without success up, in which increasing distance between verb and particle results in declining acceptability, and there is no clear point at which, and no obvious grammatical mechanism by which, one could stipulate that sentences of this sort have become ungrammatical. But this assumption is, at the very least, questionable for structures such as center embeddings of S in (1.1a), which are also difficult to process, but whose unacceptability/ungrammaticality is now definable in terms of grammatical structure: e.g. star an S with non-null material to both left and right within a containing S.

It is not sufficient to argue that sentences like (1.1a) are grammatical because they can be generated by independently motivated rules or principles of English. For there are countless ungrammatical strings that are blocked by a grammatical constraint on otherwise productive generative mechanisms, especially in the highly interactive, principle-based models of current generative theory. Nor can the "grammatical but unacceptable" status of (1.1a) be
supported by pointing out that it can be processed and understood, given
enough time, pencil and paper, etc., because this is also true of sentences that
are clearly ungrammatical. The sentence *Mary like classical music is
straightforwardly comprehensible, despite its ungrammaticality. Hence,
given the native speaker's unanalyzed rejection of (1.1a), the initial assump-
tion that must be made is that this sentence is ungrammatical. This assump-
tion will need to be revised if the independently motivated format of
grammatical descriptions can come up with no plausible way of formulating
the constraint, principle or rule in question, but in the present instance this is
not the case.

When we take a comparative linguistic perspective and look beyond the
English-only data of Chomsky (1965), the assumption of pure acceptability
becomes quite untenable. For if we assume (along with Frazier 1979a, 1985;
Frazier and Rayner 1988; Inoue and Fodor 1994; Inoue 1991; and many
others) that the fundamental principles of language processing are uni-
versal, then acceptability judgments for comparable structures of comparable
difficulty should be constant across languages, as long as grammatical rules
can be independently motivated that would generate the relevant structures
as grammatical. But judgments are not constant, as we have seen. Take the
center embedding of S in a language with basic SOV word order, as in the
Japanese sentence (1.6a). In a great many languages a sentential direct object
that follows a subject and precedes the verb is regarded as ungrammatical in
the research tradition for that language, and the sentences in question are
neither readily attested nor accepted in performance. Persian is an example
(cf. Dryer 1980):

(1.8) *s[An zan s[ke an mard sangi partab kard] mi danat]
   “the woman that the man rock threw CONT knows,” i.e.
The woman knows that the man threw a rock.

Further examples of this type are German, Yaqui and Turkish (cf. Dryer
1980). But in Japanese, the tradition of grammatical research states that such
sentences are grammatical (cf. Kuno 1973b, Dryer 1980), and they do occur
in performance, albeit with a frequency that declines as the length of the
center embedding increases. This same performance regularity can be seen
with other center-embedded phrases as well: center-embedded NPs and PPs
of increasing length are increasingly dispreferred, but still grammatical (cf.
ch. 4.1.5).

The same problem arises with self-embeddings such as (1.2) in English,
repeated here:
Introduction

(1.2) *The farmer §[that the cow §[that gave bad milk] kicked] died.

This sentence is scarcely usable, even though independently motivated rules should be able to generate it, and even when pragmatic and semantic conditions are fairly optimal, as they are in this case. Cows may indeed give bad milk, and the cow is more likely to have kicked the farmer than vice versa. But sentences with exactly this structure are usable in other languages, under these optimal conditions, and without any of the hesitations that accompany the English (1.2), e.g. in German:

(1.9) Der Bauer §[der die Kuh §[die schlechte Milch gab] schlachtete] ist krank.

"the farmer who the cow which bad milk gave killed" is sick," i.e.

The farmer who killed the cow which gave bad milk is sick.

If these sentences are all considered to be grammatical, the differences between them are unexplained. I conclude that they are not all grammatical. Sentences (1.6a) and (1.8), (1.2) and (1.9), have different statuses because the grammars of Japanese, Persian, English, and German are different. Native speakers of these languages employ the same universal processing strategies that will, in principle, assign the same level of performance difficulty to center embeddings of $\bar{S}$ and to self-embedded relative clauses respectively. What differs is the grammatical response: the center embedding of $\bar{S}$ is ungrammatical in Persian, but grammatical in Japanese; self-embedded relatives are ungrammatical in English, but grammatical in German – i.e. these sentences are blocked versus not blocked by their respective grammars. The challenge posed by these data is, of course, the one I raised in section 1.3: why the differential response? Why aren't self-embeddings ungrammatical in all languages? I will return to this matter presently.

These considerations, coupled with the proposals from the current research literature, argue against the assumption of pure acceptability and in favor of a causative role for performance in shaping grammars. With regard to Chomsky and Lasnik's (1977) rather more abstract argument that performance might gradually have shaped the human language faculty (i.e. innate UG) in its evolution, I wish to sound a cautionary note. This kind of scenario is not impossible, of course, but since processing mechanisms must also be innately determined ultimately, and since they need to be assumed anyway in order to explain how language is performed, why should performance considerations in effect be duplicated from one innate "module" into another,
namely the innate grammar? It is simpler and more economical to assume that the innate processor constrains options that are left open by the innate grammar, whose ultimate causation lies elsewhere. This alternative places the locus of a grammatical response to performance firmly within particular grammars (i.e. precisely the locus that Chomsky’s assumption of pure acceptability rejects), rather than within the innate grammar faculty (which he explicitly endorses). The alternative is supported by the data of this book. It has the advantage that it will enable us to explain a lot of cross-language variation. Either there may be alternative structural variants that are equally optimal for processing; or there may be different degrees of structural complexity, to which grammars will respond in direct proportion to the degree of difficulty.

The main problems for the processing explanations for grammars proposed hitherto are that they are insufficient (i.e. they do not account for all languages, and not all sources of difficulty appear to have structural reflexes); or else that they make an unwarranted assumption of absolute unprocessability. One way to proceed, therefore, is to focus not on proposed absolute limits on the processing system, but on a fundamental fact about performance that everyone seems to agree on, namely that processing difficulty is a matter of degree. There are lots of performance data supporting different degrees of difficulty and complexity associated with structural variants of certain types, data from psycholinguistic experiments, from discourse frequencies, and from acceptability intuitions, and we can harness this evidence and try to develop precise, quantifiable metrics for these degrees. The metric of syntactic complexity that I shall propose in ch. 2 in terms of Structural Domains is a case in point. Early Immediate Constituents, the word-order processing metric that will be used for most of the predictions of this book, is a particular instance of this complexity metric. Such talk of degrees avoids a premature commitment to some absolute limit, such as a five-to-six-word viewing window within a particular parser (etc.), and is empirically more defensible. It may also enable us, ultimately, to account for phenomena that are necessarily non-occurring across languages, by saying that the degree of processing difficulty has exceeded a certain threshold. But for the moment we don’t know what that threshold is.

With regard to the question of whether performance has or has not influenced grammars, we can then juxtapose these metrics, and performance evidence for them, with comparative data from grammars. The cross-linguistic generalizations that we should concentrate on are, correspondingly, not the absolute universals that are supposedly present in or absent from every
grammar, but variation-defining universals, such as implicational regularities, hierarchies such as Keenan and Comrie's (1977) Accessibility Hierarchy, and distributional universals determining language frequencies (cf. Hawkins 1988c for a summary of these different logical types of universals). I will argue that these grammatical data correlate very precisely with degrees of processing ease in performance, and provide evidence for the profound influence of performance on grammars, of a type that has not yet received the systematic attention it deserves. This approach can also avoid the problems of insufficiency discussed above, for we no longer expect that all languages will necessarily block some difficult structure, nor that all sources of difficulty will necessarily be avoided in any given language. Rather, the philosophy that underlies this book can be set out as follows.

If we survey all the structural variants that are attested across languages and permitted by the (ultimately innate) Universal Grammar, it is clear that any one grammar either does not need, or cannot combine, all of the options that are potentially available. Selections are accordingly made within each grammatical domain, e.g. within the set of possible relativization structures, possible movement structures, possible center embeddings and word orders, and these selections are made according to criteria of simplicity, including ease of processing. I claim that the degree of processing complexity will determine which structures are selected for grammaticalization from each domain, and the order in which they are selected in the event that a plurality is grammaticalized. Structures that are easier to process will be grammaticalized before the grammar sanctions a more difficult structure of the relevant type. This results in implicational dependencies such as "if SOV, then postpositions in PP," a word-order co-occurrence that will be argued to be optimal for processing, in contrast to SOV and prepositions. It predicts the implicational structuring of universal hierarchies such as Keenan and Comrie's (1977) hierarchy for relativization, as well as many cross-linguistic frequencies for different structural options. This approach is compatible with the fact that a difficult structure, such as the center embedding of $\bar{S}$, may sometimes be attested. But it is no longer a counter-example to some absolute prohibition. Rather, its relative infrequency is now permitted, and a vague demonstration of the absence of this structure in certain languages will be replaced by a hierarchy of increasingly complex center-embedded nodes (of e.g. NP, $\bar{VP}$, and $\bar{S}$) with precise predictions for the permitted distribution of grammaticality judgments across languages, and for the relative frequencies of the center embeddings, across grammars and in performance.3
The approach of the present book

It is plausible to assume further that languages will strive for some overall balance in their selection of more and less complex options from different grammatical domains. We see clear evidence for this in grammatical domains that are complementary to one another in serving the same expressive function: languages with very few relativization options (e.g. on subjects only, or subjects and direct objects only) have a more complex set of passive-type rules for promoting non-subjects and non-direct objects to more accessible positions (cf. Keenan and Comrie 1977, Givon 1979, Croft 1990); languages with richer and more complex morphological case systems have fewer fixed word orders for subject and object noun phrases, i.e. for the very phrases that are coded by nominative versus accusative or ergative versus absolutive case within the morphology (cf. Keenan's 1978 principle of covariation of functional equivalents); and languages with smaller vowel inventories will require more consonants in order to achieve the necessary minimum of phonological units that are required for expressive power. The predictions that can be made across domains in this way are probably of a statistical nature only; those within domains, however, seem to be extremely precise. The present book focuses more on the latter.

In addition to the strongly cross-linguistic flavor of the current work, which I believe enables us to solve some of the problems enumerated above, I will not hesitate to go beyond current psycholinguistic theorizing and propose hypotheses, derived from cross-linguistic data, for future experimental testing. Psychologists have devoted more attention to word recognition and morphology than they have to syntax, and theories of syntactic parsing and production are still in their relative infancy. One reason why the assumption of pure acceptability could survive as long as it did was because relatively little was known about performance. But once more sophisticated theories become available, and more data from language use, and once the responsiveness of grammars to performance begins to be appreciated, then the conventionalized grammatical rules and structures of different languages can become highly relevant to psycholinguistics, because they may represent a direct response to processing mechanisms. Whether this is the underlying cause in any given case must, of course, be argued for, by establishing a correlation between the grammatical data and some performance theory and its supporting data, and by negotiation with purely grammatical considerations. For example, there may be no grammatical explanation for the facts at hand, merely a grammatical stipulation; the regularity that underlies the grammatical facts or the pattern of cross-linguistic variation may greatly complicate independently motivated grammatical
mechanisms. If, under these circumstances, we can observe a correlation between grammatical regularities and performance data (e.g. involving degrees of performance difficulty), then a performance motivation becomes plausible. If the grammatical and performance data can be further shown to follow from a given theory of performance, holding for the relevant domain(s), then a performance explanation will be very strongly motivated indeed.

I will offer a number of new psycholinguistic proposals in this work. Some of these will build directly on existing work, e.g. the theory of syntactic complexity defined in ch. 2. Others will react against existing proposals such as Frazier's (1979b, 1985) explanation for the Greenbergian word-order correlations, summarized in section 1.2. Yet others will be suggested by the grammatical data themselves. Many cross-linguistic variation patterns are not adequately explained either by existing grammatical theories or by current psycholinguistic proposals. Processing theories will be offered for these, and will be supported by performance data taken either from the published literature or collected specifically for the present purpose (cf. e.g. ch. 4). In some cases I will appeal to what we might call the logical problem of language performance (to borrow a phrase from the related logical problem of language acquisition, cf. Hornstein and Lightfoot 1981). That is, we need to ask how it is possible for language users to successfully produce or comprehend a given structure in real time. How can an abstract syntactic grouping, corresponding to a high-level phrase within a constituent structure tree, actually be recognized, given the rapid phonetic input of the incoming speech stream? When we ask a question like that, we see grammatical rules and principles in a new light and we can propose hypotheses that make sense of a number of grammatical facts that have seemed arbitrary or mysterious hitherto, as in our discussion of the head of phrase generalization in ch. 6. Psychologists can then subject these hypotheses and their predictions to further experimental testing within individual languages, and are hereby invited to do so.

1.5 Towards a theory of language universals

The consequences of this performance approach to grammar and to universals are potentially far-reaching. Within Chomskyan generative grammar (cf. e.g. Chomsky 1965, 1968, 1981), the universals of grammar are explained ultimately in terms of the innate language endowment of the human species. This endowment, according to Chomsky, consists of a sophisticated and abstract form of knowledge about grammatical properties and about the
structure and organization of grammatical generalizations. It is this knowledge that structures the descriptive grammars of the world's languages, and that facilitates the acquisition of language by each successive generation within each language community.

As I see it, particular grammars are ultimately shaped by (at least) two distinct innate modules, not one: by highly general grammatical principles along the lines of generative UG; and by equally general principles of language performance that regulate the processing of all languages. I will argue that the most general and satisfactory explanation for most of the data in this book does not reduce to some principle of grammar. There are descriptive rules, of course, that generate all and only the grammatical strings of each language. But many universal regularities do not follow from general grammatical principles of a type that it is reasonable to postulate for an innate UG. Relativization would be a simpler rule to state in many languages if it did not have to be restricted to just some syntactic positions, e.g. subjects only, or subjects and direct objects. Ungrammatical center embeddings reduce the productivity of many independently motivated rules. These apparent idiosyncracies are motivated by performance, and they result in highly structured patterns, such as universal hierarchies, when we compare the relevant rule outputs across languages.

Therefore, even if we suspect that a grammatical universal is grounded in the innate capacity for language, we cannot just assume that it is innately represented as a grammatical property in some form that corresponds closely to its representation within a generative model. It may be explainable by an (ultimately innate) principle of processing.

Similarly, there are plausible accounts of syntactic structure in terms of human cognition (cf. DeLancey 1987; Croft 1990), much of which is also presumably innately grounded. Also relevant here are Keenan's (1987) logical and semantic explanations for cross-linguistic syntactic and morphological generalizations.

The general philosophy for which I would argue is accordingly one in which the particular grammars of the world's languages are structured by a number of fundamental explanatory forces: innate grammatical knowledge; innate processing mechanisms; cognitive-semantic principles that include an innate core; and presumably other functional considerations as well. What generative grammarians are primarily doing, as I see it, is formulating a set of grammatical conventions for different languages, often at a level of considerable generality. These conventions must then be explained within a modular theory of language universals, one component of which is innate grammatical
knowledge. Grammatical rules and principles, in this view, are simply the data for a more explanatory theory of language universals. Even very general grammatical conventions can no longer be automatically attributed to UG, since they may be the result of performance. Rather, UG must be abstracted out of the mass of language-particular generalizations, by negotiation and consultation with other possible causative mechanisms, such as processing, and establishing the primitives of grammar now becomes a much more complex exercise. It is important to work out the most adequate descriptive formulations of language-particular conventions that we can, and this is what different models of grammar, and different variants of individual models, are currently doing. But such quests for descriptive adequacy should not be an end in themselves: questions of competing formulations for rules, and even of competing grammatical models, should contribute to, and not distract us from, the more explanatory question of why these rules, in any of their proposed formats, have the basic properties that they do and define all and only the outputs that they do. And the most general point of which I hope to convince the reader is that much of syntax is driven by processing considerations, interacting with the other fundamental components of a general theory of language universals.
In this chapter I shall ask how a grammar could, in principle, respond to processing. Grammars have their own basic format of grammatical categories and principles, as defined within a given theory or model. So when we talk about a grammar responding to a principle of performance, or "grammaticalizing" that principle, we have to view this interaction either within the context of some given format, or (as we shall do here) within the context of general properties of many such formats, if one wishes to remain more theory-neutral. The ultimate explanation for a format or formats presumably reduces to some interaction of the innateness, cognitive-semantic and other explanatory primitives underlying universals that were summarized in ch. 1.5. Performance principles contribute to this format and to the associated grammatical principles, but they do so relative to restrictions that are imposed by the other primitives, and they will, in turn, impose restrictions of their own on other explanatory forces. As a result, the precise manner in which a given performance principle can be expressed in a grammar is complex, and the translation from performance to grammar may not always be one-to-one. I begin this chapter by outlining some general ways in which grammars appear to have responded to processing. This enumeration is not intended to be exhaustive. Rather, it serves merely as a background to the major types of grammatical responses for which I shall argue in this book.

2.1 How is a processing principle expressed in a grammar?

First, grammatical rules or principles may select some categories versus others from the total set made available by the theory of grammar, in a way that reflects processing ease. For example, the grammar may block a center-embedded S, as in the English structure (1.1a) repeated here as (2.1a), while tolerating a center-embedded NP in this environment (cf (2.1b)), on the grounds that Ss are typically much longer than NPs in performance, making the containing S more difficult to process:
The grammaticalization of processing principles

(2.1) a. *Did §[that John failed his exam] surprise Mary?
   b. Did NP[the fact] surprise Mary?

Correspondingly, extraposition rules may apply to an Š in many languages, but not to NP (cf. Dryer 1980; Hawkins 1992a). The grammar transforms structures that are regularly difficult in these cases into their preferred counterparts (cf. chs. 4.3 and 5.4).

An interesting dilemma arises when the processing difficulty in question involves terminal elements only and there is no appropriate dominating category in terms of which a grammatical rule or principle can be formulated. An example is the so-called rule of Heavy NP Shift (cf. Ross 1967) exemplified in (2.2b):

(2.2) a. I vp[gave NP[the valuable book that was extremely difficult to find] pp[to Mary]]
   b. I vp[gave pp[to Mary] NP[the valuable book that was extremely difficult to find]]

This rule cannot be stated in terms of “Move NP,” because not all NPs undergo it, for example single-word NPs such as proper names and pronouns. Nor can we propose some subcategory of “Heavy NP,” because such NPs cover a variety of structural types, from complex NPs at the one end that are structurally definable (NP[NP Š]) to those that are simply heavy in terminal elements on the other. But rules do not look at terminal strings—they apply to categories. More generally, heaviness is not a grammatical notion. Whether this “rule” applies at all depends as much on the length of the other constituents of the VP as it does on the length of the heavy NP (cf. ch. 4.3.2). The grammar cannot make a coherent response in these cases, therefore, and this is what explains a lot of confusion in the grammatical literature over how to handle alternations such as (2.2). The approach I shall adopt here is to say that (2.2a) is the basic order of the English VP. The precise status of the rearrangement in (2.2b) will be discussed in ch. 3.5.2.

Secondly, processing ease and efficiency can determine the relative ordering of categories. English has grammaticalized vp[V NP PP] within its VP, and pp[P NP] within its PP, because these are the orders that are most efficient according to the principle of Early Immediate Constituents. V is typically a one-word item, and NP is on aggregate significantly shorter than PP, in contrast to exceptional cases such as (2.2a) in which the NP is longer than the PP, which are exactly the cases that are subject to shift, thereby returning...
these VPs to the short-before-long regularity of the basic order. P before NP will receive a similar explanation. The directionality of rule-governed rearrangements from a basic order is also predicted by Early Immediate Constituents: extraposition-type rules will move an S to the right in a head-initial language, and to the left (in general) within a head-final language.

Thirdly, many mother–daughter relations in phrase structure appear to be driven by processing. If an abstract constituent structure representation is to be clearly recognized and derived from a rapid linear parse string, this imposes certain constraints. Some categories must be capable of providing evidence about higher nodes and about the phrasal groupings of words. This is, I believe, what underlies and motivates the “head of phrase” generalization. It accounts for many correlated phenomena, such as the obligatory presence of certain daughters, and the constituent domains within which relations such as government and subcategorization can hold. This processing perspective on phrase structure also makes sense of facts that are not so readily handled by the head of phrase generalization, involving predicted alternations within and across languages between different daughter categories (not all of them heads) that can serve to construct higher nodes.

Fourthly, processing leads to constraints on otherwise productive grammatical rules. Subjacency has been explained in this way in Berwick and Weinberg (1984). A rather different account is proposed in section 2.3, where it will be argued that subjacency effects constitute an implicational hierarchy across languages, for which there is a syntactic complexity correlate. Similarly, languages in which relativization is constrained in its applicability to different syntactic positions will be explained in terms of the same complexity metric. The Nested Dependency Constraint is another constraint motivated by processing (cf. ch. 1.2).

Fifthly, and related to this last point, processing considerations lead to systematic exceptions to independently motivated formulations for rules or principles of grammar. For example, there are a number of exceptions to rules moving NP in English precisely when that NP is a pronoun. Particle Movement is an example. It applies freely with non-pronominal NPs, as in (2.3), but not with pronouns, cf. (2.4):

(2.3) a. Mary called her friend up.
    b. Mary called up her friend.

(2.4) a. Mary called her up.
    b. *Mary called up her.
The grammaticalization of processing principles

This will be motivated in terms of Early Immediate Constituents in ch. 5.4.2. On other occasions, a normally optional process such as That-Placement in English subordinate clauses may become obligatory in exactly the environment(s) in which its application is required by a processing constraint, here Frazier's (1985) Impermissible Ambiguity Constraint. Recall (1.5), repeated as (2.5):

(2.5) a. That Mary is sick surprised Bill.
   b. *Mary is sick surprised Bill.

More troublesome examples of processing-motivated exceptions arise when the relevant grammatical conventions are more general. A case is discussed in ch. 5.3.2 involving prenominal adjective phrases in English. English appears to allow only single-word adjectives in prenominal position, and not adjective phrases:

(2.6) a. a yellow book
    b. *a yellow with age book
    c. a book yellow with age

It does, however, allow adverbial modifiers or specifiers of an adjective in both pre- and postnominal position:

(2.7) a. a very yellow book
    b. a book very yellow with age

By standard assumptions (cf. e.g. Jackendoff 1977) the addition of an adverb to an adjective requires us to postulate the existence of a dominating adjective phrase, and hence adjectives in both pre- and postnominal position are dominated by AdjP. But if AdjP occurs before N in the NP, all lower dominated material should also be able to occur in this position, by virtue of a general structural principle: transitivity of domination. If a node A dominates B, and B dominates C, then A dominates C. Therefore, any rule, of positioning, rearrangement, or deletion, that applies to A applies to whatever it dominates. But (2.6b) shows that a positioning rule that makes reference to A does not apply to C, where A dominates C. This troublesome exception must simply be stipulated, and once again the ungrammaticality will be motivated by the principle of Early Immediate Constituents.

Sixth, processing can be argued to constrain very general patterns of variation throughout a whole grammar, e.g. whether a language will employ raising rules or WH-movement within its syntax, whether it will assign a productive set of non-agentive θ-roles to grammatical subjects, or whether
it will have a rich morphological case system. In some additional publications (Hawkins 1992b, 1994) I give a processing motivation for the presence versus absence of these kinds of grammatical properties in languages with different basic verb positions.

In all of these examples it is important to view the processing contribution to grammars within the context of the general theory of language universals outlined in ch. 1.5. It is because processing interacts with other explanatory primitives that grammars have a partially independent format of categories and principles. Grammars define a mapping between surface forms (ultimately phonetic forms) and meanings, and while this mapping must be capable of being produced and recognised in real time, and while this imposes huge constraints on the mapping process, there are other constraints as well, as we have mentioned. One very visible consequence of this is that there is inevitably going to be some partial mismatch in the grammatical response to a processing principle or preference.

For example, the grammar may outlaw center-embedded categories such as \( \tilde{S} \) in (2.1a) since they provide the greatest processing difficulty most of the time (the needs of expressive power being met by extraposition structures such as Did it surprise Mary that John failed his exam?). A center-embedded NP does not generally produce such difficulties, because the length of NP is typically shorter than that of \( \tilde{S} \), and so (2.1b) is grammatical. But grammatical rules do not look at terminal strings. So on those occasions in performance when an NP happens to be longer than an \( \tilde{S} \) (contrary to the normal case), as in (2.1c), the grammaticality facts remain the same.

(2.1)  
\begin{align*}
\text{a. } & \text{*Did } \tilde{S}[\text{that John failed his exam}] \text{ surprise Mary?} \\
\text{b. } & \text{Did NP[this fact] surprise Mary?} \\
\text{c. } & \text{Did NP[the fact that John failed his exam] surprise Mary?}
\end{align*}

The center-embedded \( \tilde{S} \) is ungrammatical, the center-embedded NP and complex NP are not. The grammar is sensitive to the categorial status of these word groupings, but not to their length. That is, because the length of NP is typically significantly less than that of \( \tilde{S} \), the grammar does not disallow (2.1b), and structures like (2.1c) can then survive as grammatical. As a result, the grammar stands in a close, but not completely one-to-one, relationship with a functional pressure that led to a grammatical response, involving on this occasion the structural complexity of center-embedding environments.

This state of affairs has been insufficiently appreciated by generative and functional grammarians alike. It means that the grammaticality of a structure
such as (2.1c) is no argument against a processing explanation for the ungrammaticality of (2.1a). Quite the contrary, we expect there to be partial mismatches. The explanation relies instead on the observed average length differential between S and NP, and on the correlation between, on the one hand, the lower performance scores for center-embedded $\bar{S}$s versus NPs according to some processing theory and metric (cf. ch. 3), and on the other hand, the grammatical response that differentiates between these two environments by blocking the structure that is regularly worse. The explanation is disproved by attacking either these average differentials, or the processing theory that defines the putative difficulties associated with center embeddings. But the grammaticality of (2.1c) is fully compatible with this general line of reasoning.

In a similar way the relative ordering of $VP[V \ NP \ PP]$ has been conventionalized within the English VP in response to the average length differentials between these categories, operating in conjunction with the same processing theory that defines dispreferences for center embeddings, i.e. Early Immediate Constituents (cf. ch. 5.3.1). Again, there will be a minority of VPs in which NP exceeds PP in length, thereby going against the relative length differential that holds in the normal case and that led to the grammatical response, and some of these will remain in the grammaticalized (and now non-optimal) V NP PP order, while others will be subject to "Heavy NP Shift" in a manner that turns out to be highly systematic and predictable (cf. ch. 4.3.2 for precise quantitative data).

Similar cases of a partial discrepancy between processing and the grammatical response can be expected in the other types of examples discussed as well.

2.2 A theory of structural complexity

Much of the literature on the processing impact on grammars appeals to some notion of "difficulty": certain structures are difficult to use, and this brings about the grammatical response in question. I suggested in ch. 1.4 that we must think rather more broadly about the grammar-processing interface and ask not just whether some structures are difficult or not, but also how they can be successfully used at all in real time, given the limited and often degenerate nature of the speech stream – to use the analogy with the logical problem of language acquisition again. But certainly difficulty will remain a fundamental concept in this context, and I cautioned against premature
A theory of structural complexity

attempts to equate this notion with "impossible to process," arguing instead for a more gradient approach in terms of degrees of processing difficulty with correlated cross-linguistic variation phenomena, such as implicational hierarchies. This will require a theory of structural complexity that is intrinsically gradient in nature, and in this section I propose a general theory that can serve as the basis for the theory of word order, Early Immediate Constituents, to be motivated in the next chapter.

2.2.1 Structural Domains and their complexity

Just as the concept of "structure" is fundamental to syntactic processes and representations (and also to semantics, cf. section 2.4), so too it is fundamental to a theory of complexity. The basic intuition that underlies the metrics of syntactic complexity proposed in Miller and Chomsky (1963) and Frazier (1985) is that complexity involves the amount of structure in a tree diagram. I will argue against their particular formulations, but I believe that this intuition is fundamentally correct, and I will accordingly redefine it and extend it in this context.

I begin by defining the notion of a Structural Domain (SD):

(2.8) Structural Domain (SD)

A Structural Domain consists of a grammatically and/or psycholinguistically significant subset of structurally related nodes in a tree dominated by a constituent C.

Two nodes A and B will be considered to be structurally related if one dominates the other, if they are sisters, or if one is a sister of some third node that dominates the other. An SD consists of a set of these structurally related nodes dominated by a given constituent C, possibly (indeed quite regularly) a proper subset of the total number of nodes that C dominates within a given sentence. This subset will be grammatically significant if it constitutes the domain over which a grammatical operation or principle is defined. For example, the set of nodes over which WH-movement is defined will constitute an SD for movement, comprising the structural position from which movement has taken place, the landing site, and all intervening nodes defining the domination and sisterhood relations on the path from the original position of the WH-word to the landing site. Or the set of nodes in a clause (S) that define the subject relation, or the direct object relation, will constitute an SD; cf. below.
A set of nodes will be psycholinguistically significant if these nodes are referred to by some parsing or production operation. Since the grammar is constantly accessed in processing, it follows that SDs that are grammatically significant will be psycholinguistically significant as well. However, there are sets of nodes that may not be directly relevant for the statement of grammatical regularities, but that are highly significant for processing. For example, the set of terminal and non-terminal nodes that are sufficient for recognizing the immediate constituent structure of a constituent C will be referred to as a Constituent Recognition Domain. Such a domain will typically consist of a proper subset of the nodes dominated by C, and though this subset may not be referred to by grammatical regularities directly, it is significant for the parser, since it enables this latter to recognize the internal structure of C rapidly and efficiently and possibly in advance of much of the material that C dominates (cf. ch. 3).

It will often be useful to refer to the SD of a particular node X in C. This subset of nodes dominated by C will consist of those that are structurally related to X in that they define what we might call its "structural integration" within C, i.e. the dominating nodes, sister nodes, and sisters of the dominating nodes (or c-commanding nodes, cf. Reinhart 1983) related to X. Putting this in more traditional, Saussurean terminology (cf. Saussure 1916), we can think of the SD of X in terms of the paradigmatic (or containing) and syntagmatic (or combinatorial) relations that X contracts and that define its structural role in C. Let us define an SD of X, provisionally, as follows:

(2.9) **Syntactically Defined Structural Domain of a Node X in C**

The SD of a node X in C consists of the following syntactically defined nodes that structurally integrate X in C:
- all nodes dominating X within C (including C itself);
- all sisters of X;
- all sisters of all nodes dominating X within C.

We will subsequently need to take account of surface coding devices, such as case morphology, that are also relevant for structural relations in many languages, and that affect the precise composition of an SD for X in addition to the syntax.

Consider first the tree for a subject–verb–object sentence in English shown in (2.10). The SD of the subject NP₁ consists of the mother node, S, and the sister VP.
A theory of structural complexity

The SD for the direct object $\text{NP}_j$ consists of two nodes dominating $\text{NP}_j$ within S, namely S itself and VP, plus the sister of $\text{NP}_i$, V, and the sister of VP, $\text{NP}_i$. That is, the subject $\text{NP}_i$ is one of the nodes that the object $\text{NP}_j$ has within its SD and that contributes to the structural integration of the object, but the object is not one of the nodes that the subject has in its SD. The SD for the subject is accordingly properly included in that for the object.

This simple illustration reveals a fundamental principle. Higher nodes in a tree have smaller SDs than lower nodes. $\text{NP}_i$ has only two nodes in its SD, $\text{NP}_j$ has four. Correspondingly, $\text{NP}_j$ is more structurally integrated into the sentence, in the sense that more dominating and sister nodes are required in order to define an NP in this structural position, the direct object position. A subject simply has to combine with a VP under the immediate domination of S. It does not have to combine with any particular daughters of this VP, and subjects can therefore occur grammatically either with or without direct objects. Objects do require the co-occurrence of subjects. Similarly, subjects can occur with transitive verbs, intransitive verbs, be plus predicate adjectives, etc, within VPs. Verbs, on the other hand, are grammatically specified for certain types of co-occurring arguments in their subcategorization and selectional restrictions within the lexicon, and these restrictions are imposed on subject and object NPs, both of which are in the SD for V in (2.10), \{S, NP$_i$, VP, NP$_j$\}.

In languages with flatter structures for S (i.e. without a VP), there appears to be a similar structural asymmetry between subjects and objects, signalled typically by some morphological coding device such as nominative versus accusative case marking. The accusative in these languages behaves very much like a direct object in English: it occurs in transitive sentences that also require a nominative subject, whereas the nominative may occur with or without an accusative; and it behaves like a direct object from the perspective of structurally sensitive grammatical operations, such as anaphoric binding and quantifier scope (cf. section 2.4). For these languages it is natural to propose that the morphology should be capable of contributing to the
definition of the structurally related nodes that structurally integrate X within a clause. Even though two nodes may be syntactically sisters and within each other's SDs by the definition in (2.9), the morphology may reduce the SD of one of them: the accusative-marked NP will have a nominative sister in its SD, but the nominative will not have the accusative in its SD. By removing the accusative from the SD for the nominative, we will be able to make similar grammatical predictions for case-marking languages to those we make for English. An English direct object can be an anaphor to a subject antecedent contained in its SD, whereas a subject cannot be an anaphor to an object not contained within its SD (cf. section 2.4); similarly an accusative can be an anaphor to a nominative, but not vice versa. That is, antecedents will need to be in the SDs of their anaphors, according to this theory, and how these SDs are defined across languages will depend not only on syntactic configuration, but also on morphology.1

SDs may also be reduced topologically, i.e. on the basis of linear precedence, in the event that the structural relations between sister nodes are asymmetric, so that a node to the left has priority over one to the right for anaphoric binding or quantifier scope. Given two sister nodes ordered A before B, we will say that the SD for B includes A in these cases, but not vice versa, and hence that A contributes to the structural integration of B, but not vice versa.

Primus (1987, 1991a) gives an excellent discussion of the relevance of morphological coding and topological position for grammatical operations across languages (partially summarized in section 2.3). All languages have syntactic configuration to at least some degree, and this provides the boundary conditions for structural relations that serve to structurally integrate X in C. Syntactically defined SDs may then be fine-tuned by case morphology, precedence, etc., in languages that make use of these options, by removing certain nodes from these SDs. We can therefore generalize (2.9) by allowing for these possible modifications in the syntactically defined SD of a node X, as follows:

(2.11) **Structural Domain of a Node X in C**

The SD of a node X in C consists of the following nodes that structurally integrate X in C:

- all nodes dominating X within C (including C itself);
- all or some sisters of X (depending on the surface coding conventions);
- all sisters of the nodes dominating X within C (but cf. n. 1 ch. 2).
The precise manner in which the surface coding conventions of morphology and topology can impact the composition of SDs defined syntactically, when in fact they do so, appears to be quite constrained. The constraints are described in the hierarchies of Primus (1987, 1991a). Thus, a nominative can be in the SD of an accusative, but not vice versa. If nominative and accusative are sisters, the accusative may be removed from the SD for the nominative, but the nominative cannot be removed from the SD for the accusative. Similarly, a phrase to the right of X may be excluded from the SD of X, but a phrase to the left cannot be (cf. further section 2.3).

I now define structural complexity:

(2.12) **Structural Complexity**

The structural complexity of an SD is measured by counting the set of nodes within it: the complexity of SD<sub>i</sub> exceeds that of SD<sub>j</sub> iff SD<sub>i</sub> > SD<sub>j</sub>.

In other words, complexity involves the number of structural relations within different portions of a tree, measured in terms of sets of structurally related nodes, and relative complexity involves relative numbers of these nodes.

I showed in (2.10) that lower nodes in an English S, such as a direct object, have larger and more complex SDs than higher nodes, such as a subject. Similarly, the SD of the (indirect object or oblique) NP<sub>k</sub> in (2.13) properly includes that of both NP<sub>i</sub> (the subject) and NP<sub>j</sub> (the object):

(2.13)

```
        S
       / \  
  NP<sub>i</sub>  VP
       /   \  
      V   NP<sub>j</sub>  PP
        /   \   
      P   NP<sub>k</sub>
```

SD of NP<sub>i</sub> = \{S, VP\}
SD of NP<sub>j</sub> = \{S, NP<sub>i</sub>, VP, V, PP\}
SD of NP<sub>k</sub> = \{S, NP<sub>i</sub>, VP, V, NP<sub>j</sub>, PP, P\}

Possessive phrases (i.e. genitives) are associated with even greater complexity, as illustrated in (2.14) in which the SD for NP<sub>n</sub> contains eight nodes, one more than that for NP<sub>k</sub>.
The grammaticalization of processing principles

(2.14) \[
\begin{array}{c}
S \\
\downarrow \\
NP_i \quad VP \\
\downarrow \\
V \quad NP_j \\
\downarrow \\
PossP \quad N \\
\downarrow \\
NP_n \quad Poss
\end{array}
\]

Nodes within an embedded S or VP contained in S will also have progressively larger SDs, the more dominating and sister nodes are contained within their SDs.

As SDs increase in complexity, we can expect that there will be some measurable consequences for performance, and also for grammars. The predictions will need to be motivated and formulated precisely, and the whole of ch. 3 is devoted to doing this for performance and grammatical data involving different orderings of immediate constituents. In the interim let us present some data from other grammatical areas that illustrate our general philosophy. The basic prediction we wish to make for performance is that larger SDs will be associated with greater or equal processing cost compared with smaller SDs. This cost should be observable in psycholinguistic experiments, in the frequencies with which various alternative structures are selected in performance when the grammar provides a choice, and (possibly) in acceptability intuitions, though these are well known to be less reliable than the other forms of data. For grammars we predict a close correlation between these kinds of performance data and the conventionalized grammatical rules of particular grammars. For example, the syntactic environments in which rules such as Relativization or WH-movement "cut-off" in a given language should be in accordance with the complexity of these environments, cf. section 2.3.

Consider some performance data involving relativization in English. English allows structures such as (2.13) and (2.14) to be embedded as relative clause modifiers of a head noun. The NP relativized on, for example the syntactic position marked 0 in \[the \ woman_{i} \quad \text{that \ I \ saw \ 0_{i} \ today},\] defines an SD of nodes that structurally integrate it within a "Relativization Domain," defined as follows:
A theory of structural complexity

(2.15) Relativization Domain

A Relativization Domain consists of that subset of nodes within an NP immediately dominating an $\bar{S}$ and a head noun that structurally integrate the empty category or pronoun that is co-indexed with the head and in its original (d-structure) position. (For a definition of the nodes that can structurally integrate a node $X$ in $C$, cf. (2.11).)

As Relativization Domains increase in complexity, we predict a corresponding dispreference for the relevant structures in performance. More precisely, we predict a greater or equal dispreference as structural complexity increases.

All the NP positions in (2.13) and (2.14) can be relativized on in English: NP$_i$ (subjects), NP$_j$ (direct objects), NP$_k$ (indirect objects or oblique NPs), and NP$_n$ (genitives). The Relativization Domains defined by these positions involve increasing complexity. All relative clauses contain the same higher structure, i.e. the dominating NP and $\bar{S}$ nodes and the head noun. What distinguishes them crucially is the size of the SD within the relative clause itself, i.e. within $S$. In (2.13) the domain of the NP$_i$ consists of two non-terminal nodes in $S$, NP$_j$ consists of five, and NP$_k$ of seven; in (2.14) NP$_n$ consists of eight non-terminal nodes. The relative ranking illustrated by these totals will be preserved when the additional higher nodes are added to these Relativization Domains, and will be argued in section 2.3 to explain Keenan and Comrie’s (1977) Accessibility Hierarchy and its predictions for grammatical conventions across languages. In this context notice the following support from performance.

Keenan (1975) has counted the relative frequencies of relativizations on different syntactic positions in English texts, and has found a clear correlation between the position relativized on and frequency of occurrence. The aggregated frequencies from his various texts for the four NP positions shown in (2.13) and (2.14) are given in (2.16):

(2.16) Aggregated frequencies for relativizations in English texts (Keenan 1975)

\[
\begin{array}{lcccc}
SU & DO & OBL (\& IO) & GEN \\
46\% & 24\% & 15\% & 5\%
\end{array}
\]

These declining percentages are in exact correspondence with the increasing sizes of the Relativization Domains for these NP positions, within the framework presented here.

In addition, Keenan and S. Hawkins (1987) report the results of a psychological experiment that was designed specifically to test the psychological
validation of the Accessibility Hierarchy and that was performed on both adults and children. The experiment was a repetition experiment in which subjects were asked to repeat back relative clauses formed on different positions, in written form and after a certain time interval (which was filled by a memorization task involving a certain number of digits). The experiment controlled for variables such as word frequency and sentence length in the sentences used, and for sex, IQ scores (children) and educational attainment (adults) among the subjects. The crucial variable whose significance the experiment was trying to test was, therefore, the size and complexity level of different Relativization Domains.

The general expectation of Keenan and S. Hawkins was that accuracy levels in the repetition task would correlate with positions on the Keenan–Comrie hierarchy. The prediction we would make here is similar: if a Relativization Domain RD\textsubscript{j} has greater complexity than another domain RD\textsubscript{i}, then repetition accuracies for RD\textsubscript{j} should be worse than or equal to those for RD\textsubscript{i}. In other words, repetition accuracies should be worse or equal as we go down each of the positions on the hierarchy. The Keenan and S. Hawkins data are given in (2.17). It is clear that they provide strong support for these predictions (GEN-SU stands for relativization on a genitive within a subject NP; GEN-DO for relativization on a genitive within a direct object NP; cf. section 2.3 for more detailed discussion of these and other Relativization Domains):

\[(2.17)\] Repetition accuracies for relativizations (Keenan and S. Hawkins 1987)

<table>
<thead>
<tr>
<th></th>
<th>SU</th>
<th>DO</th>
<th>IO</th>
<th>OBL</th>
<th>GEN-SU</th>
<th>GEN-DO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adults</td>
<td>64%</td>
<td>62.5%</td>
<td>57%</td>
<td>52%</td>
<td>31%</td>
<td>36%</td>
</tr>
<tr>
<td>Children</td>
<td>63%</td>
<td>51%</td>
<td>50%</td>
<td>35%</td>
<td>21%</td>
<td>18%</td>
</tr>
</tbody>
</table>

2.2.2 Miller and Chomsky’s and Frazier’s metrics of complexity

The theory of structural complexity that I have just defined attaches crucial significance to the number of non-terminal nodes within certain portions of a tree that are claimed to be grammatically and/or psycholinguistically significant. It differs from Miller and Chomsky’s (1963) original proposal, which was formulated in terms of the ratio between non-terminal and terminal nodes throughout a sentence. If a sentential subject in English such as *That John failed his exam surprised Mary* is extrapolated to create *It surprised Mary that John failed his exam*, the addition of the terminal element *it* results in a
A theory of structural complexity

(very slightly) lower ratio of non-terminals to terminals. Lower ratios reduce the complexity of a sentence in their theory; hence extraposed sentences are (here correctly) predicted to be easier to process. The trouble with this metric is that there are many differences in processing complexity that it does not predict, as Frazier (1985) has shown, either because the relevant sentences have equal non-terminal to terminal ratios, or because the more difficult structures actually have lower ratios. I would also suggest that it is not generally plausible to measure complexity in this way, because the ratio between non-terminals and terminals is largely constant across sentences: in general, the more words there are, the more non-terminal nodes. Hence, the ratio per se is not going to be very revealing.

In response to this, Frazier (1985) has proposed that the ratio be calculated locally, on any three adjacent terminals, and she argues that this overcomes the problems with Miller and Chomsky's global count, which is not sensitive to the precise distribution of non-terminal to terminal ratios within a sentence. The locality assumption is an interesting innovation, but I believe that it still does not give us an adequate complexity measure, for the following reasons.

First, Frazier's revised metric makes either no predictions or incorrect predictions for relative degrees of difficulty among certain structural types for which there is now performance evidence. For example, her metric makes the wrong predictions for the relative clause data of (2.17); cf. the discussion of (2.18)-(2.20) below.

Second, while Frazier is right in asserting that Miller and Chomsky's ratio is not sensitive to the distribution of non-terminals to terminals, her own local ratio introduces problems of its own. The basic problem is that her method excludes from consideration a number of non-terminal nodes that are relevant for determining complexity, and that would be counted in the Miller-Chomsky metric. She calculates all non-terminals over a given terminal as they are constructed on-line. If, for example, an S node has already been constructed in the left-to-right parse at the time that a verb constructs VP, then the only non-terminal that is considered in the ratio for the verb is VP, making a 1/1 ratio in her terms (she also ignores all pre-terminal lexical categories such as V). Similarly, if some non-terminal node has not yet been constructed in a left-branching structure parsed bottom-up, and if this non-terminal will eventually dominate one or more of the terminals in Frazier's three-word viewing window, it also is omitted from the non-terminal to terminal ratio. The trouble with this method is that it actually removes us from the original Miller-Chomsky
insight that complexity is a function of the amount of structure that needs to be processed, for it excludes from consideration all the non-terminal nodes that are constructed outside of the three-word viewing window, many of which do ultimately contract structural relations of domination or combination with one or more of the terminals in question. These non-terminal nodes are relevant for the processing of terminals, and their number purportedly determines the degree of complexity, yet Frazier’s local measurements give us no means of counting them.

There is a third objection that can be raised to Frazier’s metric. She considers only a single set of three adjacent terminals within the sentences compared, namely the one that has the highest non-terminal to terminal ratio for that sentence, and argues that it is these peaks of complexity that are indicative of overall difficulty: the sentence with the highest peak is more complex. But for any sentence of word length \( n \) there are \( n-2 \) sequences of three adjacent terminals, and it is just as plausible, if not more plausible, to measure all of these when assessing the complexity of a whole sentence, especially since the difference between peaks is not all that great in her examples.\(^5\)

Let us test Frazier’s metric on three structures of English, namely relative clauses formed on subjects, on direct objects, and on indirect objects or oblique NPs, i.e. positions \( \text{NP}_i \), \( \text{NP}_j \) and \( \text{NP}_k \) in (2.13). In (2.18)–(2.20) I set out the non-terminals that are constructed above each terminal, using exactly the assumptions made by Frazier: the lexical category node for each word is not counted; \( S \) and \( \bar{S} \) nodes count 1\(^\frac{1}{2}\); all other nodes are counted as 1. The empty NP relativized on is also postulated in these onlines parses and is counted as 1.

\[
(2.18) \quad \begin{array}{c}
\text{S} \\
\text{NP} \\
\text{NP} \\
I \quad \text{like} \\
1 \quad 1 \\
2\frac{1}{2} \quad 3\frac{1}{2}
\end{array}
\]

\[
\begin{array}{c}
\text{NP} \\
\text{VP} \\
\text{NP} \\
\text{S} \\
\text{S} \\
\text{NP} \\
\text{VP}
\end{array}
\]

\[
\begin{array}{c}
\text{I} \\
\text{like} \\
\text{students} \\
\text{who} \\
\text{sing}
\end{array}
\]
In terms of peaks of complexity, (2.18) and (2.19) have the worst three-word sequences: *students who sing* in (2.18) has a non-terminal to terminal ratio of 7/3, and so does *whom(m) professors pass* in (2.19). The worst sequences in (2.20) are *students who professors give passes to* and *who professors give*, both with 6/3. Hence, Frazier predicts that relativizations on *su* and *do* should be worse than relativizations on *io* and *obl*; and her metric makes no distinction between *su* and *do*. In fact, the data of (2.17) and also (2.16) suggest that *io* and *obl* involve more performance difficulty than *su* and *do*, and that *do* is more difficult than *su*. The performance data therefore support a complexity metric in terms of Structural Domains, not one in terms of non-terminal to terminal ratios within a viewing window of three adjacent terminals.

If instead of examining just peaks of complexity, i.e. the worst three-word sequence within a given sentence, we were to aggregate over all three-word sequences in these sentences (which I suggested is just as plausible, if not more plausible; cf. n. 5, ch. 2), we can achieve a more fine-tuned discrimination among (2.18), (2.19), and (2.20). Ignoring the matrix subject and verb, *I like,*
which is common to all three sentences, and focusing only on the head and relative clause, (2.18) has just one three-word sequence, students who sing, with a ratio of 7/3, i.e. 2.33. Sentence (2.19) has two such sequences, students who(m) professors and whom(m) professors pass, which average out at 6.5/3, i.e. 2.17. Students who professors give passes to in (2.20) is analyzable into four three-word sequences, the aggregate for which is 5.125/3, or 1.71. In other words, using this mathematical method, complexity would again decline rather than increase down (2.18)-(2.20), and the counterfactual prediction would be made that relativization should get easier down Keenan and Comrie's Accessibility Hierarchy.

Whichever quantitative method we use, a three-word viewing window does not appear to be a grammatically or even psycholinguistically significant domain within which complexity should be measured. Complexity involves the amount of structure, i.e. dominating and sister nodes, at different points in a tree. This is what I take the original insight to be all about: the more structure, the greater the complexity. And the reason we will ultimately give for this is that the processing of more structural relations requires more simultaneous computations on-line. This insight can be captured by using the notion of a Structural Domain as defined in (2.8) and (2.11), in conjunction with the definition of structural complexity given in (2.12). A direct object NP involves the processing of more structural relations within S than a subject NP, etc. Measuring complexity in this way provides a ready explanation for the degrees for difficulty associated with different structural options, such as relative clauses formed on different positions, and for different orderings of constituents (cf. ch. 3). Some grammatical consequences of these different degrees for cross-linguistic variation patterns and implicational hierarchies are explored in the next section.

There is a further potential advantage for this approach to complexity: we can explain and collapse Frazier's (1979a, 1985) three experimentally motivated parsing preferences in temporarily ambiguous sentences, Minimal Attachment, Late Closure, and the Most Recent Filler. They all involve preferences for less complex SDs over more complex alternatives, when the parser has a choice. For Minimal Attachment sentences, e.g. Bill knew the answer was correct, this is quite transparent. The postulation of fewer nodes consistent with the grammar means that the SDs of temporarily ambiguous nodes will be smaller: the answer will have a smaller SD within the matrix S if it is minimally attached to this S. Late Closure, e.g. the low attachment of in the library in John put the book that he had been reading in the library, results in smaller Constituent Recognition Domains throughout
Complexity predictions for grammatical variation

a sentence: if *in the library* is attached to the matrix VP headed by *put*, this VP acquires a long and complex recognition domain; if it is attached to the lower VP headed by *reading*, the recognition domain for the matrix VP is significantly shorter, while that for the lower VP is no less efficient (according to our quantificational method, cf. ch. 3). This point is illustrated in greater detail in chs. 3.5.3 and 4.5. And assigning the most recent potential fillers to gaps results in “movement domains” of lesser complexity, cf. section 2.3.

2.3 Complexity predictions for grammatical variation across languages

The performance data from English involving relative clauses formed on different syntactic positions within a clause (cf. (2.16) and (2.17)) are very intriguing when considered alongside grammatical data from different languages. Keenan and Comrie (1972, 1977) have presented cross-linguistic evidence for an Accessibility Hierarchy (AH) of basic grammatical relations, defined in (2.21):[6]

(2.21) SU > DO > IO > OBL > GEN (> = more accessible than)

Among the many pieces of grammatical evidence supporting (2.21) is the following regularity (in its original 1972 formulation). Relativization can cease to apply, or “cut off,” in a particular language in a manner predicted by the hierarchy: if a language permits relativization on a low position in AH, it will permit it on all higher positions. Illustrative languages with restricted relativization strategies are given in (2.22):

(2.22) Relativization on: SU only: Malagasy, Maori
    SU & DO only: Luganda, Indonesian
    SU & DO & IO only: Basque
    SU & DO & IO & OBL only: North Frisian, Catalan
    SU & DO & IO & OBL & GEN: English, Hausa, Kera

The decreasing accessibility to relativization down AH exemplified in (2.22) corresponds exactly to the declining frequencies and ease of use in the performance data from English (a language that permits relativization all the way down the hierarchy). Hence we have a correlation between performance and grammar: the lower the position relativized on in AH, the
less frequent that relative clause type in performance, the more difficult that structure is under psycholinguistic experimental conditions, and the fewer languages there are that actually permit relativization on that position.

The cut-off prediction illustrated in (2.22) was later revised in Keenan and Comrie (1977) to account for a handful of exceptions, all West Indonesian languages such as Toba Batak, in which there is a gap in the positions relativized on (e.g. at the DO position) and an obligatory promotion of that position into a higher one (e.g. SU) that is relativizable. In n. 8 ch. 2 I argue in favor of retaining the original prediction (though in slightly revised form), and I modify the more recent predictions in terms of strategies, bringing them into line with the kinds of processing motivations considered here.

The explanation that we wish to give for (2.22) is the following. Grammars have restricted the rule of relativization to apply to some categories only in a way that reflects the complexity of the SDs for the positions relativized on: SUs are the least complex, DOs next, and so on. Moreover, if a low position on AH is relativizable (with considerable complexity in its SD), then all positions with less complex SDs, most of the nodes of which are already contained in the SDs for more complex positions, will also be relativizable. Hence all higher positions will be relativizable as well. This same basic idea was actually proposed by Keenan and Comrie in 1977: "The AH directly reflects the psychological ease of comprehension. That is, the lower a position is on the AH, the harder it is to understand relative clauses formed on that position" (p. 88; authors' italics). They did not have a complexity metric, however, with which to support this claim.

The concept of a Structural Domain, and specifically of a Relativization Domain (cf. (2.15)), does provide a metric that can be used to explain the performance data from English, and the grammatical data from many languages. In order to demonstrate that the different positions on the hierarchy are genuinely associated with increasing complexity, however, we need to have a consistent way of comparing the Relativization Domains for each position, and we need to take into account cross-linguistic differences in surface coding. For example, the SD of a direct object in English will include a PP in the structure of (2.13), but no PP in (2.10). Which alternative are we going to quantify when calculating the Relativization Domain for a direct object? Similarly, the SD for a nominative NP in a language without a VP may include an AuxP, i.e. \{S, V, AuxP\}. If this same AuxP is not also included when we quantify the SD for an accusative, i.e. \{S, V, NP-Nom\}, these two domains will have the same number of nodes and equal complexity will be predicted.
One way of achieving consistency is to make all calculations in terms of Minimal Structural Domains, defined as follows:

(2.23) The Minimal SD of a node $X$ in $C$ consists of the smallest set of structurally integrating nodes that are grammatically required to co-occur with $X$ in $C$.

This will mean that the Relativization Domain of the direct object in English will be measured without the PP. If the AuxP is optional within $S$ in a language, it will be excluded from the SDs of both the nominative and the accusative; if it is obligatory, it will be included in both; and so on.

Moreover, the omission of optional nodes does not affect the relative ranking of positions established on the basis of minimal SDs. If they were to be included consistently, this ranking would merely be reinforced. Imagine that an AuxP is added to an English tree such as (2.13). The SD of the subject NP$_i$ will then be \{S, VP, AuxP\}. But similarly the SD of the direct object NP$_j$ and of all lower positions will include the AuxP as well. That is, any additional, optional nodes in higher positions in a tree will also be included in the SDs of lower positions, whereas the converse fails. The result is that additional nodes preserve the relative rankings established by minimal SDs.

Let us now consider all the positions on Keenan and Comrie's AH (cf. (2.21)) in terms of their minimal SDs, and with reference to some of the major coding differences between languages.

The relationship between a SU and a DO has already been discussed. The minimal SD of the SU is properly included in that of the DO, both in languages that have a VP and in languages that do not, i.e. $\text{Min SD(SU)} \subseteq \text{Min SD(DO)}$. For languages with flat structures, SU and DO will generally be distinguished morphologically, as we have seen, e.g. by nominative versus accusative case marking.

The minimal SD of a DO is properly included in that of an IO, since a clause with an IO generally requires an accompanying DO and also a SU, whereas a DO can occur both with and without an IO, i.e. $\text{Min SD(DO)} \subseteq \text{Min SD(IO)}$. This is supported by the assumption in models such as Relational Grammar (cf. Blake 1990) that IO is the third argument of a three-place predicate, DO being the second. It is also supported by an examination of the co-occurrence requirements of verbs throughout a whole language. This has been done for German by Primus (1993), using a corpus of 17,500 non-compound verbs listed in Mater (1971). There is a very close correlation in German between SU, DO and IO in Relational Grammar terms and nominative, accusative, and dative case-marking respectively.
Primus shows that all transitive verbs in German take a nominative and one or more of the other cases, accusative, dative, or genitive. Within these latter, almost all (95%) of the dative-taking verbs require an accusative as well, i.e. IO generally requires a DO. But the converse fails: only 52% of accusative-taking verbs require a dative as well.

Consider now NPs in an OBL function, compared with SUs, DOs, and IOs. Whether a language has a VP or not, the minimal SD for a SU will be properly included in that for an OBL, since OBL requires an accompanying SU, but the converse fails, i.e. Min SD(SU) ⊆ Min SD(OBL).

For DO and OBL we need to examine the precise coding devices used in different language types. In an English-type language in which DO is represented by an NP, OBL by a PP, and there is a VP, the minimal SD for DO is clearly properly included in that for OBL. Without a VP there will still be a proper inclusion relation, as long as the OBL NP is contained within a PP, giving it extra structural depth compared with DO. In languages such as Finnish, however, in which there are rich case-marking systems for OBLs in addition to PPs, some OBLs will be syntactically identical to DOs in surface appearance (i.e. they will be NPs), and their minimal SDs will also be identical: each will require a SU, but a DO will not require an accompanying OBL and an OBL will not require a DO (in intransitive clauses).

Aggregating over all of these cross-linguistic differences, we appear to have a relation of inclusion between the minimal SDs of DOs and OBLs, i.e. Min SD (DO) ⊆ Min SD(OBL). Hence the minimal SD of an OBL either contains the same set of nodes as that of a DO plus at least one additional node, or it contains the same set as the DO.

The relationship between IO and OBL is complicated by the fact that many languages do not make a clear distinction between these two positions. In John gave the book to Mary, the IO to Mary has the same surface syntax as an OBL, i.e. PP. In languages that employ dative case on an NP for IO and PP for OBL, the minimal SD for IO will contain both SU and DO nodes in addition to S, VP and V, {S, SU-NP, VP, DO-NP, V}, while the minimal SD for OBL will not contain a DO but will contain PP and P, {S, SU-NP, VP, V, PP, P}. The former contains five nodes, the latter six. This time there is no relation of proper inclusion, but the minimal SD for IO is smaller than that for OBL, Min SD(IO) < Min SD(OBL). In languages of the Finnish type in which OBL arguments may be both case-marked NPs and PPs, the PP will contain six nodes in its minimal SD, and the OBL NP just four, {S, SU-NP, VP, V}, since OBLs do not require accompanying DOs. Averaging six and four produces five, which is the same size as the minimal SD for IO, i.e. Min
Complexity predictions for grammatical variation

SD(IO) = Min SD(OBL). Aggregating over different language types, we therefore have: Min SD(IO) ≤ Min SD(OBL). For languages in which both IO and OBL are PPs (as in English), the very existence of a separate IO position is doubtful. In such cases, a subset of OBLs will require accompanying DOs, and the surface constituency of OBL will typically be more structured than that of SU and DO (e.g. PP versus NP), or at least equally structured, all of which corroborates Min SD(SU) ⊂ Min SD(OBL) and Min SD(DO) ⊂ Min SD(OBL), which were established above.

For genitives, let us assume an NP-internal Possessive Phrase structure for all languages, along the lines of (2.14) above. Clearly, a genitive on each of the argument positions of AH will add nodes to the minimal SD for that position, so that we have Min SD(SU) ⊂ Min SD(GEN-SU), Min SD(DO) ⊂ Min SD(GEN-DO), etc., down to OBL. Interestingly, the complexity score for the genitive with least complexity, GEN-SU, turns out to be greater than or equal to the score for all higher positions on AH, i.e. SU, DO, IO, and OBL, thereby confirming its position below OBL on the hierarchy. Let us illustrate this for a language with an English-type syntax. The minimal SDs for SU, DO, and OBL have been assessed at two, four, and six respectively, with IO also at six if it exists as a separate position. The minimal SD for a GEN-SU is six, \{S, VP, NP-SU, NP, PossP, Poss\}, GEN-DO has eight, and GEN-OBL ten. Hence, aggregating over all genitives on all AH positions, we have the following relative ranking: Min SD(SU) ⊂ Min SD(GEN); Min SD(DO) < Min SD(GEN); Min SD(IO) ≤ Min SD(GEN); Min SD(OBL) ≤ Min SD(GEN).

Summarizing, we can say that for each adjacent position on AH, Xi,Xj, the minimal SD for Xi is always less than or equal to that for Xj, Min SD(Xi) ≤ Min SD(Xj), and for most positions there is a relation of proper inclusion or inclusion, i.e. Min SD(Xi) ⊂ Min SD(Xj) or Min SD(Xi) ⊆ Min SD (Xj). This is shown in (2.24):

\begin{align*}
(2.24) & \ a. \ Min SD(SU) \subset Min SD(DO) \\
& b. \ Min SD(SU) \subset Min SD(IO), Min SD(DO) \subset Min SD(IO) \\
& c. \ Min SD(SU) \subset Min SD(OBL), Min SD(DO) \subset Min SD(OBL), \\
& \quad Min SD(IO) \leq Min SD(OBL) \\
& d. \ Min SD(SU) \subset Min SD(GEN-SU), Min SD(DO) \subset Min \\
& \quad SD(GEN-DO), etc. \\
& e. \ Min SD(SU) \subset Min SD(GEN), Min SD(DO) < Min SD(GEN), \\
& \quad Min SD(IO) \leq Min SD(GEN), Min SD(OBL) \leq Min \\
& \quad SD(GEN)
\end{align*}
The overall pattern of additive complexity down AH seems clear, despite the fact that the grammatical relations in terms of which AH is stated can be coded in different ways. It is a basic tenet of the theory of Relational Grammar (cf. Blake 1990) that grammatical relations are primitives and that fundamental generalizations can be captured in terms of them. But this view is increasingly being attacked by those who argue that significant generalizations are actually lost in this way, and that the relevant generalizations should be stated in terms of the precise syntactic or morphological coding devices for the grammatical relations in question, especially in languages such as German, Icelandic, and Hungarian, in which both the morphology and the syntax constrain syntactic operations (cf. Primus 1987, 1991a). In fact, Primus proposes that the AH should be dispensed with altogether and should be replaced by a number of subhierarchies defined on the different coding devices for grammatical relations.

For example, for languages with nominative-accusative case morphology she proposes the following hierarchy:

\[(2.25) \text{nominative} > \text{accusative} > \text{dative} > \text{other}\]

For the syntax she proposes a syntactic hierarchy that ranks positions according to their asymmetrical c-command relations, so that an NP immediately dominated by S outranks an NP immediately dominated by VP since the former asymmetrically c-commands the latter. She also argues for a topological hierarchy whereby material to the left outranks material to the right, and for a semantic hierarchy in terms of θ-roles. All these hierarchies are supported by showing that the kinds of subject-object asymmetries involving binding and quantifier scope, etc., that receive a syntactic explanation in terms of c-command in rich configurational languages like English, hold equally for languages in which subject and object are expressed by nominative and accusative case respectively. Primus accordingly proposes that the notion of command should be generalized from the syntax to all of these hierarchically arranged entities and that a higher position on each hierarchy will asymmetrically command all lower positions. An accusative reflexive anaphor can be bound to a commanding nominative antecedent, but not vice versa, etc. The well-formedness of a particular sentence in a particular language is only guaranteed, in her model, if all the command restrictions on all hierarchies have been adhered to. (Cf. further n. 1 ch. 2.)

Primus's data and discussion provide additional support for the reality of hierarchically arranged argument positions. What is significant about her
subhierarchies from the perspective of the complexity theory proposed here is that all of them appear to conform to the same basic descriptive regularity that we have proposed for the AH itself: each higher position has a smaller minimal SD than each lower position. The morphological hierarchy (2.25) corresponds exactly to the relative ranking SU > DO > IO > Other for languages in which there is morphological coding of SU, DO, and IO, and hence the same reasoning about co-occurrence requirements holds for this morphological hierarchy as well. The syntactic hierarchy in terms of asymmetrical c-command relations has a very precise correlate with complexity: if A asymmetrically c-commands B, then the SD of B contains A but not vice versa, hence SD sizes increase down the syntactic hierarchy. For the topological hierarchy in which A outranks B if A precedes B, the SD for B will contain A, but not vice versa, i.e. the SD for B is again greater. Although considerations of space preclude a fuller discussion, I would therefore argue that the same relative complexity explanation that has been proposed for the AH will carry over to Primus's subhierarchies as well. The subhierarchies are needed because of partial conflicts between them in the definition of grammatical relations and in the application of structurally sensitive rules across languages. They are all relative complexity rankings, however.

Returning to the AH itself, there is a set of facts that was noticed by Keenan and Comrie for which we can now give a more complete explanation. Keenan and Comrie pointed out that a number of languages permit a pronoun to be retained in the position relativized on. Hebrew is an example:

(2.26) ha-isha \(s\) [she- Yon natan la et ha- sefer]\n
"the woman that John gave to-her DO the book," i.e.

the woman that John gave the book to

Keenan (1972) had already observed that such languages permit relativization in a larger set of environments than languages that do not retain pronouns, particularly in the intuitively more difficult structures such as coordinate and complex NPs (cf. Ross's 1967 distinction between copying rules that retain a pronoun and are not subject to the Co-ordinate Structure Constraint or Complex NP Constraint, and chopping rules that do not retain a pronoun and are subject to these constraints). The explanation that Keenan proposed for this fact is given in his principle of Conservation of Logical Structure (CLS): pronoun retention extends the set of relativization environments because relative clauses with pronouns correspond more closely to their logical-semantic structures than do relative
The grammaticalization of processing principles

clauses without such pronouns. The pronoun identifies the semantically appropriate position of the NP within the relative clause that is coreferential to the head, and so makes the relationship between relative clause and head more semantically transparent.

Keenan and Comrie (1977) provide data on pronoun retention in relation to the AH. Languages that retain pronouns do not necessarily retain them on all the positions of AH. Pronoun retention is favored in the lower and more complex positions, and in a manner that turns out to be extremely precise: if a language retains a pronoun on a high position, it retains it on all lower positions that the language permits relativization on. Thus, the implicational pattern of pronoun retention is the reverse of that for the cut-off predictions given in (2.22), as shown in (2.27):

(2.27) Pronoun-retaining relative clauses

| SU & DO & IO & OBL & GEN: | Urhobo |
| DO & IO & OBL & GEN:     | Arabic (Classical), Hebrew, Chinese |
| IO & OBL & GEN:          | Batak, English, Hausa |
| OBL & GEN:              | Fulani, Minang-Kabau |
| GEN:                    | Javanese, Japanese, Turkish |

The complexity theory that we have presented here provides an explanation for the data of (2.27) and for the mirror-image implicational pattern between (2.22) and (2.27). The cut-off data are explained on the grounds that relativization may cease to apply as Relativization Domains become more complex down each of the positions of the hierarchy. But pronoun retention has been independently shown to be favored in difficult syntactic environments. Hence, we expect pronoun retention to be favored in the lower positions of the AH. More precisely, if the AH–complexity correlation is as precise as we have claimed it is, and if pronoun retention is motivated by the need to relieve the difficulty of complex syntactic environments, then we should expect both that pronoun retention will be just as precisely correlated with the AH as the cut-off data are and that these two sources of data will be mirror-images of each other. Hence, we predict an implicational patterning going from low to high positions in the cut-off data, and from high to low in the pronoun retention data, and that is what we get.

Surprisingly, Keenan and Comrie's (1977) principles in terms of relative clause strategies do not actually predict the data of (2.27), and their
The proposed explanation is incomplete. Keenan and Comrie define different relative clause strategies on the basis of two properties: the order of relative clause and head noun (e.g. prenominal, postnominal); and \([\pm \text{Case}]\), i.e. does the relative clause contain a nominal element that unequivocally expresses which NP is being relativated on (pronoun retention being one kind of \([+ \text{Case}]\) strategy)? But because different strategies are defined by this combinatorial function of two properties, e.g. postnominal and \([+ \text{Case}]\) or prenominal and \([- \text{Case}]\), because \([+ \text{Case}]\) covers a variety of different strategies, and because Keenan and Comrie’s Hierarchy Constraints refer to relative clause strategies so defined, nothing is predicted about pronoun retention (or even \([+ \text{Case}]\) strategies) per se. As far as their principles are concerned, pronoun retention could be found on SU & IO & GEN relatives, on DO & GEN, and so on. Clearly, the precision of these data is not being accounted for. But by viewing AH as a complexity ranking and pronoun retention as a device that relieves the processing difficulty associated with complexity, these data can be explained. Speakers retain pronouns because relative clauses formed on complex positions are easier to process when more of the logical-semantic structure is preserved and when the relationship between the head and the relative clause is more semantically transparent. Hence, the ultimate explanation for pronoun retention is a processing one, not a semantic one. It happens to be the case that semantic transparency facilitates processing ease on this occasion. On other occasions, when semantic transparency does not facilitate processing, it will not be grammaticalized (cf. Hawkins 1988c: 11 for some relevant examples).

The hierarchically arranged grammatical data of (2.22) and (2.27) can therefore be explained in terms of the increasing complexity of the Structural Domains for relativization. I believe that there are many such universal hierarchies that can be explained in terms of increasing structural complexity. In ch. 5.5 I shall explain some universal word-order hierarchies this way. What is interesting about all these hierarchies is that there is no obvious explanation for them in purely grammatical terms, and I believe this is the reason why Keenan and Comrie’s AH has been largely ignored in the generative literature. There is no clear motivation within a grammatical model for the observed restrictions on relativization, and for pronoun retention in some positions and not others. And there is no motivation for the implicational relationships between these basic grammatical relations. These are complications as far as the grammar is concerned, but they receive a straightforward explanation in processing terms, and this explanation then
results in patterns of grammatical variation across languages correlating with
degrees of complexity and processing ease.

This correlation between the data of language performance and cross-lin-
guistic data from grammars can be referred to as the Performance–Grammar
Correlation. It predicts the following correspondences between performance
and grammar:

\[(2.28) \text{ Predicted Performance–Grammar Correspondences}\]

Given: a hierarchy, \(H\), of minimal SDs, \(X_1, X_2, X_3 \ldots X_n\);
\(X_1 \cap X_2 \cap X_3 \cap \ldots X_n \neq 0\) (i.e. the intersection of all SDs
is not null – they share at least some nodes);
increasing structural complexity down \(H\), i.e. for each
adjacent \(X_i, X_j\) on \(H\), \(\text{Min SD}(X_i) \leq \text{Min SD}(X_j)\).

Then: \textit{performance predictions}: in languages in which some plurality
of these SDs are grammatical, performance preferences for
each adjacent \(X_i, X_j\) will be such that \(X_i \geq X_j\) (\(X_i\) is more
or equally preferred to \(X_j\)); performance data can involve ease
of processing in psycholinguistic experiments, frequency of
occurrence in natural discourse, and acceptability judgments;

\textit{grammatical predictions}:
\textit{grammatical cut-off predictions}: for each adjacent \(X_i, X_j\)
on \(H\), grammatical operation \(G\) may apply to \(X_i\) and not
to \(X_j\); and if \(G\) applies to \(X_j\) (a low point), it will apply to
all higher points;
\textit{conservation of logical structure predictions}: for each adja-
cent \(X_i, X_j\) on \(H\), if \(X_i\) (a high point) is coded in surface
structure as a pronoun in a way that conserves logical
structure and makes processing easier, then all lower
positions will be coded as a pronoun, if they are gram-
matical at all.

It should be clear that the data of this section are in accordance with these
predictions.

Consider finally a brief additional example involving WH-movement. I
believe that there is a complexity correlate to the syntactic environments in
which WH-movement may cut off in those languages that permit it at all. Let
us define a Structural Domain for movement, a Movement Domain, parallel
to the Structural Domain for relativization in (2.15):
(2.29) **Movement Domain**
A Movement Domain consists of that subset of nodes within the \( S \) most immediately dominating the moved WH that structurally integrate the empty category or pronoun that is co-indexed with WH and in its original (d-structure) position. (For a definition of the nodes that can structurally integrate a node X in C, see (2.11).)

Following the logic of our relative clause discussion, we might expect to be able to rank different Movement Domains according to their relative complexity, with consequences for ease of processing and frequency of use, for restrictions on WH-movement across languages, and for pronoun retention. I have not undertaken a systematic investigation of WH-movement in performance and grammar, and there are probably a number of relevant explanatory considerations here. However, I do not believe that current models of grammar are quite succeeding in capturing the regularities in this area. In ch. 1.3 I questioned the subjacency constraint on WH-movement on the grounds that there are languages like English that provide *prima facie* counterevidence to it, since a WH-word can be separated by two or more bounding nodes from its trace. There is no obvious independent support for successive Comp-to-Comp movements, upon which the formulation of subjacency depends (viz., no more than one bounding node can be crossed at a time). And if WH-elements can be moved in a single step out of embedded clauses and into their landing site, there will be regular violations.

Consider now the following partial extraction hierarchy, comprising Movement Domains of increasing complexity:

(2.30) **Partial Extraction Hierarchy**

a. \([WH_i \ldots VP(V \overline{VP}\{ \ldots \Theta_i\})]\)  
[Who does he hope \( V \overline{VP}\{ \ldots \Theta_i\} \) to marry \( \Theta \)]

b. \([WH_i \ldots VP(V NP \overline{VP}\{ \ldots \Theta_i\})]\)  
[Who did he persuade John \( V \overline{VP}\{ \ldots \Theta_i\} \) to marry \( \Theta \)]

c. \([WH_i \ldots VP(V S\{ \ldots \Theta_i\})]\)  
[Who did he announce \( S\{ \ldots \Theta_i\} \) that Mary had married \( \Theta \)]

d. \([WH_i \ldots VP(V NP S\{ \ldots \Theta_i\})]\)  
[Who did he announce \( NP S\{ \ldots \Theta_i\} \) the fact \( S\{ \ldots \Theta_i\} \) that Mary had married \( \Theta \)]

The Movement Domain of (2.30b) contains one more node than (2.30a), namely the NP *John*. The embedded \( S \) in (2.30c) introduces greater structural complexity than the embedded VP in (2.30a) and (2.30b): \( S \) contains a VP.
plus a subject NP that is absent from VP, and also, depending on the language, finite auxiliary material and a sentence complementizer. Example (2.30d) contains an NP sister to S within a complex NP, and so introduces even more structural complexity into the Movement Domain than (2.30c).

Example (2.28) makes several predictions for the hierarchy of (2.30). For example, it makes performance predictions for the relative frequencies with which extractions will be attested out of the environments in question. For English I have undertaken a brief and informal text count which suggests the following pattern: WH-movement is most frequent out of a non-embedded environment (i.e. questions such as *Who did you see?*); for embedding environments such as (2.30a–c) there appear to be declining frequencies down the hierarchy. Extraction out of a complex NP is ungrammatical (unless a pronoun is retained, cf. (2.35) below). These facts are in accordance with our predictions, but they must be considered tentative until more systematic performance data have been collected.

Example (2.28) predicts that WH-movement may cut off across languages in accordance with this hierarchy. I have tested this on a sample of three languages that have clear (and well-studied) WH-movement rules, English, German, and Russian:

(2.31) WH-movement on (2.30a): always OK in English, German, Russian
(2.30b): OK in English and Russian, not always in German
(2.30c): OK in English, not in German (standard) or Russian
(2.30d): not OK in English, German, Russian

All three languages permit WH-movement in (2.30a). German assigns ungrammaticality to at least some instances of (2.30b). The following sentence is starred by Bierwisch (1963):

(2.32) *Die Unterlagen, s[verdächtigt man ihn Vphrase [0]] untergeschlagen
"The records suspects one him withheld zu haben"]
to have," i.e.
One suspects him of having withheld the records.

Sentences of type (2.30c) are fully grammatical in English, but not in German (the standard language at least, cf. Hawkins 1986: 87–92) or in Russian, cf. (2.33) from Comrie (1971):
Example (2.30d) is not grammatical in English, German, or Russian.

We therefore see that as we move down the hierarchy of (2.30), extraction becomes progressively more difficult, as predicted. Grammars define their cut-off points for extraction in a way that appears to be at least partially predicted by considerations of structural complexity. What we want to say in the case of English is not that WH-movement can cross only one S (or NP) node: the WH-word can cross any number of S (and S) nodes in this language in sentences such as (1.7) above, preserving grammaticality, although complexity will increase as more such nodes are crossed, and processing ease will decline. Rather, English has grammaticalized a prohibition on extraction out of complex NP environments, $NP[NP \; s[\ldots WH \; \ldots]]$ (cf. Ross's 1967 Complex NP Constraint), since this containing NP regularly contributes a highly complex set of structurally related nodes to a Movement Domain, more than a containing S alone, which in turn contributes more than an embedded VP. Russian and German, by contrast, have grammaticalized a prohibition on WH-extractions out of S, the next most complex containing node on the hierarchy of (2.30), and so on.

Grammars have responded to performance complexity, therefore, by identifying certain containing constituents, or combinations of constituents, that dominate complex sets of structurally related nodes, and by prohibiting extraction out of these constituents or configurations. What is subject to parametric variation is the degree of complexity that is tolerated within Movement Domains in different languages, English being more generous than Russian and German. This variation is implicationally structured, just like the variation on relativization within a clause, and is ultimately explained by performance. The containing constituents or configurations on this account are like the bounding nodes of subjacency theory, except that movement is now not permitted across any bounding node so identified, and some boundaries are defined by a plurality of nodes. Thus, $S$ is a bounding node in Russian, and no extraction is possible out of it; while the complex NP configuration defines a corresponding extraction boundary for English.

The parallels between relativization constraints within a clause and extraction constraints out of embedded clauses of different types become even more striking when we consider pronoun retention, for pronouns can be retained in difficult WH-movement environments across languages (cf. Ross 1967), just
as they can be in difficult relative clauses. Consider the following facts from English. English does not normally permit pronoun retention in the three highest environments of (2.30):

(2.34) a. *Who does he hope \(\overline{VP}\)[to marry her]
   b. *Who did he persuade John \(\overline{VP}\)[to marry her]
   c. *Who did he announce \(\overline{S}\)[that John had married her]

Pronoun retention within a complex NP such as (2.35a), however, is better, and it is better still if there is yet another \(\overline{S}\) node over which movement has taken place, as in (2.35b).

(2.35) a. ?Who did he announce \(\overline{NP}\)[the fact \(\overline{S}\)[that John had married her]]
   b. Who was it \(\overline{S}\)[that he announced \(\overline{NP}\)[the fact \(\overline{S}\)[that John had married her]]]

### 2.4 Structural Domains and their semantic correlates

I have argued that the syntactic complexity of a node X in C is to be defined in terms of the number of structural relations that structurally integrate X in C. The semantic interpretation of nodes is also dependent upon structure. Current work in generative grammar has shown that there is a strong correlation between the environments within which syntactic dependencies are stated (e.g. reflexive anaphors must be c-commanded by their antecedents) and the environments within which semantic dependencies such as logical scope can hold, e.g. a quantifier phrase within the scope of another quantifier phrase or of negation must be c-commanded by these latter (cf. Reinhart 1983). In fact, I believe that the structural basis of semantic interpretation is even more extensive than is currently recognized. All examples of one node semantically determining the precise interpretation of another seem to be structure-dependent: X can be semantically dependent on Y only if Y is one of the nodes that structurally integrate X, and hence only if Y is in the SD of X within some higher constituent C.

This assumption, which will need to be defended below, predicts that there will be the following semantic correlate to syntactic complexity: the larger the SD for a node X (i.e. the more complex), the more nodes there will be that can semantically determine the interpretation of X; the smaller the SD for X, the fewer such nodes, and the less semantically determined and the more independent X will be within its containing constituent.
The proposal that underlies this claim is the Principle of Structural Determination, defined as follows:

(2.36) **Principle of Structural Determination**
A node X can depend syntactically and/or semantically only on those nodes that structurally integrate it (as defined in (2.11)), i.e. on nodes within a Structural Domain of X in a higher constituent C.

Anaphoric binding is an example of a syntactic and a semantic dependency. The reflexive anaphor *himself* requires an antecedent such as *John*, and (2.36) defines a (necessary) condition on the distribution of these elements: X, the dependent anaphor, must have Y, the antecedent, in its SD within the most immediately containing S. This predicts the grammaticality of (2.37a), in which *himself* is in DO position and has *John* in its SD, and the ungrammaticality of (2.37b) where the SD for the subject anaphor *himself* does not contain *John*:

(2.37) a. John washed himself
b. *Himself washed John.

The quantifier scope examples of (2.38) pattern in a similar way:

(2.38) a. Every bright student wanted a good professor.
b. A good professor wanted every bright student.

The SD for *a good professor* in the DO position of the clause in (2.38a) contains both the SU *every bright student* and the verb *wanted*. As predicted, the semantic interpretation of *a good professor* can be determined by these elements. Example (2.38a) can mean that every bright student is such that he or she wants that there should be some (possibly non-existent) good professor for that student, i.e. with the existential quantifier of *a good professor* having narrow scope relative to *every* and the verb *wanted*. In addition, (2.38a) has further readings with more specific interpretations: it can mean that for every bright student there is some (possibly different but existing) good professor and every bright student wants that good professor (whoever it might be); i.e. with the existential having narrow scope relative to *every*, but wide scope relative to *want*. The same sentence (2.38a) can also mean that there is some highly specific good professor who is such that every bright student wants him or her, where the existential quantifier has widest scope of all. As the degree of specificity rises, the semantic interpretation of *a good professor* becomes progressively independent of, and semantically uninfluenced by, the nodes of its SD. In (2.38b), by contrast, where *a good*
professor is in SU position and neither the every phrase nor want is contained in the SD for a good professor, there is only one (non-generic) interpretation, the one corresponding to the most specific reading of (2.38a): there is some highly specific good professor who is such that he or she wants every bright student.

Example (2.38) suggests two things. First, a quantifier phrase can be semantically dependent on (i.e. under the scope of) only those nodes that are within its SD (here c-commanding nodes in Reinhart's 1983 sense). There is no reading for (2.38b) in which the existential quantifier is under the scope of the verb want or of the quantifier every. Secondly, the existential quantifier in (2.38a) does not have to be interpreted as semantically dependent on the nodes within its SD. It can escape from such semantic dependence under conditions that can be described in terms of "degrees of specificity" (cf. Fodor and Sag 1982), and which probably correlate with the amount of pragmatic information about the referent that speaker and hearer bring to the sentence interpretation process (cf. Hawkins 1978, 1991). Definite NPs favor wide scope, for example, because the referent is generally mutually known as an existing individual independently of the description used for identification. Depending on the amount of pragmatic information associated with the indefinite a good professor, the degree of specificity will also be greater. Hence, the Principle of Structural Determination (2.36) defines a necessary condition on syntactic and semantic dependencies, not a sufficient one. A node X can be dependent on the nodes Y within its SD in C, but it does not have to be.

Examples (2.37) and (2.38) support the proposition that syntactic dependencies and semantic interpretation are both regulated by language structure. Given two nodes, X and Y, where the SD for Y is properly contained within the SD for X, SD(Y) ⊆ SD(X), X can depend on and be determined by Y, but not vice versa. Hence, Y is autonomous of X. The reason is that Y is one of the nodes that structurally integrates X and that defines its structural role within the containing constituent, whereas X is not one of the nodes that does this for Y.

The descriptive facts of (2.37) and (2.38) are usually accounted for in terms of c-command asymmetries (cf. Reinhart 1983). We now see that there are more general structural principles at work. If A c-commands B, then A is one of the nodes that structurally integrates B and that defines its structural role. By analyzing these data in terms of SDs, we can achieve certain advantages over c-command.
First, there is a certain arbitrariness in the formulation of c-command. Why should branching be significant (a node A c-commands a node B just in case the first branching node above A dominates B, etc.)? This follows because if A is a branching daughter of some node W, then A necessarily has a sister node that must either be B itself, or that must ultimately dominate B. Hence, talk of branching is a disguised way of saying that A is either a sister of B, or a sister of some node that dominates B, and hence that A is in the SD of B in W. We are simply dealing here with an extension of the sisterhood concept to include nodes that dominate B, in addition to immediate sisterhood between A and B. The nodes that structurally integrate B are therefore dominating nodes and sisters either of B or of nodes that dominate B.

Secondly, by viewing c-command as just one of the structural relations that defines the structural integration of a node X in C, and by in effect turning the definition of domains around, from the nodes that a given constituent c-commands (as in Reinhart 1983) to the nodes that c-command and dominate a given constituent, we can explain something that is simply stipulated in the c-command account. Fundamental to this account is the insight that antecedents have to c-command anaphors, and that anaphors have to be c-commanded. But why isn’t it the other way round? What is it that prevents anaphors from c-commanding their antecedents? It would be just as simple for the grammar to formulate its binding restrictions as “anaphors must c-command their antecedents,” rather than “antecedents must c-command anaphors.” So why are sentences like (2.37b) *Himself washed John ungrammatical?

If we think in terms of SDs of structurally integrating nodes, we can see a reason. A direct object requires the co-occurrence of a subject, and depends on this latter for defining the structural role of a direct object, even in sentences that contain no anaphors at all. In John washed the boy, the SU, and washed, the transitive verb, in order to be a DO. But the SU just needs any set of words that can be classified as a VP or predicate, depending on the language, including a verb alone. If there is a VP, then the SU will not include daughters of this VP in its SD. If there is no VP, and SU and DO are distinguished by nominative versus accusative case marking, then the nominative will be in the SD of the accusative, but not vice versa. Even in simple sentences without anaphors, therefore, there is a structural dependency between the DO and the SU that is asymmetric, and this explains why, when the DO is an explicit anaphor requiring agreement and co-indexing between two NPs, as in (2.37), the direction of agreement goes from the DO to the SU, mirroring the direction of structural dependency.
The grammaticalization of processing principles

that exists anyway for these positions and that is captured in our concept of an SD of X in C. The Principle of Structural Determination amounts to the claim that all explicit syntactic and semantic dependencies of this sort will reflect structural asymmetries and dependencies that exist anyway, and hence it provides a motivation for what is simply stipulated by the grammar of c-command.

It follows from the Principle of Structural Determination that the larger the SD is for a given node X, the more nodes there are that can contribute to the semantic interpretation of X. Consider cases where X is a referential expression such as NP. Larger SDs will provide more nodes that can determine the reference of this NP syntactically and semantically: the SD may contain an antecedent to which the NP is a bound anaphor, or scope-bearing elements relative to which it has narrow scope, or nodes that can restrict what I shall call the "quantity of the NP's reference potential." Smaller SDs, by contrast, render an NP more referentially independent within its containing sentence. This has already been illustrated for bound anaphora and scope in (2.37) and (2.38): SUs are more referentially independent than DOs. Consider now some examples of quantity-reference restrictions. As the size of an SD for X increases, so the quantity of the objects or mass referred to by an NP in the position of X can decrease as a function of the semantic interpretation of other nodes within its SD.

There is a long-standing debate in the literature on generic reference about how universal the reference of a bare plural has to be for it to be considered generic as opposed to non-generic (cf. Chesterman 1991 for summary and discussion). Consider (2.39):

(2.39) Beavers build dams with branches.

This example shows that the quantity of entities referred to by a generic plural is systematically restricted by the nodes within its SD. How many beavers are we talking about? All beavers (or beavers in general). How many dams? As many as beavers build. How many branches? As many as beavers build dams with. That is, beavers is referentially independent of everything else in the sentence and receives a nomic and universal interpretation. Dams is interpreted relative to beavers, the SU, and build, the transitive verb: it is not a claim about all dams. And branches, which has all these other sister nodes within its SD (cf. the tree diagram in (2.13)) is interpreted relative to all of them: the only branches that are being referred to are those used by beavers for the purpose of building dams. This sentence reveals rather graphically that the compositional process of combining the meanings of
sentence parts into a whole-sentence interpretation is structure-dependent. One could imagine a communication system in which the interpretation of beavers would be made relative to that of dams in (2.39), but this is excluded by the Principle of Structural Determination.

Consider also the famous pair of examples from Anderson (1971):

(2.40) a. John loaded the wagon with the hay.
    b. John loaded the hay onto the wagon.

The wagon in DO position in (2.40a) is generally taken to refer to the whole wagon, which John is claimed to have loaded with as much of the hay as is required, i.e. the hay may refer to less than all the hay. With the hay in DO position in (2.40b) a reference to all the hay seems to be preferred (cf. Dowty 1991 for further discussion), while the wagon will be filled in proportion to the amount of hay available. In other words, the DO is referentially more independent than the NP within a PP. This latter depends upon the DO for its interpretation, just as a DO has been shown to depend upon a SU.

There is a further property of sentences (2.39) and (2.40) which Dixon (1989) draws attention to. They are complete without the PPs (Beavers build dams, John loaded the wagon, John loaded the hay). The PP provides further information that functions like a predication on complete propositions: it is with branches that beavers build dams, it is with hay that John loaded the wagon, etc. As a result, the PP almost has the status of an "afterthought," and sentences like (2.40) that have alternative ways of distributing the arguments of a single verb syntactically will have different interpretations depending on which argument is placed within the propositional core of the sentence, and which within the predication.

All of these examples suggest that semantic interpretation is a compositional function whereby the meaning of a given node X can be added to that of the nodes within its SD, and can vary its interpretation and be dependent on these nodes, whereas these latter are independent of X and cannot vary with it. As the SDs of different nodes X get larger within C, X becomes more referentially dependent and variable in its interpretation. The addition of meanings in sentence composition is rather like the addition of layers to an onion: the shape and size of each layer is determined by all lower layers, and determines but is not determined by all higher layers. One might speculate that natural language interpretations have to be asymmetrical in this way, since if they were symmetrical there would be an infinite regress into mutual delimitation, and meanings would be indeterminate (the dams that are built by the beavers that build the dams that are built by the beavers that . . . ).
The grammar must then encode the priorities between phrases, and indicate which one is in the SD of the other, and hence which one has its meaning assigned first independently of others. This can be done via syntactic configuration and c-command, or by morphological case marking, or by linear precedence, and this would appear to underlie the kinds of asymmetrical hierarchies of argument positions discussed in Primus (1987) and (1991a); cf. section 2.3.

For language processing this means that large SDs involve a more complex compositional process than small SDs. In order to determine the semantic interpretation of a node X in C, it is necessary to access all the nodes in X's SD and to compute what impact they have (if any) on the meaning of X. Clearly, the more nodes that are accessed in this way, the more simultaneous computations are made of relevance to X. Principles of syntactic dependency and of semantic determination are not only structure-dependent in the manner of (2.36), therefore, they reveal graphically how, as certain nodes are processed, the portions of a syntactic tree expand within which syntactic and semantic decisions must be made. As SDs expand, there is an increase in the number of items within current working memory and in the number of computations that must be performed on them. If grammars are in any way sensitive to processing load, it is natural to expect that they will respond to this in some way. There is now substantial evidence, of the kind presented in the last section involving SDs for relativization and for WH-movement, to suggest that they have. In the next chapter I shall focus on another grammatical area for which a processing explanation in terms of SD sizes seems particularly compelling: linear order.
3 Early immediate constituents

3.1 The basic idea

Let me begin with the basic intuition that underlies the major principle of this chapter: Early Immediate Constituents (EIC). I believe that words and constituents occur in the orders they do so that syntactic groupings and their immediate constituents (ICs) can be recognized (and produced) as rapidly and efficiently as possible in language performance. Different orderings of elements result in more or less rapid IC recognition. Consider again the Heavy NP Shift example (2.2) from English, renumbered as (3.1):

\[(3.1)\]
\[
a. \text{VP}\{\text{gave}\text{ NP}\{\text{the valuable book that was extremely difficult to find}\} \text{PP}\{\text{to Mary}\}}\]
\[
b. \text{VP}\{\text{gave}\text{ PP}\{\text{to Mary}\} \text{ NP}\{\text{the valuable book that was extremely difficult to find}\}}\]

Example (3.1b) provides a more rapid presentation of the three ICs of the VP (V, NP, and PP) than (3.1a). The verb gave is the first IC of the VP in both examples and signals to the parser that a VP should be constructed. The PP is a two-word IC here. Its positioning to the left of the lengthy NP in (3.1b) makes it possible for all three daughter ICs to be recognized within a short viewing window, since the NP can be recognized on the basis of the determiner the, occurring in leftmost position within this NP. In (3.1a), on the other hand, the viewing window extends all the way from gave to the preposition to, which constructs the PP, and the heaviness of the intervening NP delays access to this third IC. Of the 12 total words dominated by this VP, therefore, 11 need to be examined for IC recognition in (3.1a), whereas just four suffice in (3.1b).

This example suggests that the ICs of a phrase can be recognized on the basis of a proper subset of the words and categories dominated by that
phrase, and that the size of this proper subset can vary. Some orderings shorten the number of dominated elements that must be scanned for IC recognition, making this process faster and also more efficient, since the same amount of information is extracted from less input. If we think in terms of the concept of a Structural Domain, which I defined in the last chapter in (2.8), we can say that (3.1a) and (3.1b) differ in their "constituent recognition domains" (CRDs) for VP, i.e. in the subsets of nodes and words dominated by VP that are sufficient for the recognition of VP and its ICs. A CRD is a type of Structural Domain consisting of the nodes that make phrase-structure recognition possible on-line. Example (3.1a) has a more complex SD for VP recognition than (3.1b), and is accordingly dispreferred in language use (cf. ch. 4.3.2 for empirical confirmation).

I believe that this preference for (3.1b) is not something that is unique to Heavy NP Shift, or even to other rightward-moving rearrangements in English and other languages. Rather, what we are seeing here is just one manifestation of an extremely general principle of ordering that holds, I shall argue, not only for rearrangement phenomena but for basic word orders as well, not only in English but in all languages, and not only throughout the grammar but throughout language performance as well. This principle will explain many ordering conventions of generative grammar, which are simply stipulated at the present time, and will motivate many apparent exceptions to these conventions. It will also account for many pragmatic and discourse correlates of word-order arrangements and rearrangements, and will reveal that these are secondary consequences of syntactic processing and not primary generalizations in their own right. In the process I will suggest a resolution of the contradiction in the current pragmatics literature over whether given precedes new information (cf. the Prague School) or new precedes given (cf. the work of Givon and his associates).

In order to give substance to these claims, some parsing principles must first be defined.

3.2 Some principles of parsing

In discussing (3.1) I suggested that the mother node VP could have different Constituent Recognition Domains. This term can be defined as follows:

(3.2) **Constituent Recognition Domain** (CRD)

The CRD for a phrasal mother node M consists of the set of terminal and non-terminal nodes that must be parsed in order to recognize M
and all ICs of M, proceeding from the terminal node in the parse string that constructs the first IC on the left, to the terminal node that constructs the last IC on the right, and including all intervening terminal nodes and the non-terminal nodes that they construct.

The CRD for (3.1a) is accordingly initiated by the verb gave on the left, and proceeds to the eleventh word of the parse string, to, and it consists of all the terminal and non-terminal nodes dominated by VP within this stretch of the parse string. This CRD is a psycholinguistically significant subset of the total set of nodes in the tree for (3.1a), since it is on the basis of this subset that an important parsing decision can be made, namely: what is the structure of the VP? Hence, this CRD is a Structural Domain, as this term was defined in (2.8). The structural complexity of this SD (cf. (2.12)) is, moreover, extreme, since it consists of a full 32 terminal and non-terminal nodes (including the dominating VP itself), i.e. all the nodes shown between the two broken lines in (3.1'a):¹

(3.1'a)
Putting this another way, we can say that 28 terminal and non-terminal nodes have to be scanned in order to recognize the structure of the VP, this structure being represented by the VP itself, and its three ICs, V, NP, and PP.

The complexity of the corresponding CRD for (3.1b) is significantly less, just 12 terminal and non-terminal nodes, as shown in (3.1'b):

(3.1'b)

The recognition of VP structure now requires the scanning of only eight terminal and non-terminal nodes.

The concept of a CRD therefore enables us to say that different orderings of words and constituents result in larger or smaller SDs. This then provides a basis for preferring some orderings to others, namely those with smaller CRDs, since they are less complex and more efficient: they answer the same parsing question about the structure of the VP using less input material. Smaller CRDs contain fewer elements that must be held in working memory and that are relevant to a given parsing decision, and this reduces the number of phrasal nodes whose structure must be computed simultaneously. The more these other phrasal nodes can be processed after the structure of VP has been fully recognized, as in (3.1'b), the fewer demands are made on the processor. The different CRD scores for (3.1'a) and (3.1'b) appear to correspond to differences in processing load, therefore, involving the size of working memory and the number of computations performed simultaneously on elements within this working memory.

The definition in (3.2) refers to a terminal node, i.e. a word, "constructing" a constituent. This is a fundamental parsing concept which goes back to Kimball's (1973) New Nodes principle. The syntactic categories of some...
Some principles of parsing

words are unique to their respective dominating phrasal constituents, or mother nodes, and they act as unambiguous signals on-line to the parser to construct the relevant mother. Other categories do not permit reliable inferences to be drawn about dominating structure. For example, a noun can tell the parser to construct an NP; but an NP may be immediately dominated by a VP, a PP, or by another NP and does not in and of itself permit the dominating node to be reliably inferred. If a V immediately precedes NP in the linear parse string (as it does in English), then V will first construct the mother node, VP, and NP can then be attached to this mother. But in languages in which NP consistently precedes the V within VP, the P(ostposition) within PP, and the head noun within NP, the attachment of NP to its dominating node cannot be made with confidence until the relevant (rightmost) mother-node-constructing category is encountered. The parser could conceivably employ guesswork here (assigning NP arbitrarily to, say, PP on-line) or it could place NP in a look-ahead buffer for unattached constituents, along the lines of Marcus's (1980) parser, attaching it as soon as the rightmost category is encountered that constructs the mother.

Our word-order data from performance and grammar provide evidence that the latter is the course adopted by the human parser. These data are only predicted correctly on the assumption that information about higher constituent structure becomes available exactly when it is uniquely determined by the input. In any parsing device that strives for rapidity and efficiency, regular guessing is costly, requiring back-tracking and major tree surgery whenever the guesses turn out to be wrong, and it would also result in more complex CRDs in all head-final languages (cf. (3.9) below).

Moreover, notice that once a dominating node, such as VP, can be reliably postulated, it does then become possible to assign ICs to it with confidence that do not uniquely determine VP themselves, such as NP, but that are dominated by VP in the well-formedness rules of the language. Hence, reliable mother node construction is a prerequisite both for recognizing abstract syntactic structure and for attaching non-uniquely determining ICs to higher phrases, and for this reason categories that uniquely determine their mother nodes play a vital role in parsing. Such categories seem to be of two general types: closed-class items or function words (e.g. Det can construct NP as in (3.1), Comp constructs S, Aux constructs S); and lexical heads of phrases (N is unique to NP, V to VP, P(reposition) and P(ostposition) to PP, and Adj to AdjP). In more recent theories of syntax
such as Chomsky (1986), closed-class items such as Comp and Infl (i.e. Aux) are regarded as (functional) heads of the higher projections CompP and InflP respectively. Similarly Abney (1987) argues for Det as the head of a DetP rather than of NP. Heads are now consistently the categories that construct their respective mothers, and parsers that incorporate this model of grammar are referred to as “head-driven,” cf. e.g. Pritchett (in press).

I hypothesize, therefore, following Marcus, that mother nodes will be constructed only when their presence is uniquely determined by the input. If they are not so determined, the parser will employ look-ahead until higher constituent structure can be reliably inferred.

I hypothesize also, following Jerry Fodor (1983), that parsing decisions are made obligatorily and immediately upon encountering the triggering input. A mother node must be constructed when a syntactic category is recognized that uniquely determines the higher node. Other ICs will be attached as rapidly as possible to this higher node, thereby making parsing decisions about sisterhood and groupings proceed as fast and as efficiently as possible. Since linguistic material is produced item by item on-line, sister ICs cannot all be presented simultaneously. But they can be presented more or less rapidly in succession, and the principle of EIC will define a clear preference for speed in this regard.

We have, therefore, the following principle of Mother Node Construction:

(3.3) \textit{Mother Node Construction (MNC)}

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.

For non-mother-node-constructing categories we define a principle of Immediate Constituent Attachment:

(3.4) \textit{Immediate Constituent Attachment (ICA)}

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.

An adequate parser will need to incorporate many other instructions in addition to (3.3) and (3.4), and further construction principles will be defined in ch. 6.2, e.g. Sister Node Construction and Grandmother Node
Construction. MNC is the most general and universal in terms of the number of structures to which it applies, and it will be the main focus of this book. It constructs the major phrasal constituents within which linear ordering predictions can be tested.

Notice the following assumptions that underlie MNC. This principle assumes that mother-node-constructing categories (MNCCs) do indeed exist, and that they have the parsing function described, i.e. they enable the hearer to recognize and construct phrasal mother nodes on-line. I shall refer to this as the axiom of MNCC Existence, defined in (3.5):

(3.5) Axiom of MNCC Existence

For each phrasal mother node M there will be at least one daughter category C that can construct M on each occasion of use.

This axiom asserts that all phrasal mother nodes will be recognizable by virtue of some MNCC, and it allows explicitly for a possible plurality of MNCCs for any given M.

A second assumption underlying MNC is captured in the axiom of MNCC Uniqueness:

(3.6) Axiom of MNCC Uniqueness

Each MNCC will consistently construct a unique M on each occasion of use.

In other words, although a given M may be constructed by different MNCCs, a given MNCC cannot construct different Ms on different occasions of use.

There is a clear functional motivation for these axioms. If several words, categories, and constituents all belong to a common mother phrase, it has to be the case that the parser can recognize the appropriate M for these items within a parse string. M is an abstract entity, and there is no explicit phonetic material in the input that signals VP or PP, in the same way that words are signaled through phonetic correspondences with their phonological representations. So abstract syntactic information has to be inferred from the terminal material that is present in the parse string, and from the non-terminal categorical information that can be accessed as soon as a given word has been recognized. For example, the terminal element gave in (3.1) is a verb, and the knowledge that this is its syntactic category permits construction of the node VP. The axiom of MNCC Existence asserts that the inference of M's identity must always be derivable from some daughter of M, rather than from, say, a
sister, or from some more distant node in the tree. Such possibilities are not excluded in addition to MNC, but (3.5) claims that no phrasal node could be constructed only on this basis, and in ch. 6.2.1 I suggest a reason for this. By making the identity of a phrase dependent upon a daughter that is unique to it, higher constituent structure can always be readily and rapidly inferred. But having the identity of M depend entirely upon a non-dominated constituent would not be efficient, since it would regularly force the co-occurrence and/or adjacency of the constructing category and this would lead to numerous complications.

Similarly, if Ms were not unique to MNCCs, and an adjective could construct an AdjP or a VP, for example, there would be systematic ambiguity in performance over which M was intended on a particular occasion. Making the wrong selection would result in structural misanalyses, garden paths, etc., all of which are known to cause processing difficulty, and would defeat the very purpose of having MNCCs, which I regard (following Jerry Fodor 1983) as a means of providing an immediate, obligatory and reflex-like recognition of higher constituent structure on-line.

These axioms motivate the principle of Mother Node Construction (3.3) in performance. They also have a profound impact on constituent structure itself, by constraining mother–daughter relations, co-occurrence restrictions within a phrase, deletion possibilities, and morpho-syntactic processes, as illustrated in ch. 6.1.

3.3 Some preliminary evidence for EIC

With this much background let us present a preliminary sample of evidence taken from the currently available published literature that supports our linear ordering principle EIC, before defining this principle and its predictions precisely. The evidence comes from the following four sources of data: from real-time psycholinguistic experiments using alternative orderings of ICs in languages that permit such alternatives; from text-frequency counts for these alternatives; from native speaker acceptability judgments (often experimentally elicited); and from the grammaticalized constituent orders of the world’s languages. The first three of these are performance data, the fourth involves grammar.

Consider first Particle Movement sentences in English. Hunter (1981) and Hunter and Prideaux (1983) have tested native speaker acceptability judgments for sentences such as (3.7a–d) in an experimental situation:
Some preliminary evidence for EIC

They found that the further the particle is from the verb, the less acceptable the sentence becomes. The VP has three ICs in (3.7): the verb looked, the particle up, and the NP. The verb is the leftmost IC, and what varies down (3.7) is the rapidity with which the remaining two ICs can be attached to VP by Immediate Constituent Attachment (3.4). The CRD for VP is shortest and comprises fewest terminal and non-terminal nodes in (3.7a), and longest and syntactically most dense in (3.7d), and these increasing CRD sizes correlate with decreasing acceptability in the experiment. They also correlate most precisely with decreasing frequencies of occurrence in English texts, and with corresponding increases in frequency for the shifted _VP[V Part NP]_ structures (cf. ch. 4.3.1).

A similar situation obtains for VP-internal heavy NPs as in (3.1a). The longer such NPs are compared with the following PP, the larger the CRD for VP becomes, and the greater the preference for the rearranged order _VP[V PP NP]_ becomes. As preferences for rearrangements increase, acceptability judgments and frequencies of occurrence for the basic order _VP[V NP PP]_ decline (cf. ch. 4.3.2). Once again, CRDs that are as small as possible are preferred for VP.

Extraposition is similarly motivated:

(3.8) a. _S_[That Bill was frightened] _VP[surprised _NP[Mary]]_]

b. _S_[It _VP[surprised _NP[Mary]] S_[that Bill was frightened]]

In (3.8a) the lengthy sentential subject delays access to the second IC of S, the VP, constructed by the verb surprised. Within the VP two ICs are recognized on the basis of an optimal two words, surprised and Mary, but the CRD for S is very large, as shown. Extraposition in (3.8b) brings forward the VP, making a very small CRD for S. Moreover, the VP remains optimally efficient in this example, since there are now three daughter ICs, V, NP, and S,
recognized and attached within a three-word viewing window: $\text{VP}(\textit{surprised}_{\text{v}}, \textit{Mary}_{\text{NP}}, \textit{that}_{\text{s}})$. Experimental support for the greater difficulty of processing (3.8a) versus (3.8b) has been provided by Frazier and Rayner (1988), using an eye-movement experiment. Erdmann (1988) corroborates this with text-frequency counts: extraposed finite clauses outnumber their sentential subject equivalents by over 11 to 1 in his extensive corpus of English (cf. ch. 4.3.4).

A relative clause within an English subject NP also delays access to the VP and adds complexity to the CRD for S. A relative clause within a direct object NP, by contrast, does not increase the CRD for either VP or S. The S recognition domain is not increased because the direct object occurs to the right of the verb, which constructs VP. And the VP domain is not increased because the NP containing the relative clause is recognized on its left periphery, and so the right-branching relative $\tilde{S}$ is not contained within the CRD for VP. It is therefore significant that relative clauses within matrix subjects were shown to be more difficult to process than relative clauses within matrix objects, by both Wanat (1971), using an eye-movement experiment, and Levin et al. (1972), who used eye–voice span to measure processing complexity.

These examples suggest that the parser prefers structures that provide earlier and more rapid access to the ICs of VP and S. More generally, they reveal a preference for CRDs that are as small and as efficient as possible.

Consider now a head-final language like Japanese, in which Immediate Constituent Attachment regularly requires look-ahead and applies forward. Example (3.9a) is the center-embedded $\tilde{S}$ structure discussed in ch. 1 (numbered (1.6a)), and (3.9b) is its preposed alternant:6

\[ (3.9) \]

\[ a. \text{NP}[\text{Mary-ga}] \text{VP}[\text{S}_1[\text{S}_2[\text{kinoo John-ga kekkonsi-ta} \text{to}] \text{it-ta}] ] 
\]

\[ b. \text{NP}[\text{Mary-ga}] \text{VP}[\text{it-ta}] \]

The CRD for the highest $S_1$ is very long and unwieldy in (3.9a), but it can be considerably shortened by preposing $\tilde{S}$, as in (3.9b), in which $\tilde{S}$ is the first IC of the main clause, $S_2$. This $\tilde{S}$ is not constructed until the clause-final complementizer $\text{to}$ is reached, however. Until that point, the parser simply constructs an $S$, labeled here $S_1$, and does not know (and does not attempt to guess) whether this $S_1$ will eventually be a main or a subordinate clause, and if the latter a clause directly dominated by NP within a complex NP, or instead by S, or VP, and so on. Hence, the S nodes of (3.9b) are parsed
Some preliminary evidence for EIC  67

bottom-up, and once S is constructed, the remaining ICs of the matrix S₂, NP, and VP, are attached within a very short and rapid viewing window:

Frazier and Rayner (1988) cite evidence from Ueda (1984) that supports the bottom-up parsing of (3.9b). The first word *kinoo* "yesterday" is preferably attached locally to the first clause encountered, which will subsequently turn out to be a subordinate clause, and Frazier and Rayner point out that this is predicted by Minimal Attachment, but only on the assumption that the first S constructed is not necessarily the topmost S. The syntactic parsing of English and Japanese contrasts, in their view, in that the first S encountered in English is assumed to be the topmost S, whereas this assumption is not made in Japanese. Other parsing principles developed initially for English, such as the First Analysis Constraint (which includes Minimal Attachment), are claimed to be equally operative in Japanese. The top-down versus bottom-up difference in the processing of these two language types is further supported in Mazuka and Lust (1990), and with some modifications in Inoue (1991).

Evidence for a preference for minimal constituent recognition domains in Japanese comes from the fact noted by e.g. Dryer (1980: 134) that the heavier an internal S is in sentences such as (3.9a), the more frequently it is preposed in Japanese performance. Text counts from Japanese given in ch. 4.1.5 also confirm that left-branching sentence-internal ICs are preposed to the left with a frequency that matches the EIC preferences. The result is successive CRDs parsed bottom-up, each of which is short and efficient. In a parallel manner, the heavier an S or any other right-branching IC is in English, the more frequently it will be postposed. In both language types shorter CRDs are preferred over longer ones, but the different branching direction of mother-node-constructing categories, such as complementizers, means that this preference is accomplished by preposing in the one case and postposing in the other.

Notice that these performance preferences, and the different directionalities of rearrangements in English and Japanese, support the processing mechanisms proposed in the last section. The postposing of a heavy NP to the right of a shorter PP in (3.1b), or to the right of a single-word particle in (3.7), will only regularly reduce the VP recognition domain on the assumption that these NPs are recognized as NPs as soon as their uniquely determining daughter categories are encountered. If these NPs could be constructed after the point of unique determination, there would be no (or less) motivation to postpose them, since the CRD for the higher VP in a structure such as
Early immediate constituents

(3.1b) could then extend from $gave_\nu$ to some point within the heavy NP itself, making the overall length of the VP recognition domain closer to or identical with the lengthy CRDs of the untransformed structures. The very motive for postposing would thereby be undermined. Conversely, if $S$ could be recognized prior to encountering the complementizer in the Japanese (3.9b), there would be no (or less) advantage in preposing, since the CRD for the main clause could then be just as long as in the center-embedded clause (3.9a). Hence, only if ICs are constructed at exactly those points at which they are uniquely determined are these performance preferences regularly explainable as a consequence of shorter and less complex CRDs.

Moreover, if degree of heaviness (a structural consideration) is indeed what motivates the frequency and directionality of heavy IC rearrangements in performance and grammar, then it does not appear as if any contextual biases or predictions are of much use, or even desirable, in anticipating forthcoming structure. The rearrangements shorten CRDs: top-down context effects could actually lengthen them, especially in head-final languages, by constructing mother nodes (such as $S$ in (3.9b)) in advance of their bottom-up input. In head-initial languages, context effects could also potentially remove the motive for rearranging ICs: if the PP could be anticipated in advance of its appearance in (3.1a), there would be less reason to postpone the heavy NP, creating (3.1b). Since the preferences and dispreferences to be documented in this book seem to follow the dictates of structural length, regardless of the contextual predictability of the relevant ICs (cf. ch. 4.4), this suggests that the human parser is informationally encapsulated, in the sense of Jerry Fodor (1983), and prefers a first-pass scanning of the shortest possible CRDs provided by syntactic parsing procedures, regardless of context. In the words of Fodor (1983: 135): "What the context analyzer is prohibited from doing is telling the parser which line of analysis it ought to try next – i.e. semantic information can't be used predictively to guide the parse."

These preliminary examples supporting EIC primarily involve data from performance. A lot more evidence will be presented in the next chapter. The evidence from grammar will be discussed in detail in chs. 5–6. Notice in this context only that any grammar that blocks center embeddings corresponding to the Japanese (3.9a) (recall the Persian example (1.8) and the English (1.1a) in ch. 1) will be blocking structures with large CRDs. EIC will be able to explain this grammatical response, just as it explains the preference for left branching over center embedding with ICs of increasing weight in Japanese. The perceptual difficulty of center embeddings is not an isolated
fact, therefore, but is just one of many output predictions of EIC. Similarly, any grammar that blocks sentential subjects as in (3.8a) is blocking a structure with a large CRD. This structure was ungrammatical (or at least it is unattested) in all stages of the history of English up until the present, even through the Early Modern period (cf. Stockwell 1977a). The directionality of extraposition rules in English is also the result of EIC: any language with a left-peripheral complementizer can be predicted to postpone $\hat{S}$ to the right. Only in this way will CRDs be reduced, given that the complementizer constructs $\hat{S}$ on the left. Conversely, languages with right-peripheral complementizers, such as Japanese, will optimize their CRDs by preposing $\hat{S}$ to the left. In those languages in which such movements are created by grammatical rule rather than by a performance rearrangement (and the extraposition process of (3.8a) in English is certainly the result of a rule, since it is structure-changing, cf. Reinhart 1983), EIC will again have elicited a grammatical response.

3.4 Defining and measuring EIC

All the examples considered so far suggest that performance preferences correlate quite precisely with CRD sizes: as the recognition domains for VP and S get larger and contain more terminal and non-terminal nodes, so acceptability levels, ease of processing measured experimentally, and textual frequency all decline. When there are alternative orderings for one and the same set of ICs, the ordering that guarantees the smallest CRD (on account of the relative distribution of categories that construct mother and daughters respectively) is preferred. How can we best quantify this correlation?

One possibility would be to simply add up all the terminal and non-terminal nodes in a CRD. The CRD for VP in (3.1'a) contains 32 nodes, compared with 12 for (3.1'b), etc., as illustrated above. Another possibility was also briefly mentioned. We can calculate how many terminal and non-terminal nodes there are in the CRD apart from the nodes representing the mother and the ICs. This separation of higher-level nodes from lower-level and terminal nodes enables us to quantify the number of nodes that need to be processed in order to recognize the structure of the VP, and we could express this as a ratio of higher to lower nodes. Moreover, since all phrases have a mother, and since the number of ICs may differ across constituents (and across languages), the crucial ratio seems to be that between ICs and non-ICs: how many non-ICs need to be processed in order to recognize the ICs (at
Early immediate constituents

least one of which will construct the mother)? For (3.1'\(a\)) this produces a ratio of 3/28, compared with 3/8 for (3.1'\(b\)). In terms of percentages, 3/28 = 10.7\%, 3/8 = 37.5\%, and these percentages clearly quantify the greater efficiency of (3.1'\(b\)).

Consider this method applied to (3.7c) and its preferred counterpart with a postposed NP, given as (3.10a) and (3.10b) respectively:

\[(3.10)\]

a. \[S\]
   \[NP\]
   \[VP\]
   \[N\]
   \[V\]
   \[NP\]
   \[Det\]
   \[N\]
   \[PP\]
   \[Part\]
   \[the\]
   \[number\]
   \[Pr\]
   \[NP\]
   \[of\]
   \[Det\]
   \[N\]
   \[ticket\]

b. \[S\]
   \[NP\]
   \[VP\]
   \[N\]
   \[V\]
   \[Part\]
   \[NP\]
   \[Det\]
   \[N\]
   \[PP\]
   \[the\]
   \[number\]
   \[Pr\]
   \[NP\]
   \[of\]
   \[Det\]
   \[N\]
   \[ticket\]
The CRDs for VP are as shown. There are 14 non-ICs required for the recognition of the three ICs of VP in (3.10a), i.e. $3/14 = 21.4\%$, compared with just four for the same ICs in (3.10b), i.e. $3/4 = 75\%$. Again this captures the relative efficiency difference between the two and the preference for (3.10b). As the CRD for a VP with an unshifted NP grows, as in (3.7d) which contains a complex NP (cf. the tree structure in (3.11)), its IC-to-non-IC ratio will be worse than that of (3.10a) (= (3.7c)), namely $3/19 = 15.8\%$:

At the same time, the shifted counterpart of (3.11), *Joe looked up the number that Mary had forgotten*, will have the same $3/4$ ratio as (3.10b), and so will all the shifted structures with VP[V Part NP], because NPs are recognized on their left peripheries. The degree of preference for a shifted NP can therefore be quantified using the differential between IC-to-non-IC ratios for unshifted and shifted counterparts. The greater this differential, the greater the preference for the shifted structure and the more frequently we can expect the shift to take place in performance.

Calculating IC-to-non-IC ratios in this manner seems to provide us with a good index of the size and efficiency of CRDs. It also has further advantages compared with the simple addition of all nodes in a CRD. Consider languages with flatter constituent structures, e.g. with no VP node within S (cf. Hale 1983). The S node in such languages will contain more ICs than
in English, and hence more total nodes within the CRD for S. Many nodes dominated by VP that are within the CRD for this node in English will be in the CRD for S in languages with flatter structures. We do not wish to say that languages with flatter structures are necessarily less efficient, however. To do so would ignore the fact that their S nodes are structurally richer, comprising constituents that are distributed between two nodes, S and VP, in English. By making our relative complexity calculations in terms of ratios of ICs-to-non-ICs, we render it structurally sensitive and compatible with cross-linguistic variation, and we count efficiency relative to whatever number of ICs there happens to be within a given phrase.

This kind of sensitivity is required even within a language. A transformation like Extrapolation in English changes constituent structure. The VP in (3.8a) has two ICs prior to Extrapolation; the VP in (3.8b), after Extrapolation, has three. A comparison of the two VPs must be sensitive to this structural difference and must take it into account: more nodes need to be processed, but they lead to the recognition of more ICs. Both the S and the VP domains must be considered when calculating the EIC benefits of Extrapolation, since material is removed from the CRD for S and added to the CRD for VP. This can be done by aggregating the IC-to-non-IC percentage for the VP CRD with that for S in each of (3.8a) and (3.8b), and then comparing aggregates (cf. below). This aggregation will need to be done anyway when we calculate the EIC efficiency ratio for a whole sentence, which can be expressed as the aggregate of the IC-to-non-IC percentage scores for all CRDs (i.e. for all phrasal nodes) within the sentence. Since different phrases may contain different numbers of ICs, the use of ratios once again provides a more structurally sensitive and revealing index of complexity and efficiency.

The discussion so far raises two major issues, one theoretical, the other practical. In order to quantify nodes within a CRD, using any mathematical procedure, we need to know precisely what the constituent structure of a given phrase is. Yet grammatical models differ in this regard, sometimes radically, as do competing analyses within one and the same model. Faced with such differences, my strategy will be to be conservative and to maximize the common ground, so that my performance theory will not depend upon just one model. I shall also try to limit the grammatical entities I assume to those for which the empirical evidence is strongest, and which are observable in surface structure in some form. Every theory needs to accept the reality of major phrasal groupings of words, such as NP, AdjP, S, etc., even though the labeling of these may differ (e.g. NP or DetP, cf. Abney 1987). Theories vary,
however, in the richness of the internal structure postulated within these phrasal groupings, for example in the number of projections of the head dominated by the maximal projection within an X-bar theory (compare e.g. Jackendoff 1977 with other versions of X-bar theory, cf. Kornai and Pullum 1990). Some theories insist on binary branching ICs for all nodes, others do not. Theories differ over whether every phrase has a head, and if so which IC is the head (cf. Dryer 1992 for a contrastive summary of different theories in this regard). And theories differ also in the number of empty categories that are postulated in the syntax (Government-Binding Theory being the most generous in this regard, cf. Chomsky 1981, 1982, 1986).

The theoretical entity that will be crucial in my constituency assumptions is the highest phrasal category for a given phrase, i.e. the maximal projection in X-bar terms. It will not matter how this category is labeled as long as it ultimately dominates the same set of terminal nodes. I will assume a distinction between lexical categories, N, V, Adj, and P, and functional categories, Comp, Aux (or Infl), and Det, though nothing depends on this, and both will be distinguished from phrasal categories. The ICs of these phrasal categories will be defined as those categories separated by no other nodes from the mother phrase (in a non-X-bar theory), or as those that are dominated only by projections of the head (in an X-bar theory) but excluding the projections themselves, cf. (3.12) below. I reject the assumption of consistent binary branching, which I regard as empirically unsupported and theoretically unmotivated. My approach will be compatible with, but will not require, a head of phrase generalization, and in ch. 6.1 I will argue that many grammatical regularities that are currently stated in terms of this notion can be explained within a performance theory as consequences of node construction on-line, and that this explanation is equally (often more) compatible with grammars that do not make use of heads at all. I will assume also that empty categories, such as the traces of WH-movement, may be represented in surface structure, but they will not be counted for the purpose of EIC ratios. That is, empty categories are ignored.

My working assumptions are summarized in (3.12):

(3.12) **Working assumptions about constituent structure**

The ICs of each phrasal node M will be all and only the phonetically realized lexical, functional, and phrasal categories most immediately dominated by M in surface structure, excluding any phrasal categories that are non-maximal projections of the head.
The term "immediate constituent" is therefore used in its literal sense here whenever there is no intervening dominating node between the IC and M. For theories that postulate non-maximal projections, we can identify a set of ICs largely identical to those in other theories by appealing to the notion of "most immediate domination" and by excluding non-maximal projections as ICs. In this way we define comparable sets of Ms, comparable ICs for each M, and derivatively comparable non-ICs for these Ms (since this procedure will apply to all Ms contained within a higher M), across different grammatical models.

There is a related practical problem in all of this. Any serious empirical testing of EIC is going to require examination of a broad range of constructions in a wide range of languages. Yet even if one operates within the grammatical assumptions of a single model, there are still going to be numerous constructions in numerous languages that haven't been examined yet, or whose constituent structure is unclear, or subject to conflicting analyses, especially if one wishes to collect large numbers of sentences in texts, as we do. Plus, the sheer practical difficulty of drawing a constituent structure tree for every sentence in the data from each language would be daunting, even if the theoretical basis for every syntactic decision were secure.

For these reasons we need a simple and practical methodology with which to test EIC in a variety of languages, at least on a preliminary basis, to see if it is correct. What I propose is that we identify a given set of ICs for some phrase, following the working assumptions of (3.12), e.g. the three ICs of VP, \{V, NP, PP\}. ICs can be quite small and manageable in number, often just two or three constituents. The non-ICs, on the other hand, are much more numerous, and it is these non-ICs that lead to most of the practical problems. Even for English there are many different analyses for sentences (3.1) and (3.7), most of them involving non-ICs within the CRD for VP. Yet these differences are largely irrelevant to the point of EIC: large CRDs with large numbers of non-ICs are dispreferred, irrespective of how these constituents are labeled and irrespective (within reason) of how much phrase-internal structure is actually postulated in different analyses – such structure will be constant across all alternative orderings and across all phrases. It is not crucial to our EIC theory to get the precise non-IC structure right; what is crucial is to quantify more versus less non-IC structure for a given set of ICs.

One way to do this is to focus on observables, and count the relative numbers of words within a CRD. More words means more structure: each new word in a domain adds one terminal node, plus a pre-terminal category node, and possibly more non-terminal nodes besides. Relative numbers of
words should correlate, therefore, with relative numbers of dominating nodes. As a result there should be a close correspondence between the relative ranking among the IC-to-non-IC ratios for different CRDs of a given M, and the relative ranking for these same CRDs measured in terms of ratios of ICs to words. If this correspondence holds, we can simply restrict our attention to ICs and to words, and avoid a commitment to, as well as a quantification of, all other aspects of constituent structure.

We can test this proposed correspondence on the English structures considered hitherto, namely (3.1) and (3.7). Table 3.1 juxtaposes the different EIC calculation procedures on the two VP CRDs of (3.1) and on the four VP CRDs of (3.7). They do indeed define the same relative ranking. Naturally, the absolute scores for IC-to-word ratios are higher, because the amount of non-IC structure that is being quantified is less, just the terminal elements; hence the ratio of IC to non-IC material is higher when words alone are considered. But since it is the relative size of CRDs that matters when defining preferences and derivative ordering predictions, this is not significant. Given this correspondence, I conclude that IC-to-word ratios can be used as a shorthand for IC-to-non-IC ratios.  

It might be objected that a calculation procedure using IC-to-word ratios represents a return to the Miller and Chomsky (1963) complexity metric in terms of ratios of non-terminal to terminal nodes, which was criticized in ch. 2.2.2. But this is not so. Miller and Chomsky’s metric was defined on a whole sentence; ours is relativized to CRDs. Complexity in our terms means low ratios within the CRD (lots of non-ICs compared to ICs, i.e. low IC-to-non-IC ratios); for Miller and Chomsky it means high ratios for the whole sentence (lots of non-terminals compared to terminals, i.e. high non-terminal to terminal ratios). Their metric fails, I argued, because the ratio between

<table>
<thead>
<tr>
<th>Sentence</th>
<th>IC-to-word ratio</th>
<th>IC-to-non-IC ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3.1)</td>
<td>3/11 = 27.3%</td>
<td>3/28 = 10.7%</td>
</tr>
<tr>
<td>(3.1a)</td>
<td>3/4 = 75%</td>
<td>3/8 = 37.5%</td>
</tr>
<tr>
<td>(3.1b)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3.7)</td>
<td>3/3 = 100%</td>
<td>3/4 = 75%</td>
</tr>
<tr>
<td>(3.7a)</td>
<td>3/4 = 75%</td>
<td>3/7 = 42.9%</td>
</tr>
<tr>
<td>(3.7b)</td>
<td>3/7 = 42.9%</td>
<td>3/14 = 21.4%</td>
</tr>
<tr>
<td>(3.7c)</td>
<td>3/8 = 37.5%</td>
<td>3/19 = 15.8%</td>
</tr>
<tr>
<td>(3.7d)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
terminals and non.terminals is largely a constant function: the more terminals
in a sentence, the more non-terminal structure. We now see within this
chapter that their metric cannot discriminate between (3.1'a) and (3.1'b),
or between (3.10a) and (3.10b), both of which have identical terminal and
non-terminal nodes throughout the tree and in which the performance pre-
ference is uniquely a consequence of the different ordering. This preference is
captured in our theory by distinguishing different CRDs and by relativizing
IC-to-non-IC or IC-to-word ratios to CRDs. Moreover, when we make an
IC-to-word calculation, we use words as an indicator both of the terminal
nodes and of the non-terminal structure within a CRD that intervenes between
the ICs and the words. A high ratio means that there is not much additional
structure in the CRD apart from the ICs; a low ratio that there is a lot of
additional structure. This additional structure would be counted under the
non-terminal nodes in Miller and Chomsky's metric, when defined over the
sentence as a whole. By relativizing our ratios to CRDs, we can capture
significant differences between CRDs with differently ordered ICs, and
make predictions which the global ratio of Miller and Chomsky cannot. We
also have a different theory of the local domains within which complexity
should be computed compared with Frazier (1985), who proposed a three-
word viewing window plus the phrasal nodes constructed by these three
words. This, I believe, is not a grammatically or psycholinguistically signifi-
cant domain (cf. the critique in ch. 2.2.2). Despite these differences, all of these
theories are built on the same insight that complexity is a function of the
amount of abstract structure that needs to be derived from the terminal
elements of the parse string. They differ in the domains over which the com-
plexity function is defined: the whole sentence for Miller and Chomsky; three
adjacent words plus the nodes that they construct in Frazier (1985); and a
Constituent Recognition Domain here. I claim that the local domain captured
by the CRD concept is the grammatically and psycholinguistically significant
domain upon which the complexity function should be defined for all prefer-
ences relating to linear order; and more generally that Structural Domains as
defined in (2.8) are the appropriate domains for assessing the relative complex-
ity of all linguistic operations and structures.

To summarize, relative CRD sizes and efficiency levels will be quantified
by calculating ratios of ICs to non-ICs or to words alone, as follows:

(3.13) Calculating IC-to-non-IC ratios
The IC-to-non-IC ratio for a CRD is calculated by dividing the
number of ICs in the domain by the total number of non-ICs (or
Defining and measuring EIC

We can now define the principle of Early Immediate Constituents:

(3.14)  *Early Immediate Constituents* (EIC)

The human parser prefers linear orders that maximize the IC-to-non-IC ratios of constituent recognition domains.

When IC-to-non-IC ratios are maximized, the size of a CRD will be as small and efficient as it can possibly be for the recognition of a mother node and its immediate constituents. This predicts the preference for (3.1'b) over (3.1'a), and for (3.10b) over (3.10a). It remains, however, to specify the precise mathematics of this preference.

On the one hand EIC clearly prefers CRDs whose IC-to-non-IC ratios are as high as possible. This leads us to predict that an ordering or orderings with the best ratio will be selected from each set of alternatives in the unmarked case, both in performance and in grammars. The best ratio may not always be very high in absolute terms, if a CRD contains a plurality of large ICs. EIC therefore amounts to a preference for the most optimal ordering(s) possible within each set. At the same time EIC's preferences are a matter of degree: the bigger the EIC score for one alternative CRD over another, the greater the preference. Hence the preference for the shifted \[ VP[looked \ up \ NP[the \ . . . ]] \] over the unshifted \[ VP[looked \ NP[the \ . . . ] \ up] \ increases from (3.7b) to (3.7c) and from (3.7c) to (3.7d). The shifted structure has an IC-to-non-IC ratio of \(3/4 = 75\%\) and an IC-to-word ratio of \(3/3 = 100\%\) in all of these sentences. The unshifted counterparts have progressively lower ratios, and there is a progressively larger differential between optimal and non-optimal CRDs (cf. table 3.1). We accordingly expect performance reflexes of these different degrees, in the form of acceptability differences, relative frequencies of occurrence for shifted and unshifted counterparts, etc. Similarly, for the Heavy NP Shift structure of (3.1) we define a degree of preference for \( VP[V \ PP \ NP] \) over \( VP[V \ NP \ PP] \) in proportion to the size difference of \( NP > PP \).

EIC's preferences amount, therefore, to two interrelated claims and predictions. First there is the preference for the most optimal order(s), i.e. the one(s) with the highest possible EIC scores compared with other possible orders. EIC is not exceptionless: it holds in general, or in the unmarked case. But since exceptions are permitted, we want to be able to say something
about them and to make predictions for when they will be expected and when not. EIC’s second claim does this by quantifying the degree of difference between non-optimal orders and their optimal counterparts. Some of these relative degrees are large, others are smaller. Since I claim that EIC is the major determinant of IC ordering in general, and since it is an intrinsically gradient principle, it is reasonable to expect that it will continue to predict linear ordering preferences even when ratios are not completely optimal. This leads us to predict that exceptions should generally be found in cases where IC-to-non-IC ratios are still high, but not optimal. It enables us to make the implicational prediction that if exceptional orderings with large ratio contrasts to their optimal alternative(s) are attested at all, there will also be exceptions with smaller contrasts; and conversely if there are no exceptions when the ratio contrasts are small, there will be none when they are larger either.

Logically these two claims of EIC are independent of one another. The overall preference for the most optimal CRDs in the unmarked case could hold at the same time that exceptional orderings were not correctly predicted by the relative strength of their EIC scores; i.e. the exceptions could be unsystematic from the point of view of EIC, without jeopardizing the truth of the general claim. Conversely, the predicted correlation between relative EIC scores and ordering preferences could hold, without it having to be the case that optimal ratios occurred in the unmarked case, for example 90% of the time. The correlation could be satisfied by a simple relative frequency ranking, in which declining CRD scores for competing IC orders were matched by relative frequencies of occurrence such as 45% > 30% > 15% > 10%, etc. Only the two claims together lead to a 90% prediction for the IC order(s) most preferred by EIC, plus high EIC scores for the remaining 10%. Moreover, these two claims both follow from the same explanatory insight: the best possible EIC scores for all orderings of all ICs; plus a tolerance for non-optimal scores only when the difference between non-optimal and optimal scores is small. The attested proportion of these marked-case, non-optimal orders should ultimately reflect the number of performance instances in which not much is at stake.

We can therefore amplify our definition of EIC, as follows:

(3.14) \( EIC \) (Expanded)

The human parser prefers linear orders that maximize the IC-to-non-IC ratios of constituent recognition domains. Orders with the most optimal ratios will be preferred over their non-optimal counterparts
Defining and measuring EIC

in the unmarked case; orders with non-optimal ratios will be more or equally preferred in direct proportion to the magnitude of their ratios. For finer discriminations, IC-to-non-IC ratios can be measured left-to-right.

The last sentence will be clarified below.

Consider now the effects of Extrapolation in (3.8), repeated here:

\[(3.8) \quad \begin{array}{l}
\text{a. } s_{\text{S}}[\text{That Bill was frightened}]_{\text{VP}}[\text{surprised }_{\text{NP}}[\text{Mary}]]
\\
\text{b. } s_{\text{S}}[\text{It }_{\text{VP}}[\text{surprised }_{\text{NP}}[\text{Mary}]]_{\text{S}}[\text{that Bill was frightened}]]
\end{array}\]

This time we have to consider the ratios for both the S and VP CRDs. The lengthy sentential subject in (3.8a) gives the S domain a low IC-to-word ratio of \(2/5 = 40\%\). The VP is \(2/2 = 100\%\), and the aggregate of the two scores is accordingly \(70\%\). The removal of the sentential subject in (3.8b) and its replacement by the dummy subject \(it\) means that there are two ICs of S recognized by two adjacent words, i.e. \(2/2 = 100\%\). The VP now has three ICs recognized in this example by three adjacent words, i.e. \(3/3 = 100\%\). The aggregate for (3.8b) is therefore \(100\%\). The differential between (3.8a) and (3.8b) can be quantified by these aggregates for the two CRDs that are affected by the transformation. The ratios for all the other phrasal nodes in these sentences remain constant (and optimal), and do not need to be considered.

Notice that the VP in (3.8b) would not have an optimal IC-to-word ratio if the NP were a longer phrase such as \(the \text{ clever student}\), rather than the single word \(Mary\). The CRD for VP would then comprise five words and the same three ICs, \(\text{VP}('surprised'_{NP}, ['the'_{NP} 'clever student'] 'that'_{S})\), i.e. \(3/5 = 60\%\). Meanwhile, this extra length would not affect the CRD for VP in (3.8a), since that VP comprises only two ICs, the second of which, NP, will still be recognized on its left periphery (by Det here) regardless of how much NP-dominated material occurs to the right of the category that constructs NP. The S and VP aggregate for \(That \text{ Bill was frightened surprised the clever student}\) will therefore be exactly the same as for (3.8a), \(70\%\). The aggregate for \(It \text{ surprised the clever student that Bill was frightened}\) will now be only \(80\%\). The degree of preference for the extraposed version is much less, and EIC accordingly predicts that Extrapolation should occur less frequently the larger the CRD for VP becomes in the extraposed structure. On the other hand, the longer the sentential subject in
the untransformed structure (3.8a), the worse the ratio for the S domain will be, and the greater can be the potential improvement brought about by Extraposition. The frequency of Extraposition in language use is therefore predicted by EIC to be a function of the degree of difference between the aggregated ratios for S and VP CRDs in the untransformed structure (3.8a), and those in the transformed counterpart (3.8b).

Corresponding to the postposing of the S in English, Japanese preposes it, as we saw in (3.9), repeated here with both matrix S and VP CRDs indicated:

\[(3.9) \quad \text{a. } S_1[\text{NP[Mary-ga]} \text{ VP}[S_2[\text{kinoo John-ga kekkonsi-ta to]} \text{ it-ta]}] \]
\[\text{b. } S_2[S_1[\text{Kinoo John-ga kekkonsi-ta to]} \text{ NP[Mary-ga]} \text{ VP[it-ta]}] \]

The IC-to-word ratio for the matrix S CRD in (3.9a), i.e. $S_1$, is $2/6 = 33.3\%$. The VP is more efficient: two daughter ICs, V and S, are recognized on the basis of two adjacent words, $\text{VP}(\text{to}_S, \text{it-ta}_V)$, i.e. $2/2 = 100\%$. This gives an aggregate of 66.7\%. The preposing of S in (3.9b) shortens the CRD for the matrix S, now $S_2$. On one analysis, $S_2$ consists of three ICs, $S$, the subject NP, and VP, and these three ICs are recognized on the basis of an optimally efficient three words: $s_2(\text{to}_S, \text{Mary-ga}_{NP}, \text{it-ta}_V)$, i.e. $3/3 = 100\%$. The VP on this analysis will consist of just a single word, $\text{VP}(\text{it-ta}_V)$, i.e. $1/1 = 100\%$. The aggregate for the preposed (3.9b) is accordingly 100\%, significantly greater than 66.7\% for the center-embedded (3.9a).

There is another analysis for (3.9b). Japanese is a language with considerable freedom of word order, apart from the position of the verb which is strictly clause-final. It is not obvious, therefore, that the mechanism that orders S in surface structure is any different from that which orders all other constituents, i.e. there may be no structure-changing grammatical rule relating (3.9a) and (3.9b), not even a Scrambling rule or Move-alpha in the sense of Saito (1985), in contrast to Extraposition in English. Rather, I shall argue in section 3.5.2 that linear orderings in free-word-order languages are arranged by a performance module containing EIC, and not by the grammar. The grammar builds constituent structure, and EIC predicts orders in performance based on the IC-to-non-IC ratios of CRDs. The performance module cannot change constituent structure, however, so one of the consequences of this is extensive discontinuous structure in free-word-order languages, and even in many (rearranged) structures of English (cf.
McCawley 1982, 1987 and O'Grady 1987 for further motivation). If this is the correct way to view S preposing in Japanese, then (3.9a) and (3.9b) will have the same constituent structure, and the S of (3.9b) will be a discontinuous daughter of the VP, separated by the subject Mary-ga. This will affect the CRDs for both the matrix S (i.e. S₂) and VP. The first IC of S₂ to be encountered in the parse string is the subject NP, Mary-ga, the second is the VP, constructed by it-ta, i.e. S₂(Mary-gaNP, it-taVP). This gives an IC-to-word ratio of 2/2 = 100%, which is the same score as in the analysis in which S is an IC of S₂, even though there are now fewer ICs. For the VP, the first IC constructed will be S, by the complementizer to as usual. The second IC will be the verb it-ta. Intervening between them will be the non-VP constituent Mary-ga. This constituent adds to the complexity of the CRD for VP, since it increases the distance from the word that constructs the first IC of the domain to the word that constructs the last IC. Even though it is not a daughter of VP, our definition of a CRD in (3.2) quite deliberately does not insist that such intervening material must be dominated by VP. Hence, the IC-to-word ratio for the VP domain in this analysis, VP(toS, Mary-ga, it-taVP), is 2/3 = 66.7%. Aggregating this with the 100% score for the S domain produces 83.4%, which is still significantly higher than the 66.7% aggregate for (3.9a). On both analyses, therefore, the preposing of S is motivated by EIC, though the degree of preference is less in the discontinuous VP analysis. (Cf. further chs. 4.1.5 and 5.4.1.)

One very general prediction that is made by EIC, for all languages, is that any discontinuous structures will be progressively dispreferred in performance in proportion to the amount of non-dominated material that intervenes between ICs. This is because CRDs become progressively longer as a result. I see this as the explanation for the fact that adjacency between ICs is generally preferred, unless this results in unacceptably high IC-to-word ratios within some higher recognition domain, as in the case of a center embedding. The advantages of preposing or postposing the center-embedded constituent for the higher CRD must then be offset against the cost of possibly separating two ICs within a lower CRD, and the larger the distance of the separation, the worse the IC-to-non-IC ratios will be. Some non-adjacency predictions of this sort will be tested in the next chapter (cf. especially ch. 4.1.5 and 4.3.5).

Finally in this section, let us return to the definition of EIC given in (3.14'). The third sentence of this definition asserts that relative preferences among CRDs can also be calculated on a left-to-right basis, and that finer discriminations can be achieved this way. Here is the intuition behind this.
CRDs with high IC-to-non-IC ratios have smaller Structural Domains than those with lower ratios (cf. ch. 2.2.1). Their ICs will be recognized more efficiently and more rapidly, because there are fewer non-ICs to process. But we will sometimes want to discriminate further between CRDs whose ratios are identical or similar, especially when these ratios are non-optimal. Imagine, for example, that there are three ICs of a constituent M; X, Y, and Z; and that some orderings of X, Y, and Z can be recognized on the basis of three adjacent words, whereas other orderings always require more. Imagine further that X is always a single word (e.g. N), and that Y is also a single word (e.g. Adj), whereas Z is a more complex Ŝ, whose average length is four words and which is recognized and constructed on its right periphery by a constituent C, such as a complementizer (cf. chs. 5.2.1 and 6.3.4). Clearly, in such a case orderings such as \( \text{NP}[S[S C] \text{ Adj} N] \) or \( \text{NP}[S[S C] \text{ N Adj}] \) will be optimal because three adjacent words suffice to recognize three ICs. But there are many different non-optimal orderings of X, Y, and Z which require a six-word viewing window, e.g. \( \text{NP}[S[S C] \text{ Adj} N] \), \( \text{NP}[\text{Adj} \ S[S C] \text{ N}] \), \( \text{NP}[\text{Adj} \ N \ S[S C]] \), and \( \text{NP}[\text{N Adj} \ S[S C]] \), all of which have the same IC-to-word ratio of \( 3/6 = 50\% \). Among these options it is within the spirit of our efficiency metric to suggest that linear orderings will be preferred that place the shorter ICs in a leftward position within CRDs, because they then provide the earliest possible temporal access to a significant subset of the daughter ICs of M. If the longer IC is center-embedded, as in \( \text{NP}[N[S[S C] \text{ Adj}]] \), then two of the three ICs will not be recognizable until the last two words of the phrase. Similarly, if the constructing category C is left-peripheral and Ŝ occurs initially in NP, \( \text{NP}[S[C S] \text{ Adj} N] \), there will be the same delay. But if the two single-word ICs are leftmost, then two-thirds of the ICs of M will be recognizable within the first two words of the CRD, making that portion of the CRD highly efficient. Considerations of rapidity and efficiency in processing lead us to expect a leftward skewing in these cases, albeit in structures that are already dispreferred in terms of their IC-to-non-IC ratios. Nonetheless, there is evidence from both performance and grammar to suggest that this skewing is real.

We can make the appropriate discriminations among non-optimal CRDs by calculating our IC-to-non-IC ratios left to right, as defined in (3.15):

\[(3.15) \text{Calculating left-to-right IC-to-non-IC ratios}^{10} \]

The L-to-R IC-to-non-IC ratio for a non-optimal CRD is measured by first counting the ICs in the domain from left to right (starting from 1), and then counting the non-ICs (or words alone) in the
domain (again starting from 1). The first IC is then divided by the total number of non-ICs that it dominates (e.g. 1/2); the second IC is divided by the highest total for the non-ICs that it dominates (e.g. if this IC dominates the third through seventh non-IC in the domain, then 2/7 is the ratio for the second IC); and so on for all subsequent ICs. The ratio for each IC is expressed as a percentage, and these percentages are then aggregated to achieve a score for the whole CRD.

Consider the following three sample CRDs for NP, where $\bar{S}$ dominates four words:

\[(3.16)\]

- a. N Adj $\bar{S}$[S C]
- b. N $\bar{S}$[S C] Adj
- c. $\bar{S}$[C S] Adj N

All of these CRDs have a non-optimal IC-to-word ratio of $3/6 = 50\%$. The calculation procedure of (3.15) discriminates between them, however, ranking them as shown in \((3.16')\) on the basis of their L-to-R IC-to-word ratios:

\[(3.16')\]

- a. N Adj [S C]
  1/1 2/2 3/6
  100% 100% 50%
  Aggregate = 83.3%

- b. N [S C] Adj
  1/1 2/5 3/6
  100% 40% 50%
  Aggregate = 63.3%

- c. [C S] Adj N
  1/4 2/5 3/6
  25% 40% 50%
  Aggregate = 38.3%

### 3.5 EIC predictions for performance

Having defined EIC in \((3.14)\) and \((3.14')\), and having motivated the associated quantification metrics in terms of IC-to-non-IC ratios \((3.13)\) and L-to-R IC-to-non-IC ratios \((3.15)\), I can now formulate EIC's predictions. I begin with the predictions for performance, followed by those for grammar. These predictions will be tested in subsequent chapters.
3.5.1 Predictions for linear ordering

I mentioned in section 3.3 that there are three major types of performance data on which EIC can be tested, and I gave some preliminary support from each: psycholinguistic experiments testing for structural difficulty in language use; relative frequencies of occurrence in text counts; and native speaker acceptability judgments. Following the logic of the formulation of EIC given in (3.14'), I make two interrelated predictions for linear ordering alternations in performance, an unmarked case prediction, and a marked case one:

(3.17) EIC Performance Ordering Prediction

Given: alternative grammatical orderings of ICs within a mother phrase M contained within a sentence, where each IC dominates a set of non-ICs;

Then: for all possible sets of non-ICs, the orderings of ICs that provide the most optimal EIC ratios for the sentence will be preferred in performance over their non-optimal counterparts in the unmarked case; orderings with non-optimal EIC ratios will be more or equally preferred in direct proportion to the magnitude of their ratios.

The ICs of a given M will vary in length and complexity in performance, as different terminal elements are selected for expressive reasons. The NP and the PP within a VP, for example, can each be structurally simple and short, or complex and long. In (3.17) there is an unmarked case prediction for all assignments of non-ICs to ICs: the grammatically permitted orderings that have the most optimal IC-to-non-IC ratios will be the ones that are preferred, in general, in all sources of performance data. Orderings with non-optimal ratios will be preferred in proportion to their ratios. In other words, we expect that non-optimal orderings will be a distinct minority, and that the quantity of these non-optimal cases relative to their optimal counterparts will decline as EIC scores decline.

The rationale behind this fine-tuned prediction is that there is a constant effort in language use to maximize IC-to-non-IC ratios, thereby making syntactic recognition more rapid and more efficient. This explains the unmarked case prediction. The marked case prediction follows because, according to our metric, the speaker should avoid (or dislike) highly dispreferred orderings in favor of their preferred counterparts, to a greater extent than he/she avoids (or dislikes) less dispreferred orderings in relation to their
counterparts. If the ratios for two alternative IC orders are close, there is little basis for choosing between them, and not much is at stake if the slightly less preferred order is selected. We therefore expect non-optimal ordering selections to be more frequent in these cases. The speaker will have less reason for choosing the order that is completely optimal, and one can imagine him/her expending less effort or making more errors on-line when trying to calculate the most advantageous order from among a set of alternatives whose scores are very close. Larger differentials should be more readily perceptible and calculable, however. And if there are any other, non-EIC, reasons for ordering a given set of ICs, EIC itself will be more susceptible to being overridden precisely in the cases where its own degree of preference is slight.

Methodologically, we can proceed to test (3.17) as follows. Pick a given set of ICs for some mother phrase, and divide their performance occurrences into sets of non-ICs for each IC, based on the number of terminal and non-terminal nodes dominated by each IC. For simplicity, let word quantities alone stand for non-IC nodes. For example, pick a VP comprising the three ICs, \( \text{VP}\{V, NP, PP\} \), and let us say that the V always dominates a single word, whereas NP and PP vary. We can then distinguish different subsets of this set of ICs, based on the number of words assigned to NP and PP respectively. There will be one subset where \( NP = 2 \) and \( PP = 3 \) words, another where \( NP = 2 \) \( PP = 4 \), others where \( NP = 2 \) \( PP = 5 \), \( NP = 3 \) \( PP = 4 \), \( NP = 4 \) \( PP = 3 \), and so on. Since this results in very large numbers of subsets, I shall generally collapse them into e.g. all sets in which \( PP > NP \) by one word, by two words, etc., in order to simplify the exposition and tabulation of the data.

We can then test to see whether the unmarked case prediction holds for all these different sets of word-total assignments, by adding up the total number of orderings within text counts that are most optimal for each subset of word-total assignments, and comparing them with the total number of non-optimal orderings. For English, for example, this would mean adding up the total number of \( \text{VP}\{V NP PP\} \) versus \( \text{VP}\{V PP NP\} \) orders for each set of word-total assignments where \( PP > NP \) by one word, by two words, etc., and where \( NP > PP \) by one word, two words, etc. If \( PP > NP \), then \( \text{VP}\{V NP PP\} \) will be the optimal order; for \( NP > PP \), \( \text{VP}\{V PP NP\} \) will be optimal; if \( NP = PP \), then both will be optimal. We expect the overall total for all optimal orders to be significantly greater than the total for all non-optimal orders and to be attested in the unmarked case. Additionally, we can conduct psycholinguistic experiments to see whether the optimal orders are systematically preferred in tests of perfor-
performance difficulty, for all sets of word-total assignments. The optimal orders should always yield the best experimental scores (reaction times, success rates at the task at hand, etc.), when everything else is held constant. We can also make use of acceptability intuitions, though it must be conceded that these intuitions are the least reliable form of performance data, since they involve off-line introspection about an intrinsically on-line activity. The best we can hope for is that the most optimal orders will either be preferred to, or equal to, non-optimal orders in acceptability. At the extremes of an EIC dispreference, intuitions can probably register a judgment of unacceptability. But it is not clear that they can accurately reflect optimal versus somewhat less optimal distinctions, and for this reason a weaker prediction is required for these data.

For the marked case prediction, we expect the EIC orders with non-optimal ratios to be more or equally frequent relative to their preferred counterparts, in proportion to their ratios. The "more or equal" formulation is now motivated by the fact that many non-optimal orderings are not attested at all, and so we cannot expect a non-optimal order with a relatively good ratio to be more frequently attested than one that is worse in the event that the better one already stands at zero frequency of occurrence. For psycholinguistic experiments that test for performance difficulty we predict that non-optimal orders will correlate with experimental scores in proportion to the degree of EIC's preference. Acceptability intuitions should also be more or equally preferred, in proportion to EIC's preferences. Both the unmarked and the marked case predictions for acceptability are of the more-or-equal kind, therefore. For text frequencies and psycholinguistic experiments, however, both of which are spontaneous on-line instances of language use, I predict a clear quantitative distinction between the most optimal EIC orders and all non-optimal ones, and within the non-optimal orders a more or equal prediction is made.\textsuperscript{11}

The particular performance predictions that I shall test in this book involve word-order frequencies in texts. Some additional evidence of relevance to (3.17), involving other forms of performance data and taken from the current research literature, has already been illustrated in section 3.3. The specific predictions made by (3.17) for text frequencies are set out in (3.18):

(3.18) \textit{EIC Text Frequency Prediction}\textsuperscript{12}

Given: alternative grammatical orderings of ICs within a mother phrase M contained with a sentence, where each IC dominates a set of non-ICs;
Then: for all different sets of non-ICs (or words alone), the orderings of ICs that provide the most optimal EIC ratios for the sentence will be the ones that are found in text samples, in the unmarked case; orderings with non-optimal EIC ratios will be more or equally frequent relative to their preferred counterparts, in direct proportion to the magnitude of their ratios.

3.5.2 The performance–grammar interaction

The ordering predictions of (3.17) and (3.18) define performance preferences among IC orderings that are sanctioned by the grammar, i.e. they are all grammatical. The precise manner in which EIC interacts with the grammar is subtle, however, and we need to outline some different possibilities.

In the most extreme case, we can imagine a grammar that generates constituent structures only, i.e. structural relations of motherhood, daughterhood, and sisterhood, but without assigning any linear ordering to the ICs. All logically possible orderings of ICs will therefore be grammatical. It is reasonable to propose that orders are actually assigned in such cases within a performance module that models language use by speakers and hearers and that includes a grammar as one of its components. The grammar delivers unordered structural representations to the performance module, and this latter then arranges items in a linear string, and assigns to each grammatically possible ordering a preference ranking in accordance with EIC. This ranking then predicts the data of language performance, as in (3.18).

Second, the grammar could assign a fixed position to one or more of the ICs within a phrase and deliver ordered structural representations to the performance module. The verb might be fixed in right-peripheral position in the VP, with the other daughters of VP being freely ordered, as in Japanese. Or the verb could be positioned to the left of a direct object NP, which is in turn to the left of a PP, as is commonly assumed for English. For the subsets of grammatically possible IC orderings that are not actually positioned by the grammar in these cases, performance can again arrange and rank the logical options, as it does for phrases whose IC ordering is totally free. In fixed-order languages, grammars will often incorporate rules of rearrangement, especially for ICs of the sentential nodes, S and VP (cf. Extraposition, Particle Movement, etc.), much less commonly for the ICs of the (non-sentential) NP and PP. Most of these rearrangements appear to be motivated by the same kinds of performance advantages that one finds in
languages whose IC orders are free. They transform a subset of initially assigned, or basic, orders, precisely in those cases where EIC ratios would be non-optimal in the basic order but optimal in the rearranged alternative. The difference is that IC ordering is now both arranged and rearranged within the grammar, by rules or principles. Within Chomsky's (1965) Standard Theory of Transformational Grammar, basic orders were assigned by the PS rules generating deep structures, and rearrangements were effected by transformations deriving surface structures from deep structures. The role of a performance module in relation to such a grammar is less extensive than it is when no ordering is assigned by grammatical conventions. The structural representations delivered by the grammar are now ordered, and the role of performance is simply to rank the grammatically generated alternative orders and select the one(s) for the unmarked-case prediction, etc., as a function of their EIC ratios.

A variant of this approach is to have the grammar generate both basic and rearranged orders directly, by the same grammatical mechanism, e.g. by separate serialization rules applying to the outputs of unordered (cross-categorial) PS rules, as advocated in Hawkins (1983: 201–205). Such a separation of serialization from constituency has been proposed in a number of grammatical frameworks, and leads typically to the elimination of a basic/non-basic order distinction. One apparent advantage for these approaches is that it is sometimes difficult to decide which of two competing orders is basic. In this book I shall nonetheless maintain the traditional assumption of a basic word order, for those structures in which there are conventionalized rules of grammatical ordering. There are typically grammatical arguments for basicness, even in difficult cases such as verb-particle constructions (cf. Emonds 1976). The data of language performance provide further arguments, of a type whose significance has not yet been appreciated in the grammatical literature. If [V Part NP] were the basic order in English, we predict that there would never be any grammatical rearrangements to [V NP Part], because the former is already optimal for EIC. If the latter is basic, however, we predict both the existence of two orders in performance and more rearrangements to [V Part NP] the greater the length and complexity of NP. This is exactly what we find: cf. ch. 4.3.1 (I return to a discussion of Particle Movement below.)

According to the first type of performance–grammar interaction, therefore, the performance module arranges and ranks ICs whose relative order is grammatically free. In the second type, performance ranks IC orderings that are grammatically positioned and/or grammatically rearranged. In
both cases the rankings lead to selections among the alternatives, and to predictions for performance frequencies. There is a third possibility. If the performance module can arrange items provided by the grammar of a free-word-order language, there is no reason why it cannot also rearrange items provided to it by a grammar that contains linear ordering conventions. Even in languages with highly fixed word orders, such as English, there are some IC orderings that are demonstrably free. The performance module must do the arranging in these cases. But if it arranges, it can also potentially rearrange and convert a linearly ordered structural representation provided by the grammar into one with a different linear ordering. This raises the question of whether rearrangements are brought about by rules or by performance. The same question arises for basic orders: are they positioned by grammatical rules, or are they simply the most frequent arrangements of free word orders because they have the best EIC ratios?

In ch. 2.1 I argued that Heavy NP Shift is not a coherent grammatical rule of English. The class of NPs that undergo it is not grammatically definable: many do, but single-word names and unstressed pronouns do not. And whether it applies or not depends exclusively on non-IC structure, and on the relative weight and complexity of the NP compared to the PP within VP[VP NP PP]. This is exactly what motivates the preferred performance arrangements of ICs to which no basic order has been assigned to begin with. Since it is not obvious that Heavy NP Shift is a rule of grammar, and since there are at least some IC orderings that need to be arranged by the performance module even in English, considerations of generality suggest that this performance module may also be able to rearrange IC orders within the structures provided by the grammar, even those to which an order has already been assigned. If this reasoning is valid, a number of proposed rearrangement rules may not be rules of grammar at all, but performance readjustments of a grammaticalized basic order; likewise a number of proposed basic order rules, in English and other languages, may simply be the result of performance arrangements and their ranked selections.

The difficulty in establishing the existence of grammatical conventions here results from the basic premise of this book: ordering in the grammar and ordering in performance are ultimately determined by the same principle. For example, the order [NP PP] is significantly more frequent than the reverse in the English VP. This could be taken as evidence for a basic order rule, which would be explained in this context by observing that direct object NPs in English are significantly shorter than PPs: the average size of this NP in my text counts is 2.5 words, compared with 4.6 words for PP; and 71% of PPs are
longer than NP within the VP, while 85% are longer or equal. Hence, the optimal ordering has been grammaticalized. Performance rearrangements would then be predicted to apply within the minority of cases (14%) in which NP is longer than PP. On the other hand, the EIC performance prediction alone would predict significantly more instances of [NP PP] than of [PP NP], given these average length differentials, whether there was grammaticalization or not.

Grammatical arguments must be adduced in such cases to motivate a grammatically positioned basic order. Performance data can also provide evidence, for if we assume a grammaticalization of [NP PP] in the English VP, we explain an interesting performance asymmetry. In the 14% of structures in which NP > PP, only a minority of 30% undergo conversion to the EIC-preferred [V PP NP], predominantly those in which NP > PP by four or more words, i.e. where the differential between these two variable ICs is large. There are no instances of [V PP NP] in my data when NP > PP by just one word, and only two instances when NP > PP by two words (which represents only a 13% conversion rate). The conversion rate for NP > PP by three words is 43%, or less than half; for NP > PP by four words it becomes a majority of 71%; and for NP > PP by five or more words the conversion rate is 100%. We can explain this pattern of results by saying that the (EIC-induced) grammaticalized basic order [NP PP] is systematically retained not only in the 86% of cases where PP > NP that motivated the grammatical response, but also in those cases of NP > PP where the EIC differential is not large and there is little motivation for undoing the basic order. As the differential grows, however, so does the need for rearrangement. By contrast, if there were no grammaticalization of [NP PP], we would not expect this delayed activation of EIC in performance. ICs with one and two-word differentials should, all things being equal, immediately follow EIC's performance predictions. The almost total absence of the preferred word orders in these subsets of word-total assignments is readily explained by grammaticalization, and specifically by a grammaticalization of the EIC-preferred ordering that is motivated for the great majority of NPs and PPs. This is a further example of the partial mismatch between performance and the grammatical response to performance which I discussed in ch. 2.1.

For rearrangements, the existence of a rule of grammar must again be argued for, and the argument is not going to be straightforward, precisely because performance itself would produce a very similar set of data if there were no grammatical rule at all. Performance rearrangements from a grammaticalized basic order, such as Heavy NP Shift, are made because there is a
minority of NPs and PPs whose relative length and complexity goes against the aggregate for the great majority that has led to the grammaticalized basic order. Rearrangement rules exist for the same reason. Extraposition applies to minority instances of certain structural types whose EIC ratios are systematically improved by applying the rule. Subjects, for example, are not typically sentential in English and other languages; when they are, EIC ratios are generally improved by Extraposition.

The argument for a rule of rearrangement must show, first of all, that the rule can be properly formulated within some format for grammatical description. The argument against Heavy NP Shift was made on the grounds that heaviness is not a coherent grammatical category. S, on the other hand, is a coherent category, and the extraposition of this category into right-peripheral position within VP is something that a rule of grammar could legitimately do. A second way to decide the matter involves constituent structure. Rules of rearrangement can potentially change constituent structure, though they might also preserve it, rearranging items in such a way that all mother-daughter relations remain intact. It is reasonable to propose, however, that performance arrangements and rearrangements do not change constituent structure. All the major structural properties of a sentence are determined by the grammar and are delivered to the performance module for linear ordering assignments and/or rankings in accordance with EIC. Hence, if the rearrangement process changes constituent structure, it must be a rule of grammar. Extraposition of S is a rearrangement rule by this criterion, cf. (3.8). Thirdly, if the rearrangement is accompanied by any other grammatical changes in addition to the reordering itself, such as the insertion of a pronoun in the position from which Extraposition has taken place, then this is also evidence for a rule of grammar. Finally, there may be distinctions in grammaticality that are not captured on the pure performance account and that need to be built into a grammaticalized rule. Relevant examples will be discussed below.

The EIC performance-ordering predictions of (3.17) and (3.18) apply to all structures provided by the grammar. Some will have been assigned basic and/or rearranged orders within the grammar, others will not, depending on the language and also on the structure. Predictions for linear order in performance should hold regardless of the status of order in the grammar. If order is grammatically free, then the unmarked and marked case predictions will be defined on all logically possible IC orderings that are grammatically compatible with the structure of the relevant phrase, and on different assignments of non-IC structure and words to these ICs. If order is fixed and/or
Early immediate constituents

grammatically rearranged, then the linear orderings whose EIC ratios are calculated will include grammaticalized ones, and the unmarked and marked-case predictions will be defined on basic orders and their rearranged counterparts. If the rearrangements are effected by grammatical rule, (3.17) and (3.18) will amount to a prediction for the frequency with which the rearrangement rule applies in performance.

The basic ordering \([V \ NP \ PP]\) in English, for example, will be most optimal in all 86% of instances in which \(PP \geq NP\). Even if it were never rearranged in performance (the statistical effect of rearrangement being to make a full 90% of \([V \ \{NP \ PP\}\] structures optimal) the unmarked case prediction would still be fulfilled. Most grammaticalized basic orders are optimal in this way, I claim, but occasionally we encounter some that are not, e.g. \([V \ NP \ Part]\). Emonds (1976) argues that \([V \ NP \ Part]\) is the basic order for these ICs, with Particle Movement deriving \([V \ Part \ NP]\) counterparts. Since both \(V\) and \(Part\) are single words, the transformed structure should be preferred whenever \(NP = 2 +\). 70% of the NPs in my data contain two or more words. Hence, the unmarked case prediction requires that a majority of these structures should undergo rearrangement by Particle Movement, and they do. The unmarked case prediction is now satisfied by a combination of basic and transformed orders, with a full 86% of instances of \([V \ \{NP, \ Part\}\] being optimal in one or the other order in my text counts (cf. ch. 4.3.1).

From the perspective of EIC, the grammaticalization of \([V \ NP \ Part]\) as a basic order of English is anomalous, and it is predicted to be cross-linguistically infrequent and historically transient. It is inefficient and unmotivated to conventionalize a basic order that requires rearrangement in a majority of instances, and it is this kind of discrepancy between grammar and performance that leads, I believe, to grammatical reanalysis and historical change. When performance-driven rearrangements outnumber the retention instances for a basic order, it is more efficient either to abandon a basic order altogether, or to make the transformed structure basic. This is not because I assume that language users are necessarily running through the derivational process in production and comprehension, first generating a basic order and then transforming it, along the lines of the Derivational Theory of Complexity (cf. Prideaux 1984 for a summary). It is simply because I see grammaticalized basic orders as conventionalizations of the most optimal and most frequent orders in performance; and if, because of changes elsewhere in the grammar, a basic order ceases to enjoy this performance optimality, then the grammar no longer has a motive for maintaining it as basic. The changing verb position in the history of English can profitably be viewed
EIC predictions for performance

from this perspective. [V NP Part] is a relic of the earlier, more productive, rule positioning verbs and verbal dependents finally in the VP (cf. ch. 6.3.3). The rearrangement instances now outnumber the retentions, and the performance motivation for the basic order no longer exists (cf. further ch. 5.3.1).

For structures that involve no fixed ordering or reordering in English, such as the relative positioning of two post-verbal PPs, for example, the unmarked case prediction applies directly to unordered tree structures. A total of 89% turn out to be optimal in this case (cf. ch. 4.1.1).

3.5.3 Predictions for preferred IC attachments in temporary ambiguities

There is an important detail in the formulation of (3.17) and (3.18) which I have not yet commented on explicitly. The IC orderings that are predicted in the unmarked case are those that provide the most optimal EIC ratios for the relevant sentence, i.e. for all the phrasal nodes that it dominates. The reasons for formulating the prediction this way, rather than in terms of optimal ratios for a given phrase such as VP or NP, are the following. First, some ordering alternations affect more than one phrasal node, as we saw for Extraposition of a sentential subject in English, cf. (3.8). The only way to assess the relative efficiency of the different orders here is to aggregate the scores for all nodes involved within the sentence, and compare the aggregations. Secondly, by thinking globally in terms of the sentence, we make an interesting and surprising prediction. Sometimes the ratio for one node, e.g. VP, will be improved if the ratio for another contained within it, e.g. NP, becomes worse. Imagine that we separate the ICs of the NP, making them discontinuous, and that we shorten in the process the CRD for the containing VP. It could then happen that the degree of improvement brought about in the VP would more than offset the decline in EIC scores within the NP resulting from the separation. In this case, the benefits for the sentence as a whole would lead us to predict that the NP should be separated, and hence that this particular phrase should preferably occur with less than optimal ratios. Such a prediction cannot be made if we insist that all phrasal nodes have the most optimal orderings all the time. It can be made if we calculate EIC ratios for whole sentences, and there is evidence to suggest that exactly this prediction should be made in cases of this kind (cf. ch. 4.3.5). Thirdly, EIC improvements in any one phrasal category will always be reflected in the aggregations for S, when everything else remains constant. Hence, our performance-ordering predictions in terms of EIC ratios for whole sentences will also cover the simplest cases in which a reordering affects only the material within a single
phrase. (For practical purposes, we can continue to focus just on the local phrases involved in any reordering, without performing EIC calculations on other phrases and ICs whose orderings are held constant.)

We also make another prediction when we think globally in terms of the benefits for a whole sentence. Sometimes the grammar is compatible with alternative possible attachments of one and the same IC to different mother nodes when the IC occurs in a given position, and the result is temporary ambiguity and a possible garden path. Example (3.19) shows the two possible attachment sites for the PP in the library in the English sentence John read the book that he had selected in the library:

\[
\text{(3.19) a. John } \text{VP}_1[\text{read NP}[\text{the book } \text{that he had } \text{VP}_2[\text{selected } \text{PP}[\text{in the library}]]]] \\
\text{b. John } \text{VP}_1[\text{read NP}[\text{the book } \text{that he had } \text{VP}_2[\text{Selected}]] \text{PP}[\text{in the library}]]
\]

In (3.19a) the PP is attached low, to VP₂; in (3.19b) it is attached high, to VP₁. The CRDs for each VP in the two analyses are shown. VP₁ in (3.19a) comprises two ICs, V and NP, recognized in an optimally efficient two words, \( \text{VP}_1(\text{read}_V, \text{the}_\text{NP}) \). VP₂ also comprises two ICs, V and PP, again recognized by just two words, \( \text{VP}_2(\text{selected}_V, \text{in}_{\text{PP}}) \). But in (3.19b) the high attachment of PP means that VP₁ now consists of three ICs and its CRD spans all eight words from read to in, making a highly inefficient IC-to-non-IC ratio. VP₂ is still optimal, since it consists of one IC, V, recognized by one word, selected. Aggregating over both VPs, it is clear that (3.19a) provides a much improved EIC ratio overall. Hence, we predict that VP₂ should be the preferred attachment site in language performance.

This prediction is formulated in (3.20):

\[
\text{(3.20) EIC Performance Attachment Prediction}
\]

Given: alternative grammatical attachments of an IC to different dominating phrases within a sentence;

Then: attachments will be preferred in performance that provide the most optimal EIC ratios for the sentence, in the unmarked case; attachments with non-optimal EIC ratios will be more or equally preferred in direct proportion to the magnitude of their EIC ratios.
We will see in ch. 4.5 that EIC can account for a number of temporary ambiguity preferences and garden paths that have been attributed to Late Closure by Frazier (1979a). I argued in ch. 2.2.2 that Frazier's other principles (Minimal Attachment and the Most Recent Filler) seem to reduce to a preference for less complex Structural Domains, as this term was defined in (2.8) and (2.9). A CRD as defined in this chapter (cf. (3.2)) is simply a type of SD, and a high EIC ratio means low complexity for that CRD and efficient recognition of IC structure within the relevant mother node. Hence, all Frazier's principles involving temporary ambiguity preferences appear to reduce to a preference for less complex SDs, when the parser has a choice.

3.6 EIC predictions for the grammar

I turn now to the predictions made by EIC for conventionalized rules or principles of the grammar, beginning with the predictions for basic orders.

3.6.1 Basic order predictions

When discussing the grammaticalization of \([V \text{ NP PP}]\) as a basic order in English (in section 3.5.2), I pointed out that this order produced an optimal EIC ratio in 86% of the instances of \([V \{NP \text{ PP}\}]\) in my text counts, and that it did so because the PP in this structure is significantly longer than the NP on average, 4.6 words to 2.5 words. Since the verb is typically a single word, the left-to-right ordering of ICs in the English VP is quite systematic in going from shorter to longer items. I want to argue that \([V \text{ NP PP}]\) is typical of grammaticalized basic orders, not just in English, but in all the structures of all languages in which order is assigned by the grammar at all. Grammaticalized basic orders are conventionalizations of the most optimal orders in performance. In a free-word-order language EIC makes the unmarked case prediction that the optimal orderings will be most frequent in performance. If such a language changes into a fixed-order language, it makes sense that the ordering that is fixed by grammatical convention will be the one that occurred in the unmarked case anyway, prior to the acquisition of the ordering convention. This gives us a diachronically plausible account of the acquisition of word-order conventions. If there is a strong correlation between performance and grammar, as I argued in ch. 2.3, then we will expect to see grammars conventionalizing the output predictions of the EIC Performance Ordering Prediction (3.17): optimal orders should be grammaticalized in the unmarked case; and the frequency of any non-optimal orders
across languages should be in proportion to the magnitude of their EIC scores.

This is captured in the following prediction:

(3.21) *EIC Basic Order Prediction*

The basic orders assigned to the ICs of phrasal categories by grammatical rules or principles will be those that have the most optimal ratios for ICs of aggregate length, in the unmarked case; any basic orders whose EIC ratios are not optimal will be more or equally frequent across languages in direct proportion to their EIC ratios.

For example, consider what this predicts for a PP embedded within a VP, $\text{VP}\{V_{PP}\{P, NP\}\}$. There are four logically possible orderings of these categories, shown in (3.22):

(3.22)  

a. $\text{VP}\{V_{PP}\{P, NP\}\}$  

b. $\text{VP}\{PP\{NP, P\} V\}$  

c. $\text{VP}\{V_{PP}\{NP, P\}\}$  

d. $\text{VP}\{PP\{P, NP\} V\}$  

The (a) and (b) versions are both optimal for VP recognition, with CRDs as shown. The verb is the first IC in (a) and P occurs immediately to its right, where it can construct PP, the second IC, immediately after recognition of the first. In (b) the verb and P are again adjacent, though this time the verb is final in VP, and the P is a postposition. A bottom-up parse takes place in these languages, and the PP is not constructed until P is encountered, whereupon the V is recognized immediately after that. If we think in terms of IC-to-word ratios, (a) and (b) will have an optimal ratio of $2/2 = 100\%$. On the other hand, (c) and (d) have much longer CRDs for VP, spanning all the categories dominated by the VP, and as a result they both have non-optimal ratios. Accordingly, (3.21) predicts that the optimal orders will be those that are found grammaticalized in the world’s languages in the unmarked case: languages with verb-initial order in the VP (i.e. VO languages) should co-occur with prepositions; verb-final (or OV) languages should have postpositions; and the frequency of these two optimal types combined should far exceed the combined frequency of the marked and non-optimal orders, (c) and (d). We will see that this prediction is fulfilled (ch. 5.1.2). Moreover, we will also define a slight preference for (c) over (d) when we calculate the IC-to-non-IC ratios for these non-optimal CRDs on a left-to-right basis, in accordance with (3.15). The leftward positioning of the single-word verb
as opposed to the multi-word PP improves the ratio for VP, and we shall see that in this, as in other comparable cases, there is a consistent skewing towards the (c) structure. The skewing is slight because both (c) and (d) are already predicted to be minority types by the unmarked case prediction, but it seems to be empirically real.

The prediction of (3.21) refers to ICs of "aggregate length," as a way of assessing which orderings of a given set of ICs will be optimal according to EIC. The greater aggregate length of PP compared with NP in the VP leads to the prediction for [V NP PP] ordering in English. Similarly, an Š has greater aggregate length than a VP or an NP, since it can and typically does dominate more non-terminal and terminal nodes. If Š occurs center-embedded within a containing S, this S will have a much worse EIC ratio than if VP or NP are center-embedded. We accordingly predict that Š will be the least frequent center-embedded category in performance (if it is permitted at all, cf. ch. 2.1), and correspondingly that it will be the least sanctioned center-embedding category across grammars. Aggregates enable us to calculate and predict optimal orderings, and to rank different orderings according to their degree of optimality.

But how do we know what the aggregate lengths of different ICs are in different languages? The answer is that we have to examine large numbers of them in performance samples such as texts, and calculate them. But since different languages could conceivably have different aggregates for common ICs, this would appear to pose a problem for cross-language predictions of grammaticalization. Many of these will be tested on hundreds of grammars, yet for only a handful of these do we have extensive text counts (cf. ch. 4).

It does, in fact, pose a problem, but not an insuperable one. For binary branching structures such as V_P{V_P{P NP}}, we can assume (fairly safely in most cases) that V and P each contain one word, and we do not need to know the precise aggregate for NP in order to assert that only two orders are optimal within the VP recognition domain, namely (3.22a) and (b) above. V will construct VP, and P will construct PP, and whatever the length of NP is, it will delay access to the PP in (3.22c) and to the V in (3.22d). Any word content at all within NP therefore produces non-optimal VP orders in these latter cases, and similar considerations apply to all binary branching structures.

The aggregates become more critical for multiple branching structures, though even here many predictions do not depend on precise figures. Consider the following three daughter ICs of NP: _NP{N Adj Š{C S}}_,
Early immediate constituents

where C is a category such as a complementizer that constructs S. Assume that N, Adj, and C are all single-word items. Then only four of the 12 logically possible arrangements of these categories can have optimal EIC scores: \( NP[\text{N Adj} \, S]\), \( NP[\text{Adj N} \, S]\), \( NP[\, S \, C, \text{Adj} \, N]\), and \( NP[\, S \, C \, \text{N Adj}]\). These are, in fact, exactly the four orders that occur productively across languages, cf. ch. 5.2.1. The prediction is possible in this case because two of the three ICs of NP are single-words items, N and Adj, and they must be adjacent to C within S to achieve 100% scores, making it irrelevant what the aggregate for S is.

What I shall do in this book is concentrate on predictions that are not rendered problematic by any cross-linguistic differences in aggregate lengths for common categories. It is, of course, the relative differences between aggregates that are important, not the absolute scores. The aggregate for a sentential node, S or VP, is always going to be greater than that for a referential node such as NP, since the former properly includes the latter. Relative positioning predictions can be made with confidence for these cases as well, therefore, even when absolute aggregates are known to differ from language to language.

When calculating the predictions made by (3.21) for a given set of ICs, I shall assign non-ICs (or words alone) to these ICs in accordance with the following procedure. First assign to each IC the most minimal set of (overt) non-ICs compatible with the grammar. For example, for \{V, NP, PP\} in English, the V must consists of a single terminal element or word, the NP must comprise at least the category noun and a terminal element, and the PP must contain a P and an NP, which must minimally dominate a noun and a terminal element again. These minimal assignments will suffice in this case to predict the highest IC-to-non-IC ratio for the ordering [V NP PP] within the VP recognition domain (namely 3/5 = 60%, compared with 3/8 = 38% for [V PP NP]). If words alone are used in place of non-ICs, then the minimum word content of V will be one, it will also be one for NP, and two for PP. These minimum assignments will then need to be adjusted to reflect the fact that NP can consist of more than one word and is typically longer than V. Hence, add one to the direct object NP and to the NP dominated by PP. This gives the following adjusted minimum word assignments: V = 1; NP = 2; PP = 3. With these assignments, the ordering [V NP PP] has the highest IC-to-word ratio of 3/4 = 75%; [V PP NP] has 3/5 = 60% (assuming the NP is still constructed on its left periphery, by a Det or N).

These minimum assignments of non-IC structure or adjusted minimum word assignments to ICs seem to accurately reflect the relative differences
between ICs in structural composition and word length in performance, in those languages for which I have extensive performance evidence. They do not predict the absolute word-length aggregates. For V, NP, and PP in English, these are 1, 2.5, and 4.6 respectively. But this is not necessary either. The prediction for grammars is based on the relative complexity and length differences among ICs, and the use of minimally required non-IC nodes and adjusted word minimum appears to be capable of capturing this relativity and so provides us with a simple mathematical procedure for setting up basic order predictions.

3.6.2 Grammatical rearrangement predictions

The Basic Order Prediction (3.21) asserts that grammars generally conventionalize linear orders that are most optimal for EIC in most performance instances in which the relevant ICs occur. In a binary branching structure such as \( vP \{ V \ PP \{ P \ NP \} \} \), the optimal orders (3.22a) and (b) can each be optimal 100% of the time in performance, and there will be no instances in which the (c) or (d) versions could be preferred. This is because V and P are each single-word items, and their adjacency is always preferred. But for any structure containing two or more ICs of variable complexity and length, there will typically be a minority of instances in which the basic and normally preferred ordering is, in fact, dispreferred. \([V \ PP \ NP]\) will be preferred by EIC in 14% of instances (all those in which NP > PP) over the basic order \([V \ NP \ PP]\) in English performance.

The logic of our performance theory of grammar incorporates an inherent prediction for these cases. If basic orders are indeed grammatical responses to EIC, it follows that those subsets of performance instances for which a basic order is not optimal should also respond to EIC, and they should do so in direct proportion to the degree of EIC dispreference for the basic order. We accordingly predict rearrangements from the basic order, in order to reinstate the EIC advantages that led to a basic order in the first place. These rearrangements may be effected by a grammatical rule, such as Extraposition; alternatively they may be pure performance rearrangements, as we suggested for Heavy NP Shift in section 3.5.2.

If there is a rule of rearrangement, it can be predicted to have certain grammatical properties, namely those that will guarantee that the rearrangement does regularly improve EIC ratios for the relevant sentence. Some testable predictions are formulated in (3.23):
(3.23) **EIC Grammatical Rearrangement Prediction**

If the grammar encodes a rearrangement rule in response to EIC, this rule will apply only to those grammatical categories and only in that (leftward or rightward) direction that will potentially increase the EIC ratio for a sentence; if there is a rule for rearranging a category with a non-optimal but relatively good EIC ratio, there will be a rule (possibly the same rule) for all categories with a worse ratio, and any conditions of application on the latter (e.g. obligatoriness) will result in more or equal applications; no such rule will apply to those (sub)categories or in a direction that necessarily results in no EIC increase in every application.

The following predictions are made in (3.23) for extraposition-type rules in different languages. First of all, the category or categories to which such rules apply, and the structures in which they are contained in their basic orders, must be such that there will be regular EIC increases resulting from the rearrangement. \( \tilde{S} \) is a category that guarantees such improvements on a regular basis; hence this is a category to which a grammaticalized rearrangement rule can be expected to apply quite generally across languages. Secondly, if extraposition applies to an embedded VP (abbreviated here \( \tilde{VP} \)), which is structurally less complex than an \( \tilde{S} \), it will also apply to \( \tilde{S} \), but the converse may fail, and the conditions of application on \( \tilde{S} \) extraposition will be stronger or equal. With regard to the directionality of movement, we make a third prediction. If the \( \tilde{S} \) is constructed by a left-peripheral complementizer, i.e. \( \text{§}\{\text{Comp} \ S\} \), it will be reordered to (or will remain to) the right of its sister ICs; if it is constructed by a right-peripheral complementizer, \( \text{§}\{\text{S Comp}\} \), it will preferably be reordered to (or will remain to) the left.

(3.23) predicts regular rearrangements not just for heavy categories such as \( \tilde{S} \), but also for very light categories such as single-word particles or pronouns. The rule of Particle Movement in English, converting \([\text{V NP Part}] \) into \([\text{V Part NP}] \), fronts a single-word particle next to the single-word verb in the VP. It results in improved EIC ratios whenever \( \text{NP} \geq 2 \) words in length. It is no accident, as I see it, that the rule blocks when the NP is a pronoun, since the EIC ratio for the basic order will always be optimal, cf. \( \text{I looked it up} \neq *\text{I looked up it} \), and no application of Particle Movement would result in any EIC increase. Similar limitations involving pronouns pervade the grammar of English, cf. \( \text{I } \text{vp[gave it to my wife]} \neq *\text{I } \text{vp[gave my wife it]} \) (Dative Movement), where there would be an EIC decrease from
a three-word to a four-word VP recognition domain; and *down the hill S[VP[ran]] ≠ down the hill S[VP[ran] NP[he]], where there would be no EIC increase in the S domain, in contrast to down the hill S[VP[ran]] ⇒ down the hill S[VP[ran] NP[the little boy]], where there would be.

Such limitations on the applicability of the relevant rules are predicted by the requirement that EIC-motivated rearrangements will not apply to those categories (or grammatically specifiable subcategories such as pronouns) that necessarily result in no EIC increase in every application of the rule. And this requirement is motivated in turn by the fact that if a rearrangement rule owes its raison d'être to EIC, it cannot be expected to apply in those grammatically definable cases where there is never an EIC increase. Notice how this automatically discriminates between pronouns and another subcategory of NP that regularly dominates only a single word, namely Name. Rearrangements are possible with names, cf. e.g. look John up ⇒ look up John, and down the hill John ran ⇒ down the hill ran John. But names are not confined to single words, in contrast to pronouns (cf. John Smith, John Fitzgerald Kennedy), and so do not necessarily result in no EIC increase in every application of these rules. This kind of grammatical fine-tuning is what we expect when a grammatical rule is conventionalized in response to a performance need for rearrangement.

### 3.6.3 Hierarchy predictions

The Basic Order Prediction of (3.21) predicts optimal EIC ratios for grammaticalized basic orders in the unmarked case. In the marked case, non-optimal grammaticalizations are predicted with a frequency that is directly proportional to their EIC ratios. We can strengthen this prediction to an implicational one, following the logic of the complexity predictions for cross-linguistic variation in ch. 2.3. A grammar should only resort to a non-optimal basic order with a very low EIC ratio for a given CRD if it already tolerates low scores for that CRD in general. More precisely, a very low ratio for a given set of ICs within this domain implies that the grammar will tolerate all relevant intermediate degrees of complexity for other ordered ICs. For example, a center-embedded S will generally give its containing node a worse ratio than a center-embedded VP, because of the greater length and complexity of the former. Hence, any grammar that tolerates the former should tolerate the latter, though not necessarily vice versa.

CRDs with very low scores are often those in which there is/are some center-embedded constituent(s) separating the node that constructs
the first IC from the one that constructs the last. This leads us to predict a hierarchy of center-embedded categories, with different languages cutting off down the hierarchy and either preventing the generation of the relevant center embeddings within the structure-building component (e.g. the PS rules), or else generating and obligatorily transforming them. The net result will be the same: an implicational hierarchy of non-optimal basic orders of increasing complexity. One way to formulate this prediction is as follows:

(3.24) **EIC Grammatical Hierarchy Prediction**
Given: an IC position P center-embedded in a CRD for a phrasal node D (i.e. there is non-null material to both left and right of P within D);

a set of alternative categories \{C\} that could occur in position P according to the independently motivated and unordered PS-rules of the grammar;

Then: if a category \(C_i\) from \{C\} produces low EIC ratios for D and is grammatical, then all the other categories from \{C\} that produce improved ratios will also be grammatical.

3.7 **EIC's predictions for typologically different structures**

The predictions made by EIC for performance and grammar depend crucially on the structure of the ICs upon which they are defined. Of paramount importance is the question: when is the IC constructed and recognized in the on-line parse? An S in English is constructed by a left-peripheral Comp, i.e. \([\text{Comp} \ S]\); VP is constructed by a left-peripheral V; and NP is constructed by a left-peripheral Det or N, in advance of a phrasal (PP) genitive or relative clause modifier. In Japanese, these ICs are constructed on their right peripheries: by a right-peripheral Comp in \(\tilde{S}\); by a right-peripheral V in the VP; and by a Det or N that follows phrasal genitives and relative clauses within NP. The CRDs for mother nodes will comprise very different sets of terminal and non-terminal dominated nodes, depending on where the first and last IC are constructed, and IC-to-non-IC ratios will vary accordingly. The result will be different predictions for linear ordering in different language types. English-style \(\tilde{S}\) structures will preferably be postposed within their mother phrases; Japanese-style \(\tilde{S}\)s will preferably be preposed; etc.
It is useful to summarize in tabular form what EIC's predictions are for the major structural types that I shall examine in this work. The predictions for binary branching structures have already been discussed in section 3.6.1; cf. example (3.22) illustrating the different orderings of $v_p{\{V \ p_p\{P \ N_P}\}}$. For multiple branching structures I shall generally focus on phrases in which there are just two variably positioned ICs, with the category that constructs the mother node (abbreviated here as $C$) held constant. This renders both the empirical search and the mathematics more manageable. We want there to be significant numbers of performance instances matching the structural templates in question within our text samples, for example, and there are not many instances of phrases with more ICs than this in language use. The mathematically possible linear orderings also increase exponentially with the addition of each new variably positioned IC: for two variable ICs there are two logically possible orders; for three there are six; for four the total jumps to 24. As the mathematics increases, textual frequency declines, however, making it necessary to survey an impractically large amount of text in order to gain a significant $n$ for each logically possible ordering.

Table 3.2 summarizes the predictions. $C$ stands for the category that constructs the mother node within the relevant containing phrase. $m$IC represents a sister IC to $C$ which is constructed and recognized on its left periphery, such as a prepositional phrase. IC$_m$ stands for a sister IC to $C$ constructed and recognized on its right periphery, like a postpositional phrase. The "$m$" subscript accordingly indicates the position in which ICs are constructed. If the IC dominates just a single word, of course, then that word will have to construct the IC, and the distinction between $m$IC and IC$_m$ is neutralized.

The predictions for binary branching structures are summarized in Table 3.2(1). C, the category that constructs its mother, is preferably adjacent to the category that constructs the single sister IC, as in table 3.2(1 a) and (b). Within the dispreferred orders, (lc) is preferred to (d) by IC-to-word ratios calculated left-to-right, cf. (3.15).

Three major types of multiple branching structures are distinguished: in table 3.2(2) the two variably positioned ICs are both of the type $m$IC; in table 3.2(3) they are both IC$_m$; and in table 3.2(4) there is one of each. In (2a), C is positioned leftmost (for example a leftmost verb within VP), and (2a) asserts that EIC prefers a shorter $m$IC before a longer $m$IC, in direct proportion to the word-total difference. By positioning the longer $m$IC to the right, the CRD for the containing node becomes smaller. It will contain all the nodes dominated by the shorter $m$IC, as well as the constructing category for the longer $m$IC, but all the other nodes dominated by this latter are outside the
Early immediate constituents

CRD. In the reverse ordering, all the nodes dominated by the longer \( m \text{IC} \) are in the CRD, the shorter \( m \text{IC} \) is rightmost, and the IC-to-word ratio is lower. When two \( m \text{IC} \)s are of equal length, either one can be designated by \( \text{IC}_1 \) or \( \text{IC}_2 \) in this template, and so either order is predicted; longer versus shorter ICs have their values fixed as \( \text{IC}_2 \) and \( \text{IC}_1 \) respectively.

Table 3.2(2b) is a structure that is actually dispreferred by EIC, since it has \( C \) constructing its mother on a right periphery at the same time as its sister ICs are constructed on their left peripheries (but cf. ch. 6.3.3). This is a marked and infrequent ordering cross-linguistically, exemplified by the VP

Table 3.2  

<table>
<thead>
<tr>
<th>EIC's predictions for typologically different structures</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Binary branching structures</strong></td>
</tr>
<tr>
<td>(1) a. ([C_m \text{IC}])</td>
</tr>
<tr>
<td>c. ([C \text{IC}_m])</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Multiple branching structures</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2) a. ([C {m \text{IC}_1 \text{IC}_2}]) if ( \text{IC}_2 \geq \text{IC}_1 ), then ([C_m \text{IC}_1 \text{IC}_2]) is preferred in direct proportion to the word total difference (i.e. short before long);</td>
</tr>
<tr>
<td>b. ([{m \text{IC}_1 \text{IC}_2} C]) if ( \text{IC}_2 \geq \text{IC}_1 ), then ([m \text{IC}_1 \text{IC}_2 C]) is slightly preferred by IC-to-word ratios calculated L-to-R, in proportion to the word total difference (i.e. short before long);</td>
</tr>
<tr>
<td>(3) ([{m \text{IC}_2 \text{IC}_m} C]) if ( \text{IC}_2 \geq \text{IC}_1 ), then ([2 \text{IC}_m \text{IC}_1 C]) is preferred in direct proportion to the word total difference (i.e. long before short);</td>
</tr>
<tr>
<td>(4) a. ([{m \text{IC}_m \text{IC}} C]) for any length of ICs, ([\text{IC}_m \text{IC} C]) is strongly preferred (short, equal or long ( \text{IC}_m ) before long, equal or short ( m \text{IC} ));</td>
</tr>
<tr>
<td>b. ([C {m \text{IC}_m \text{IC}}]) if ( \text{IC}_m \geq \text{IC}_m ), then ([C \text{IC}_m \text{IC}_m]) is strongly preferred in direct proportion to the word-total difference (i.e. short or equal ( \text{IC}_m ) before long or equal ( m \text{IC} ));</td>
</tr>
<tr>
<td>if ( \text{IC}_m &gt; m \text{IC} ), then ([C_m \text{IC} \text{IC}_m]) is gradually preferred by IC-to-word ratios calculated L-to-R, in proportion to the word-total difference (i.e. short ( m \text{IC} ) before long ( IC_m )).</td>
</tr>
</tbody>
</table>
in German where V stands to the right and is preceded by left-constructed ICs such as NPs and PPs. Within this dispreference there is still a preference for the relative ordering of the _mICs: short before long is slightly preferred by IC-to-word ratios calculated left to right. Similarly, notice that if yet another IC occurs to the right of _mIC_1 and _mIC_2 in table 3.2(2a), then this IC will be the final IC of the domain, and both _mIC_1 and _mIC_2 will be entirely within the domain. The short-before-long preference will again be defined by the left-to-right calculation procedure.

For consistently right-peripheral ICs, as in table 3.2(3), the preference defined by EIC is the mirror-image of that in (2), long before short. By preposing a long IC_m, as in Japanese, most of the nodes dominated by this latter are outside the higher CRD, since this node is parsed bottom-up and is not constructed and recognized until the right-peripheral constructing category is reached. If the shorter IC_m precedes, all the nodes dominated by the longer IC_m are within the CRD, and the IC-to-non-IC ratio becomes worse. Again, IC_m's of equal length are predicted to occur in either order.

The predictions are more complex for languages that combine both left- and right-mother-node-constructing categories, i.e. mixed languages such as Hungarian that have both V-initial and V-final VPs, postpositions within PP, and NPs that can be constructed by a left-peripheral Det or N (i.e. mNP) as well as NPs with right-peripheral construction (NP_m), for example when dominating a prenominal genitive phrase. In table 3.2(4) I set out the predictions for two variably positioned ICs in such a language, one of which is IC_m and the other _mIC_. Table 3.2(4a) positions C to the right. The prediction here is very consistent. IC_m will always be preferred in leftmost position whatever its length and whatever the length of _mIC_. The reason is the same as the one we gave for the preference for long initial IC_m's in Japanese: because IC_m is parsed bottom-up, most of the nodes that it dominates can be excluded from the relevant CRD if it is the first IC within this domain; but the motivation for leftward positioning is now all the greater because the other IC is left-mother-node-constructing, and if it stands to the left, every single node dominated by both _mIC and IC_m will be within the CRD. IC_m before _mIC will always be the preferred order, therefore, for any non-IC or word assignments to these ICs.

Table 3.2(4b) defines the predictions for a C that stands to the left of IC_m and _mIC, for example when the verb in Hungarian precedes PP_m and a left-recognized mNP. Since the domain is now initiated by the verb, the predictions revert essentially to the short-before-long preference characteristic of English-type languages. If _mIC ≥ IC_m, then [IC_m _mIC] is strongly preferred,
with the longer or equal \( m_{IC} \) to the right. \( m_{IC} \) is now the last IC in the domain and all nodes that it dominates are outside the higher CRD apart from those that construct it. If \( IC_{m} > m_{IC} \), it turns out, interestingly, that \([m_{IC} \ IC_{m}]\) is gradually preferred in proportion to the word-total difference, again with the shorter \( m_{IC} \) before the longer \( IC_{m} \). This is predicted by IC-to-word ratios calculated on a left-to-right basis.

3.8 The relationship of EIC to other principles

Having now presented the EIC theory and its predictions for performance and grammar, we need to consider other principles of performance and grammar that are currently being pursued in the research literature on word order. How does EIC relate to, and interact with, these other proposals?

Notice first that EIC is a theory of syntactic processing. It claims that what drives the ordering of elements is the need to recognize abstract syntactic structure in a rapid and efficient manner on-line. Syntactic recognition is a necessary precondition for the recognition of semantic structure and for the assignment of grammatical meaning to a sentence, i.e. truth-conditional meaning and aspects of pragmatics that are grammatically conventionalized, such as conventional implicatures (cf. Grice 1975 and Levinson 1983). These conventionalized meanings then serve as input to the inferential processes of discourse that result in different kinds of conversational implicatures (cf. Grice 1975; Levinson 1983; Sperber and Wilson 1986). Hence, EIC-preferred orders are those that facilitate the recognition of phrase structure and the interpretation of this structure in a rapid and efficient manner. If EIC is a fundamental principle of performance, and I believe it is, it is because structure recognition is a vital prerequisite for semantic and pragmatic processes, and hence for successful and efficient communication.

Notice, secondly, that EIC is a unified theory of performance and grammar. It predicts ordering in performance for structures and languages whose order is grammatically free; and it makes predictions for the ordering conventions in grammars that have them, and derivatively for their reflexes in performance. It claims that there is a strong correlation between performance and grammar, and derivatively between order in free-word-order languages and order in fixed-word-order languages.

This is not the current view of word order that one finds in the linguistics literature. Free-word-order languages are generally considered to be regulated by a very different set of performance principles from this, involving pragmatic information status and the “givenness,” “newness,”
"predictability," and "importance" of entities that are referred to in discourse. For fixed-order languages, different grammatical models have their own formal proposals for generating some constituent orders and not others, none of which makes reference to any ultimate explanation in terms of syntactic processing.

3.8.1 Principles of generative grammar

In the model of Chomsky (1965), order is assigned by the same set of rules that build constituent structure, PS rules, and is then possibly rearranged by transformations. In Gazdar et al. (1985), linear precedence rules are kept distinct from immediate domination or structure-building rules, just as the serialization rules of Hawkins (1983: 201–205) were kept distinct from the PS rules. Within Government-Binding theory (cf. Chomsky 1981, 1982), a number of parameters are envisaged for cross-linguistic variation, and the parameters that have been proposed for word order (in Travis 1984, 1989 and Koopman 1984) are: head ordering; directionality of case assignment; and directionality of θ-role assignment.

There is nothing irreconcilable between EIC and these kinds of formal proposals. Quite the contrary. I see EIC as the explanation for why these rules and principles actually generate the orders they do, in different language types. These generative mechanisms are primarily descriptive: they describe the ordering data of given languages by generating only the observed orders. But it is only in a limited sense that they can be said to provide an explanation for the orders they generate. The Standard Theory approach of Chomsky (1965) was the most stipulative of all: just about any ordering conventions could have been written into the PS rules as they were then conceived, without offending independent principles or rendering the grammar less highly valued. Hence, the implicit claim was made that all logically possible ordering conventions were equally good, and could be equally expected across languages.

The situation changed with the development of X-bar Theory. Since the PS rules still defined both constituency and linear ordering in Jackendoff (1977) and Lightfoot (1979), the formulation of cross-categorial generalizations for constituency led to a more general set of statements for linear ordering. Grammars were now more highly valued if phrases at comparable bar levels had not only comparable daughter categories, but similarly ordered daughter categories as well (cf. Hawkins 1983: 183–205 for discussion). Thus, the order [V NP] within the VP was predicted to co-occur with [P NP] in the PP. A
mixture of [NP V] and [P NP] destroys cross-category parallelism, not only for ordering but for the very definition of phrase structure within a model in which PS rules define both constituency and linear ordering. For this reason the proposal was made in Hawkins (1983) to separate cross-categorial PS rules from cross-categorial serialization rules, and it was predicted that the most frequently occurring language types would be those with the most general and least complex serialization rules.¹⁴

This approach can constrain a number of word-order combinations throughout a grammar and make predictions for their exemplification across languages. It leaves a large number of possibilities unconstrained, however, for which EIC provides a more restrictive theory and explanation. Cross-categorial rules are more highly valued if complements or restrictive modifiers of the head occur in the same order across phrasal categories. But there are many mathematically possible orderings of these dependents of the head relative to one another that would be equally good in terms of cross-category generalizations and rule simplification, yet most of them are unattested or highly infrequent. For example, within the set of possible complements of X’, Jackendoff (1977) orders an NP before a PP before an Š, to produce orders such as v[V NP PP Š], n'[N PP Š] and p'[P NP PP] in English. But his rules would be just as highly valued with any of the six logically possible orders of these three complements, as long as the same ordering was observed within each phrase, i.e. [X PP NP Š], [X Š NP PP], [X NP Š PP], etc. EIC, by contrast, can motivate the choice of [X NP PP Š] in languages such as English (cf. ch. 5.3.3). Similarly, there are many mathematically possible orderings of restrictive modifiers relative to a head, whose positioning is not constrained by cross-categorial rule schemas beyond the requirement that the selection hold across phrasal categories. Nothing will predict that [N Adj Š] and [Adj N Š] are highly productive orders in a head-initial language, whereas [N Š Adj] is completely unattested; or that [Š Adj N] and [Š N Adj] are highly productive in head-final languages, and [Adj Š N] highly infrequent. These facts are predicted by EIC, however (cf. ch. 5.2.1). Cross-categorial rules seem to be most successful when defining the positioning of heads in relation to non-heads, since head positioning has been claimed to be largely consistent across categories (cf. ch. 6.1). But even this generalization fails in a number of interesting cases (cf. Dryer 1992, Hawkins 1993). Overall, linear order can certainly be described by cross-categorial rule mechanisms, especially those that separate the task of serialization from the definition of phrase structure, but the unrestrictiveness of this approach means that many linear orders are not predicted or explained.
Even when a prediction is made, as with [V NP] and [P NP], it is not obvious that cross-categorial rule simplification can qualify as an explanation rather than a description, given the failure of this explanation to apply to many other cases. That is to say, the correct prediction here is correct only accidentally.

Similar considerations apply to more recent work in the principles and parameters approach (cf. Koopman 1984; Travis 1984, 1989). The parameters of headedness and of directionality of case and θ-role assignment are again cross-categorial in nature. By allowing each of these independent grammatical primitives, the head of phrase generalization, case assignment and θ-role assignment, to be intrinsically directional, this approach attempts, in effect, to relativize the set of cross-categorial ordering parallelisms to those that are definable in terms of one or more of these independent grammatical properties. Thus, Japanese is head-final and assigns both case and θ-roles to the left. English is head-initial and is claimed to assign case and θ-roles consistently rightwards, though the positioning of the subject provides *prima facie* counterevidence to this. A language like Chinese, on the other hand, is head-final and assigns case to the left, but θ-roles to the right, while Kpelle is head-initial and assigns case to the left and θ-roles to the right. German is even more of a mixture, assigning case and θ-roles to the left in some categories (the VP), to the right in others (the PP), and being head-final in the former and head-initial in the latter. This approach seems to be primarily motivated by the need to account for mixed languages of these types, and is claimed to be successful in relation to them.

The success is purely descriptive, however. Languages of all these types clearly exist, and grammars need to be written for them, possibly in these terms. But this approach still fails to constrain many of the relative complement orders discussed in connection with Jackendoff (1977), which would be equally case- and/or θ-marked, or the different relative orderings of restrictive modifiers, which would be neither. Moreover, there is nothing explanatory about these principles. Case and θ-roles are assigned under tightly specified structural conditions and serve, in effect, to define the relation of co-constituency between assigner and assignee. These processes are not intrinsically directional, and adding a dimension of direction to what is primarily a structural relation is tantamount to having PS rules define both linear precedence and immediate domination. This may (or may not) correctly describe the ordering facts of a given language or group of languages, but it doesn’t follow from anything, and there is no independent reason why the grammar should be set up this way.
A related problem of explanation, but not of description, concerns the fact that while the parameters approach does exclude certain combinations of parameter settings, all of the combinations exemplified by Japanese, English, Chinese, Kpelle, and German appear to be equally possible. Yet the first two language types are highly productive and frequent, the last three highly unproductive and infrequent. To allow all combinations is descriptively defensible, since all five types exist. But at an explanatory level, productivity and frequency differences do not appear to follow from these principles; nor do numerous additional ordering constraints; nor is there any reason why structure-defining operations should be directional in the first place.

By contrast, EIC does make the appropriate discriminations between productive and non-productive or non-existent language types; it does account for these additional constraints; and there is a reason why it makes the predictions that it does, having to do with the size and complexity of alternative constituent recognition domains, the preferences among which follow in turn from a general theory of structural complexity.

From the perspective adopted here, therefore, EIC is a possible explanation for the descriptive rules or principles of particular grammars. It is important to establish the most adequate descriptive statements we can in relation to the data of individual languages. But it is becoming increasingly clear that there are many notationally equivalent or near-equivalent ways of doing this, most of which are stipulations about observed word orders that do not follow from more fundamental properties of the grammar. The reason for this is that order is not ultimately a grammatical phenomenon, it is explained by performance. The conventionalized orders of particular grammars are the result of general processing mechanisms that are innate, in conjunction with general efficiency and complexity considerations applied to the structural representations that they produce in language use.

This view has two immediate consequences. First, linear order is not regulated by an innate UG. There are no innate and parameterized universals of directionality for case and θ-role assignment. The innate grammar quite plausibly contains general principles of linguistic structure, but it makes no reference to order. Instead, performance, in the form of EIC, motivates the orderings that become conventionalized in particular languages. Secondly, this explains the arbitrary and tangential connection between grammatical principles of order and principles of structure. Since order is not primarily determined by fundamental grammatical principles and yet is clearly a constrained output of many grammars, it has to be added to
The relationship of EIC to other principles

some component of the grammar in some way. How one chooses to do this may be as much a consequence of the individual linguist's theoretical taste as it is of any fundamental reality. It is also likely that different particular grammars will lend themselves more or less readily to the different formal proposals of various models, some being thoroughly compatible with Jackendoff-style cross-categorial rules, others requiring separate specifications for directionality of case assignment versus θ-role assignment, and so on. Many of these generative mechanisms may be descriptively useful, therefore, but I would contend that none of them are providing us with an answer to the why question, because linear order is not ultimately a grammatical phenomenon. See ch. 7.2 for further discussion.

3.8.2 Pragmatic principles

The prevailing view in the literature is that the selection among truth-conditionally equivalent word orders permitted by the grammar in languages with considerable word-order freedom is determined primarily by considerations of information structure (cf. e.g. Givon 1983, 1988; Thompson 1978; Mithun 1992; Payne 1992). The theory proposed here casts doubt on this view. I believe that the major determinant of word-order variation, and indeed of all word order, is syntactic weight, while informational notions play only a subsidiary role. Clearly, this position will be controversial. How can it even be reasonably entertained in the light of all the evidence adduced by discourse theorists in favor of pragmatic conditioning?

The first cause for scepticism comes from an examination of the pragmatic theories themselves. There are two major research traditions: that of the Prague School (cf. e.g. Firbas 1964, 1966); and that of Givon and his associates (cf. Givon 1983, 1988). The Prague School asserts that a sentence increases its "communicative dynamism" in its left-to-right presentation, with new information in the discourse following given information. The same idea had been proposed earlier by Behaghel (1932). Givon's ordering principle is summarized in the phrase "attend first to the most urgent task." His claim is that items are sequenced in such a way that either unpredictable precedes predictable information in discourse, or important precedes unimportant information. Predictable information is information that is given or old in some sense (cf. Hawkins 1978 for a detailed discussion of the types and degrees of givenness that need to be distinguished in the area of definite reference); unpredictable information is generally new. Part of Givon's theory therefore asserts just the opposite of the Prague School's
Early immediate constituents

given-before-new principle, namely new before given. Even allowing for the importance factor, however, it cannot be the case that both of these theories are correct!

There is a certain confusion, therefore, over basic principles that makes it difficult to evaluate what the role of pragmatics is supposed to be in linear ordering. Is discourse sequenced in such a way that the speaker places previously mentioned or inferrable items before new ones? Or does he/she do it the other way round? Not only is this issue unclear, but if we add considerations of heaviness to pragmatic ones, then there is almost no imaginable word-order variant in any language that cannot be motivated by some principle or other in the published literature. Clearly this situation is unsatisfactory. To explain all logical possibilities is to explain nothing. And to those such as myself this suggests that word-order variation in performance may not be primarily determined by pragmatics; that both pragmatic ordering types do occur, albeit perhaps in different language types and structures; and that the true motive for ordering alternations is to be found elsewhere.

Secondly, neither Givon nor other discourse theorists pay sufficient attention to the interacting considerations of syntax. It is readily acknowledged that grammaticalized orders exist and that these will override pragmatic considerations in the languages that possess them (cf. Thompson 1978 and Mithun 1992). But there is no systematic investigation of what the grammatical structure actually is of the sentences and phrases of languages whose ordering is claimed to be driven by discourse alone. Even if discourse principles are operative, these languages still have phrase structure, and there may even be grammatical arguments for a conventionalized basic order despite the word-order freedom, as Hale (1992) has convincingly shown for Papago. Such grammatical structure has to be recognized in language performance, preferably in an efficient manner, and hence the predictions of EIC are just as applicable to languages with lots of word-order freedom as they are to English. This means that pragmatic predictions will need to be juxtaposed with the predictions made by syntactic processing, and this in turn means that explicit assumptions are going to have to be made about the grammatical structure that is being recognized on-line in languages with word-order freedom. Unless this is done, it will be unclear how the data of performance are to be interpreted, and what they prove.

This is especially urgent because there are, in fact, correlations between information status and syntactic weight. Syntactically shorter noun phrases (such as pronouns and proper names) are generally more definite than longer ones, because less information is needed to identify referents that are
"mutually known" (cf. Clark and Marshall 1981) or "mutually manifest" (cf. Sperber and Wilson 1986), compared with those that are being introduced to the hearer and described for the first time. So if definite items generally precede indefinite ones, this could be because short items are predicted to precede longer ones in that structure and language (cf. table 3.2), or because of their information status. If the reverse ordering is found, this could be because long items are predicted to precede shorter ones, or because of the opposite informational theory. In other words, certain observed tendencies in the data of performance can be derivative and secondary consequences of other more fundamental determinants. And separating the true determinants from the epiphenomena cannot be done without considering the precise grammatical structure of the items in question, the ordering predictions made by syntactic processing, and the ordering predictions made by different pragmatic theories. Discourse theorists, however, typically ignore syntactic analyses altogether, so their findings are inconclusive.  

In the next chapter I shall explicitly juxtapose the textual predictions of syntactic processing (cf. (3.18)) with the results of the quantified text methodology that Givon has developed for the testing of his task urgency theory. In his methodology, degree of (un)predictability is measured by counting how far back within the last 20 clauses of the text an entity was mentioned, while degree of importance is measured by counting how many times an entity is mentioned in the subsequent 10 clauses. This comparison of methodologies will use exactly the same textual material from different language types. The comparison shows that there are indeed correlations between syntactic weight and information status, stronger in some languages than in others. It shows that both given-before-new and new-before-given orders are productive, as expected. It shows also that EIC's predictions are well supported, whereas Givon's are not. Most interestingly perhaps, it shows that if we take EIC as the primary determinant of order, we can predict, within reason, when the one pragmatic ordering will be found, and when the other, as a function of the structural type in question and its short-before-long or long-before-short ordering (cf. table 3.2), and of the degree of strength for EIC's predictions in individual structural instances.

These findings must be considered tentative until they have been confirmed or refuted by other researchers with access to more languages than I and my collaborators could examine. Nonetheless, I believe that a syntactic processing approach of the type advocated here can potentially resolve current conflicts in the pragmatics literature. It also provides a unifying principle for both free and fixed word order structures. The former respond
immediately to EIC, based on the syntactic weight that each IC happens to have on each occasion in performance. The latter are grammaticalizations of EIC based on aggregate weights. In the transition from free to fixed word order, the most frequent orderings in performance become grammaticalized; in the reverse shift, the hitherto grammaticalized order will still be the most frequent in performance. Syntactic change therefore receives a natural explanation in this theory.

The suspicion that there is a single principle underlying order in free- and fixed-word-order languages is heightened by the fact that there are, despite first appearances, a great many similarities between them. In addition to a number of free word orders, English has many rearrangement possibilities, of the type illustrated in this chapter, whose effects are fundamentally similar to those one sees in languages without many grammaticalized basic orders. It can even split up its NPs by processes such as Extraposition from NP. Its Heavy NP Shift applies even though there is a grammaticalized basic order, and so on. Conversely, I will argue that there are actually more grammaticalized basic orders and English-type rearrangements in the predominantly free-order structures and languages than has been appreciated hitherto, for example in Japanese and Korean (recall also Hale’s 1992 arguments for basic ordering in Papago, cf. n. 15 ch. 3). All of these ordering phenomena, namely basic orders, rearrangements of basic orders, and free-order arrangements, are driven by EIC, as I see it, and what varies across languages is simply the degree of grammaticalization; i.e. the number of structures in which a basic order has been conventionalized, in accordance with aggregate weights.\(^\text{17}\)

I do not have textual data from the languages with the freest word orders of all, those of Australia. I take it as axiomatic, however, that these languages are like all others in having constituent structure in the syntax, and that any differences are a matter of degree only, and I would predict that NP discontinuities and other free-word-order phenomena will adhere to EIC’s performance predictions in the same way they do in all the free-word-order structures that I have examined. I take heart from the fact that there is a growing sentiment among Australianists that the selection among free-word-order options in these languages appears to have little to do with pragmatics (Stephen Levinson, personal communication). Cf. also n. 16 ch. 3.

There is one big proviso to all of this. Many languages appear to have “discourse-configurational” nodes in their syntactic representations, i.e. nodes for topic phrases and focus phrases. Hungarian is an example (cf ch. 4.1.2 and Kiss 1987). These nodes may have a pragmatic function and
correlate; alternatively there may be a more semantic basis to these nodes, with clear truth-conditional differences between the sentences that position items there and those that position them elsewhere. Either way, the selection of a syntactic category for positioning into one of these nodes in a given sentence will not be driven by considerations of syntactic weight – it will be driven by whatever pragmatic and/or semantic content the category has within the relevant sentence. As a result, linear order may be driven by pragmatic and/or semantic considerations, just in case the language in question has a grammaticalized node and position for the relevant pragmatic and/or semantic functions. But pragmatics and/or semantics do not, it seems, determine the positioning of ICs whose relative ordering is left free by the grammar.

Moreover, the grammatically fixed positions of discourse-configurational nodes are plausibly explained by EIC, just as other grammaticalized orders are, even though the selection of items to fill these nodes in each performance instance is governed by the discourse and/or semantics, as argued by Primus (1991b). Many languages have a grammaticalized topic node, most typically in left-peripheral position in the matrix clause (cf. Gundel 1988, Primus 1991b). A number of discourse and cognitive explanations have been offered for this “left-right asymmetry” in topic positioning (cf. e.g. Sasse 1987). But EIC also provides an explanation in terms of the syntactic weight distribution of topic and comment nodes respectively, cf. ch. 5.6.3. In addition, Primus argues that the EIC account generalizes to explain the differential positioning of focus constituents within and across languages, for which there is no coherent discourse or cognitive explanation. In other words, EIC provides a unified theory of the positioning of discourse-configurational nodes, and this theory is simply a by-product of EIC’s explanation for the positioning of purely syntactic nodes.

Finally, there is a more general and philosophical point that needs to be made about pragmatic theories of free word order. These theories are motivated by the desire to find some constraints on, and reasons for, word-order selections in performance when the grammar allows many possibilities. The basic premise is that these selections will be functional, and will advance successful and efficient communication. I agree with this basic premise. I disagree with the philosophy that underlies their proposals about what makes communication more efficient and functional in this context. Their functionalism is essentially non- or extra-grammatical, attaching the significance that it does to general knowledge, previous knowledge of the text, the situation of utterance, and conversational inferencing. These are the factors
that distinguish between givenness and newness, and that are claimed to
guide the selection of different orders. But before any of these pragmatic
knowledge sources and inferencing strategies can be activated in process-
ing, the parse string has to be phonologically, morphologically, and syntac-
tically decoded. The recognition of linguistic form, i.e. sound units, word
units and larger syntactic units, logically precedes the assignment of mean-
ings to these units, because one cannot begin to assign a meaning to a given
form, activating in the process one’s grammatical knowledge of form–mean-
ing pairings, unless one knows what form it is that is to be interpreted. Hence,
form processing must, in general, have priority over content, given that both
semantic and pragmatic processing require prior access to form. As I see it,
the primary function of word order is to make form recognition, specifically
syntactic structure, more efficient, rather than to facilitate the processing of
pragmatic information.

Consider word recognition. It is the incoming physical signal that gradu-
ally provides enough phonological and morphological information for the
hearer to be able to access the intended lexical item and to distinguish it from
other members of the phonologically activated “cohort” (cf. Marslen-Wilson
and Tyler 1980; Hawkins and Cutler 1988). But so too in the syntax. The
linear sequence of words and morphemes in the parse string gradually enables
the hearer to recognize structural groupings of words, i.e. mother, daughter,
and sister relations, via parsing principles such as Mother Node Construction
(cf. (3.3)) and Immediate Constituent Attachment (cf. (3.4)), and to distin-
guish the intended structure from other possibilities. These structural group-
ings are then vital prerequisites for semantic interpretation, which is a highly
structure-dependent process (cf. the principle of Structural Determination
(2.34)), and for pragmatic processing.

The EIC principle results in more efficient structure recognition, therefore,
which I would argue is a fundamental functional requirement on any human
communication system. Ordering items according to their discourse status is
simply less fundamental, and less helpful to the hearer. The degree of
predictability of an entity will be known to the hearer anyway, based on
previous mention or the lack of it, etc., whether this status is explicitly
flagged or not. And my text counts will show future importance to be
irrelevant (cf. ch. 4.4). It is not clear what the precise functional advantages
of pragmatic ordering are, therefore; and whatever they are, the exigencies
of rapid and efficient grammar recognition are arguably much
greater. The Prague School theory of given-before-new ordering seems to
be particularly non-functional: why should each sentence begin with what
The relationship of EIC to other principles

has already been presented, delaying the newest information till the very end? There are plausible cognitive explanations for the positioning of a topic before a comment (cf. Sasse 1987), but the theory of communicative dynamism is more general than this, claiming that given before new holds for all adjacent items throughout a sentence. But why should this be? Givon’s task-urgency theory advocates a functional advantage for new before given, and this seems more plausible, in general. Yet how functional can any pragmatic serialization principle be unless one considers it alongside the grammatical information that is available on-line at different points in the parse string, whose processing is a prerequisite for accessing all semantic and pragmatic information? And empirically both pragmatic orderings will be shown later to be productive, within and across languages.

For all of these reasons, I believe that a different approach to performance is needed, and EIC provides one.

3.8.3 Weight-based principles

EIC builds on the insight that syntactic weight or length is relevant for the ordering of elements. There seems to be general agreement about this in the literature, although the extent of this relevance is usually seen as being limited to a handful of structures that are particularly difficult for processing, such as center-embedded finite clause complements (cf. Grosu and Thompson 1977, Dryer 1980). In this book it is argued that weight is a far more fundamental determinant of order than is currently appreciated. Our theory provides some answers to the following general questions: what exactly is/are the principle(s) of weight? Why is weight relevant to ordering? And how extensive is weight in relation to, and in comparison with, other explanatory principles of ordering? I believe that discussions of weight hitherto have been too Eurocentric in the languages and structures that have been examined and in the inferences that have been drawn about its effects. I provide an explanation for these effects in terms of structural complexity. And I argue that this explanation subsumes numerous other principles, of grammar and of pragmatics, that have been presented as primitive and independent hitherto. Hence, the number of basic entities can be reduced, and some principles can be derived from others and can be eliminated.

The effects of weight have received a lot of attention in English linguistics. The name “Heavy NP Shift” is a recognition of the role of weight in bringing about these particular rearrangements. The rule of Extraposition postponing S has been recognized as an instance of the “heavier element principle” by
Stockwell (1977b: 68) and its application in performance has been shown by Erdmann (1988) to be driven by weight. For German, Behaghel proposed his famous Gesetz der wachsenden Glieder, which states that a shorter element precedes a longer one wherever possible (cf. e.g. Behaghel 1932: 6), and he also pointed out that this principle applies to other languages as well. Gil (1986) has since generalized this idea into a short-before-long principle for all languages.

It should be clear from this chapter, and especially from table 3.2, that short before long is not the correct universal formulation of the weight principle. It works for languages like English in which higher phrases are consistently recognized on a left periphery, i.e. in which constituents are of the type $m\text{IC}$; and it works for some structures in languages whose constituents are predominantly recognized on a right periphery, i.e. $\text{IC}_m$. But the general preference in languages like Japanese and Korean is just the opposite of English: long before short. The reason is that constituent recognition domains are consistently shortened in English-type languages when short precedes long, but lengthened when long precedes short; CRDs in Japanese and Korean are consistently shortened when long precedes short (because of the right-peripheral recognition of the long constituent), but lengthened when short precedes long.

The mirror-image effect here follows from one and the same principle, EIC. Moreover, this principle is in turn contained within, and follows from, a larger theory of complexity and processing efficiency, and this theory provides an explanation for what is otherwise an arbitrary and stipulated principle of weight, and for the variation in its manifestations across languages. This same explanation can then be shown to generalize to other, supposedly separate principles advocated by different scholars, and on occasion by one and the same scholar.

Consider Behaghel again. The principle that he proposes as the most significant determinant of ordering, das oberste Gesetz, states that “das geistig eng Zusammengehörige auch eng zusammengestellt wird” (Behaghel 1932: 4), i.e. what belongs together mentally is placed close together. From the examples he gives, it is clear that he intends this to mean that syntactic phrases should remain intact in surface structure. A possessive genitive should be “bound to” its nominal head (within NP), an adjective to a noun (again within NP), and an adverb to an adjective (within AdjP). Putting this another way: discontinuities of daughter ICs are dispreferred. But this is exactly what EIC predicts as well. The relevant phrases all have worse ratios when non-dominated material intervenes from a different phrase, making the
daughters discontinuous. EIC subsumes Behaghel's *oberste Gesetz*. Moreover, Behaghel's principle is actually incorrect on some occasions. There are structures in which it is advantageous to have a discontinuity, even though the EIC score for one phrase becomes worse, because the score for a higher containing phrase is thereby improved, and the improvement offsets the deterioration in the lower phrase and results in a higher ratio for the sentence overall (cf. ch. 4.3.5 for examples from German). EIC therefore predicts both the cases to which Behaghel's principle applies and those to which it does not, so his *oberste Gesetz* can be eliminated.

So far we have seen that the *Gesetz der wachsenden Glieder* and the *oberste Gesetz* both reduce to EIC. So does a third principle, involving pragmatics. This is essentially the given-before-new principle ("es stehen die alten Begriffe vor den neuen," Behaghel 1932: 4), which I discussed in the last section. The pragmatic effects that appear to motivate a given-before-new principle in German and other languages (but new before given in yet other languages and structures) will be derived as secondary consequences from EIC in ch. 4.4.

Therefore, at least three of Behaghel's laws – his principle of weight, an adjacency principle that is strongly grammaticalized, and a pragmatic principle – can apparently be reduced to one. Clearly, any such reduction is to be preferred to a multiplicity of basic entities. Less general statements should give way to more general ones, especially when, as in the present case, the less general statements are only partially correct, in ways that are predicted by the more general theory, EIC.

Behaghel's *Gesetz der wachsenden Glieder* reappears in more recent and more explicit form in Dik's (1978, 1989) Functional Grammar, as the LIPOC principle, i.e. the Language-Independent Preferred Order of Constituents. The current formulation (cf. Dik 1989: 351) is:

\[\text{(3.25) LIPOC}\]

Other things being equal, constituents prefer to be placed in an order of increasing complexity, where the complexity of constituents is defined as follows:

- (i) clitic > pronoun > noun phrase > adpositional phrase > subordinate clause;
- (ii) for any category X: X > X co X;
- (iii) for any categories X and Y: X > X [sub Y]

"\(>\)" means "precedes" in this context. (ii) refers to conjoined constituents of the same category, and (iii) refers to the addition of subordinating material.
LIPOC is clearly a principle of increasing weight, and is appropriate for those languages and structures in which there is a short-before-long preference within CRDs. It does not account for the CRDs with long-before-short preferences, however, but derives the relevant orders from other principles with which LIPOC interacts and to which it is claimed to be subordinate. The Principle of Head Proximity, for example, states that "constituent ordering rules conspire to keep the heads of different domains as close together as possible," and this brings about the mirror-image orderings between head-initial and head-final languages (cf. ch. 5.1), and the resulting long-before-short orders in the latter language type.

But both LIPOC and Head Proximity follow from EIC. So does Dik's principle of Domain Integrity, which is his version of Behaghel's *oberste Gesetz*: "constituents prefer to remain within their proper domain; domains prefer not to be interrupted by constituents from other domains." So too do his Relator principles, on the assumption that Relators are constructing categories in our sense. There are too many basic entities here, which appear to be reducible to one and the same principle.
4 Testing EIC's performance predictions

In this chapter I shall test the predictions made by EIC for language performance formulated in ch. 3.5. Some preliminary evidence taken from the published literature and involving psycholinguistic experiments, acceptability intuitions, and text frequency counts has already been summarized in ch. 3.3. This evidence did not test EIC directly, however, and was limited in the range of languages considered, just English and Japanese. In this chapter performance data will be considered from the following ten languages: English, German, Greek, Polish, Rumanian; Finnish, Hungarian; Turkish; Japanese and Korean. These languages are typologically diverse and exemplify all the structural possibilities set out in table 3.2. They also represent four distinct genetic stocks (in the classification of Ruhlen 1991: 380): the first five are Indo-European and cover four subfamilies (Germanic, Greek, Slavic, and Romance); Finnish and Hungarian are from the Finno-Ugric branch of Uralic; Turkish is from the Turkic branch of Altaic; and Japanese and Korean belong to the Japanese-Korean-Ainu grouping which Ruhlen now considers separate from Altaic. These languages exhibit considerable freedom of word order, though they also have many grammaticalized orders as well, especially English.

The EIC was explicitly tested on nine of these languages (all but Greek), both by myself and by collaborators at the University of Southern California and in the Constituent Order Group of the European Science Foundation Programme in Language Typology. The selection of languages was partly the result of the expertise of willing collaborators and partly the result of an attempt on my part to maximize typological and genetic diversity within the constraints of available personnel. The methodology I have developed for testing EIC requires fine-tuned syntactic and also pragmatic analysis of textual data, and this can only be done by trained linguists with considerable practical and theoretical knowledge of the language in question.
The particular prediction to be tested is the EIC Text Frequency Prediction of (3.18), which is just one of the predictions of the Performance Ordering Prediction in (3.17). I repeat (3.18) for convenience:

(3.18) **EIC Text Frequency Prediction**

Given: alternative grammatical orderings of ICs within a mother phrase M contained within a sentence, where each IC dominates a set of non-ICs;

Then: for all different sets of non-ICs (or words alone), the orderings of ICs that provide the most optimal EIC ratios for the sentence will be the ones that are found in text samples, in the unmarked case; orderings with non-optimal EIC ratios will be more or equally frequent relative to their preferred counterparts, in direct proportion to the magnitude of their ratios.

I first test (3.18) on a set of multiple branching structures exhibiting complete or considerable word-order freedom, i.e. for which current grammatical analyses generally propose no grammaticalized order (cf. section 4.1). These structures were selected because they are clearly identifiable in actual texts by independent investigators. In section 4.2 I consider some binary branching structures. Section 4.3 examines rearrangements of basic word orders in structures for which there are grammatical arguments for ordering conventions. In section 4.4 I juxtapose the results of EIC's predictions with pragmatic predictions, using text samples from four different languages, and in section 4.5 I consider EIC's predictions for temporary ambiguity preferences, in English and Japanese.

It is my hope that other researchers will be encouraged to test (3.18) on structures and languages beyond those considered here, and that other types of performance data (especially experimental data) will also be gathered that specifically test EIC's general performance prediction (3.17).

### 4.1 Arrangements of free word orders in multiple branching structures

I begin with an examination of multiple branching structures of the type (2a) in table 3.2, i.e. [C {mIC1 mIC2}]. Multiple branching structures have more potential for free ordering than binary branching structures, {C IC}, since the positioning of C, the category that constructs the mother node, is regularly fixed by the grammar. As a result, the position of a single accompanying sister phrase is fixed as well. When there are two or more sisters, however,
these may regularly be freely ordered. I accordingly test the multiple branching predictions first, and the binary branching structures later, once the methodology has been illustrated using the more productive cases of free ordering.

4.1.1 English [NP V {PP, PP}]

Consider the relative ordering of two PPs in English, i.e. alternants such as [The raven flew [through the barn] [into the open air]] and [The raven flew [into the open air] [through the barn]]. Both orders are grammatical, and indeed sentences of this type seem to be regularly permutable in this way. The two PPs will not be permutable if the second is embedded within an NP in the first, as in flew [through the barn [of the farm]], or flew [through the barn [by the farmhouse]]. The former case is not permutable at all (cf. *flew [of the farm] [through the barn]); the latter is permutable only with a change in structure and meaning: flew [by the farmhouse] [through the barn] may be false, whereas flew [through the barn [by the farmhouse]] could be true. If we take all cases of two PPs in an English text that are both permutable and truth-conditionally equivalent, therefore, we can test prediction (3.18) in relation to them.¹

Notice that there are still different possible syntactic analyses for two adjacent PPs, even when they are permutable and not embedded one in the other. Both may be immediately dominated by VP; or by S; or the first may be dominated by VP and the second by S; or (if we allow discontinuity of structure) the first may be immediately dominated by S and the second by VP. These four possibilities are shown in (4.1):

\[(4.1) \quad \begin{align*}
  a. & \quad s[NP_v[V PP_i PP_j]] \\
  b. & \quad s[NP_v[V] PP_i PP_j] \\
  c. & \quad s[NP_v[V PP_i] PP_j] \\
  d. & \quad s[NP_v[V PP_i] VP[PP_j]]
\end{align*}\]

It turns out that EIC's preferences are the same in all four structures: a shorter PP is always preferred before a longer one, and is progressively more preferred the greater the word-total difference between them.

Table 4.1 illustrates this by calculating the IC-to-word ratios for the S and VP recognition domains under each analysis, in accordance with the procedure defined in (3.13) in the last chapter. The illustration compares PPs
Testing EIC's performance predictions

Table 4.1  *IC-to-word ratios for English [NP V {PP_1, PP_2}]*

Assume NP = 2 words, V = 1.

<table>
<thead>
<tr>
<th>I</th>
<th>PP_1 = 3</th>
<th>PP_2 = 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[NP V PP_1 PP_2]</td>
<td>[NP V PP_2 PP_1]</td>
</tr>
<tr>
<td>a.</td>
<td>s[NP vP[V PP_1 PP_2]]</td>
<td>s[NP vP[V PP_2 PP_1]]</td>
</tr>
<tr>
<td></td>
<td>S CRD: 2/3 = 67%</td>
<td>S CRD: 2/3 = 67%</td>
</tr>
<tr>
<td></td>
<td>VP CRD: 3/5 = 60%</td>
<td>VP CRD: 3/7 = 43%</td>
</tr>
<tr>
<td></td>
<td>Aggregate = 63.5</td>
<td>Aggregate = 55</td>
</tr>
<tr>
<td>b.</td>
<td>s[NP vP[V] PP_1 PP_2]</td>
<td>s[NP vP[V] PP_2 PP_1]</td>
</tr>
<tr>
<td></td>
<td>S CRD: 4/7 = 57%</td>
<td>S CRD: 4/9 = 44%</td>
</tr>
<tr>
<td></td>
<td>VP CRD: 1/1 = 100%</td>
<td>VP CRD: 1/1 = 100%</td>
</tr>
<tr>
<td></td>
<td>Aggregate = 78.5</td>
<td>Aggregate = 72</td>
</tr>
<tr>
<td>c.</td>
<td>s[NP vP[V PP_1] PP_2]</td>
<td>s[NP vP[V PP_2] PP_1]</td>
</tr>
<tr>
<td></td>
<td>S CRD: 3/7 = 43%</td>
<td>S CRD: 3/9 = 33%</td>
</tr>
<tr>
<td></td>
<td>VP CRD: 2/2 = 100%</td>
<td>VP CRD: 2/2 = 100%</td>
</tr>
<tr>
<td></td>
<td>Aggregate = 71.5</td>
<td>Aggregate = 66.5</td>
</tr>
<tr>
<td>d.</td>
<td>s[NP vP[V] PP_1 PP_2]</td>
<td>s[NP vP[V] PP_2 PP_1]</td>
</tr>
<tr>
<td></td>
<td>S CRD: 3/4 = 75%</td>
<td>S CRD: 3/4 = 75%</td>
</tr>
<tr>
<td></td>
<td>VP CRD: 2/5 = 40%</td>
<td>VP CRD: 2/7 = 29%</td>
</tr>
<tr>
<td></td>
<td>Aggregate = 57.5</td>
<td>Aggregate = 52</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II</th>
<th>PP_1 = 3</th>
<th>PP_2 = 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[NP V PP_1 PP_2]</td>
<td>[NP V PP_2 PP_1]</td>
</tr>
<tr>
<td>a.</td>
<td>Aggregate = 63.5</td>
<td>Aggregate = 52.5</td>
</tr>
<tr>
<td></td>
<td>VP differential = 22%</td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>Aggregate = 78.5</td>
<td>Aggregate = 70</td>
</tr>
<tr>
<td></td>
<td>S differential = 17%</td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>Aggregate = 71.5</td>
<td>Aggregate = 65</td>
</tr>
<tr>
<td></td>
<td>S differential = 13%</td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td>Aggregate = 57.5</td>
<td>Aggregate = 50</td>
</tr>
<tr>
<td></td>
<td>VP differential = 15%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III</th>
<th>PP_1 = 3</th>
<th>PP_2 = 7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[NP V PP_1 PP_2]</td>
<td>[NP V PP_2 PP_1]</td>
</tr>
<tr>
<td>a.</td>
<td>Aggregate = 63.5</td>
<td>Aggregate = 50</td>
</tr>
<tr>
<td></td>
<td>VP differential = 27%</td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>Aggregate = 78.5</td>
<td>Aggregate = 68</td>
</tr>
<tr>
<td></td>
<td>S differential = 21%</td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>Aggregate = 71.5</td>
<td>Aggregate = 63.5</td>
</tr>
<tr>
<td></td>
<td>S differential = 16%</td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td>Aggregate = 57.5</td>
<td>Aggregate = 48.5</td>
</tr>
<tr>
<td></td>
<td>VP differential = 18%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IV</th>
<th>PP_1 = 3</th>
<th>PP_2 = 10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[NP V PP_1 PP_2]</td>
<td>[NP V PP_2 PP_1]</td>
</tr>
<tr>
<td>a.</td>
<td>Aggregate = 63.5</td>
<td>Aggregate = 46</td>
</tr>
<tr>
<td></td>
<td>VP differential = 35%</td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>Aggregate = 78.5</td>
<td>Aggregate = 64.5</td>
</tr>
<tr>
<td></td>
<td>S differential = 28%</td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>Aggregate = 71.5</td>
<td>Aggregate = 60.5</td>
</tr>
<tr>
<td></td>
<td>S differential = 22%</td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td>Aggregate = 57.5</td>
<td>Aggregate = 46</td>
</tr>
<tr>
<td></td>
<td>VP differential = 23%</td>
<td></td>
</tr>
</tbody>
</table>
Arrangements of free word orders in multiple branching structures

doing by two words in length (PP1 = 3 PP2 = 5), by three words (PP1 = 3 PP2 = 6), by four words (PP1 = 3 PP2 = 7), etc., in each of the orders [NP V PP1 PP2] and [NP V PP2 PP1]. The lengths of the subject NP and of V are held constant and are assigned two words and one word respectively. In all these cases words are used as an indicator of non-IC structure, following the discussion in ch. 3.4 (cf. table 3.1). Thus, in the (4.1a) analysis the S domain comprises two ICs, NP and VP, the latter being constructed on its left periphery by V, and the VP comprises three ICs, V, PP1, and PP2. The CRD for S therefore has an IC total of two, and a word total of three (two of them dominated by NP, the third by V), making an IC-to-word total of 2/3 = 67%, in all instances of this structure (cf. Ia–IVa in table 4.1). The ratio for the VP domain, however, varies with the relative length and complexity of the PPs. The CRD for VP proceeds from V to the preposition that constructs PP2, i.e. from the word that constructs the first IC of the domain to the word that constructs the last IC, following the definition of a CRD in (3.2). When PP2 is the smaller PP (i.e. PP1 in table 4.1) and consists of three words, the word total for the VP CRD is five, one for V, three for PP1, and one for the P of PP2, making an IC-to-word ratio of 3/5 = 60%. When PP1 is the larger PP2 with five words, the ratio is 3/7 = 43%, which makes a differential for the VP of 17%. Since the S ratio is constant, the positioning of the shorter before the longer PP improves the ratio for VP, and with it the aggregate for the sentence as a whole, represented here by the S and VP nodes alone (the only nodes that are potentially involved in this ordering alternation).

In the (4.1b) analysis, the two PPs are S-dominated, and so the CRD for S consists of four ICs, while the VP comprises just V, the category that constructs it. The relative advantage of ordering short before long is now captured in the S domain and is carried forward into the sentence aggregate. For the (4.1c) analysis there are three ICs of S and two of VP. If PP1 is shorter than PP2, the VP that contains the shorter PP1 will be shorter, and the distance from NP to the preposition constructing PP2 will be shorter as well, meaning that the short-before-long preference is again captured in the S domain. The VP domain will remain optimal, with V and PP1 consistently recognized by two adjacent words.

In (4.1d), the discontinuous analysis, the S domain improves as a result of the discontinuity compared with (4.1c), because the third IC of S, PP3, is encountered earlier than it would be if PP1 were adjacent to V. However, the VP becomes progressively worse as the distance increases between its discontinuous ICs, V and PP2. Since the S domain remains constant in this
Testing EIC's performance predictions

analysis (a PP\(_1\) of any length is constructed on its left periphery), the short-before-long preference is reflected in the VP domain, where many of the words in the CRD for VP are S-dominated, and is again carried forward into the sentence aggregate.

In all four structures, therefore, a two-word differential between PP\(_1\) and PP\(_2\) results in a significant preference for the shorter PP before the longer one. There is one circumstance, however, in which EIC could define a long-before-short preference for English, depending on the analysis assigned to individual sentences. Notice that in table 4.1 the aggregate for the Ic structure s[NP VP[VP PP\(_2\)] PP\(_1\)] is 66.5, i.e. with the longer PP\(_2\) dominated by VP and preceding the shorter PP\(_1\) dominated by S. This aggregate is lower than the 71.5 figure for the corresponding PP\(_1\) before PP\(_2\). But structures in which PP\(_2\) is, for independent reasons, VP-dominated are not necessarily going to undergo a constituency change as well as an ordering change, for reasons of EIC, positioning the shorter PP\(_1\) under VP and PP\(_2\) under S. It may be possible for PP\(_2\) to be rearranged out of the VP and into S, following an S-dominated PP\(_1\), as in Ib s[NP VP[V PP\(_2\)] PP\(_1\)], which would increase the aggregate from 66.5 to 78.5. This kind of constituency reassignment has been proposed for a number of rightward-moving rules in the literature. But there is then no EIC motivation, and no independent support, for reattaching the shorter PP\(_1\) under VP at the same time. And, depending on one's theory, it may not be legitimate to change the constituency of PP\(_2\) either. s[NP VP[V PP\(_2\)] PP\(_1\)] would then alternate with the discontinuous structure Id s[NP VP[V PP\(_1\) VP[PP\(_2\)]]], and on this occasion the discontinuity results in a lower aggregate of 57.5. Hence, if this is the only alternative to Ic, the longer PP\(_2\) is predicted to remain before the shorter S-dominated PP\(_1\).

By presenting the EIC calculations in the manner of table 4.1, we see a clear preference for short before long PPs across parallel structures with identical constituents, and a clear preference for short before long in the event that a VP-dominated PP can be reattached into S. Only if we allow discontinuous VPs does EIC prefer some long before short ICs (in the non-discontinuous structure). In what follows I shall formulate EIC's predictions on the assumption that there is no discontinuity in these English VPs, even though we shall see strong evidence for discontinuity in other structures and languages later. Even if there are some discontinuous VPs involving two PPs in English, their number is evidently insufficient to have any significant impact on EIC's predictions.

As the word differential increases between PPs, so does the differential in IC-to-word scores for either the VP or S domains, and with it the aggregate
differential. This can be seen by comparing all the scores for each structural analysis down I–IV in table 4.1.

Relevant data taken from 200 pages of written English texts are tabulated in table 4.2. The precise structural template used allowed for an optional particle to intervene between V and the first PP, i.e. it allowed for verb-particle constructions comprising two words in addition to single-word verbs alone. But no other ICs were permitted in the relevant clause. If an adverbial was also present, or a third PP, or a direct object NP, the clause was excluded from consideration. The motivation for this was simply to be able to focus exclusively on the particular prediction I was testing, namely the relative ordering of two ICs of the type mIC with a left-peripheral constructing category C. Nothing hinges on this, however, since the same relative-length prediction would be made for two English PPs whether additional ICs were present or not, and in other cases I will allow for certain limited and precisely specified additional possibilities, in order to increase the number of relevant instances. But by limiting our data in this way, it becomes easier for independent investigators to identify the relevant structures in texts and to replicate or disconfirm the findings, and we avoid having to consider questions of structural analysis (e.g. VP- or S-domination) for constituents that are immaterial to the test at hand. In short, while we limit our n, we hold everything (or nearly everything) constant except for the precise variation that we are interested in.

The manner in which the data are tabulated requires comment. I first give the quantity of instances in which PP₁ = PP₂, namely 35. Thereafter I list the quantities in which PP₂ > PP₁ by one word, two to three words, four words, and five or more words, in each of the orders [PP₁ PP₂] (labeled here X) and [PP₂ PP₁] (labeled Y). The ratio of Y to X is then calculated as a percentage for each word-content differential. In this way we quantify the proportion of non-optimal to optimal orders for each set of word-total assignments to the two PPs.

By presenting things this way, in terms of the difference between PP₂ and PP₁ we actually collapse a number of different possibilities. PP₂ will exceed PP₁ by one word when PP₂ = 3 PP₁ = 2, or PP₂ = 5 PP₁ = 4, or PP₂ = 9 PP₁ = 8, and similarly for the two-word and higher differentials. The motivation for doing this is that it gives us larger and potentially more significant subsets of word-assignment types. It does, at the same time, produce a slight distortion, which needs to be commented on explicitly. When PP₁ = 3 and PP₂ = 5 in analysis (4.1a), there is a significant difference in IC-to-word ratios within the VP CRD when ordering these items [V PP₁ PP₂], rather
than [V PP₂ PP₁], namely 60% versus 43%, i.e. a difference of 17% (cf. Ia in table 4.1). When PP₁ = 3 and PP₂ = 4, with just a one-word differential, there would still be a significant EIC difference: the preferred ordering has a ratio of 60%, the dispreferred one 50%, i.e. there is a difference of 10%. But when PP₁ and PP₂ are both very long, the EIC difference between the two orders is not very significant. If PP₁ = 8 and PP₂ = 10, the ratio for the VP CRD in the order [V PP₁ PP₂] is 3/10 = 30%, and in the order [V PP₂ PP₁] it is 3/12 = 25%. This 5% difference is much less than the 17% difference for the same two-word differential when PP₁ = 3 PP₂ = 5, and is even less than the 10% difference associated with the one-word differential for PP₁ = 3 PP₂ = 4. These figures capture the intuition that there isn’t a whole lot to choose between [V PP₁ PP₂] and [V PP₂ PP₁] when both are very long and differ in small amounts, whereas even a small word-total differential can make a big difference when the number of words assigned is much smaller. As a result, some two-word differentials have lower EIC differences than one-word differentials. The issue is somewhat academic, however, because VPs containing two very long PPs like this are exceedingly rare, and for this reason we can tolerate the distortion, here and in subsequent tables, without skewing our predictions unduly.

EIC makes two predictions for the data of table 4.2: an unmarked-case prediction which asserts that orderings of PP₁ and PP₂ that are most optimal for EIC will be the ones that are generally found, i.e. for all possible word-total assignments to these PPs; and a marked case prediction which asserts that orders with non-optimal ratios will be more or equally frequent relative to their optimal counterparts, the better their ratios are, i.e. the smaller the word-total difference between PP₁ and PP₂.

The unmarked case prediction is tested by adding up the quantities of preferred [PP₁ PP₂] orders on the top line in which PP₂ is longer, plus the total for PP₁ = PP₂. These latter are most optimal in either order and cannot possibly be improved, so they need to be added to the cases where only one order is most optimal, in order to quantify how many textual instances overall are the best they could possibly be, compared with those whose EIC scores could be improved by being ordered differently. We therefore calculate the proportion of the text total in which PP₂ > PP₁ in the order [V PP₁ PP₂]. For the marked case prediction, we expect that the non-optimal structures with PP₂ > PP₁ in the order [V PP₂ PP₁] will generally be those with small word-total differences, and more precisely will decline in number as we move from left to right across word-total differences of increasing length. As these totals increase, so does the degree of difference between
Table 4.2  English \([NP \; V \{PP_1 \; PP_2\}]\)


\[
[\text{NP} \; V \text{ (Part)} \{PP_1 \; PP_2\}] \text{ (no other ICs in the clause)}
\]

<table>
<thead>
<tr>
<th></th>
<th>(PP_1 = PP_2)</th>
<th>(PP_2 &gt; PP_1 : 1) word</th>
<th>:2–3</th>
<th>:4</th>
<th>:5 +</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X)</td>
<td>([PP_1 ; PP_2])</td>
<td>35</td>
<td>27</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>(Y)</td>
<td>([PP_2 ; PP_1])</td>
<td>12</td>
<td>5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Ratio of (Y/X):</td>
<td></td>
<td>31%</td>
<td>14%</td>
<td>9%</td>
<td>0%</td>
</tr>
</tbody>
</table>

**EIC predictions** (assuming no VP discontinuity; consistent short-before-long predictions)

Unmarked case: \(143/161 = \text{most optimal } (X + =), \text{ i.e. } 88.8\% \text{ Correct}

Marked case: \(\text{ratio of } Y/X \text{ for } PP_2 > PP_1 : 1 \geq 2–3 \geq 4 \geq 5 + \text{ All correct}

the non-optimal EIC score and that of its optimal counterpart in the preferred order. Hence, relatively fewer non-optimal orders are predicted.

These predictions are quite strikingly confirmed. The orderings most optimal for EIC account for 88.8% of the total, and within the 11.2% of non-optimal orders, those with just a one-word differential have the highest rate of occurrence (12 in absolute numbers, or 31% of all orderings with one-word differentials), the orderings with two-to-three word differentials have the next highest rate of occurrence, and so on. The preliminary conclusion to be drawn from these results is that language performance is extremely fine-tuned to considerations of syntactic length and complexity, and specifically to EIC.

### 4.1.2  Hungarian \([V \{_{m}NP_1 \; _{m}NP_2\}]\)

A similar prediction can be tested on data from a typologically different language, Hungarian, involving two NPs in postverbal position. These data have been collected and analyzed by Stephen Matthews.

The NPs selected for this test were those that are recognized and constructed on their left peripheries, for example by a head noun or determiner, and that involve no left-branching genitive NPs or relative clauses. An example sentence would be (4.2):
Testing EIC's performance predictions

(4.2) a. [Döngetik NP[facipőink] NP[az utcakat]]
   "batter wooden shoes-lpl the streets-ACC," i.e.
   Our wooden shoes batter the streets.

b. [Döngetik NP[az utcakat] NP[facipőink]]

NPs of this type are left-mother-node-constructing in our terms, i.e. \( m_1 \)NP. NPs with left-branching genitives or relatives are \( m_2 \)NP, since the highest dominating NP is not constructed until the head noun is encountered, and our theory makes a distinction between these two types in Hungarian (cf. section 4.1.8).

Both orderings of (4.2) are fully grammatical and truth-conditionally equivalent. This is because the postverbal domain in Hungarian is generally agreed to be a free-order domain. For constituents appearing before the verb, Kiss (1991) has proposed a configurational structure comprising a Topic node immediately attached to S, and a Focus node immediately attached to VP, as shown in (4.3) (cf. also Kiss 1987):

(4.3)

\[
\begin{array}{c}
S \\
\text{[Topic]} \\
\text{[Focus]} \\
V' \\
V \\
XP^* \\
\end{array}
\]

In addition, quantifier phrases can be iteratively attached to VP. The structure of (4.2), in this analysis, would be immediately dominated by \( V' \). Whether this is the correct node label or not, the crucial point for our purposes is that material to the right of \( V \) constitutes a freely ordered flat structure, whereas material to the left is hierarchically organized and can be expected to be sensitive to the pragmatic-semantic content of these "discourse-configurational" nodes. From the point of view of ordering, the explanatory question posed by the grammatical structure of (4.3) is: why the left-right asymmetry of Topic before VP and Focus before \( V' \)? Primus (1991b) argues that EIC provides a natural explanation, but since this is an explanation for grammaticalized positioning, and since the relevant nodes have pragmatic and/or semantic content, we must expect that the selection of items to fill these nodes in performance will be made on the basis of this content and not on the basis of EIC, even though EIC explains the gram-
maticalization (cf. chs. 3.8.2 and 5.6.3). The true test for the applicability of EIC to Hungarian word order in performance is therefore in the postverbal domain.

Relevant data are set out in table 4.4. Given the structure in example (4.3), there are now no competing structural analyses for the attachment of $m_{NP_1}$ and $m_{NP_2}$. The EIC preferences for different word-total differentials are therefore very much as they are for the two PPs of English in the (4.1a) analysis in which both PPs are dominated by VP (cf. table 4.1). The precise structural template used permits there to be an optional adverbial phrase preceding V. Since this occurs outside the domain we are interested in, it is immaterial whether it is present or not. The template also permits a single additional constituent XP to occur optionally and to the right of both $m_{NP_1}$ and $m_{NP_2}$. In those instances in which there is such an IC, it will be the rightmost IC within the CRD for V'. This will not interfere with EIC's predictions either. If XP is present, the CRD for V' will proceed from V to XP, and IC-to-word ratios can then be calculated on a left-to-right basis (cf. (3.15) in ch. 3.4), defining the same short-before-long preference for two $m_{NPs}$ as before. This is illustrated in table 4.3, which shows consistent preferences for a longer $m_{NP}$ to the right of a shorter $m_{NP}$, and an increasing preference from a two- to a four- to a six-word differential.

The data in table 4.4 are again very much in accordance with EIC's predictions. Of these orders, 91.4% are as optimal as they can be, and all the marked case predictions are correct. In fact, the only non-optimal orders that are productively attested at all are those with just a one-word differential.

These results support EIC. They also confirm the syntactic analysis of the postverbal domain in Hungarian as a free-order domain. Support for a more structured analysis of the pre-verbal domain, as shown in (4.3), comes from additional data collected by Stephen Matthews. He examined the relative length of two $m_{NPs}$ when one preceded V and the other followed, i.e. in the template $[(\text{AdvP}) m_{NP} V \, m_{NP} \, (\text{XP})]$, and when both preceded, $[(\text{AdvP}) \, m_{NP} \, m_{NP} \, V \, (\text{XP})]$. Using the same text as in table 4.4, he found that there were now many instances of longer NPs preceding shorter ones, which means that the conditioning factor governing their order is no longer their relative length. This is compatible with an analysis in which semantic-pragmatic status determines positioning in the pre-verbal domain. It remains the case, however, that a preverbal NP in Hungarian will generally be shorter than its sister node, VP or V' in (4.3) – which, according to Primus (1991b), led to the grammaticalization of these left-right asymmetries in Hungarian and other languages.
Table 4.3  *Left-to-right IC-to-word ratios for Hungarian* \([V\{mNP_1 \ mNP_2\} \times P]\)

Assume XP is constructed on its left periphery by the first word encountered, i.e. \(mXP\) (the same predictions go through if \(XP = XP_m\)); assume \(V = 1\) word.

<table>
<thead>
<tr>
<th></th>
<th>(mNP_1 = 2)</th>
<th>(mNP_2 = 4)</th>
<th>(V{mNP_1 \ mNP_2} \times P)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1/1)</td>
<td>(2/3)</td>
<td>(3/7)</td>
</tr>
<tr>
<td>Aggregate % for CRD</td>
<td>(65%)</td>
<td>Aggregate % for CRD</td>
<td>(58%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>(mNP_1 = 2)</th>
<th>(mNP_2 = 6)</th>
<th>(V{mNP_1 \ mNP_2} \times P)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1/1)</td>
<td>(2/3)</td>
<td>(3/9)</td>
</tr>
<tr>
<td>Aggregate % for CRD</td>
<td>(60%)</td>
<td>Aggregate % for CRD</td>
<td>(50.5%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>(mNP_1 = 2)</th>
<th>(mNP_2 = 8)</th>
<th>(V{mNP_1 \ mNP_2} \times P)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1/1)</td>
<td>(2/3)</td>
<td>(3/11)</td>
</tr>
<tr>
<td>Aggregate % for CRD</td>
<td>(57%)</td>
<td>Aggregate % for CRD</td>
<td>(45.5%)</td>
</tr>
</tbody>
</table>
Arrangements of free word orders in multiple branching structures

Table 4.4 Hungarian $[V \{nNP_1, nNP_2\}]$

112 pages of data: first 112 pp. from Andor Gelleri *Jamaica Rum*, Szépirodalmi Könyvkiadó, Budapest.

$s[(\text{AdvP}) V \{mNP_1, mNP_2\} (XP)]$ (no other ICs in the clause)

where $mNP = \text{any NP constructed on its left periphery, i.e. with no left-branching genitive NP or relative S};$

$(\text{AdvP}) = \text{an optional adverb phrase in sentence-initial position};$

$(XP) = \text{an optional single constituent of any category in sentence-final position}$

\[
\begin{array}{cccc}
\text{n} = 116 & mNP_1 = mNP_2 & mNP_2 > mNP_1 & \text{1 word :2 :3+} \\
X [mNP_1, mNP_2] & 50 & 27 & 8 \\
Y [mNP_2, mNP_1] & 21 & 9 & 1 \\
\text{Ratio of Y/X} & 15\% & 4\% & 0\%
\end{array}
\]

*EIC Predictions*

Unmarked case: $106/116 = \text{most optimal (X+ =), i.e. 91.4\% Correct}$

Marked case: $\text{ratio of Y/X for } mNP_2 > mNP_1 : 1 \geq 2 \geq 3+$ $\text{All correct}$

4.1.3 Rumanian $[V \{mIC_1, mIC_2\}]$

As in Hungarian, the postverbal domain in Rumanian exhibits considerable freedom of word order (cf. Mallinson 1986). Also as in Hungarian, there is a Topic slot preceding the verb which can be filled by a variety of constituents, and not just the nominative-marked grammatical subject (cf. Mallinson 1986; Kiss and Primus 1991). The positioning of constituents following the verb can therefore provide another test of EIC’s predictions for free-word-order arrangements. These data have been collected and analyzed by Beatrice Primus.

Primus examined the word length of any two adjacent ICs immediately following V in Rumanian. These ICs could be two NPs, as in (4.4) (taken from Mallinson 1986: 90).

(4.4) Maria a [dat pre§identelui un buchet]

“Maria Perf-3Sg given president-Det-DAT a-Masc bouquet,” i.e.

Maria has given the president a bouquet

or an NP and a prepositional phrase, or two PPs, or a non-sentential constituent plus an S (recognized by a left-peripheral Comp), etc. These ICs will be overwhelmingly of the type $mIC$, i.e. recognized and constructed on their left peripheries in this language. Kiss and Primus (1991) suggest that there
Testing EIC's performance predictions may be some length-based grammaticalization of certain constituents, e.g. of \( \tilde{S} \) in clause-final position. If this is a grammaticalized order, its right-peripheral positioning is exactly what EIC predicts, given the aggregate length of \( \tilde{S} \) compared with other ICs and the fact that it is constructed on its left flank (cf. ch. 5.2). Since these aggregate lengths should be reflected in arbitrarily chosen text samples, and since there is otherwise an extreme degree of freedom in this domain, the structure \([V \{mIC_1 mIC_2\}]\) can count as a free-word-order structure. EIC makes the same prediction for performance whether or not there has been some grammaticalization. Primus collected 216 instances of \([V IC IC]\) from three texts. A total of \(193/216 = 89\%\) were most optimal, which satisfies EIC's unmarked case prediction. The proportion of dispreferred orders with one-word differentials was \(8/40 = 20\%\); for two-to-three-word differentials it was \(13/58 = 22\%\); and for four or more it was \(2/90 = 2\%\). The difference between 20\% and 22\% will scarcely be significant in a sample of this size; the difference between both and 2\% is much more significant, and this decline as the word-total difference increases among non-optimal orders is in accordance with the marked-case prediction. Again EIC is well supported.

4.1.4 Turkish \([\{mNP_1 mNP_2\} V]\)

Let us consider now an example of structure type (2b) in table (3.2), \([\{mIC_1 mIC_2\} C]\). Turkish provides relevant examples within its VP. It is a predominantly verb-final language, except for finite \(\bar{S}\) complements (constructed by the \(\bar{S}\)-initial complementizer \(ki\)) which are obligatorily postposed (cf. Dryer 1980: 131). Its NP structure is not unlike that of Hungarian, with genitives and relative clauses occurring on a left branch and rendering these NPs recognizable on their right peripheries, i.e. \(mNP\) in our notation, while other NPs appear to be constructed initially, by Det or N for example, i.e. \(mNP\). The PP is postpositional, \(PP_m\), making this constituent fundamentally similar to \(NP_m\) from the perspective of parsing, rather than to \(mNP\). The data to be presented in this section involve two adjacent non-subject \(mNPs\), both immediately preceding \(V\), and with an optional subject NP preceding both. An example is (4.5), with the subject omitted:

(4.5) a. \([NP[\tilde{on}\-\tilde{un}\-\tilde{e}] NP[\tilde{beyaz}\-\tilde{onlug}\-\tilde{un}\-\tilde{i}] tak-ti] \]
   \("front\-GEN\-DAT white\-apron\-GEN\-ACC tie-PAST,\" i.e. \n   X tied a white apron onto his/her front.

b. \([NP[\tilde{beyaz}\-\tilde{onlug}\-\tilde{un}\-\tilde{i}] NP[\tilde{on}\-\tilde{un}\-\tilde{e}] tak-ti]\)
Both orders of (4.5) are fully grammatical. Data of this type were collected and analyzed by Charlotte Reinholtz.

The data are set out in table 4.6. There is now a total of 81% of these orders that are most optimal. This is lower than the 89% figure for English [V PP PP] sequences in table 4.2 and the 91% figure for Hungarian [V mNP mNP] in table 4.4. There may be a reason for this. The short-before-long preference that EIC defines for structures of the type \([C mIC mIC]\) is stronger than the preference for \([mIC mIC C]\). In the former case, C constructs the relevant mother node on the left, and the second \(mIC\) can be the last IC of its domain. All the nodes and words dominated by the last \(mIC\) are thereby excluded from the domain except for the left-peripheral node that constructs it. CRDs can therefore be reduced quite significantly in size when a shorter \(mIC\) is positioned before a longer one. But in \([mIC mIC C]\) structures such as (4.5) where C constructs its mother on the right, the CRD necessarily proceeds from the first \(mIC\) on the left to C, and whatever the word length of each \(mIC\), all the words of each IC will be in the domain in both possible orders. The advantage of short before long is now a weaker advantage that comes exclusively from the left-to-right distribution of the same non-ICs and words within the domain, and not from any absolute size reductions in the CRD.

Consider the case in which \(mNP_1 = 2\) words and \(mNP_2 = 4\), i.e. where there is a two-word differential, and both NPs are VP-dominated. The CRD for VP measured in terms of IC-to-word ratios (cf. (3.14)) will be \(3/7 = 43\%\) in both orders. The left-to-right measurement yields a slight advantage of 8\% for the short-before-long ordering, as shown in table 4.5. Compare, by contrast, the relative ordering of two \(mPPs\) in the English VP when there is the same two-word difference (cf. Ia in table 4.1). The absolute size of the VP CRD is reduced from \(3/7 = 43\%\) to \(3/5 = 60\%\) when the shorter precedes the longer \(mPP\), giving a greater advantage of 17\% over the long-before-short ordering (and an almost identical 17.4\% advantage when the calculations are done left to right).

The preference for short before long Turkish structures of the type \(\{mIC_1 mIC_2\} V\) is therefore predicted by EIC, but it is a weaker prediction for languages of this type than it is for languages in which V is leftmost in its domain and CRDs can potentially vary in absolute size. Correspondingly, the unmarked case prediction is satisfied by a smaller plurality, and the ratios for non-optimal to optimal orders are higher in the marked case predictions.

Notice finally that I have assumed these non-subject \(mNPs\) of Turkish to be VP-dominated in the calculations of table 4.5. Turkish is a case-marking
Table 4.5 *Left-to-right IC-to-word ratios for Turkish* \(\{mNP_1, mNP_2\} V\)

Assume all ICs = VP-dominated (the same predictions go through for all possible constituency attachments); assume V = 1 word.

\[
\begin{array}{cccc}
\text{v}_P[mNP_1 & mNP_2 & V] & \text{v}_P[mNP_2 & mNP_1 & V] \\
1/2 & 2/6 & 3/7 & 1/4 & 2/6 & 3/7 \\
50\% & 33\% & 43\% & 25\% & 33\% & 43\%
\end{array}
\]

Aggregate % for CRD = 42\%  
Aggregate % for CRD = 34\%  
Differential = 8\%

\[
\begin{array}{cccc}
\text{v}_P[mNP_1 & mNP_2 & V] & \text{v}_P[mNP_2 & mNP_1 & V] \\
1/2 & 2/8 & 3/9 & 1/6 & 2/8 & 3/9 \\
50\% & 25\% & 33\% & 17\% & 25\% & 33\%
\end{array}
\]

Aggregate % for CRD = 36\%  
Aggregate % for CRD = 25\%  
Differential = 11\%
Table 4.6 *Turkish* \( \{ \text{mNP}_1 \text{mNP}_2 \} V \)

91 pages of data: pp. 7–97 from Nazli Eray *Ah BayIm Ahl*, Bilgi YayInevi, Ankara.

\[ s[(\text{NP-Subject})\{\text{mNP}_1 \text{mNP}_2 \} V] \text{ (no other ICs in the clause)} \]

where \( \text{mNP} = \) a non-subject NP constructed on its left periphery, i.e. with no left-branching genitive or relative S.

\[
\begin{array}{cccc}
\text{n} & \text{mNP}_1 = \text{mNP}_2 & \text{mNP}_2 > \text{mNP}_1 : 1 \text{ word} & 2-3 \ 4+ \\
X \{\text{mNP}_1 \text{mNP}_2\} & 102 & 40 & 17 \ 2 \\
Y \{\text{mNP}_2 \text{mNP}_1\} & 27 & 12 & 1 \\
\text{Ratio of Y/X} & 40\% & \ 41\% \ 33\% \\
\end{array}
\]

*EIC Predictions*

Unmarked case: 161/198 = most optimal \((X+)\), i.e. 81.3\% Correct
Marked case: ratio of Y/X for \( \text{mNP}_2 > \text{mNP}_1 : 1 \geq 2-3 \geq 4+ \) Correct

language, and these cases are assigned by the verb. The \( \text{mNP} \)s of the text sample were overwhelmingly case-marked (accusative, dative, instrumental, etc.), and the assumption that they occur within the immediate phrase in which case is assigned is therefore not unreasonable. Nothing hinges on this in the present context, since the same left-to-right short-before-long preference will be defined by EIC whether the two \( \text{mNP} \)s are VP-dominated or S-dominated, and whether one is attached to VP and the other to S, just as this preference is consistently defined for English in table 4.1. The existence of a VP in Turkish will be supported by additional data to be considered in section 4.1.7.

4.1.5 *Japanese* \( \{ \text{IC}_m \ 2\text{IC}_m \} V \)

Consider structures of type (3) in table 3.2, i.e. \( \{\text{IC}_m \ 2\text{IC}_m \} C \). For this we need languages that are consistently head-final, or in our terms consistently right-mother-node-constructing. Japanese and Korean fall into this category, and we shall consider both, starting with Japanese. The data and analysis in this section come from Kaoru Horie.

The predictions defined by EIC for type (3) structures are the exact opposite of those for head-initial languages like English: we now expect long ICs to precede short ones. The reason is that if ICs are constructed on their right peripheries, for example if \( S \) is constructed by a rightmost complementizer,
then positioning a long IC leftmost in the domain shortens the distance from
the construction of the first IC to the last, C. If a short IC\textsubscript{m} precedes $\bar{S}$\[S Comp] then the distance from the first IC to the last will be greater, IC-to-
word ratios will be worse, and the $\bar{S}$ will be center-embedded. I discussed and
illustrated a case of this sort in ch. 3.3. Since the preposing of $\bar{S}$ constituents is
a well-established fact of Japanese, and since these $\bar{S}$s are always complement-
tizer-final and regularly much heavier than other constituents, I decided to
test EIC on some more critical data involving NPs and PPs, where the relative
length differences are smaller and relative ordering is highly variable. If EIC
can make correct predictions for these more referential nodes, then not only
will we have established its compatibility with typological diversity
(specifically its ability to make mirror-image predictions for different lan-
guage types), but we will have further motivated its relevance to constituents
whose positioning is generally explained in exclusively discourse-pragmatic
terms (cf. ch. 3.8.2). We will also have succeeded in showing that the percep-
tual difficulty of so-called “center embeddings,” by which is almost univers-
sally understood the center embedding of S, is just one of many types of
structures for which EIC defines dispreferences, and should not be singled
out for special treatment as if its explanation were somehow independent.
NPs and PPs can also be center-embedded, and when they are the degree of
dispreference rises with their length.

The NPs that I shall focus on are those that are marked with the particles
-ga, -o, and -ni. Most typically these indicate the subject, the object, and the
indirect object respectively. Examples are as follows (from Kuno 1978:
68–69):

(4.6) [Tanako-ga] [sono hon -o] katta
   “Tanako \textit{subj} that \textit{book obj} bought,” i.e.
   Tanako bought the book.

(4.7) [Taroo-ga] [Hanako-ni] [sono tegami-o] yon-da
   “Taroo \textit{subj Hanako ind-obj that letter obj} read-Past,” i.e.
   Taroo read that letter to Hanako.

As is well known, the relative positioning of these NPs is variable, while the
verb is fixed in the rightmost position of the clause. Example (4.6) is equally
grammatical in the order [\textit{son hon-o}] [Tanako-ga] \textit{katta}, and (4.7) can occur in
all six mathematically possible permutations of the three NPs. NPs with the
Topic marker -\textit{wa} were not considered, since they also have a largely fixed
(left-peripheral) position.
Because of the extreme freedom of positioning for NPs, and because the verb is final and does not enable one to anticipate the argument structure of the clause in any way, the role of case particles is crucial to the interpretation and processing of Japanese sentences. The surface syntax of NPs is fundamentally similar in appearance to that of PPs, which consist of a postposition and a left-branching NP, and to possessive phrases, consisting of the genitive particle -no preceded by an NP (example from Kuno 1978: 79):

(4.8) \[ NP[\text{PossP}[\text{Taroo-no} \ otoosan-ga]} \ PP[\text{Amerika-e} \ itta]

"Taroo Poss father subj America to went," i.e.

Taroo's father went to America.

In all these phrases it is the right-peripheral element that gives the crucial information about what type of phrase it is and that constructs it in our terms: the postposition \( e \) constructs a PP; the possessive particle constructs a PossP; and the case particles determine the type of NP that we are dealing with, which we might represent as feature-marked categories \( \text{NP} \), \( \text{Nom} \), \( \text{NP} \), \( \text{Acc} \), and \( \text{NP} \), \( \text{Dat} \) for \(-ga\), \(-o\), and \(-ni\) phrases respectively. Topic phrases marked with \(-wa\) suggest a similar analysis: a variety of phrasal types can immediately precede \(-wa\) and function as topics, and it is crucially the particle that identifies the phrase as a topic phrase (cf. Hinds 1986: 157–163). The result is consistent left-branching and consistent right-peripheral construction of the dominating phrase in all these cases, as shown in (4.9) (where Poss = no, Nom = ga, Acc = o, Dat = ni and Top = wa).

(4.9) a. \( \text{PP[NP P]} \)

b. \( \text{PossP[NP Poss]} \)

c. \( \text{NP[NP Nom], NP[NP Acc], NP[NP Dat]} \)

\( [+\text{Nom}] \quad [+\text{Acc}] \quad [+\text{Dat}] \)

d. \( \text{TopP[XP Top]} \)

The left-branching NP or XP will itself consist of left-branching ICs (such as prenominal relative clauses) in the event that there is any complexity in these nodes, which will result in favorable IC-to-non-IC ratios for these phrases in conjunction with a bottom-up parse.

I am also going to assume that there is a VP in Japanese, based on the arguments of Saito (1985) and Saito and Hoji (1983). Our data provide support for this assumption, and as a result the precise constituency assignments for NPs and PPs will need to be taken into account, with EIC ratios calculated accordingly. This poses some analytical and methodological problems for a textual investigation, since it is not always certain what the precise
constituency of each IC is in every structure of the text. Fortunately, it turns out that EIC generally makes a consistent prediction for the different attachment possibilities of two ICs preceding V: the longer IC is preferred before the shorter one across all parallel structures, just as the shorter IC is preferred before the longer IC in English (cf. table 4.1); and the degree of preference increases as the length differential increases, again as in English.

The logical possibilities are set out in (4.10):

\[(4.10) \quad \begin{align*}
a. & \text{s} [i_{IC_m} j_{IC m} v_p[V]] \\
b. & \text{v}_p [i_{IC_m} j_{IC m} V] \\
c. & \text{s} [i_{IC_m} v_p[i_{IC_m} V]] \\
d. & \text{s} [v_p[i_{IC_m} j_{IC_m} v_p[V]]]
\end{align*}\]

Both ICs may be S-dominated (4.10a), or both may be VP-dominated (4.10b). The first may be S-dominated and the second VP-dominated (4.10c). Finally, the first may be (discontinuously) VP-dominated and the second S-dominated (4.10d).

In table 4.7 I give illustrative IC-to-word ratios for all four analyses, first when there is a two-word differential \((1_{IC_m} = 3 \quad 2_{IC_m} = 5)\), and then for a four-word differential \((1_{IC_m} = 3 \quad 2_{IC_m} = 7)\). In all four analyses there is an improvement in either S or VP when the longer IC is positioned before the shorter one, and this improvement is consistently increased from the two-word to the four-word differential.

Let us consider first a Japanese structure consisting of two case-marked NPs of any of the types enumerated in (4.9c), i.e. \([NP_m \text{NP}_m V]\). NPs marked with -ga will typically be S-dominated, those marked with -o and -ni VP-dominated. Because there are several different syntactic analyses in the literature for the constituent structures assigned to alternating pairs such as \([NPga \text{NP}_ni V]\) and \([NP_ ni \text{NP}_ga V]\) I shall proceed to test EIC in two steps. First, I shall see whether there is evidence for the long-before-short patterning predicted in table 4.7, regardless of the precise constituency of each NP. That is, this first test will be theory-neutral since EIC does, in general, define a long-before-short preference for Japanese. Second, I shall examine one pair of NPs in detail, namely \([\{NPga \text{NP}_o\} V]\), testing EIC's predictions against each of the competing syntactic analyses that have been proposed for this alternation. This will result in a more precise test, with interesting theoretical consequences for the grammatical analyses involved.
Table 4.7 *IC*-to-*word* ratios for Japanese $\{IC_m, IC_m\} V$\(^6\)

Assume: $V$ constructs VP; 
$V = 1$ word

<table>
<thead>
<tr>
<th></th>
<th>$IC_m = 3$</th>
<th>$IC_m = 5$</th>
<th>$IC_m = 7$</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>$[2IC_m, IC_m V]$</td>
<td>$[1IC_m, 2IC_m V]$</td>
<td>$[1IC_m, 2IC_m V]$</td>
</tr>
<tr>
<td>a.</td>
<td>$s[2IC_m, IC_m V]$</td>
<td>$s[1IC_m, 2IC_m V]$</td>
<td>$s[1IC_m, 2IC_m V]$</td>
</tr>
<tr>
<td>S CRD:</td>
<td>3/5 = 60%</td>
<td>S CRD: 3/7 = 43%</td>
<td>S differential = 17%</td>
</tr>
<tr>
<td>VP CRD:</td>
<td>1/1 = 100%</td>
<td>VP CRD: 1/1 = 100%</td>
<td></td>
</tr>
<tr>
<td>Aggregate = 80</td>
<td>Aggregate = 71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>$vp[2IC_m, IC_m V]$</td>
<td>$vp[1IC_m, 2IC_m V]$</td>
<td></td>
</tr>
<tr>
<td>VP CRD:</td>
<td>3/5 = 60%</td>
<td>VP CRD: 3/7 = 43%</td>
<td>VP differential = 17%</td>
</tr>
<tr>
<td>S CRD:</td>
<td>2/5 = 40%</td>
<td>S CRD: 2/7 = 29%</td>
<td>S differential = 11%</td>
</tr>
<tr>
<td>VP CRD:</td>
<td>2/2 = 100%</td>
<td>VP CRD: 2/2 = 100%</td>
<td></td>
</tr>
<tr>
<td>Aggregate = 70</td>
<td>Aggregate = 64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>$s[1IC_m, vp[1IC_m V]]$</td>
<td>$s[1IC_m, vp[2IC_m V]]$</td>
<td></td>
</tr>
<tr>
<td>S CRD:</td>
<td>2/5 = 40%</td>
<td>S CRD: 2/7 = 29%</td>
<td>S differential = 11%</td>
</tr>
<tr>
<td>VP CRD:</td>
<td>2/2 = 100%</td>
<td>VP CRD: 2/2 = 100%</td>
<td></td>
</tr>
<tr>
<td>Aggregate = 70</td>
<td>Aggregate = 64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td>$s[vp[2IC_m, IC_m V]]$</td>
<td>$s[vp[1IC_m, 2IC_m V]]$</td>
<td></td>
</tr>
<tr>
<td>S CRD:</td>
<td>2/2 = 100%</td>
<td>S CRD: 2/7 = 29%</td>
<td>S differential = 11%</td>
</tr>
<tr>
<td>VP CRD:</td>
<td>2/5 = 40%</td>
<td>VP CRD: 2/7 = 29%</td>
<td>VP differential = 17%</td>
</tr>
<tr>
<td>Aggregate = 70</td>
<td>Aggregate = 64</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data on the relative positioning of any two case-marked NP\(_m\)s are given in table 4.8 section I. The structural template used requires comment. This template allows for an optional Topic phrase on the left. It also allows for an optional NP\(_ga\) to the left of the two NPs whose relative positioning is being considered. So, in the structure in example (4.7) the -o and the -ni phrases would be the two NPs immediately preceding the verb, and Taroo-ga would be the optional subject to their left. In example (4.6) the two NPs immediately preceding V would be Tanako-ga and sono hon-o, i.e. this -ga...
Table 4.8 Japanese $\{_{1}NP_{m}^{2}NP_{m}\} V$

150 pages of data: first 70 pp. from Juugo Kuroiwa *Yami No Hada*, Koobunsha Tokyo; first 40 pp. from Yasuyoshi Kobakura *Okinawa Monogatari*, Shinchoosha Tokyo; first 40 pp. from Isamu Togawa *Konoe Fumimaro To Juushintachi*, Kodansha Tokyo.

$[(\text{TopP})\,\{\{S \geq 1\} (NP_{ga})\} \,\{_{1}NP_{m}^{2}NP_{m}\} \,V]\}$ (no other ICs in the clause apart from optional adverbials in any position)

where $NP_{m} = \text{an NP with a final } -ga, -o, \text{ or } -ni \text{ particle}$

$\{_{1}NP_{m}^{2}NP_{m}\} \,V\} \text{ for NPs with any combination of particles}$

<table>
<thead>
<tr>
<th></th>
<th>$n = 121$</th>
<th>$1NP_{m} = 2NP_{m}$</th>
<th>$2NP_{m} &gt; 1NP_{m}$</th>
<th>:1–2 words</th>
<th>:3–6</th>
<th>:7+</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>$[2NP_{m} ,1NP_{m}]$</td>
<td>23</td>
<td>15</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>$[1NP_{m} ,2NP_{m}]$</td>
<td>51</td>
<td>19</td>
<td>5</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Ratio of $Y/X$:
- EIC predictions (assuming no VP discontinuity; consistent long-before-short predictions)
  - Unmarked case: $97/121 = \text{most optimal } (X = +)$, i.e. 80.2% Correct
  - Marked case: ratio of $Y/X$ for $2NP_{m} > 1NP_{m}$ :1–2 $\geq$ 3–6 $\geq$ 7+ All correct
II [{NP\text{ga}} NP\text{o}] V

n = 85

<table>
<thead>
<tr>
<th></th>
<th>NP\text{o} &gt; NP\text{ga} :1–4 words</th>
<th>NP\text{ga} = NP\text{o}</th>
<th>NP\text{ga} &gt; NP\text{o} :1–4</th>
<th>:5+</th>
</tr>
</thead>
<tbody>
<tr>
<td>X [NP\text{ga} NP\text{o}]</td>
<td>15</td>
<td>36</td>
<td>18</td>
<td>12</td>
</tr>
<tr>
<td>Y [NP\text{o} NP\text{ga}]</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Ratio of Y/X: 17% 3% 0% 0%

\textit{EIC predictions (1)}: assume s[NP\text{ga} NP\text{o} V] alternates with s[NP\text{o} NP\text{ga} V]
Optimal orders: X where NP\text{ga} \geq NP\text{o}
Y where NP\text{o} \geq NP\text{ga}
Unmarked case: 70/85 = most optimal, i.e. 82%
Marked case: ratio of Y/X for NP\text{ga} > NP\text{o} :1–4 \geq 5+
       ratio of X/Y for NP\text{o} > NP\text{ga} :1–4 (=83%) \geq ratio
       of Y/X for NP\text{ga} > NP\text{o} :5+ (=0%)
Correct

\textit{EIC predictions (2)}: assume s[NP\text{ga} vP[NP\text{o} V]] alternates with s[vP[NP\text{o}] NP\text{ga} vP[V]]
Optimal orders: X where NP\text{ga} \geq NP\text{o} and NP\text{o} > NP\text{ga} :1–8 words
Y where NP\text{o} > NP\text{ga} :9+ words
Unmarked case: 81/85 = most optimal, i.e. 95%
Marked case: ratio of Y/X for NP\text{o} > NP\text{ga} :1–4 \geq (NP\text{ga} = NP\text{o})
       \geq (NP\text{ga} > NP\text{o})
       All correct

\textit{EIC predictions (3)}: assume s[NP\text{ga} vP[NP\text{o} V]] alternates with s[NP\text{o} NP\text{ga} vP[V]]
Optimal orders: X where NP\text{ga} > NP\text{o}
Y where NP\text{o} \geq NP\text{ga}
Unmarked case: 34/85 = most optimal, i.e. 40%
Marked case: e.g. ratio of X/Y for NP\text{ga} = NP\text{o} (=97%) \geq (NP\text{o} > NP\text{ga}) (=83%)
       ratio of Y/X for NP\text{ga} > NP\text{o} :1–4 \geq ratio
       of X/Y for NP\text{o} > NP\text{ga} :1–4
Correct

\textit{EIC predictions (4)}: assume s[NP\text{ga} vP[NP\text{o} V]] alternates with s_1[NP\text{o} s_2[NP\text{ga} vP[V]]]
Optimal orders and predictions as for EIC predictions (3).
Testing EIC's performance predictions

phrase would now be one of the NPs designated as \( _1 \text{NP}_m \) or \( _2 \text{NP}_m \), and an additional possible NP\text{(a)} preceding them would be empty. The template also permits one or more embedded clauses to precede the two NPs in our test, as well as optional adverbials in any position.

Each of the particles -\text{ga}, -\text{o}, -\text{ni}, and -\text{no} was counted as a separate word for the purpose of IC-to-word ratios, on account of the surface parallelism with postpositions, cf. (4.9). It was also decided to make some greater concession to the existence of grammatical structure within the verb and adjective as well, by counting tense morphemes such as -\text{da} in (4.7) as separate words. Taroo-ga was therefore counted as two words and so was yon-da. All other grammatical particles, such as conjunction particles and the sentence-final question particle -\text{ka}, were also counted as single words.

The unmarked case prediction is supported in table 4.8 section I with a score of 80.2%. All the marked case predictions are also supported: the tolerance for non-optimal (short before long) orders with a one-to-two-word differential is 45%; for three to six words this drops to 25%; and for seven or more words there are no non-optimal orders. These data reveal a clear preference for long-before-short ordering, especially for length differentials in excess of three words: 23/28 of these structures (or 82%) have long-before-short ordering, while 100% of the orders with differentials in excess of seven words have long before short. This pattern is the exact opposite of the English one exemplified in table 4.2.

These figures support EIC. However, the overall success rate of 80% for the unmarked case predictions is lower than the results of previous sections, and the 45% tolerance for short-before-long orders with one-to-two-word differentials is higher than we should expect from a consistent long-before-short prediction. This could be for reasons of grammaticalization. For example, [NP\text{ga} NP\text{o}] could be the basic order of Japanese, with NP\text{o} preposing having the status of a rearrangement rather than an arrangement (cf ch. 3.5.2). I will show in section 4.3 that such grammaticalizations are visible in performance data precisely when there are small length differentials between ICs, and when the basic order is retained in the face of weak contrary EIC preferences for a rearrangement. That is, short-before-long [NP\text{ga} NP\text{o}] sequences would be retained even when EIC preferred the reverse. The trouble with this explanation is that the overall success rate for performance data should still be higher than 80% even when there is a basic grammatical order. In other clear cases of basic ordering (e.g. [V NP PP] in the English VP alternating with a rearranged [V PP NP], cf. section 4.3.2), the overall number of optimal EIC orders in performance is higher.
A second possibility is to take seriously the VP assumption for Japanese. Notice that if there is no VP, then all the NP_m's in table 4.8 section I will be contained in a flat structure, s[{NPga NPo} V], s[{NPo NPni} V], etc.; there will be a consistent long-before-short prediction for all NP_m pairs; and we will be back to good, but not great, EIC predictions. On the other hand, with the VP, the structures will be different: an NPga preceding NPo within a transitive sentence will have the structure s[NPga VP[NPo V]]; NPo and NPni will both be VP-dominated; and so on. EIC's predictions will differ accordingly, and will even include (on one analysis) a preference for short-before-long orders in a small but important subset of the data.

Table 4.8 section II sets out the data for sequences of NPga and NPo. What is at first surprising is that there is a massive preference for SOV over OSV: 81/85 (or 95%) have [NPga NPo V], and just four have the reverse. EIC's predictions for this distribution are then tested on the basis of four distinct sets of structural assumptions. Prediction (1) assumes a flat S structure for both orders. Predictions (2)-(4) assume a VP. Prediction (2) assumes no constituency change when NPo is preposed, i.e. s[NPga VP[NPo V]] alternates with a discontinuous VP, s[VP[NPo] NPga VP[V]]. This discontinuous analysis has been proposed for the corresponding sentences of Korean by O'Grady (1987), who argues that the fronted direct object still behaves like a VP constituent from the point of view of anaphoric binding and quantifier scope. Prediction (3) assumes that the fronted NPo is not discontinuous but is attached to S, i.e. s[NPo NPga VP[V]]; while prediction (4) follows Saito (1985) in Chomsky-adjoining NPo to the original S dominating NPga and VP, i.e. s1[NPo s2[NPga VP[V]]].

Under each prediction I define the orders that are optimal for EIC according to that structural analysis. For prediction (1) the optimal orders are consistently long before short or equals. The success rate for the unmarked-case predictions is a respectable 82%, and the marked case predictions are correct as well. Assessing the optimal orders in predictions (2)-(4) is more complicated, and we need to run some more EIC calculations to supplement those of table 4.7. This is done in table 4.71.

On the left-hand side of table 4.71 I give the IC-to-word ratios for SOV structures, assuming a VP and with different word totals assigned to NPga and NPo. On the right-hand side I give the ratio for the corresponding OSV, in each of the three structures that have been proposed, labeled (2), (3), and (4) respectively, corresponding to the predictions of table 4.8 section II. Since the ratios for both the S and VP domains are now affected across these competing orders, all EIC comparisons are made in terms of aggregate
Table 4.7: IC-to-word ratios for Japanese \[{NPga \ NPo} V\]^7

Assume: -\textit{ga} constructs \(NPga\) and \(S\);
- \textit{o} constructs \(NPo\) and \(VP\) whenever \(VP\) dominates \(NPo\);
\(V = 1\) word

<table>
<thead>
<tr>
<th></th>
<th>(NPga)</th>
<th>(NPo)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>(s[NPga \ V])</td>
<td>(s[NPo \ NPga \ V])</td>
</tr>
<tr>
<td></td>
<td>S CRD: 2/3 = 67%</td>
<td>S CRD: 2/11 = 18%</td>
</tr>
<tr>
<td></td>
<td>VP CRD: 2/2 = 100%</td>
<td>VP CRD: 2/12 = 17%</td>
</tr>
<tr>
<td></td>
<td>Aggregate = 83.5</td>
<td>Aggregate = 17.5</td>
</tr>
<tr>
<td></td>
<td>Agg. differential = (-)66</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(s[NPo \ NPga \ V])</td>
<td>(s[NPga \ NPo \ V])</td>
</tr>
<tr>
<td></td>
<td>S CRD: 3/12 = 25%</td>
<td>S CRD: 2/11 = 18%</td>
</tr>
<tr>
<td></td>
<td>VP CRD: 1/1 = 100%</td>
<td>VP CRD: 2/2 = 100%</td>
</tr>
<tr>
<td></td>
<td>Aggregate = 62.5</td>
<td>Aggregate = 73</td>
</tr>
<tr>
<td></td>
<td>Agg. differential = (-)21</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Aggregate = 83.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2) Aggregate = 27</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3) Aggregate = 69</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4) Aggregate = 76</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Agg. differential = (-)56.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Agg. differential = (-)14.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Agg. differential = (-)7.5</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Aggregate = 83.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2) Aggregate = 36.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3) Aggregate = 75</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4) Aggregate = 80</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Agg. differential = (-)47</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Agg. differential = (-)8.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Agg. differential = (-)3.5</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Aggregate = 83.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2) Aggregate = 58.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3) Aggregate = 87.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4) Aggregate = 89</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Agg. differential = (-)25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Agg. differential = (+)4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Agg. differential = (+)5.5</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Aggregate = 70</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2) Aggregate = 58.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3) Aggregate = 87.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4) Aggregate = 89</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Agg. differential = (-)11.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Agg. differential = (+)17.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Agg. differential = (+)19</td>
<td></td>
</tr>
<tr>
<td>VI</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Aggregate = 64.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2) Aggregate = 58.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3) Aggregate = 87.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4) Aggregate = 89</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Agg. differential = (-)36</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Agg. differential = (+)23</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Agg. differential = (+)24.5</td>
<td></td>
</tr>
<tr>
<td>VII</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Aggregate = 61</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2) Aggregate = 58.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3) Aggregate = 87.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4) Aggregate = 89</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Agg. differential = (-)2.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Agg. differential = (+)26.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Agg. differential = (+)28</td>
<td></td>
</tr>
<tr>
<td>VIII</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Aggregate = 59</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2) Aggregate = 58.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3) Aggregate = 87.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4) Aggregate = 89</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Agg. differential = (-)30.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Agg. differential = (+)28.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Agg. differential = (+)30</td>
<td></td>
</tr>
<tr>
<td>IX</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Aggregate = 57.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2) Aggregate = 58.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3) Aggregate = 87.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4) Aggregate = 89</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Agg. differential = (+)1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Agg. differential = (+)30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Agg. differential = (+)31.5</td>
<td></td>
</tr>
</tbody>
</table>
percentages for these nodes. In the case of prediction (4), there are two S
nodes to be considered, plus VP; hence the aggregate is now derived from
three nodes. A minus sign in the aggregate differential column, e.g. (−) 21,
means that the OSV structure on the right has a lower aggregate than SOV on
the left; a plus means that it has a higher aggregate.

Proceeding down the table, the subject is first assigned significantly more
words than the object. This difference is gradually decreased until both are
equal (see IV), and the direct object is then gradually assigned more words
than the subject. There are two striking patterns in table 4.71. The first is that
the two OSV structures of predictions (3) and (4), \( s[NPo \, NPga \, VP[V]] \) and
\( s[NPo \, s2[NPga \, VP[V]] \), have almost identical scores, i.e. it doesn’t make
much difference, from the point of view of EIC, which one we assume.
They both define a consistent preference for the longer NP\textsubscript{m} to the left,
and they even define a small preference for OSV when the two NP\textsubscript{m}s are
of equal length. The second is that the discontinuous VP analysis of predic-
tion (2) makes a very different prediction. It defines a much greater prefer-
ence for SOV than the other two analyses when the subject is longer than the
object. It still defines a strong preference for SOV when both are equal. And
even when the object is longer than the subject by up to eight words, SOV is
still preferred over OSV, albeit with ever-diminishing differentials (V–VIII).
Only when the object is more than eight words longer than the subject is OSV
preferred (IX).

Throughout these calculations each CRD proceeds from the first IC domi-
nated by the mother node to the last, as usual. The case particles are crucial in
indicating which type of NP we are dealing with, whether a subject or an
object, i.e. they construct the appropriate feature-marked NP (cf. (4.9c)),
this NP being immediately dominated by S or VP. In an S domain consist-
ing of \([NPga \, VP] \), the CRD for S proceeds from the word that constructs
the first IC, namely \( ga \), to the word that constructs the VP. If the VP
consists of \( VP[NPo \, VI] \), the CRD for VP proceeds from \( o \), which constructs
the first IC, to \( V \).

The manner in which VP is constructed requires explicit comment. In table
4.7 and in n. 7 ch. 4 attached to table 4.71, I ran the calculations on the
assumption that V alone constructs VP, just as it does in English. This is
undoubtedly one possibility, and is arguably the only possibility in a linear
sequence such as \( s[s[S \, Comp] \, NPga \, V] \). An \( S \) can be dominated by a variety
of nodes and does not permit the parser to assume any higher dominating
structure on-line. NP\textsubscript{ga} will be S-dominated, and I will argue in ch. 6.2.2 that
\( ga \) will construct not just the subject NP (by MNC), but S as well, by the
principle of Grandmother Node Construction. V then constructs VP, and this will give an optimal IC-to-word ratio for the S domain. If the subcategorization frame for V requires VP domination for the S, the resulting discontinuity $s[VP][S\ Comp][NP\ ga\ VP][V]$ will have a better EIC ratio than the reverse $s[NP\ ga\ VP][S\ Comp][V]$, with a center-embedded S (cf. ch. 5.4.1 for precise figures). But whereas a complementizer does not uniquely identify a dominating VP over S, an $o$ particle can be argued to function like $ga$: it constructs an NP, by MNC; and it constructs a VP above that, by Grandmother Node Construction. In syntactic analyses in which NP0 is always VP-dominated (i.e. in analyses permitting VP discontinuity), NP0 will always construct a higher VP. In other analyses, it may (or may not) construct VP on-line, whenever it is VP-dominated, e.g. in the order SOV.

For the purpose of the calculations in table 4.7, I have assumed that $o$ does construct VP when it is VP-dominated. This has the effect of slightly improving the SOV scores, because the VP is now recognized one word in advance of V. So, in table 4.7 section I ($NP_{ga} = 10\ NP_{o} = 2$) the SOV aggregate is 83.5 on this assumption; it would be 75 if V were to construct VP within the S domain (i.e. S CRD: 2/4 = 50%, VP CRD: 2/2 = 100%). It also has the effect of lowering the scores in the discontinuous VP structure, compared to VP construction by V alone, since the S domain now proceeds from $o$ to $ga$, being initiated by a fronted NP0 that constructs VP. Other smaller differences in the IC-to-word ratios for the structures of table 4.7, assuming VP construction by V alone, are illustrated in n. 7 ch. 4. The important point is, however, that these different assumptions do not substantially affect EIC's predictions. If the VP is constructed by V alone, the predictions made by analyses (3) and (4) in table 4.8 section II turn out to be worse. The discontinuous analysis (2) remains most optimal, though its success rate is now equal to that of the flat structure (1), with 82% of the orders being correctly predicted in the unmarked case (i.e. 70/85 are most optimal). The analysis and assumptions that achieve the best EIC results (of 95%) are, therefore, the discontinuous analysis in conjunction with the assumption that $o$ can construct VP, cf. table 4.8 section II.

The reason why the discontinuous analysis makes a different set of predictions from analyses (3) and (4) is the following. All the CRDs are going to be improved, in general, when a longer ICm precedes a shorter one. But the discontinuity of the VP means that the preference for OSV will only assert itself when there is a very long object and a very short subject. Consider first section I in table 4.7. In the SOV order, the two ICs of VP are recognized in an optimal two words, and the CRD for S proceeds from $ga$ to $o$, which gives
a score of 67% when NPo = 2. For OSV, however, the CRD for S proceeds from o, when o constructs VP in prediction (2), to ga, which constructs the subject NPga, i.e. 2/11 = 18%; and the discontinuous VP proceeds from o to V, i.e. 2/12 = 17%. Both domains have poor scores when the subject is long, therefore. But when the object is very long and the subject short, as in section IX, the SOV order will have a very poor ratio for the CRD for S (15%), and even though the VP ratio remains optimal, the combined scores for the S domain and the discontinuous VP in the OSV order will be reasonable (58.5), and sufficient to just offset the SOV score (57.5). Hence the discontinuous analysis predicts that there will be a subset of these word-total assignments to ICs in which a shorter NPga precedes a longer NPo, namely for NPo > NPga by one to eight words, according to the illustrative calculations of table 4.7.

The data of table 4.8 section II reveal that a full 15 of the 24 “exceptional” Y structures of section I, i.e. with a shorter NPm before a longer one, are accounted for in this way: they involve [NPga NPo V] where NPo exceeds NPga by one to four words. Hence, SOV order is actually predicted by EIC in these cases, in conjunction with the discontinuity analysis. The full predictions of this analysis are set out under EIC predictions (2). The unmarked case scores 95%, which is significantly higher than the 82% of prediction (1), assuming a flat structure. The marked case predictions of the discontinuous analysis are also correct. Predictions (3) and (4), by contrast, do not score at all well. The unmarked case predictions score only 40%, and the marked case predictions are mixed.

I conclude that the alternating [NPga NPo V]/[NPo NPga V] orders of Japanese provide further support for EIC’s predictions. They also provide performance evidence for grammatical structure: they support the existence of a VP; and they support an analysis in which a preposed NPo is still discontinuously attached to this VP. They do not support analyses in which the fronted NPo is attached to S, with or without Chomsky-adjunction. Flat structures are reasonably compatible with these data, but analyses that assume a VP plus discontinuity (plus on-line construction of the VP by case particles) fare best of all.

Table 4.9 presents data involving two PPms immediately preceding a verb, with an optional NPga and other optional constituents in the positions indicated in the structural template. These PPs will either be VP- or S-dominated, but no attempt was made to distinguish the two in these data. If both PPs are S-dominated, or both are VP-dominated, EIC defines a consistent long-before-short prediction. If one is S-dominated and the other VP-domi-
nated, the predominant preference will again be long before short, but there will be some structures in which short before long is preferred, if we assume VP discontinuity. Given \( s[1PP_m \ VP[2PP_m V]] \), an alternative \( s[2PP_m 1PP_m \ VP[V]] \) with the longer \( 2PP_m \) raised out of the VP and into S preceding the shorter \( 1PP_m \) will have a better EIC ratio. But if the alternative is \( s[VP[2PP_m] 1PP_m \ VP[V]] \), the original continuous VP will have the better ratio, with the shorter \( 1PP_m \) before the longer \( 2PP_m \), for the same reason that SOV is sometimes preferred to OSV under these circumstances. Despite this possibility, a full 85% of the orders in table 4.9 are most optimal, and the marked case predictions are correct as well.

Table 4.10 considers the relative positioning of an \( NP_m \) and a \( PP_m \). The \( NP_m \)s selected for this test involve those that are case-marked with \( o \), and hence VP-dominated. The \( PP_m \)s may again be potentially S- or VP-dominated. In section I of table 4.10 I collapse \( NP_o \) and \( PP_m \) into \( 2IC_m \) or \( iIC_m \), depending on their relative size (\( 2IC_m \geq 1IC_m \), as usual, cf. n. 2. ch. 4), and test EIC’s long-before-short prediction, assuming no discontinuity. The unmarked case prediction is satisfied in 82.4% of all instances, and all the marked case predictions are correct.

Section II gives a more fine-tuned syntactic analysis and presents the word-total differentials for \( NP_o \) and \( PP_m \) separately, without collapsing them. There are two possible structural alternations in these data. If the \( PP_m \) is VP-dominated, then we are dealing with a VP-internal alternation of \( VP[PP_m NP_o V] \) with \( VP[NP_o PP_m V] \), both of which I shall assume to be flat

Table 4.10

<table>
<thead>
<tr>
<th>( NP_o = PP[NP P] )</th>
<th>( PP_m )</th>
<th>( 2PP_m &gt; 1PP_m )</th>
<th>( :2-5 )</th>
<th>( :6+ )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X ) ( 2PP_m 1PP_m )</td>
<td>4</td>
<td>18</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>( Y ) ( 1PP_m 2PP_m )</td>
<td>2</td>
<td>8</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Ratio of ( Y/X )</td>
<td>33%</td>
<td>31%</td>
<td>0%</td>
<td></td>
</tr>
</tbody>
</table>

EIC predictions (assuming no VP discontinuity; consistent long-before-short predictions)

Unmarked case: \( 56/66 = \text{most optimal (X+ =), i.e. 84.8%} \) Correct
Marked case: ratio of \( Y/X \) for \( 2PP_m > 1PP_m :1 \geq 2-5 \geq 6+ \) All correct
Structures. If the PP\textsubscript{m} is S-dominated, then the basic order of Japanese will be\textsc{s}\[PP\textsubscript{m} \text{vp} [NP\textsubscript{o} V]], given the rightward positioning of the VP. Because of the evidence for VP discontinuity in the data of table 4.8 when NP\textsubscript{o} is preposed before NP\textsubscript{ga}, I shall assume discontinuity in this context as well when NP\textsubscript{o} precedes an S-dominated PP\textsubscript{m}, i.e. \textsc{s}\[vp[NP\textsubscript{o}] PP\textsubscript{m} vp[V]]. These two structural alternations are labeled (A) and (B) respectively.

EIC's preferences are similar, but not identical, across the two alternating orders. When PP\textsubscript{m} > NP\textsubscript{o}, X [PP\textsubscript{m} NP\textsubscript{o}] is consistently preferred in both (A) and (B), and this preference increases with larger differentials. When PP\textsubscript{m} = NP\textsubscript{o}, the orders X [PP\textsubscript{m} NP\textsubscript{o}] and Y [NP\textsubscript{o} PP\textsubscript{m}] are equally preferred in the (A) alternation, and X is preferred in the (B) alternation (recall the structurally identical data of table 4.7 section IV). For NP\textsubscript{o} > PP\textsubscript{m}, the preferences in the two alternations are partially at odds with each other: the Y order is consistently preferred in (A), this preference increasing with larger differentials; X is preferred in (B) where NP\textsubscript{o} > PP\textsubscript{m}: 1–8, the preference decreasing with larger differentials; and (the discontinuous) Y is preferred where NP\textsubscript{o} > PP\textsubscript{m}: 9+, the preference increasing with larger differentials.

Because alternations (A) and (B) involve partially different preferences when NP\textsubscript{o} > PP\textsubscript{m}, we need to first remove these structures (53 in all) from the data. We can then test EIC's quantitative predictions on the remaining 191, in which PP\textsubscript{m} > NP\textsubscript{o}, and for which the predictions are clear. For PP\textsubscript{m} > NP\textsubscript{o} both alternations prefer X. For PP\textsubscript{m} = NP\textsubscript{o} X or Y are equally preferred in (A), while X is preferred in (B). I do not know how many of the 91 instances of PP\textsubscript{m} = NP\textsubscript{o} in the data involve alternation (A) or (B). But since both sets of orders are optimal for EIC in analysis (A), and since all of these data can potentially be optimal even in analysis (B) (in the event that the (B) structures have the non-discontinuous order X), I shall make the simplifying assumption that all of these PP\textsubscript{m} = NP\textsubscript{o} orders are as optimal as they could possibly be, and that they join the X structures where PP\textsubscript{m} > NP\textsubscript{o} in defining EIC's unmarked case predictions. With this assumption, EIC scores a full 91% correct. All the marked case predictions are correct as well for PP\textsubscript{m} > NP\textsubscript{o}.

The remaining NP\textsubscript{o} > PP\textsubscript{m} structures involve partially different ordering preference in alternations (A) and (B), as we have seen. Nonetheless, there is an interesting, and testable, complementarity in their interaction: for both alternations the relative EIC score for Y [NP\textsubscript{o} PP\textsubscript{m}] increases with larger differentials, and that for X [PP\textsubscript{m} NP\textsubscript{o}] increases with smaller differentials. This means that we should see a progressive decline in the ratio of Y/X from
Table 4.10 *Japanese* \( \langle \{NPo \ PP_m \} \ V \rangle \)

150 pages of data: cf. table 4.8

\[(\text{TopP}) \: [(s \geq 1) \: \langle \{NPo \ PP_m \} \ V \rangle] \quad \text{(no other ICs in the clause apart from optional adverbials in any position)}\]

where \( PP_m = \text{PP[NP P]} \)

<table>
<thead>
<tr>
<th></th>
<th>( {2IC_m \ 1IC_m } \ V )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n = 244 )</td>
<td>( \text{NPo} = PP_m )</td>
</tr>
<tr>
<td><strong>X [2IC_m 1IC_m]</strong></td>
<td>91</td>
</tr>
<tr>
<td><strong>Y [1IC_m 2IC_m]</strong></td>
<td>30</td>
</tr>
<tr>
<td>Ratio of Y/X</td>
<td>34%</td>
</tr>
</tbody>
</table>

*EIC predictions* (assuming no VP discontinuity; consistent long-before-short predictions)

Unmarked case: \( 201/244 = \text{most optimal (X+ =)}, i.e. 82.4\% \quad \text{Correct} \)

Marked case: ratio of Y/X for \( 2IC_m > 1IC_m : 1-2 \geq 3-4 \geq 5-8 \geq 9+ \quad \text{All correct} \)
II \(\{PP_m \ NPo\} \ V\)

\[
\begin{array}{cccccccc}
\text{n = 244} & \text{NPo > PP}_m & :5+ & \text{words} & :3-4 & :1-2 & \text{PP}_m = \text{NPo} & \text{PP}_m > \text{NPo} :1-2 & :3-8 & :9+ \\
\hline
\text{X} [PP_m \ NPo] & 3 & 5 & 18 & 60 & 48 & 26 & 9 \\
\text{Y} [NPo PP_m] & 11 & 5 & 11 & 31 & 12 & 5 & 0 \\
\hline
\end{array}
\]

Ratio of \(Y/\ X\):

- 79% for \(X\)
- 50% for \(Y\)
- 38% for \(X\)
- 34% for \(Y\)
- 20% for \(X\)
- 16% for \(Y\)
- 0% for \(Y\)

Assume: VP; VP dominates NPo; possible VP discontinuity; 2 possible structural alternations;
(A) \(v_p[PP_m \ NPo \ V] (= X)\) alternates with \(v_p[NPo \ PP_m \ V] (= Y)\)
(B) \(s[PP_m \ VP[NPo \ V]] (= X)\) alternates with \(s[VP[NPo \ PP_m \ VP[V]] (= Y)\)

**EIC's preferences**
- \(PP_m > NPo\): \(X\) in (A) and (B), increasing with larger differentials;
- \(PP_m = NPo\): \(X\) or \(Y\) in (A), \(X\) in (B);
- \(NPo > PP_m\): \(Y\) in (A), increasing with larger differentials;
  - \(X\) in (B) where \(NPo > PP_m :1-8\), decreasing with larger differentials;
  - \(Y\) in (B) where \(NPo > PP_m :9+\), increasing with larger differentials;
- i.e. \(Y\) preference increases with larger differentials, \(X\) with smaller.

**EIC's predictions**
- \(PP_m \geq NPo\)
  - Unmarked case: \(174/191 = \text{most optimal (X+ =), i.e. 91%}\) Correct
  - Marked case: ratio of \(Y/X\) for \(PP_m > NPo :1-2 \geq 3-8 \geq 9+\) All correct
- \(NPo > PP_m\)
  - Ratio of \(Y/X\) for \(NPo > PP_m :5+ \geq 3-4 \geq 1-2\) All correct
left to right, i.e. from $\text{NPO} > \text{PP}_m$: 5+ to 3–4 to 1–2, and we do. Since the $X$ order is progressively preferred for $\text{PP}_m > \text{NPO}$ as well, the ratio of $Y/X$ continues to decline throughout the chart.

Despite the difficulties of dealing with a structural ambiguity in this case, the two possible structural alternations make either similar or complementary predictions in conjunction with EIC, and EIC's success rate seems to be good. Once again both VP and VP discontinuity are supported by the productivity of the short-before-long structures where $\text{NPO} > \text{PP}_m$. These (X) structures can be explained on the grounds that they resist conversion to a discontinuous VP structure (Y) when the word-total differential is insufficient to motivate the discontinuity. These exceptions to the predominant long-before-short pattern are not matched by correspondingly productive exceptions when $\text{PP}_m > \text{NPO}$, i.e. by structures of type $Y$, because long-before-short structures (X) are now preferred in both analyses without having to assume any VP discontinuity. Short before long is never preferred, therefore.

Finally, notice that instead of structural alternation (A), involving a flat VP structure, it might be argued that the direct object forms a constituent with the verb within the VP, and that the $X$ order should be represented as $\text{VP}[\text{PP}_m v'[\text{NPO} \text{ V}]]$. If $\text{NPO}$ continues to be dominated by $V'$ when preposed before $\text{PP}_m$, this alternation would also involve a discontinuous structure, i.e. $\text{VP}[v'[\text{NPO} \text{ PP}_m v'[\text{V}]]]$. Alternatively, $\text{NPO}$ might be raised out of $V'$, i.e. $\text{VP}[\text{NPO PP}_m v'[\text{V}]]$.

It is reasonable to propose, based on all the evidence we have considered so far, that Japanese has discontinuous structures and that $\text{NPO}$ consistently constructs its immediate mother node, VP or possibly $V'$, regardless of positioning. If structural alternation (A) does not involve a flat structure, however, but rather $\text{VP}[\text{PP}_m v'[\text{NPO} \text{ V}]]$ alternating with a discontinuous $V'$, then both structural alternations would involve discontinuity. EIC would then make the same predictions for both alternations. In particular, $X [\text{PP}_m \text{ NPO}]$ would be preferred in both analyses when $\text{NPO} > \text{PP}_m$: 1–8, and $Y$ thereafter. But nothing would then motivate the relative productivity of the $Y$ structures, which is predicted by the flat structure analysis. I calculate that the total distribution of $Y/X$ when $\text{NPO} > \text{PP}_m$: 1–8 is $27/53 = 51\%$. The distribution of $Y/X$ when $\text{PP}_m > \text{NPO}$ is just $17/100$, or $17\%$. This skewing in favor of $Y$ is what we would expect if one structural alternation favors $Y$, while the other favors $X$, and provides support for the flat structure analysis of VP, which we have assumed hitherto.
Korean \[\{1IC_m, 2IC_m\} V\]

Korean is similar to Japanese in its word-order permutation possibilities, and in the fixed final position of its verb. In a sentence such as (4.11), containing a three-place predicate, the verb is final and all six mathematically possible relative orderings of NPs are grammatical, just as they are in the corresponding sentence type (4.7) in Japanese (example from N.-K. Kim 1987: 894):

\[(4.11) \ [John \ i] \ [Mary \ eke] \ [chæk \ ël] \ cu-əs'-ta \ \\
"John \ NOM \ Mary \ DAT \ book \ ACC \ gave," \ i.e.
John gave a book to Mary.

Also as in Japanese, Korean employs the case particles illustrated in (4.11), which make its NP structure parallel to PP, PossP and TopP, all of which are constructed on their right peripheries. The arguments given in the last section for assigning the structures of (4.9) to Japanese apply equally to Korean. Hence, all of these phrases are of the type IC_m.

Korean also provides evidence for a VP and for VP discontinuity when direct objects precede subjects. Data corresponding to Japanese \([\{NPga \ NPo\} \ V]\) (cf. table 4.8 section II) were collected and analyzed for Korean by Charles Kim. The results are presented in table 4.11 and are strikingly similar to those for Japanese. EIC's predictions are tested on the basis of the same four theoretical analyses of the structural alternation between \([NPi \ NPu/1 \ V]\) and \([NPu/1 \ NPi \ V]\). The flat structure analysis of prediction (1) does less well than its Japanese counterpart, scoring just 69% for the unmarked case predictions (compared with 82%), and achieving mixed results in the marked structures. Prediction (2), assuming a VP and discontinuity, scores an identical 95% in the unmarked case. Predictions (3) and (4), assuming S-attachment of NPu/1, do just as badly as in Japanese. These Korean data, like their Japanese counterparts, provide independent support for the structural assumptions upon which prediction (2) is based. In addition, the flat structure now receives 13% less support than in Japanese, putting its overall success rate down to a level (69%) at which we can no longer say that the optimal structures are attested in the unmarked case. Korean therefore provides an argument against an analysis that is still possible, but not preferred, in Japanese.

Korean also appears to provide evidence for the grammaticalization of its direct and indirect object orders (accusative and dative-marked respectively) in VP. The order indirect object before direct object, as in (4.11), is much preferred to the reverse. With the ordering of (4.11), indirect objects were
Table 4.11 *Korean* \{NPi NPål} \(V\)

<table>
<thead>
<tr>
<th>n = 182</th>
<th>NPål &gt; NPi :9+ words</th>
<th>:1–8</th>
<th>NPi = NPål</th>
<th>NPi &gt; NPål :1–4</th>
<th>:5+</th>
</tr>
</thead>
<tbody>
<tr>
<td>X [NPi NPål]</td>
<td>5</td>
<td>49</td>
<td>59</td>
<td>45</td>
<td>18</td>
</tr>
<tr>
<td>Y [NPål NPi]</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

Ratio of Y/X: 29% 4% 0% 4% 0%

EIC predictions (1): assume \(s[NPi NPål V]\) alternates with \(s[NPål NPi V]\)

Optimal orders:
- X where NPål ≥ NPål
- Y where NPi ≥ NPål

Unmarked case: 126/182 = most optimal, i.e. 69% Incorrect

Marked case: e.g. ratio of Y/X for NPi > NPål :1–4 ≥ 5+
- ratio of Y/X for NPi > NPål :1–4 (=4%) ≥ ratio of X/Y for NPål > NPi :9+ (=71%) Correct

EIC predictions (2): assume \(s[NPi VP[NPål V]]\) alternates with \(s[NPål NPi VP[V]]\)

Optimal orders:
- X where NPål ≥ NPål and NPål > NPi :1–8 words
- Y where NPål > NPi :9+ words

Unmarked case: 173/182 = most optimal, i.e. 95% Correct

Marked case: ratio of Y/X for NPål > NPi :1–8 (=4%) ≥ (NPål ≥ NPi) (=2%) Correct

EIC predictions (3): assume \(s[NPi VP[NPål V]]\) alternates with \(s[NPål NPi VP[V]]\)

Optimal orders:
- X where NPål > NPål
- Y where NPål ≥ NPi

Unmarked case: 67/182 = most optimal, i.e. 37% Incorrect

Marked case: e.g. ratio of X/Y for NPi = NPål (=100%) ≥ (NPål ≥ NPi) (=93%) Correct
- ratio of Y/X for NPi > NPål :1–4 (=4%) ≥ ratio of X/Y for NPål > NPål :9+ (=71%) Incorrect

EIC predictions (4): assume \(s[NPi VP[NPål V]]\) alternates with \(s_1[NPål s_2[NPi VP[V]]\)

Optimal orders and predictions as for EIC predictions (3).
found to be greater than, equal to, or even shorter than, direct objects; but the only direct objects preceding indirect objects were longer than the indirect objects. This pattern is fundamentally similar to what we find in clear cases of transformational rearrangements from a basic order in other languages (cf. section 4.3.6 for precise figures).

Data were also collected (by Charles Kim) for the relative ordering of two PPₘs, corresponding to the Japanese data of table 4.9. The Korean orders showed a strong long-before-short preference, correlating with the size of the word-content differential. For example, in the 150 pages of text from Wungdam Pwuin referenced in table 4.11, dispreferred short PPₘs before long PPₘs accounted for 31% of the data with one-to-three-word differentials (9/29 instances), 20% for four to five words (1/5), and 0% for 6+ words (0/4). This is what EIC predicts.

In table 4.12 I present Korean data (collected by Charles Kim) corresponding to the Japanese data of table 4.10, and involving a direct object NPₖ plus a PPₘ. I first test (in section I) the consistent long-before-short prediction, assuming no VP discontinuity. It turns out that 82% of the data are most optimal, and the marked case predictions are correct. These are exactly the results that we found in Japanese. I then test EIC's more structurally sensitive predictions (in section II), on the assumption that the two orders of \([NPₖ PPₘ] V\) involve alternation (A) or (B). The unmarked case prediction for those orders in which PPₘ > NPₖ (which account for most of the data) scores 92%, exactly as in Japanese again, and the marked case predictions are also correct. For NPₖ > PPₘ, the declining ratio of Y to X from left to right in the chart is also predicted.

Overall, the pattern of results from Korean is fundamentally similar to that in Japanese.

4.1.7 Turkish \([\{NPₘₗNP\} V]\)

The consistently right-peripheral nature of Mother Node Construction for NP, PP, PossP, and TopP that we have hypothesized for Japanese and Korean does not extend to all verb-final languages. In section 4.1.4 I suggested that NPs in Turkish could be of the NPₘ type in the event that the head noun was preceded by a left-branching relative clause or genitive NP. Otherwise they would be NPₗ. The former require a bottom-up parse, as in Japanese and Korean; the latter can provide immediate recognition of NP by some left-peripheral constructing category such as Det or N (cf. ch. 6.3.4). This same distinction is evident in many other languages. English has both prenominal
Table 4.12 *Korean* [{NPūl PPₘ} V]


\[(\text{TopP})\[[\{(\bar{S} \geq 1) (\text{NP})\} \{\text{NPūl PP}_{ₘ}\} V]\] (no other ICs in the clause apart from optional adverbials in any position)

where PPₘ = PP[NP P]

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2ICₘ &gt; 1ICₘ : 1 word</th>
<th>:2</th>
<th>:3 +</th>
</tr>
</thead>
<tbody>
<tr>
<td>X [[2ICₘ 1ICₘ]]</td>
<td>15</td>
<td>36</td>
<td>8</td>
<td>14</td>
</tr>
<tr>
<td>Y [[1ICₘ 2ICₘ]]</td>
<td>12</td>
<td></td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Ratio of Y/X  
|   | 25% | 20% | 13% |

*EIC predictions* (assuming no VP discontinuity; consistent long before short predictions)

Unmarked case: 73/89 = most optimal (X + =), i.e. 82% Correct

Marked case: ratio of Y/X for 2ICₘ > 1ICₘ : 1 ≥ 2 ≥ 3+ All correct
II \([\{PP_m \text{ NPůl}\} V]\)

<table>
<thead>
<tr>
<th></th>
<th>(n = 89)</th>
<th>(NPůl &gt; PP_m :3 + \text{ words})</th>
<th>:2</th>
<th>:1</th>
<th>(PP_m = \text{ NPůl})</th>
<th>(PP_m &gt; \text{ NPůl} :1)</th>
<th>:2</th>
<th>:3 +</th>
</tr>
</thead>
<tbody>
<tr>
<td>X ([PP_m \text{ NPůl}])</td>
<td>2</td>
<td>1</td>
<td>7</td>
<td>9</td>
<td>32</td>
<td>7</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Y ([NPůl PP_m])</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Ratio of \(Y/X\): 60\% 50\% 36\% 40\% 14\% 13\% 0\%

Assume: VP; VP dominates \(NPůl\); possible VP discontinuity; two possible structural alternations;

(A) \(\text{VP} [PP_m \text{ NPůl} V] (= X)\) alternates with \(\text{VP} [NPůl PP_m V] (= Y)\)

(B) \(s[PP_m \text{ VP} [NPůl V]] (= X)\) alternates with \(s[VP [NPůl] PP_m \text{ VP} [V]] (= Y)\)

**EIC's preferences**

- \(PP_m > \text{ NPůl}\): X in (A) and (B), increasing with larger differentials;
- \(PP_m = \text{ NPůl}\): X or Y in (A), X in (B);
- \(NPůl > PP_m\): Y in (A), increasing with larger differentials;
- X in (B) where \(NPůl > PP_m :1-8\), decreasing with larger differentials;
- Y in (B) where \(NPůl > PP_m :9 +\), increasing with larger differentials;

I.e. Y preference increases with larger differentials, X with smaller.

**EIC's predictions**

- \(PP_m \geq NPůl\)
  - Unmarked case (\(X+ =\)): 65/71 = 92\%  Correct
  - Marked case: ratio of \(Y/X\) for \(PP_m > NPůl :1 \geq 2 \geq 3 +\) Correct
  - \(NPůl > PP_m\)
    - Ratio of \(Y/X\) for \(NPůl > PP_m :3 + \geq 2 \geq 1\) Correct
and postnominal genitives: the king's protector is of the type NP\textsubscript{m}, i.e. NP\textsubscript{Pposs}[the king's] protector; the protector of the king is a \textsubscript{m}NP, NP[the protector \textsubscript{pp}[of the king]]. Hungarian and Finnish have a productive \textsubscript{m}NP/NP\textsubscript{m} distinction as well.

This distinction is of considerable significance for EIC, since it can affect either the size of a CRD or the order and speed with which constituents within CRDs can be recognized or both, and different predictions may follow as a consequence. For [IC\textsubscript{m} IC\textsubscript{m} V] structures, EIC defines a long-before-short preference (in the absence of discontinuity), as we have just seen; for [\textsubscript{m}IC IC V] it defines the opposite preference, which was illustrated for Turkish data involving two \textsubscript{m}NPs in section 4.1.4. In this section we shall consider combinations of IC\textsubscript{m} and \textsubscript{m}IC in a verb-final structure, again in Turkish, followed by similar combinations in verb-initial structures in Hungarian and Finnish. The structural type to be considered first is therefore (4a) in table 3.2, i.e. [IC\textsubscript{m} IC\textsubscript{m} C].

Consider the structure [{NP\textsubscript{m} mNP} V], where both NPs are non-subjects. An example is (4.12):

\begin{enumerate}
\item[(4.12) a.] NP\textsubscript{m}[Salon-da-ki biitiin biblo-lar-I]
\begin{itemize}
\item "living room-LOC-GER all trinket-PLUR-ACC"
\item mNP[camekan-a] kaldIr-dlm
\item showcase-DAT stay-CAUS-PAST," i.e.
\item X kept all the trinkets that were in the living room in a showcase.
\end{itemize}
\item[(4.12) b.] mNP[camekan-a] NP\textsubscript{m}[salon-da-ki biitiin biblo-lar-I] kaldIr-dlm
\end{enumerate}

The relevant data were collected and analyzed by Charlotte Reinholtz using the sample of table 4.6.

The prediction made by EIC in these cases is that the NP\textsubscript{m} should always be positioned to the left of \textsubscript{m}NP, regardless of the length difference. This is because the sequence [NP\textsubscript{m} mNP V] involves a shorter overall domain than [\textsubscript{m}NP NP\textsubscript{m} V], as illustrated by the lengths of the underlinings: if the first IC is constructed on its right periphery, then all the other words and constituents dominated by this IC\textsubscript{m} are excluded from the higher domain; but if the first IC is constructed on its left periphery, then everything dominated by this IC\textsubscript{m} is within the higher domain, as is everything dominated by the second, IC\textsubscript{m}. The only way to exclude any NP-dominated material from the higher domain is to order NP\textsubscript{m} before \textsubscript{m}NP, and this will improve IC-to-word ratios whatever the relative lengths of the two ICs are.
The predictions of EIC are clear, therefore, so it was surprising to find that the results were not as expected, in contrast to the consistent \( \{mIC_1 \ mIC_2\} \) and \( \{IC_m_1 \ IC_m_2\} \) sequences considered hitherto, and in contrast to the mixtures of \( IC_m \) and \( mIC \) within a postverbal domain, which we shall consider in Hungarian and Finnish. In fact, the results were random: out of a total of 89 structures satisfying the template \( S[(NP-Subject) \ \{NP \ m \ mNP\} \ V] \), 42 (or 47\%) had the order \( [NP \ m \ mNP \ V] \) and 45 (53\%) had the reverse \( [mNP \ NP \ mNP \ V] \).

A closer investigation of the data, and also of the structure of Turkish, led to a hypothesis that can explain these data and that even predicts random positioning. Let us pursue the parsing assumption that case marking on NPs can provide on-line evidence about VP and S structure, for which Japanese and Korean have already provided supporting evidence. The basic idea is that particles, or even morphological parts of words, have the capacity to construct a grandmother node immediately dominating the mother node (cf. the principle of Grandmother Node Construction in ch. 6.2.2). The major consequence of this assumption for a verb-final language is that the VP will be recognizable well in advance of the verb, and this shortens the CRD for S.

In the case of Turkish, the verb assigns cases such as accusative, dative, and instrumental to NPs that it governs. Let us assume that these are all VP-dominated, and that the nominative is S-dominated. A more detailed parsing investigation of this language will need to establish whether these grandmother nodes are always predictable on-line, but for the moment let us make the simplifying assumption that they are. Even if this assumption is only partially correct, in other words even if only some cases can act as unique indicators of VP or S, we see immediately that the relative ordering of two NPs within the VP can now potentially affect both the VP recognition domain and the recognition domain for S.

Take the case of two NPs of the type \( mNP \). Assume that they are both VP-dominated and case marked and that each can construct VP. Then if the shorter precedes the longer, \( VP[mNP_1 \ mNP_2 \ V] \), not only is this preferable for the VP domain, as shown by the EIC calculations of table 4.5: it is also preferable for the S domain, since the VP of the S domain will be recognized sooner in the left-to-right parse than it would be if the longer \( mNP_2 \) preceded the shorter \( mNP_1 \). The NPs themselves can be constructed by whatever left-peripheral word is unique to NP; but the grandmother VP can only be constructed by case marking on the head noun, and regardless of where the head noun is situated within the NP, the positioning of case within the parse string will generally be further to the left when the shorter \( mNP_1 \)
precedes \( mNP_2 \). Both the VP and S domains benefit from short-before-long positioning, therefore.

The same is not true for mixtures of \( NP_m \) and \( mNP \). The EIC benefits for the two domains are now in partial conflict, and the precise manner of their conflict results in a testable prediction. Consider first the cases where \( NP_m > mNP \), which account for the majority of the data (63/89 = 71\%). This is expected, since \( NP_m \) always dominates a left-branching phrasal node, whereas \( mNP \) will often dominate just one or more non-phrasal nodes, Det, Adj, N, etc., though sometimes \( mNP \) may be more complex (as in the Persian-derived relative clause constructions of the type \( \text{NP}\[\text{bir adam }\{\text{ki söz dinlemez}\}\] "a man who word non-listens," i.e. a man who doesn't listen to advice). Let us abbreviate this longer \( NP_m \) as \( 2NP_m \), and the shorter \( mNP \) as \( mNP_1 \). For the VP domain, the order \( \text{vp}[2NP_m \ mNP_1 \ V] \) will be preferred, since all \( NP_m \) structures are preferred in phrase-initial position. But for the S domain, ordering the shorter \( mNP_1 \) before \( 2NP_m \) will result in earlier recognition of the VP, i.e. \( S[\text{NP} \ \text{vp}[mNP_1 \ 2NP_m \ V]] \) will be preferred. It becomes necessary to calculate, therefore, what the precise quantitative effect of the two orderings is on the two domains. This is done in table 4.13 for both a two-word and a four-word differential, i.e. the same word-content assignments that were made for the Turkish data of table 4.5.

It turns out that the benefits of the \( [2NP_m \ mNP_1] \) ordering for the VP are almost exactly offset by the benefits of \( [mNP_1 \ 2NP_m] \) for S. The aggregates for the two domains differ by only 2.5 for a two-word differential and by only two for a four-word differential (cf. sections I and II in table 4.13). Hence, EIC predicts that both should be productively occurring, and they are. Of the 63 instances in which \( NP_m > mNP \), 28 have \( [2NP_m \ mNP_1] \) (or 44\%), and 35 have \( [mNP_1 \ 2NP_m] \) (56\%).

Matters are very different, however, when \( mNP > NP_m \), as shown in sections III and IV of the table. The VP still exhibits a preference for \( [1NP_m \ mNP_2] \) over \( [mNP_2 \ 1NP_m] \) (the preference is slighter now on account of the fact that a short \( 1NP_m \) will not be excluding many words and non-ICs from the VP domain, and hence its first or second positioning will make little difference, e.g. a 50\% versus 43\% ratio for VP in section III), but the crucial difference is that the S now exerts no counterpressure in favor of a conversion into \( [mNP_2 \ 1NP_m] \), i.e. into long before short. Both the VP and the S domains are better off with the order \( [1NP_m \ mNP_2] \), and this results in very large aggregate differentials of 17 and 21.5 for the word assignments of sections III and IV.
Table 4.13 *IC-to-word ratios for Turkish* $\mathcal S [NP-Subj \, v_P \{mNP \, NP_m \} \, V]$

Assume NP-Subj = 2 words and is of the type NP$_m$; assume V = 1 word; assume that each of the VP-dominated NPs can construct VP via case marking but that all the words dominated by the relevant VP need to be scanned for VP construction. (I.e. the most conservative assumption.)

<table>
<thead>
<tr>
<th>I</th>
<th>mNP$_1$ = 2  , 2NP$_m$ = 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>S[NP-Subj v_P{mNP$_1$ 2NP$_m$ V}]</td>
<td>S[NP-Subj v_P{mNP$_m$ NP$_1$ V}]</td>
</tr>
<tr>
<td>S CRD: 2/3 = 67%</td>
<td>S CRD: 2/5 = 40%</td>
</tr>
<tr>
<td>VP CRD: 3/7 = 43%</td>
<td>VP CRD: 3/4 = 75%</td>
</tr>
<tr>
<td>Aggregate = 55</td>
<td>Aggregate = 57.5</td>
</tr>
<tr>
<td>Agg. differential = 2.5</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II</th>
<th>mNP$_1$ = 2  , 2NP$_m$ = 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>S[NP-Subj v_P{mNP$_1$ 2NP$_m$ V}]</td>
<td>S[NP-Subj v_P{mNP$_m$ NP$_1$ V}]</td>
</tr>
<tr>
<td>S CRD: 2/3 = 67%</td>
<td>S CRD: 2/7 = 29%</td>
</tr>
<tr>
<td>VP CRD: 3/9 = 33%</td>
<td>VP CRD: 3/4 = 75%</td>
</tr>
<tr>
<td>Aggregate = 50</td>
<td>Aggregate = 52</td>
</tr>
<tr>
<td>Agg. differential = 2</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III</th>
<th>mNP$_1$ = 2  , mNP$_2$ = 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>S[NP-Subj v_P{mNP$_m$ NP$_1$ V}]</td>
<td>S[NP-Subj v_P{mNP$_2$ mNP$_m$ V}]</td>
</tr>
<tr>
<td>S CRD: 2/3 = 67%</td>
<td>S CRD: 2/5 = 40%</td>
</tr>
<tr>
<td>VP CRD: 3/6 = 50%</td>
<td>VP CRD: 3/7 = 43%</td>
</tr>
<tr>
<td>Aggregate = 58.5</td>
<td>Aggregate = 41.5</td>
</tr>
<tr>
<td>Agg. differential = 17</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IV</th>
<th>mNP$_1$ = 2  , mNP$_2$ = 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>S[NP-Subj v_P{mNP$_m$ NP$_2$ V}]</td>
<td>S[NP-Subj v_P{mNP$_m$ NP$_1$ V}]</td>
</tr>
<tr>
<td>S CRD: 2/3 = 67%</td>
<td>S CRD: 2/7 = 29%</td>
</tr>
<tr>
<td>VP CRD: 3/8 = 38%</td>
<td>VP CRD: 3/9 = 33%</td>
</tr>
<tr>
<td>Aggregate = 52.5</td>
<td>Aggregate = 31</td>
</tr>
<tr>
<td>Agg. differential = 21.5</td>
<td></td>
</tr>
</tbody>
</table>
Testing EIC's performance predictions

Putting this another way, if the order that is consistently preferred by the VP, \([NP_m mNP V]\), is of the form long before short (\([lNP_m mNP]\)), then the S will prefer short before long (\([mNP 1 NP]\)), the advantages for each phrase will outweigh each other, and both orders are expected. But if the VP-preferred order also happens to be short before long (\([NP_m mNP]\)), then only this order should be productive.

There are eight instances in the data in which \(NP_2 > NP_m\). A full seven of these (88%) are as predicted, with the order \([NP_1 NP mNP_2]\). These numbers are small, but this appears to be a very different pattern from the random distribution of orders when \(NP_m > NP_m\), and it is predicted by EIC. More data will need to be collected of these types, and the analysis will need to plot the precise location and nature of case marking in the parse string.

4.1.8 Hungarian \([V \{PP_m mNP}\])

Let us consider now an example of structure type (4b) in table 3.2, with the same mix of IC and IC that we considered in Turkish, but with category C on the left, i.e. \([C \{IC_m mIC}\]). Hungarian affords a ready example. Assuming the tree structure in example (4.3) above, the verb initiates a domain in which NPs and PPs are freely ordered. PPs are clearly of the type PP, i.e. postpositional, and NPs are either NP or NP, depending on whether a left-branching genitive or relative clause is present or not. By concentrating on NPs of the type NP, as we did in section 4.1.2, but in combination with PP, we can test EIC's predictions on table 3.2's structure (4b). The data of this section were collected and analyzed by Stephen Matthews.

An example of the structure to be considered is (4.13):

(4.13) a. Kidugta \(_mNP_[a fejét] PP_m[a dunyha alól]

"thrust the head the quilt from-under," i.e.

X thrust (her) head out from under the quilt.

b. Kidugta PP_m[a dunyha alól] \(_mNP_[a fejét]

Both orders are fully grammatical and productively attested: of the 55 instances of this structure in the data, 17 have the ordering in (a) and 38 have that in (b). This appears to be an uninteresting distribution; it becomes highly principled, however, when we consider the relative lengths of NP and PP and the resulting recognition domains from a processing perspective.

Consider first the cases in which \(NP_m > PP_m\). If an NP is recognized early and is larger than, or equal to, PP, EIC defines a strong preference for
Table 4.14 IC-to-word ratios for Hungarian \([V \{PP_m mNP\}]\) where \(mNP \geq PP_m\)

Assume \(V = 1\) word.

<table>
<thead>
<tr>
<th>Case</th>
<th>(mNP)</th>
<th>(PP_m)</th>
<th>[(V mNP PP_m)]</th>
<th>CRD: 3/5 = 60%</th>
<th>CRD: 3/4 = 75%</th>
<th>Differential = 15%</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>2</td>
<td>2</td>
<td>[(V PP_m mNP)]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>3</td>
<td>2</td>
<td>CRD: 3/6 = 50%</td>
<td>CRD: 3/4 = 75%</td>
<td></td>
<td>Differential = 25%</td>
</tr>
<tr>
<td>III</td>
<td>4</td>
<td>2</td>
<td>CRD: 3/7 = 43%</td>
<td>CRD: 3/4 = 75%</td>
<td></td>
<td>Differential = 32%</td>
</tr>
</tbody>
</table>

postposing \(mNP\), i.e. [\(V PP_m mNP\)], thereby shortening the recognition domain and excluding many NP-dominated constituents from it. With the reverse ordering, [\(V mNP PP_m\)], the PP is not constructed until the right-peripheral postposition is encountered, and the recognition domain will proceed all the way from \(V\) to the postposition. EIC’s preferences increase the larger the word-content differential becomes, as shown in table 4.14. These preferences hold, whether they are defined as in table 4.14 or left-to-right.

If \(PP_m > mNP\), however, matters are more subtle. The order [\(PP_m mNP\)] can still potentially shorten the CRD, but the benefits of doing so will be quite minimal for NPs that do not dominate many words to begin with, and they turn out to be offset by the left-to-right advantage of ordering a shorter before a longer IC within this domain. Left-to-right ratios are illustrated in table 4.14 for word-content differentials of one, two, three, and four words, both when \(mNP\) is short (two words in length) and when it is longer (four words).

The basic pattern to emerge from table 4.14 is that [\(V mNP PP_m\)] is gradually preferred over [\(V PP_m mNP\)] as the word-total difference between \(PP_m\) and \(mNP\) gets bigger; in effect a short \(mNP\) is preferred before a long \(PP_m\) in proportion to the word-total difference. Compare the expanding differentials down I–IV, for example. However, as \(mNP\) gets larger, the degree of this preference diminishes, as shown down I'–IV', in which \(mNP\)
### Table 4.14: Left-to-right IC-to-word ratios for Hungarian $[V\{PP_m, mNP\}]$

where $PP_m > mNP$

Assume $V = 1$ word

<table>
<thead>
<tr>
<th>II</th>
<th>$PP_m = 4$</th>
<th>$mNP = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$[V\ PP_m, mNP]$</td>
<td>$[V\ mNP\ PP_m]$</td>
<td></td>
</tr>
<tr>
<td>1/1 2/5 3/6</td>
<td>1/1 2/3 3/7</td>
<td></td>
</tr>
<tr>
<td>100% 40% 50%</td>
<td>100% 67% 43%</td>
<td></td>
</tr>
<tr>
<td>Aggregate % for CRD = 63.3%</td>
<td>Aggregate % for CRD = 70%</td>
<td>Differential = 6.7%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>I'</th>
<th>$PP_m = 5$</th>
<th>$mNP = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$[V\ PP_m, mNP]$</td>
<td>$[V\ mNP\ PP_m]$</td>
<td></td>
</tr>
<tr>
<td>1/1 2/6 3/7</td>
<td>1/1 2/3 3/8</td>
<td></td>
</tr>
<tr>
<td>100% 33% 43%</td>
<td>100% 67% 38%</td>
<td></td>
</tr>
<tr>
<td>Aggregate % for CRD = 58.7%</td>
<td>Aggregate % for CRD = 68.3%</td>
<td>Differential = 9.6%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>I''</th>
<th>$PP_m = 6$</th>
<th>$mNP = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$[V\ PP_m, mNP]$</td>
<td>$[V\ mNP\ PP_m]$</td>
<td></td>
</tr>
<tr>
<td>1/1 2/7 3/8</td>
<td>1/1 2/3 3/9</td>
<td></td>
</tr>
<tr>
<td>100% 29% 38%</td>
<td>100% 67% 33%</td>
<td></td>
</tr>
<tr>
<td>Aggregate % for CRD = 55.7%</td>
<td>Aggregate % for CRD = 66.7%</td>
<td>Differential = 11%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III'</th>
<th>$PP_m = 7$</th>
<th>$mNP = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$[V\ PP_m, mNP]$</td>
<td>$[V\ mNP\ PP_m]$</td>
<td></td>
</tr>
<tr>
<td>1/1 2/8 3/9</td>
<td>1/1 2/5 3/11</td>
<td></td>
</tr>
<tr>
<td>100% 25% 33%</td>
<td>100% 40% 27%</td>
<td></td>
</tr>
<tr>
<td>Aggregate % for CRD = 52.7%</td>
<td>Aggregate % for CRD = 55.7%</td>
<td>Differential = 0%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IV'</th>
<th>$PP_m = 8$</th>
<th>$mNP = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$[V\ PP_m, mNP]$</td>
<td>$[V\ mNP\ PP_m]$</td>
<td></td>
</tr>
<tr>
<td>1/1 2/9 3/10</td>
<td>1/1 2/5 3/13</td>
<td></td>
</tr>
<tr>
<td>100% 22% 30%</td>
<td>100% 40% 23%</td>
<td></td>
</tr>
<tr>
<td>Aggregate % for CRD = 50.7%</td>
<td>Aggregate % for CRD = 54.3%</td>
<td>Differential = 3.6%</td>
</tr>
</tbody>
</table>
is assigned four words. This is because the relative advantage of ordering a shorter IC before a longer one is much less if each IC is quite long, according to the EIC calculation procedure. When there is just a one-word differential and $PP_m = 5$ (see I), there is even a slight preference the other way round, for long before short. When $PP_m = 6$ (see II), the two orders have identical scores, and thereafter the relative ordering preferences are as for III–IV, though with diminished absolute scores.

Assuming that the word-total assignments and differentials of table 4.14 cover the major options in the data, we can derive the following predictions from them: if $PP_m > mNP$ by one word, then either order can be expected; if $PP_m > mNP$ by two words, then either there will be a preference for $[V mNP PP_m]$ (when $mNP$ is short), or both orders can be expected (when $mNP$ is longer); if $PP_m > mNP$ by three or more words, then $[V mNP PP_m]$ is preferred with a strength that is directly proportional to the word-total difference between $PP_m$ and $mNP$, and that decreases with the size of the latter.

In setting out the data, we need to distinguish the two sets of predictions that our processing theory makes: the strong and increasing preference for $[V PP_m mNP]$ when $mNP > PP_m$; and the weaker preference for the reverse $[V mNP PP_m]$ when $PP_m > mNP$ by at least two words. This is done in table 4.15.

Perhaps surprisingly, these predictions turn out to be correct! The data show a solid preference for $[V PP_m mNP]$ when $mNP \geq PP_m$ (see section I), exceptionlessly so when there is a one-or-more-word differential. The only instances of the non-optimal orders $[V mNP PP_m]$ are when $mNP = PP_m$, i.e. just the case where EIC's dispreference is weakest (cf. table 4.14). Both the unmarked and marked case predictions are straightforwardly satisfied relative to this subset of the data, therefore.

When $PP_m > mNP$, the predictions are also correct (see section II): both $[V mNP PP_m]$ and $[V PP_m mNP]$ are productive when there is a one-word differential; the former is more frequent than the latter when the differential is two words; and for three or more words the only order attested is the preferred $[V mNP PP_m]$. All of these orders are as optimal as they can possibly be, except perhaps for the two instances of $[V PP_m mNP]$ where $PP_m > mNP$ by two words. The unmarked and marked case predictions are again correct in relation to this data subset, and if we conflate the two subsets we see that the unmarked case prediction is fulfilled in at least 91% of the total data.
Table 4.15 Hungarian \([V\{PP_m \text{m}NP\}]\)

112 pages of data: cf. table 4.4.

\[s[\ldots V\{PP_m \text{m}NP\} \ldots]\] (the only other ICs permitted in the clause are AdvP and/or NP-Nominative in the positions marked "\ldots")

where \(mNP = \) any NP constructed on its left periphery, i.e. with no left-branching genitive NP or relative S;

\[PP_m = pp[NP P]\]

<table>
<thead>
<tr>
<th></th>
<th>(PP_m &gt; mNP :3^+)</th>
<th>:2</th>
<th>:1</th>
<th>(mNP = PP_m)</th>
<th>(mNP &gt; PP_m :1)</th>
<th>:2</th>
</tr>
</thead>
<tbody>
<tr>
<td>X ([PP_m mNP])</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>10</td>
<td>17</td>
<td>3</td>
</tr>
<tr>
<td>Y ([mNP PP_m])</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Ratio of Y/X: 100% 71% 54% 23% 0% 0%

I \(EIC\) predictions: \(mNP \geq PP_m\)

- Unmarked case: 30/33 = most optimal (X), i.e. 90.9% Correct
- Marked case: ratio of Y/X for \(mNP = PP_m \geq (mNP > PP_m :1) \geq 2\) All correct

II \(EIC\) predictions: \(PP_m > mNP\)

- \(PP_m > mNP :3^+\) words: Y only Correct
- \(PP_m > mNP :2\) words: \(Y \geq X\) Correct
- \(PP_m > mNP :1\) word: either X or Y Correct

- Unmarked case: at least 20/22 = most optimal, i.e. 90.9% Correct
- Marked case: ratio of X/Y for \(PP_m > mNP :2 (= 29\%) \geq 3^+ (= 0\%)\) Correct
4.1.9 Finnish (V {NP IC})

Finnish has free ordering of constituents in the postverbal domain, and discourse-configurational nodes preceding the verb, much as in Hungarian. Vilkuna (1989) argues for a Topic position immediately preceding the verb, which is normally filled by the subject, and a "K" position preceding that, which is obligatorily filled by (for example) "WH" phrases. The details of these preverbal nodes differ between the two languages, but material after the verb seems to be freely positioned in both, and so Finnish is another language on which the arrangement predictions of EIC can be tested. The data of this section were collected and analyzed by Maria Vilkuna and are presented in Vilkuna (1991).

She examined the relative order of a direct object NP and some other constituent X in the postverbal domain. X could either be an adverbial phrase or an oblique (non-subject and non-object) complement of the verb. An example is (4.14):

(4.14) a. Se [anto lahjan armaalleen]
   "He gave present-ACC sweetheart-3-ALLATIVE," i.e.
   He gave a present to his sweetheart.

b. Se [anto armaalleen lahjan]

Both (4.14a) and (b) are fully grammatical.

Vilkuna's data are drawn from an earlier corpus collected by Hakulinen et al. (cf. their 1980 publication), which codes for phrasal category and word length, among other variables. However, her analysis does not distinguish between left-recognized (mIC) and right-recognized (ICm) constituents. Finnish is like Hungarian in having both processing types, mNP and NPm, together with postpositions, PPm. The structures that she tests and on which she presents statistical information could be either of type (2a) in table 3.2 (i.e. [V {mIC1 mIC2}]), or (4b) ([V {ICm mIC}]), or even [V {ICm 2ICm}]). This last will be the rarest: in our Hungarian data, for example, a combination of an NPm with PPm was much less frequently attested than mNP and PPm and than mNP and mNPm, because of the added complexity of NPm.

Although this will prevent us from conducting the kind of fine-tuned testing on Finnish that we have performed on Hungarian, especially for the relative positioning of mNP and PPm, we can still make use of her data. For notice that if [V {NP IC}] is of type (2a) [V {mIC1 mIC2}], then there will be a consistent preference for short or equal before long or equal. If it is of type (4b) [V {ICm mIC}], there will be a preference for a short ICm before a long
Testing EIC's performance predictions

For a short \( mIC \) before a long \( ICm \) in general (cf. table 4.14), and for \( V ICm mIC \) ordering when both are equal. If it is a \( V ICm ICm \), then there will again be a consistent preference for short or equal before long or equal. In none of these cases is there a consistent long-before-short preference, therefore; and only in the case of weight equality between the two ICs (\( ICm = mIC \)) will there be one ordering that is not most optimal for EIC, namely \( V mIC ICm \) (cf. table 4.14 section I). Since I have no means of knowing which order actually occurs in these cases, and since all the examples of weight equality can in principle be optimal across all of these different structural types, we can make the simplifying assumption that EIC defines a consistent preference for short or equal before long or equal in these data and test its predictions accordingly.

Vilkuna's data are set out in tables 4.16 and 4.17. Table 4.16 gives her figures first for \( V NP IC \) orders taken from the whole corpus (section I), and then for \( V IC NP \) taken from roughly half the corpus (section II). The unmarked case prediction is satisfied in 84% and 93% of the data respectively. Since precise length differentials are not given, EIC's marked case predictions cannot be tested.

Table 4.16 Finnish \( V \{NP IC\} \)

5,016 sentences of written text (mainly newspaper and magazine articles, some encyclopaedic and textbook material); cf. Hakulinen et al. 1980. Data collection and analysis from Vilkuna 1991.

\( V \{NP IC\} \) (in the order \( V NP IC \) the IC is the last constituent of the clause; in the order \( V (...) IC NP \) an optional additional IC was permitted and the length of the IC immediately preceding NP was the one that was counted)

where NP = a direct object NP (of the type \( mNP \) or \( NPm \));

IC = an adverbial phrase or oblique (non-subject non-object) complement of the verb (of the type \( mIC \) or \( ICm \)); this IC is non-clausal, i.e. it does not dominate S.

<table>
<thead>
<tr>
<th></th>
<th>( V NP IC ) n = 239</th>
<th>( V IC NP ) n = 138 (half the corpus)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP = IC</td>
<td>48 (20.1%)</td>
<td>NP = IC: 23 (16.7%)</td>
</tr>
<tr>
<td>IC &gt; NP</td>
<td>153 (64.0%)</td>
<td>NP &gt; IC: 105 (76.1%)</td>
</tr>
<tr>
<td>NP &gt; IC</td>
<td>38 (15.9%)</td>
<td>IC &gt; NP: 10 (7.2%)</td>
</tr>
<tr>
<td>Mean length of NP</td>
<td>1.91 words</td>
<td>Mean length of IC = 1.47 words</td>
</tr>
<tr>
<td>Mean length of IC</td>
<td>2.58 words</td>
<td>Mean length of NP = 3.16 words</td>
</tr>
</tbody>
</table>

**EIC predictions** (short before long preferred; assume all NP = IC structures are optimal)

Unmarked case: 201/239 = most optimal in I, i.e. 84.1% Correct
Unmarked case: 128/138 = most optimal in II, i.e. 92.8% Correct
Table 4.17 *Finnish* $\{V \{NP_i, NP_j\}\}$

Data from Vilkuna (1991: 104)

$\{V \{NP_i, NP_j\}\}$ (cf. (4.14) in the main text)

where $V = \text{antaa} \; \text{"to give"}$;  
one $NP = \text{the direct object of } \text{antaa}$; the other $NP = \text{the oblique (recipient) argument of } \text{antaa}$; both NPs can be either $mNP$ or $NP_m$.

<table>
<thead>
<tr>
<th></th>
<th>$n = 197$</th>
<th>$NP_i = NP_j$: 48 (24.5%)</th>
<th>$NP_j &gt; NP_i$: 136 (69.4%)</th>
<th>$NP_i &gt; NP_j$: 13 (6.1%)</th>
</tr>
</thead>
</table>

**EIC predictions** (short before long preferred; assume all $NP_i = NP_j$ structures are optimal)

Unmarked case: $184/197 = \text{most optimal, i.e. 93.4\% Correct}$

Table 4.17 examines the relative positioning of a direct object and oblique complement of just one verb, *antaa* "to give" (illustrated in (4.14)). The unmarked case prediction is satisfied in 93% of the data. These results are strongly supportive of EIC.

4.2 Binary branching structures

I turn now to EIC's predictions for binary branching structures of the type $\{C \text{IC}_m\}$ and $\{C \text{IC}_m\}$, cf. (1) in table 3.2. These structures are more difficult to test in the present context, because they are not generally freely ordered. In order for there to be free ordering, at least one of the mother-node-constructing categories has to be free, either $C$ or the category that constructs the IC. In the latter case we will expect alternations such as $[C \text{IC}_m]$ and $[C \text{IC}_m]$ (e.g. a V-initial VP co-occurring with both $\text{ppf}[P NP]$ and $\text{ppf}[NP P]$); in the former case, alternations like $[C \text{IC}_m]$ and $[C \text{IC}_m]$ (e.g. $\text{ppf}[P NP]$ within both a V-initial and V-final VP). If both constructing categories are freely ordered, we will get all four logical possibilities (1a–d) in table 3.2. But since grammars generally fix the positioning of constructing categories (cf. ch. 5.1), the incidence of free ordering here is much rarer than it is in multiple branching structures, where a plurality of sister ICs can be freely positioned in relation to a constructing category whose order is fixed. Binary branching structures will be of much greater interest in the context of grammaticalized word orders,
therefore, and here I will merely illustrate the fact that there can be productive ordering alternations for which EIC makes interesting predictions. The data I have are less extensive than they are for multiple branching structures, and are not in fact sufficient to distinguish free-order arrangements from rearrangements of a grammaticalized order. Sections 4.2.1–3 can therefore serve as a transition to the rearrangement predictions of section 4.3.

4.2.1 Hungarian \([V_f \overline{VP}\{V_{nf} mNP\}]\)

I begin with some data from Hungarian collected by Stephen Matthews and involving an alternation of the type \([C mIC]\) and \([C IC_m]\). The category represented as IC will be an embedded VP, symbolized here as \(\overline{VP}\), containing a non-finite verb, \(V_{nf}\), and one other daughter, specifically an NP of the type \(mNP\). Category C will be a finite verb, \(V_f\), which will be presumed to initiate the free-order domain, as discussed in section 4.1.2. An example is (4.15):

(4.15) Szeretném \(\overline{VP}[kifizetni mNP[a \text{ cechjét}]\]

"like-COND-1sg pay-INF the bill-3sg-ACC," i.e.

I would like to pay the bill.

(4.15) has the order \([V_f \overline{VP}\{V_{nf} mNP\}]\), but \([V_f \overline{VP}[mNP V_{nf}]\] is equally grammatical. I assume that \(V_{nf}\) constructs \(\overline{VP}\), being unique to it, and hence that \([V_f \overline{VP}]\) in (4.15) is a structure of type \([C mIC]\), whereas the reverse ordering of \(mNP\) and \(V_{nf}\) will be \([C IC_m]\). Clearly, EIC defines a preference for the former, in which the recognition of the branching IC takes place immediately after that of C.

There is a total of 67 instances of \([V_f \overline{VP}\{V_{nf} mNP\}]\) in the data. However, there are also 51 additional instances of these same categories occurring in different orders: three have a preposed \(\overline{VP}\) with a left-peripheral \(mNP\), \([\overline{VP}[mNP V_{nf}] V_f]\); 24 have what is possibly a discontinuous structure with a preposed \(V_{nf}\), \([\overline{VP}[V_{nf}] V_f \overline{VP}[mNP]\], or alternatively \(V_{nf}\) may have been raised into one of the c-commanding nodes outside of the free-order domain, i.e. \([V_{nf} [V_f \overline{VP}[mNP]]]\); another 24 have a preposed \(mNP\) for which the same two analyses are available, \([\overline{VP}[mNP] V_f \overline{VP}[V_{nf}]\] or \([mNP [V_f \overline{VP}[V_{nf}]]]\). The evaluation of these 51 cases is complicated by the existence of Topic and Focus nodes in Hungarian (cf. section 4.1.2), and for this reason we have to limit our attention to the 67 instances in which all the \(\overline{VP}\)-dominated material occurs to the right of \(V_f\). There are, in fact, clear length effects in the fronted \(mNP\) structures: they decline in frequency relative to other ordering alterna-
tives as the length of $mNP$ increases. Nonetheless, the existence of discourse-configurational nodes in the preverbal domain is a confounding factor that is best avoided, and that can be avoided in the postverbal domain.

The data are set out in Table 4.18. Since these data cover only just over half the total instances of $\{V_f \ V_P\{V_{nf \ mNP}\}\}$, it is not meaningful to test EIC’s unmarked case prediction for all possible word-total assignments in relation to all possible IC orderings. But we can test some of the marked-case predictions: specifically the X structure $[V_f \ [V_{nf \ mNP}]]$ will always be optimal and will be progressively less optimal the greater the length of an intervening $mNP$ in the $[V_f \ [mNP \ V_{nf}]]$ alternant. Hence, we predict that the former will be more or equally frequent, as the length of $mNP$ increases; and equivalently that the latter will be more or equally frequent as the length of $mNP$ decreases. This prediction is correct.

The data of Table 4.18 suggest a clear grammaticalization effect within the VP. If ordering were completely free within this constituent, we would expect an immediate and unmarked preference for $[V_f \ [V_{nf \ mNP}]]$; as it is there are significant numbers of the dispreferred $[V_f \ [mNP \ V_{nf}]]$. When $mNP = 1$ word, the latter order is even in the majority, becoming less frequent as $mNP$ increases in size. This suggests a grammaticalization of $V_P[NP \ V]$, with increasing rearrangements to $V_P[V NP]$ in proportion to the size of the NP.

Table 4.18 Hungarian $\{V_f \ V_P\{V_{nf \ mNP}\}\}$

112 pages of data: cf. Table 4.4.

$s((NP-Nom) \ [V_f \ V_P\{V_{nf \ mNP}\} \ldots ]]$ (the only other ICs permitted in this template must occur in the position marked “...”)

where (NP-Nom) = an optional subject NP;
$\overline{VP}$ = a VP containing a non-finite verb;
$mNP$ = an NP constructed on its left periphery, i.e. with no left-branching genitive NP or relative $\overline{S}$.

<table>
<thead>
<tr>
<th></th>
<th>$mNP = 1$ word</th>
<th>:2</th>
<th>:3</th>
<th>:4+</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>10</td>
<td>13</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>Y</td>
<td>12</td>
<td>10</td>
<td>6</td>
<td>1</td>
</tr>
</tbody>
</table>

Ratio of $Y/X$: 55% 43% 35% 20%

**EIC predictions**

Marked case: ratio of $Y/X$ for $mNP = 1 \geq 2 \geq 3 \geq 4+$ All correct
Consider next the relative positioning of the binary branching nodes NP and VP within S, and of V and NP within VP, in a language such as Polish in which there is considerable ordering variation. Relevant data from Polish have been collected by Anna Siewierska and are reported in Siewierska (1991b). Siewierska examines the relative frequencies of the six logically possible orders of subject, verb, and object in her texts and measures the average word length of subject and object in each of the six orders. She also performs a pragmatic analysis on these NPs, measuring average referential distance to a previous mention (if any), using the methodology of Givon (1983, 1988) (cf. section 4.4.1 below).

There are many predictions made by EIC for the ordering of these ICs in a language with constituents of the type $mIC$, just as there are for languages of the opposite type with $IC_m$ such as Japanese and Korean (cf. tables 4.8 and 4.11). The fixed position of the verb in these languages reduced the set of variants to just two, SOV and OSV; but all six are attested in Polish. Ideally we need to know the weight of each single category within each sentence in order to test our unmarked and marked-case predictions, but since Siewierska gives only the aggregate for each category in each order, we will have to make do with these aggregates and define our performance predictions accordingly. As usual, it is vital to make certain assumptions about the precise constituent structure for these three meaningful elements. Different analyses will result in partially different predictions, and if the constituent structure is unclear or ambiguous it will often be impossible to make predictions.

The constituent structures for three of the six orderings of the S, V, and O seem reasonably clear and are set out in (4.16a–c). Possible analyses for the remaining three orders are given in (4.16d–k).

\[(4.16)\]
\[
\begin{align*}
\text{a. } & [mS_{VP}[V_mO]] & \text{SVO} \\
\text{b. } & [vp[V_mO]_mS] & \text{VOS} \\
\text{c. } & [mS_{VP}[mOV]] & \text{SOV} \\
\text{d. } & [V_mS_mO] & \text{VSO} \\
\text{e. } & [vp[V]_mS_{VP[mO]}] & \text{VSO} \\
\text{f. } & [mO_mS_V] & \text{OSV} \\
\text{g. } & [vp[mO]_mS_{VP[mO]}] & \text{OSV} \\
\text{h. } & [mO_{[mS_{VP[V]}]}] & \text{OSV} \\
\text{i. } & [vp[mO]_mS_{VP[V]}] & \text{OSV} \\
\text{j. } & [mO_V_mS] & \text{OVS} \\
\text{k. } & [mO_{[V_mS]}] & \text{OVS}
\end{align*}
\]
I assume that Polish has a VP and that this VP remains intact in SVO, VOS and SOV orders, as shown. VSO is compatible with (at least) two analyses: a flat structure analysis in which V, S, and O are ICs of the sentence (4.16d); and a discontinuous VP analysis in which the subject intervenes between a fronted verb and a direct object that remains VP-dominated (4.16e). Fortunately both analyses result in a number of similar predictions.

The remaining two orders, OSV and OVS, are more problematic because they are each compatible with at least three different analyses, with significantly different consequences for EIC’s predictions. OSV may have a flat structure (4.16f) or a discontinuous VP (4.16g). It may also involve topicalization of the direct object into a sentence-initial topic slot perhaps immediately dominated by E (4.16h). If this is a discourse-configurational node as in Hungarian (cf. section 4.1.2) and Finnish (cf. section 4.1.9), with specific pragmatic and/or semantic correlates (cf. ch. 3.8.2), then we will not expect preposing to be driven by weight in performance, even though the grammaticalization of this discourse-configurational node in left-peripheral position is explainable by EIC (cf. Primus 1991b). For OVS there are also three alternatives, as shown (4.16i–k).

For each of the six orders we can formulate predictions for the aggregate weights of Siewierska’s data. These predictions can be of two kinds: within-structure predictions; and across-structure predictions. For example, in the [S [VO]] order EIC defines a preference for short (or equal) before long (or equal) in Polish. At the sentence level we therefore expect the VP to be longer than or equal to the subject; in the VP the object should be longer than or equal to the verb. If [S [VO]] is inverted to give [[VO] S] we expect the subject NP to be longer than or equal to the VP, while the relative weight of verb and object will be the same as in SVO. These are all within-structure predictions, with weight aggregates reflecting weight distributions that are predicted by EIC’s unmarked and marked case predictions for individual sentences within a text. Long subject NPs exceeding VP in weight will be most optimal in the order [VP NP] and will be expected in this order in the unmarked case. As a result, the aggregate for the NP should be higher than that for the VP, and conversely in the order [NP VP].

But equally the selection of different orderings from among the set of grammatical alternatives is determined, according to our theory, by constituent weights, and this enables us to make predictions for relative weight aggregates across these different orderings. Thus, a subject NP in the order [VP NP] should not only be longer than or equal to its sister VP in length; it should also be longer than or equal to the aggregate for a subject NP in the
order [NP VP]. This follows because if a subject is longer than or equal to VP in [VP NP], and if VP is longer than or equal to a subject in [NP VP], then all things being equal postposed subjects should be longer than or equal to preposed subjects, and a VP that follows a subject should be longer than or equal to one that precedes. All things might not be equal, of course, in the event that the [VP NP] structure happened to be limited, for example, to exceedingly short VPs, with the result that the following subject NP could also be quite short and still satisfy the within-structure prediction of NP ≥ VP, even though it might not satisfy the across-structure prediction. In what follows I shall accordingly separate the (more precise) within-structure predictions from the (more approximate) across-structure predictions.

In what follows I shall accordingly separate the (more precise) within-structure predictions from the (more approximate) across-structure predictions. In table 4.19 I enumerate all the predictions that can reasonably be made, assuming the structural analyses of (4.16). Under each order I enumerate first the within-structure predictions and then the across-structure predictions, giving the weight aggregates in parentheses under each. For the latter predictions, I compare the relevant category, e.g. subject or object in the relevant order, with the corresponding categories in other orders that are predicted to have a lesser or equal weight aggregate. These orders are identified by their structural analysis, given in (4.16). Thus, VP ≥ VP in (b) listed under SVO means that the VP in the SVO order is predicted to have a greater or equal aggregate to the VP in structure (4.16b), which is the (unique) structure for VOS. S ≥ S in (4.16d & e) asserts that the subject in an SVO order should be greater than or equal to the subject in VSO. Imagine that the subject is two words in length. Then whatever the length of the object, the aggregate for the sentence and VP domains will be 83.5 in structure (a), and 75 in (e). In the flat structure (d) the ratio for the sentence domain will also be 75%. The SVO ratio turns out to be systematically higher than it is for (d) and (e) in the VSO order, for all word-total assignments to the subject except for structures in which the subject is a single word. In these cases, SVO has an aggregate of 100, analysis (e) scores 83.5, and (d) is 100%. This justifies the S ≥ S in the (4.16d & e) prediction. The other across-structure predictions are all motivated on the basis of extensive EIC calculations of this sort.

There are some additional predictions that might have been expected in table 4.19 and that are absent. Sometimes one set of predictions will be opposed to another. For example, the right-branching VP in [S [OV]] should have a longer or equal aggregate compared with the left-branching VP of [VO] S, just as the VP of [S [VO]] does. However, within the VP, the V ≥ O prediction in [S [OV]] precludes the co-occurrence of a large O, and this means that the VP is not necessarily expected to be greater than or equal
Table 4.19 Polish $s\{NP \_VP\{V \_NP\}\}^{11}$

Data from Siewierska 1991b

<table>
<thead>
<tr>
<th>Order:</th>
<th>$S$</th>
<th>$V$</th>
<th>$O$</th>
<th>$V$</th>
<th>$O$</th>
<th>$S$</th>
<th>$S$</th>
<th>$O$</th>
<th>$V$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean lengths:</td>
<td>2.5</td>
<td>1</td>
<td>3.7</td>
<td>1</td>
<td>1.3</td>
<td>5</td>
<td>1.6</td>
<td>1.6</td>
<td>1</td>
</tr>
<tr>
<td>Frequency:</td>
<td>69.1%</td>
<td>10.7%</td>
<td>2.2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structure(s):</td>
<td>(a) [S [V O]]</td>
<td>(b) [V O] [S]</td>
<td>(c) [S [O V]]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EIC predicts:</td>
<td>(i) $VP \geq S$</td>
<td>(ii) $O \geq V$</td>
<td>(iii) $S \geq VP$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$[4.7 &gt; 2.5]$</td>
<td>$[3.7 &gt; 1]$</td>
<td>$[2.5 &gt; 1.1]$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(iv) $S \geq S$ in (d&amp;c)</td>
<td>(v) $O \geq O$ in (b)</td>
<td>(vi) $S \geq S$ in (f&amp;g&amp;h)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$[6.6 &gt; 1.3]$</td>
<td>$[3.7 &gt; 1.6]$</td>
<td>$[5 &gt; 1.1]$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Order:</th>
<th>$V$</th>
<th>$O$</th>
<th>$S$</th>
<th>$O$</th>
<th>$S$</th>
<th>$V$</th>
<th>$O$</th>
<th>$V$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean lengths:</td>
<td>1</td>
<td>1.1</td>
<td>6.6</td>
<td>3</td>
<td>1.1</td>
<td>1</td>
<td>3.3</td>
<td>1</td>
</tr>
<tr>
<td>Frequency:</td>
<td>7.1%</td>
<td>1.6%</td>
<td>9%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structure(s):</td>
<td>(d) [VSO] or</td>
<td>(f) (g) or (h)</td>
<td>(j) or (k)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(e) [[V] S [O]]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EIC predicts:</td>
<td>(i) $O \geq S$</td>
<td>(ii) $S \geq V$</td>
<td>(iii) $O \geq O$ in (b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$[6.6 &gt; 1]$</td>
<td>$[1.1 &gt; 1]$</td>
<td>$[6.6 &gt; 1.3]$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(iv) $O \geq O$ in (c)</td>
<td>(v) $V \geq S$</td>
<td>(vi) $S \geq S$ in (f&amp;g&amp;h)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$[6.6 &gt; 1.6]$</td>
<td>$[1 &lt; 1.1]$</td>
<td>$[3.7 &gt; 1]$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Results: Within-structure predictions: 8/10 = correct; 2/10 = close
Across-structure predictions: 10/10 = correct
Total: 18/20 = correct; 2/20 = close

Notes
The mean lengths refer to the average number of words for each constituent in the relevant order.
Frequency refers to the overall proportion of the relevant orders in Siewierska’s sample.
Structure(s) refers to the structures enumerated in (4.16) in the main text.
to the VP in [[VO] S]. On other occasions the ambiguity or unclarity of the correct structural analysis rules out possible predictions, especially for OSV and OVS.

EIC makes a total of 10 within-structure and 10 across-structure predictions for these aggregates. Of these 20 predictions, 18 are correct and two are close. The prediction that V should be longer than or equal to O in the SOV order (and hence that O should be a one-word NP such as a pronoun) is sometimes incorrect: the mean for V is 1, the mean for O is 1.6. The $V \geq S$ prediction for OSV is also incorrect by 0.1! Otherwise EIC does well with these data. A more detailed syntactic investigation of the OSV and OVS structures will make it possible for performance predictions to be tested on these orders as well, and the use of individual phrasal weights within sentences rather than aggregates will strengthen the force of the across-structure predictions by permitting a precise quantification of the set of optimal structures and of the degree of EIC's dispreference in the marked case.

4.2.3 Greek $S\{NP \ V_P \{V \ NP\}\}$

Lascaratou (1989) provides a textual analysis of subject, verb, object orders in Greek which is very similar to that done by Siewierska on Polish. Word lengths are calculated for each NP in each order, and the relative frequency for each order is calculated as well. EIC therefore makes exactly the same predictions for these Greek data as it does for Polish, and we can accordingly set out and test the predictions in table 4.20, in the manner of table 4.19.

Of the 20 total predictions, 17 are correct, and a further two are extremely close: the two predictions under SOV. To save space I shall make no further comment about these Greek data except to point out that the absolute sizes of the Greek subjects and objects are much larger than those of their Polish counterparts. This presumably reflects differences between the types of texts sampled in the two languages. What is interesting is that the two languages and the two sets of text samples are very similar to each other in the relative size differences between their respective phrasal categories. For example, the aggregate for the subject in the Greek VOS order (9 words) is over three times as long as the aggregate for the object (2.8 words), and twice as long as the subject in SVO (4.5 words). Similarly in Polish the subject aggregate in VOS (5 words) is over three times as long as the object aggregate (1.3 words), and twice as long as the subject aggregate in SVO (2.5 words). The relative differentials are the same or similar, despite the absolute size differences.
Table 4.20 *Greek* $S\{NP \; v_p\{V \; NP\}\}$

Data from Lascaratou 1989

<table>
<thead>
<tr>
<th>Order:</th>
<th>$S$</th>
<th>$V$</th>
<th>$O$</th>
<th>$V$</th>
<th>$O$</th>
<th>$S$</th>
<th>$S$</th>
<th>$O$</th>
<th>$V$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean lengths:</td>
<td>4.5</td>
<td>1</td>
<td>11.2</td>
<td>1</td>
<td>2.8</td>
<td>9</td>
<td>2.16</td>
<td>1.11</td>
<td>1</td>
</tr>
<tr>
<td>Frequency:</td>
<td>80.7%</td>
<td>1.2%</td>
<td>1.2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structure(s):</td>
<td>(a) $[S ; [V ; O]]$</td>
<td>(b) $[[V ; O] ; S]$</td>
<td>(c) $[S ; [O ; V]]$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EIC predicts:</td>
<td>(i) $VP \geq S$</td>
<td>(i) $S \geq VP$</td>
<td>(i) $VP \geq S$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$[12.2 &gt; 4.5]$</td>
<td>$[9 &gt; 3.8]$</td>
<td>$[2.11 &lt; 2.16]$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(ii) $O \geq V$</td>
<td>(ii) $0 \geq V$</td>
<td>(ii) $V \geq O$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$[11.2 &gt; 1]$</td>
<td>$[2.8 &gt; 1]$</td>
<td>$[1 &lt; 1.11]$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(iii) $VP \geq VP$ in (b)</td>
<td>(iii) $S \geq S$ in (a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$[12.2 &gt; 3.8]$</td>
<td>$[9.4 &gt; 4.5]$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(iv) $S \geq S$ in (d&amp;e)</td>
<td>(iv) $S \geq S$ in (c)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$[4.5 &gt; 2.4]$</td>
<td>$[9 &gt; 2.16]$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(v) $O \geq O$ in (b)</td>
<td>(v) $S \geq S$ in (d&amp;e)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$[11.2 &gt; 2.8]$</td>
<td>$[9 &gt; 2.4]$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(vi) $O \geq O$ in (c)</td>
<td>(vi) $S \geq S$ in (f&amp;g&amp;h)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$[11.2 &gt; 1.11]$</td>
<td>$[9 &gt; 2.4]$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Order:</th>
<th>$V$</th>
<th>$S$</th>
<th>$O$</th>
<th>$O$</th>
<th>$S$</th>
<th>$V$</th>
<th>$O$</th>
<th>$V$</th>
<th>$S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean lengths:</td>
<td>1</td>
<td>2.4</td>
<td>13.6</td>
<td>1.7</td>
<td>2.4</td>
<td>1</td>
<td>3.6</td>
<td>1</td>
<td>6.5</td>
</tr>
<tr>
<td>Frequency:</td>
<td>1.8%</td>
<td>0.7%</td>
<td>14.4%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structure(s):</td>
<td>(d) $[VSO]$ or</td>
<td>(f) $[g]$ or (h)</td>
<td>(i) $(j)$ or (k)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(e) $[[V] ; S ; [O]]$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EIC predicts:</td>
<td>(i) $O \geq S$</td>
<td>(i) $V \geq S$</td>
<td>(i) $S \geq V$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$[13.6 &gt; 2.4]$</td>
<td>$[1 &lt; 2.4]$</td>
<td>$[6.5 &gt; 1]$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(ii) $S \geq V$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$[2.4 &gt; 1]$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(iii) $O \geq O$ in (b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$[13.6 &gt; 2.8]$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(iv) $O \geq O$ in (c)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$[13.6 &gt; 1.11]$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Results:** Within-structure predictions: 7/10 = correct; 2/10 = close
Across-structure predictions: 10/10 = correct;
**Total:** 17/20 = correct; 2/20 = close

**Notes**
The *mean lengths* refer to the average number of words for each constituent in the relevant order.

*Frequency* refers to the overall proportion of the relevant orders in Lascaratou’s sample.

*Structure(s)* refers to the structures enumerated in (4.16) in the main text.
4.3 Rearrangements of basic word orders in performance

I turn now to rearrangements of basic word orders for which there are clear grammatical arguments in favour of basicness. These grammaticalizations are visible in our data, since they result in a productive retention of the grammaticalized order in certain subsets of the word-total assignments for which EIC would predict a rearrangement. Either EIC may define an equal preference for two orders [A B] and [B A], yet only one occurs at all or is productive, namely the grammaticalized one; or EIC may define a preference for [A B] over [B A], yet only [B A] occurs at all or is productive. It is predicted that such retentions of basic orders will be found only in cases of weak contrary EIC preferences, because the data from all word-total assignments are still subject to EIC's predictions for performance, despite the grammaticalization. If basic orders could be retained despite both strong and weak contrary preferences, there would be too many counterexamples in performance and EIC's unmarked case prediction would not be fulfilled. Grammaticalization can lead speakers to resist an EIC-motivated rearrangement only when there are weak contrary preferences, therefore, and these basic orders must also be productively retained in all the cases that are preferred by EIC and that led, in our theory, to the grammaticalization of a basic order in the first place. As a result, performance data can provide evidence for grammaticalized ordering conventions, just as they provide evidence for one constituent-structure analysis over another.

In what follows I will exemplify a number of cases of this sort, primarily from languages and structures for which the basic-order arguments have been most thoroughly researched. Our focus will be on the nature of, and evidence for, the basic order; it will be less significant in this context whether the rearrangement is effected by a rule or principle of grammar, or by performance alone, as this distinction was discussed in ch. 3.5.2.

4.3.1 English Particle Movement

This rearrangement was extensively discussed in ch. 3.3, 3.4, 3.5.2, and 3.6.2. I assume (following Emonds 1976) that it converts structures of the form \( \text{vp}[V \ NP \ Part] \) into \( \text{vp}[V \ Part \ NP] \), e.g. Joe looked the number up into Joe looked up the number. Here I simply present the text-frequency data from our English sample, in table 4.21.

The data show a clear adherence to EIC's predictions. A full 86% of the instances of these two structures are as optimal as they can be, and tolerance
for the non-optimal [V NP Part] declines gradually as the size of NP increases, in accordance with the marked case prediction. For NPs of five words or more, there are no instances at all of the dispreferred structure.

The grammaticalization of [V NP Part], despite the EIC motivation for [V Part NP] (cf. ch. 3.5.2), is clearly reflected in the NP = Part(1) column: 94% of instances retain the basic order here, even though both orders would be equally optimal for EIC. If ordering were free, we would expect a random distribution. The single-word NPs of English include mass nouns (water), plural count nouns (apples), names (Mary), and pronouns (him). Pre-particle positioning has been grammaticalized for pronouns and is one of the arguments for the basicness of [V NP Part] (cf. I looked him up/*I looked up him). But both orders are grammatical for the other one-word NPs, so the strong performance skewing in favor of the basic order in these cases as well (by a margin of ten to three, or 77%) is a further reflection of grammaticalization. The anomaly of this basic order in Modern English has already been commented on: a majority of instances have to be rearranged in order to comply with EIC (103/179 = 58%); if one excludes the pronominal NPs, the rearrangement ratio for those NPs that can be grammatically rearranged is even higher (103/138 = 75%).

Table 4.21 English Particle Movement

200 pages of data: cf. table 4.2.

vp[V NP Part . . . ] ⇒ vp[V Part NP . . . ] (optional additional material permitted in the position marked “. . .”)

| Basic Order [V NP Part] | 76 |
| Rearranged [V Part NP] | 103 |

<table>
<thead>
<tr>
<th>n = 179</th>
<th>NP = Part (1)</th>
<th>NP &gt; Part :1</th>
<th>:2</th>
<th>:3</th>
<th>:4+</th>
</tr>
</thead>
<tbody>
<tr>
<td>X [V NP Part]</td>
<td>51</td>
<td>21</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Y [V Part NP]</td>
<td>3</td>
<td>45</td>
<td>13</td>
<td>13</td>
<td>29</td>
</tr>
</tbody>
</table>

| Ratio of X/Y: | 94% | 32% | 19% | 7% | 0% |
| Rearrangement to Y: | (6%) | (68%) | (81%) | (93%) | (100%) |

EIC predictions
Unmarked case: 154/179 = most optimal (Y + =), i.e. 86% Correct
Marked case: ratio of X/Y for NP > Part :1 ≥ 2 ≥ 3 ≥ 4+ All correct
Testing EIC's performance predictions

Notice finally that stressed pronouns are not subject to the grammaticalization effect and can occur in either order, *he looked THOSE up*/*he looked up THOSE*, as can all other stressed NPs, *he looked the NUMBER up*/*he looked up the NUMBER*.

### 4.3.2 English Heavy NP Shift

This rearrangement has also been extensively discussed, cf. ch. 3.1, 3.2, and 3.5.2. The text-frequency data are presented in table 4.22.

EIC's unmarked-case prediction is supported in 90.2% of all orders. The grammaticalization of [V NP PP] as a basic order comes through clearly both when NP = PP and when NP > PP by one to three words. In the former case the basic order is still retained in 100% of cases, and so it is when NP > PP by one word. If NP and PP were freely ordered in the VP, we would expect random distribution when NP = PP, and a majority of [V PP NP] as soon as NP > PP by any margin at all, as we saw in the data involving two PPs in table 4.2. Even when NP > PP by two and three words, the grammaticalized order is still retained in the majority of instances, becoming a minority for a differential of four words, and then non-existent for five or more.

EIC's marked case predictions are also well supported, except for one case in which grammaticalization leads to a counterexample. The predictions are correct for the relative proportions of X [V NP PP] to Y [V PP NP] structures when NP > PP. They are also correct for the proportion of Z [V PP NP] to W [V NP PP] when PP > NP. But according to EIC, the non-optimal Z structures in which PP > NP : 1–2 words should also be better and more (or equally) frequently attested than the non-optimal X structures in which NP > PP : 3–4 words. They are not more frequently attested, of course, because X is the grammaticalized basic order which is still productively maintained even when NP > PP by up to four words, and because the Y order requires a relatively heavy NP in order to depart from the basic order.

The interesting contrast between the Particle Movement data and Heavy NP Shift is that the basic order is now clearly aligned with the EIC-preferred order for the vast majority of these structures, and the rearrangements apply to a minority of the word-assignment subsets, those for which EIC defines a particularly strong dispreference. There are just three unexpected examples in which a rearrangement to [V PP NP] takes place, even though the PP is heavier than the NP (two of them by just two words, one of them by four). This represents less than 1% of the total number in which PP > NP.
Table 4.22  *English Heavy NP Shift*\(^\text{12}\)

200 pages of data: cf. table 4.2.

\[
\text{v}_p[V \text{ NP PP}] \Rightarrow \text{v}_p[V \text{ PP NP}]
\]

(including optional verbal particles in VP, but no other ICs to the right of V within the clause)

Basic Order \([V \text{ NP PP}] = 458\)
Rearranged \([V \text{ PP NP}] = 22\)

<table>
<thead>
<tr>
<th>n = 131</th>
<th>NP = PP</th>
<th>NP &gt; PP</th>
<th>:1 word</th>
<th>:2</th>
<th>:3</th>
<th>:4</th>
<th>:5+</th>
</tr>
</thead>
<tbody>
<tr>
<td>X ([V \text{ NP PP}])</td>
<td>68</td>
<td>25</td>
<td>13</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Y ([V \text{ PP NP}])</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>

Ratio of X/Y:
- 100% 100%
- 77% 57%
- 29% 1%
- 0% 0%

Rearrangement to Y:
- (0%) (0%)
- (23%) (43%)
- (71%) (100%)

<table>
<thead>
<tr>
<th>n = 349</th>
<th>PP &gt; NP</th>
<th>:1-2 words</th>
<th>:3-7</th>
<th>:8+</th>
</tr>
</thead>
<tbody>
<tr>
<td>W ([V \text{ NP PP}])</td>
<td>209</td>
<td>103</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>Z ([V \text{ PP NP}])</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Ratio of Z/W:
- 1% 1%
- 0%

Mean length of all NPs = 2.5 words
Mean length of all PPs = 4.6 words
Mean lengths in basic order \([V \text{ NP PP}]\): NP = 2.2; PP = 4.7;
Mean lengths in rearrangement \([V \text{ PP NP}]\): PP = 3.0; NP = 9.0.

*EIC predictions*

Unmarked case: \(433/480 = \text{most optimal (Y + W + =)}\), i.e. 90.2%  
Correct

Marked case: \(\text{ratio of X/Y for NP > PP :1} \geq 2 \geq 3 \geq 3 \geq 4 \geq 5+\)  
All correct

\(\text{ratio of Z/W for PP > NP :1-2} \geq 3-7 \geq 8+\)  
All correct

\(\text{ratio of Z/W for PP > NP :1-2} \geq \text{ratio of X/Y for NP > PP :3-4}\)  
Incorrect

(Reason: \(X = \text{grammaticalized}\))
The highly systematic nature of these Heavy NP Shift rearrangements is of a type that we should now be coming to expect. These rearrangements exist, according to our theory, in order to facilitate the on-line recognition of constituent structure. It is therefore interesting to compare this view with a rather different proposal, involving the informational content of these rearrangements. Just as there are discourse-pragmatic theories of free ordering (cf. ch. 3.8.2, and 4.4 below), so it has also been argued, most systematically by Rochemont (1978, 1986) and Rochemont and Culicover (1990), that a number of rearrangements including Heavy NP Shift are associated with particular pragmatic-semantic properties and belong to a class of "focus" constructions. These authors document a number of important properties of these structures, which are of relevance to their primary concern: description within a generative model. But in the process they pay no systematic attention to what we consider here to be the primary motive for these rearrangements, namely syntactic weight, and I believe that as a result they misrepresent this structure as a pragmatically motivated one.

Notice first that data of the kind exhibited in table 4.22 immediately call into question almost any account that assigns priority to linguistic content over form. The distribution of the two orders correlates in such a precise way with a quantitative measure based on linguistic form that whatever informational theory is proposed as the driving force behind these rearrangements will need to provide informational notions that can also exhibit a correlation with this distribution and with syntactic weight. But it is highly implausible to suppose that particular meanings will be chosen for expression, or not chosen for expression, or chosen with greater frequency versus lesser frequency, based on the syntactic weights of the forms carrying these meanings. Of course, if the language provides several structures for a given meaning, such as "focus on NP," then it is certainly possible for the selection of one of these to be determined by an independent factor, such as syntactic weight, so that expressive and syntactic considerations operate together in accounting for language performance. But then we need to establish which one has priority, and data such as those of table 4.22 suggest that whatever pragmatic-semantic content Heavy NP Shift structures have has no impact on their distribution in performance. Form alone tells the whole story.

The logic of the focus claim can therefore be set out as follows. If Heavy NP Shift has a certain focusing property, it cannot be the only structure in English that carries it: if it were, form would be driving content entirely, which is implausible. However, if it shares its focusing property with at least one other structure, then this property does not appear to determine
Rearrangements of basic word orders in performance 185

its distribution. This suggests that Heavy NP Shift exists primarily for reasons of syntactic weight. This in turn raises the question of whether it is necessarily associated with any focusing properties at all.

Notice that whatever focusing properties it has are not unique to these rearrangements. Rochemont and Culicover (1990: 24) point out, for example, that the shifted structure (4.17) can be an appropriate response to the WH-question in (4.18a) but not to (4.18b):

(4.17) John gave to Mary a very valuable book.

(4.18) a. What did John give to Mary?
   b. To whom did John give a very valuable book?

This is correct: but the unshifted (4.19) is an equally appropriate response to (4.18a),

(4.19) John gave a very valuable book to Mary.

and (4.19) is also an appropriate response to (4.18b). Therefore (4.17) is actually more limited in its focus options than the basic structure from which it is derived. It is also more limited in other ways: you cannot extract from the center-embedded PP in a Heavy NP Shift structure (cf. the "freezing principle" of Culicover 1976: 298–301), whereas you can extract the center-embedded NP in the basic order:

(4.20) a. Who did John give a very valuable book to Mary?
   b. *Who did John give to Mary a very valuable book?
   c. What did John give Mary a very valuable book?

These are syntactic and semantic restrictions that need to be accounted for (and the kind of performance account that I outlined for Heavy NP Shift in ch. 3.5.2 may or may not have the potential to explain them, in conjunction with a host of other considerations and assumptions which it is not my goal to explore in the present context). But they provide no motive for converting (4.19) into (4.17) in performance – on the contrary, they will block the rearrangements in many cases. Based on these considerations, Heavy NP Shift could just as well be called a "restricted focus construction" or a "restricted extraction construction."

Do these restrictions force an obligatory focus interpretation on the NP, as Rochemont and Culicover claim? I see no evidence that they do. The matter is hard to decide, given the slipperiness of the fundamental concept, focus, which is never adequately defined. But unless something more precise is said
about "assertion focus," it will be possible to classify under this rubric a whole variety of constructions whose true explanation lies elsewhere, because no one is really sure what focus means.

Recall the proposed focus position in Hungarian (section 4.1.2). There are problems of definition in this language too, but at least we see clear evidence in our data for something other than EIC forcing constituents into this position (cf. n. 3 ch. 4). For Heavy NP Shift we do not.

Rochemont and Culicover provide one criterion that is helpful, however. They argue that a sufficient (but not necessary) condition for being a focus constituent is that this constituent should not be "context-construable" (cf. Rochemont 1986), which is their definition of new information in discourse. In other words, "if α is not c-construable (= under discussion), then α is a focus" (Rochemont and Culicover 1990: 20). For Heavy NP Shift this predicts that the PP will not be associated with new information, since this is not a focus constituent, but that the postposed NP may be.

Using the quantified text methodology of Givon (1983, 1988) (cf. section 4.4), in which 20 clauses are examined prior to each referential constituent, I tested to see whether either the direct object NP, or the NP in the PP of a Heavy NP Shift Structure, could be new information in the data of table 4.22. This is not a precise test of "context-construability," but then this latter is not a precise term, and I therefore went through my data a second time to see whether NPs classified as "unpredictable" or new in Givon's terms could nonetheless be "context-construable" in any reasonable sense. Only in one marginal case did this second pass result in a possible change in the classification. This test can give us at least an approximate indication of whether Heavy NP Shift is being driven by information structure.

The 22 rearranged structures in my data were distributed as follows. In nine of the \([V \text{ PP NP}]\) orders (possibly eight because of the one marginal case), both the PP and the NP were equally new: in these cases, therefore, the PP is associated with new information, which goes against the prediction. In three further instances, the PP is new and the NP is previously mentioned. These go against the prediction even more clearly: a focused NP can either be new or given in their theory, but the PP should not be new. In the remaining 10 (possibly 11), the PP is previously mentioned and the NP is new. These cases are in accordance with the prediction.

These figures do not support the view that there are any obligatory and necessary informational properties associated with Heavy NP Shift. These structures are more limited than their basic-order counterparts in the constituents that can be WH-questioned semantically and WHmoved syntacti-
Rearrangements of basic word orders in performance

cally. But this actually prevents them from occurring, and restricts their frequency of occurrence in performance: they appear to have no additional positively specified properties that predict when they will be used, other than their improved EIC ratios, and it is these that determine the basic structures to which Heavy NP Shift applies, and the frequency with which it does so.

There is, however, one other factor that can potentially affect matters: stress. The one convincing counterexample to EIC in our data in which the PP exceeds the NP by four words in the order [V PP NP] has a strong contrastive stress on the NP. The example is given in (4.21):\(^{13}\)

(4.21) \ldots if the totality of electrical signals represents in some other frame of reference "the mind," \ldots

This example is explicitly contrasting the mind with the brain, and the stress is indicated orthographically. Similar examples are not hard to concoct: I gave to my father these BOOKS; and even I gave to my father THESE. But once again the contrastive stress options of Heavy NP Shift are a proper subset of those found in the corresponding basic orders:

(4.22) a. I gave these books to my father (not to my mother)
    b. I gave these books to my father (not these records).

(4.23) a. *I gave to my father these books (not to my mother)
    b. I gave to my father these books (not these records).

The rearrangements are restricted contrastive stress structures, therefore. Such stress is always possible on constituents in right-peripheral position in the VP, but may fall on other constituents in the basic order. This is a formal and structural property of these constructions, with consequences for the (more limited) expressive power of the rearrangements. It is not a necessary property, but because structures such as (4.23b) are grammatically possible, they can be used on occasions of contrastive stress, even in the absence of an EIC motivation. The statistics of my sample suggest that this option is used relatively rarely: just one of the 22 rearranged structures could be explained this way. The same contrastive stress motive also exists for Particle Movement structures, as we showed in the last section.

In conclusion, while more remains to be said about the grammar of Heavy NP Shift (and the work of Rochemont and Culicover is exemplary in the way it lays out the issues and proposes solutions within the context of a formal grammar) I believe it is misleading to characterize this structure as a focus construction in any interesting sense.
4.3.3 German NP–PP rearrangements in the middle field

A similar rearrangement of a direct object NP and a PP can take place in the middle field (Mittelfeld) of German, and so it is instructive to compare this language with Heavy NP Shift structures in English. German does not permit postposing of a direct-object NP to the right of a final verb or verbal dependent, in the written language at least (cf. Hawkins 1986: 147–150 for a summary of the facts), but there are productive alternations of [NP PP V] and [PP NP V]. The NPs to be examined here are accusative-marked. Since PPs in German are typically longer than NPs, just as they are in English, and since both categories are typically recognized on their left peripheries (German being predominantly prepositional), EIC defines a preference for [NP PP], whether PP is the last element in the clause, or whether a V follows, in which case the preference is defined on a left-to-right basis (cf. the discussion of [{mIC1 mIC2} V] in Turkish, section 4.1.4). Predictably, [NP PP] is much more frequent than [PP NP] in performance, and is widely considered to be the basic, grammaticalized order (cf. e.g. Lenerz 1977). When the NP dominates a pronoun, the pronoun almost always comes first, again in accordance with EIC. Pronouns are short single-word items, and the grammar of German has positioned them before other categories of the Mittelfeld that typically exceed one word in length (cf. chs. 5.4.2 and 6.3.3). In table 4.23 I present data on the distribution of an accusative-marked non-pronominal NP (including proper names in accusative function) and a PP in the structural template indicated. (4.24) is an example:

    “the woman has the-ACC documentary in-the television seen,” i.e.
The woman has seen the documentary on television.


The unmarked case prediction is fulfilled with a score of 84.3%, and all the marked case predictions that can be tested on these data are correct as well. The grammaticalization of [NP PP] as a basic order is clearly evident in the 86% skewing when NP = PP, and also in the fact that this order retains its majority status even when NP > PP by one or two words. There is no such skewing when PP > NP: the order [NP PP] immediately enjoys an 85%
Rearrangements of basic word orders in performance

100 pages of data: first 100 pp. from P. Handke *Die linkshändige Frau*, Suhrkamp

<table>
<thead>
<tr>
<th></th>
<th>NP = PP</th>
<th>NP &gt; PP</th>
<th>:1–2 words</th>
<th>:3–7</th>
<th>:8+</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>18</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratio of X/Y:</td>
<td>86%</td>
<td>56%</td>
<td>33%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>Rearrangement to Y:</td>
<td>(14%)</td>
<td>(44%)</td>
<td>(67%)</td>
<td>(100%)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>PP &gt; NP</th>
<th>:1–4 words</th>
<th>:5+</th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>40</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Z</td>
<td>7</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Ratio of Z/W:</td>
<td>15%</td>
<td>0%</td>
<td></td>
</tr>
</tbody>
</table>

**EIC predictions**

Unmarked case: 75/89 = most optimal (Y + W + =), i.e. 84.3% Correct

Marked case: ratio of Y/X for NP > PP :1–2 ≥ 3–7 ≥ 8+ All correct

ratio of Z/W for PP > NP :1–4 ≥ 5+ Correct

advantage for a one-to-four-word differential, rising to 100% for five or more words.

With regard to the informational status of the rearranged (4.24b), Lenerz (1977) argues that this sentence has the same kind of restricted focus capability as Heavy NP Shift in English, so that (4.24a) can be an answer to the question “What has the woman seen on television?” and “Where has the woman seen the documentary?”, whereas (4.24b) can only answer the former. This claim has been disputed, however, by Gadler (1982) and Lötscher (1981), who argue that things are not so clear-cut. In any case,
Testing EIC’s performance predictions

as in the English structure, I see the rearrangement of (4.24) as being driven by EIC, for processing reasons, and the data support this.

Notice also that although Standard Written German does not permit postposing of the accusative NP to the right of V, forcing what would seem to be an intolerably long VP recognition domain, it does have a very productive means of shortening this domain, namely Extrapolation from NP, which is applied with considerable frequency to long middle-field constituents (cf. section 4.3.5).

4.3.4 English Extrapolation

The English rule of Extrapolation has been discussed in chs. 3.3 and 3.4. Here I shall summarize some very extensive performance data collected by Peter Erdmann and presented in Erdmann (1988). These data come from a corpus of written British English comprising 15 novels, 15 book-length popular scientific writings, 15 editions of the Daily Telegraph, and 15 editions of the Spectator. The data enable us to test EIC’s unmarked case prediction as well as some more particular predictions for these structures, but since Erdmann does not give the word length for each constituent in each individual sentence, it is not possible to test the marked case predictions.

Erdmann examines both the extrapolation of a finite S and the extrapolation of an infinitival VP, abbreviated here as VP. The S and VP are always grammatical subjects, either in the maintained basic order of surface structure or in the underlying basic order from which the extrapolations are derived. As for the matrix VP, he selects instances that contain the verb be plus a predicate adjective, and he distinguishes carefully between adjectives that are unmodified (or modified by a pre-adjectival adverb), which he describes as "light"; and those that have some additional modifying constituent following the adjective, which appears from his examples to be most typically a PP, and which renders the adjective "heavy." The sentence types in his data are exemplified by (4.25) and (4.26):

(4.25) a. That their time should not be wasted is important (for us all).
   b. It is important (for us all) that their time should not be wasted.

(4.26) a. To dine in a good restaurant is delightful (for us all).
   b. It is delightful (for us all) to dine in a good restaurant.

The structural alternation that we are dealing with in (4.25) can therefore be represented as s[S VP[is ADJP[Adj]]] versus s[is VP[ADJP[Adj] S]], in which
AdjP dominates a single-word adjective sometimes modified by a preceding adverb; and as $s[S\ VP[is\ Adj\ PP]]$ versus $s[\it\ VP[is\ Adj\ PP]\ S]$, for adjectives modified by a PP. In (4.26) we have the same alternations with a $\overline{VP}$ subject. EIC's preferences are a function of the IC-to-word ratios for both the S and VP domains, since both are affected by the rearrangement. By removing a heavy constituent, either S or $\overline{VP}$, from the subject position and hence from the S domain, Extraposition greatly improves the ratio for S. If the VP, prior to Extraposition, comprises only is and $\overline{Adj\ PP}$, i.e. with an IC-to-word ratio of $2.2 = 100\%$, then the addition of an $\overline{S}$ will simply maintain this optimality, since $\overline{S}$ is constructed by a left-peripheral Comp and there are now three ICs in the $\overline{VP}$, i.e. $3/3 = 100\%$. I assume that $\overline{VP}$ is constructed either by the infinitival to or by the infinitival verb itself, both of which are to the left of all other $\overline{VP}$-dominated material, and so the VP will either be optimal or considerably improved relative to the unextraposed orders in these cases as well. With just a single-word adjective phrase in the VP, therefore, the basic order will always have worse ratios than the extraposed order (even with a pre-adjectival adverb), and the size of the differential will correlate directly with the size of $\overline{S}$ and $\overline{VP}$: the longer they are, the worse the ratios of the basic order and the bigger the improvement brought about by Extraposition. This is illustrated in table 4.24 sections I and II. As $\overline{S}$ increases from four words to six, the ratio for the $\overline{S}$ domain declines in the basic order, while both the $\overline{S}$ and VP domains have optimal ratios in the transformed structure. EIC's preference for the latter increases accordingly.

If there is an AdjP dominating a PP, or indeed any other post-adjectival phrasal constituent, then whatever length this constituent has in the basic order $\overline{VP[is\ Adj\ PP]}$ will not diminish the optimal $2/2$ ratio for the VP domain, since Adj constructs AdjP. But if an extraposed $\overline{S}$ or $\overline{VP}$ is attached to the right of this AdjP within VP (following the analysis of Reinhart 1983 and others), then $\overline{S}$ or $\overline{VP}$ will be the rightmost IC of the VP, and the length of the post-adjectival PP will substantially affect the ratio for VP. The longer it is, the worse the VP will become, and this will affect the overall aggregate for a sentence, to the point where it can offset any advantage in applying Extraposition.

In table 4.24 sections III–VII I calculate the scores for an AdjP comprising a PP, for all three logical possibilities: $\overline{S} > PP$ (which is the most frequent case); $PP > \overline{S}$; and $\overline{S} = PP$. When the $\overline{S}$ is four words longer than the PP (III), there is a significant differential of 15.5 between the basic and extraposed orders, though this figure is only half the size of the differential for the
Testing EIC's performance predictions

<table>
<thead>
<tr>
<th>Table 4.24</th>
<th>IC-to-word ratios for English Extraposition</th>
</tr>
</thead>
<tbody>
<tr>
<td>s[(\bar{S} \text{ vp}{is, Adjp[Adj]}]</td>
<td>s[(it \text{ vp}{is, Adjp[Adj] \bar{S}}]</td>
</tr>
<tr>
<td>I</td>
<td>(\bar{S} = 4)</td>
</tr>
<tr>
<td></td>
<td>VP CRD: (2/2 = 100%)</td>
</tr>
<tr>
<td>II</td>
<td>(\bar{S} = 6)</td>
</tr>
<tr>
<td></td>
<td>VP CRD: (2/2 = 100%)</td>
</tr>
<tr>
<td>III</td>
<td>(\bar{S} = 6) PP = 2</td>
</tr>
<tr>
<td></td>
<td>VP CRD: (2/2 = 100%)</td>
</tr>
<tr>
<td>IV</td>
<td>(\bar{S} = 4) PP = 2</td>
</tr>
<tr>
<td></td>
<td>VP CRD: (2/2 = 100%)</td>
</tr>
<tr>
<td>V</td>
<td>(\bar{S} = 4) PP = 4</td>
</tr>
<tr>
<td></td>
<td>VP CRD: (2/2 = 100%)</td>
</tr>
<tr>
<td>VI</td>
<td>(\bar{S} = 4) PP = 6</td>
</tr>
<tr>
<td></td>
<td>VP CRD: (2/2 = 100%)</td>
</tr>
<tr>
<td>VII</td>
<td>(\bar{S} = 4) PP = 8</td>
</tr>
<tr>
<td></td>
<td>VP CRD: (2/2 = 100%)</td>
</tr>
</tbody>
</table>
Rearrangements of basic word orders in performance

193 cases exemplified in I and II where there is no PP. For a two-word difference (IV), the differential of 10 is still substantial. When \( S = PP \) (V), the aggregates are almost identical, i.e. there is now little incentive to extrapose. And when the PP is longer than \( S \), it is the basic order that is increasingly preferred (VI–VII). Extraposition would produce a worse aggregate in these cases and so is predicted not to occur.

There are two predictions that we can derive from table 4.24 and test on the quantitative data of Erdmann's corpus. First, for alternations of types I and II, i.e. without a post-adjectival PP or other modifier within AdjP, the extraposed structure will always be preferred, for all lengths of \( S \) and \( VP \) (though cf. n. 14 ch. 4), and hence we can make the usual prediction that extraposed structures will be found in the unmarked case. Second, while we do not have figures for the relative sizes of \( S \) and PP in each performance instance, it is clear from a comparison of III–VII with I–II that the benefits of Extraposition are always less when a post-adjectival PP is present, and that the basic order is even preferred when \( PP > S \). Hence, we expect, all things being equal, that the proportion of these structures that undergo Extraposition will be less than it is for adjectives without post-modification. These predictions are tested in table 4.25.

Both predictions are confirmed. Extraposition applies to 95% of the structures with an \( S \) subject and an adjective without post-modification, and to 96% of those with a \( VP \) subject. For the adjective phrases with post-modification, the rearrangement ratios are significantly less: 76% for \( S \), and 87% for \( VP \).

Erdmann (1988) gives a third set of data that is also very interesting, involving sentential subjects in passive constructions, i.e. alternations such as That there was any wrongdoing was denied and It was denied that there was any wrongdoing. As in the other cases, the extraposition structure accounts for 96.1% of the data (268/279). But Erdmann then goes on to examine which of these structures retains the agent by-phrase, i.e. That there was any wrongdoing was denied by all parties versus It was denied by all parties that there was any wrongdoing. What he finds is that the by-phrase greatly prefers the basic structure to its extraposed counterpart. A total of 23/34 basic structures retain the by-phrase (68%), whereas only 9/277 extraposition structures do (3%).

It is most plausible to assume that this by-phrase is a PP immediately attached to VP rather than to AdjP, and hence that the structural alternation is now \( S \) \( VP[is \ Participle \ PP[by \ NP]] \) versus \( S \) \( VP[is \ Participle \ PP[by \ NP] S] \), with the VP comprising an extra IC compared with the structures of table 4.24. It turns out that it makes very little difference for EIC whether the PP is
Testing EIC's performance predictions

Table 4.25 English Extraposition

Data from Erdmann 1988.

I  S Extraposition

<table>
<thead>
<tr>
<th></th>
<th>n = 385</th>
<th>n = 67</th>
</tr>
</thead>
<tbody>
<tr>
<td>X [S [is Adj]]</td>
<td>20</td>
<td>W [S [is [Adj PP]]]</td>
</tr>
<tr>
<td>Y [it [is Adj S]]</td>
<td>365</td>
<td>Z [it [is [Adj PP] S]]</td>
</tr>
<tr>
<td>Ratio of X/Y:</td>
<td>5.2%</td>
<td>Ratio of W/Z:</td>
</tr>
<tr>
<td>Rearrangement to Y:</td>
<td>(94.8%)</td>
<td>Rearrangement to Z:</td>
</tr>
</tbody>
</table>

EIC predictions

Unmarked case: of X and Y, Y = most optimal for any length of S; 365/385 = most optimal, i.e. 94.8% Correct

When PP > S, the basic order W is preferred to the extraposed Z; the overall ratio of W/Z for all word-total assignments should therefore exceed the overall ratio of X/Y, and rearrangements to Z should be relatively less frequent than rearrangements to Y. Correct

immediately attached to AdjP or to VP, either in Passive structures or in structures such as those in examples (4.25) and (4.26). In the basic order the VP will be either an optimal 2/2 or an optimal 3/3, and the S ratio will reflect the length of S. In the extraposed order, a long PP will still add significant non-IC structure to the VP domain and the length of S will be irrelevant. The ratios for VP will be slightly higher in the extraposed structure because the word total is now divided by four ICs rather than by three. For

II  VP Extraposition

<table>
<thead>
<tr>
<th></th>
<th>n = 571</th>
<th>n = 78</th>
</tr>
</thead>
<tbody>
<tr>
<td>X [VP [is Adj]]</td>
<td>22</td>
<td>W [VP [is [Adj PP]]]</td>
</tr>
<tr>
<td>Y [it [is Adj VP]]</td>
<td>549</td>
<td>Z [it [is [Adj PP] VP]]</td>
</tr>
<tr>
<td>Ratio of X/Y:</td>
<td>3.9%</td>
<td>Ratio of W/Z:</td>
</tr>
<tr>
<td>Rearrangement to Y:</td>
<td>(96.1%)</td>
<td>Rearrangement to Z:</td>
</tr>
</tbody>
</table>

EIC predictions

Unmarked case: of X and Y, Y = most optimal for any length of VP; 549/571 = most optimal, i.e. 96.1% Correct

when PP > VP, the basic order W is generally preferred to the extraposed Z; the overall ratio of W/Z for all word-total assignments should therefore exceed the overall ratio of X/Y, and rearrangements to Z should be relatively less frequent than rearrangements to Y. Correct
example, the VP CRD would be $4/5 = 80\%$ in sections III and IV of table 4.24 rather than $3/5 = 60\%$, making the extraposed aggregate 90 rather than 80 and the differentials correspondingly larger.

The retention of an agentive by-phrase in non-extraposed structures makes sense in our terms because the VP can be completely optimal with or without the PP: $\text{VP}[\text{is Participle PP}]$ has a $3/3$ ratio, and $\text{VP}[\text{is Participle}]$ has $2/2$. Hence, within this predictably limited set of basic structures we predict, correctly, that a retained by-phrase can be productive. But for the larger set of extraposed structures, an otherwise optimal VP without a PP, $\text{VP}[\text{is Participle } \bar{S}]$ (with a $3/3$ ratio), is no longer optimal when a PP intervenes, $\text{VP}[\text{is Participle PP } \bar{S}]$, and becomes progressively worse the greater the length of PP. Since the mean length of all VP-dominated PPs in English is a quite substantial 4.6 words, the EIC score for VP will be significantly non-optimal if the length of agentive by-phrases is anywhere close to this mean. For these reasons it is predicted that by-phrases should not be productively retained in extraposed structures, and they are not.

Notice also that passive by-phrases differ from most of the post-adjectival modifiers in table 4.25 in that they are associated with a large degree of optionality from the point of view of language expression. The structural alternations that need to be considered in this context, therefore, are not just \text{S}[\text{VP}[\text{is Participle PP}]] versus \text{S}[\text{VP}[\text{is Participle PP } \bar{S}]]$, but also the corresponding structures without PP at all.

Let us reflect finally on the grammaticalization issue posed by these extraposition data: what evidence is there for saying that the sentential subjects in table 4.25 (i.e. structures X and W) are in the basic order, when the great majority of them need to be transformed? As in the case of Particle Movement, there are more rearrangements than basic orders in language performance.

In fact, the evidence for grammaticalization is very clear and is exactly what we have seen in the other cases hitherto: the basic order is retained in a subset of the cases for which EIC defines a contrary preference, namely all the X structures and presumably some of the W structures too. It cannot be retained to a degree that would go against the unmarked case prediction, however, and I would predict that the cases in which it is retained involve relatively weak contrary preferences only. Moreover, if the extraposition structures were the grammaticalized basic orders, we would not in general expect to see any sentential subjects at all, since the rearrangements would result in worse EIC ratios, except for those rare cases in which the PP exceeded the $\bar{S}$ by well in excess of four words, i.e. sufficient to make the rearrangement worthwhile. This may be the correct analysis for all stages of
the history of English up to the present, since it is only in Modern English that sentential subjects begin to be recorded at all (cf. Stockwell 1977a).

Why then are sentential subjects now the basic order? It seems to me that there is an interesting contrary tension operating here to which we can appeal in order to explain the diachronic variation. If we count sentential subjects as part of the larger category of subjects, then the aggregate size of a subject is not sufficient to offset the basic regularity that we saw evidence for in Polish and Greek: the subject is typically shorter than the VP, and is accordingly positioned to its left. A language with a strongly grammaticalized subject relation, such as English (cf. Hawkins 1986 for both syntactic and semantic documentation of this fact, in comparison with German and also Old English), will automatically include sentential subjects under all the generalizations that hold for non-sentential subjects, unless there are independent reasons to the contrary. There are independent reasons to the contrary, defined by EIC, when the subject is especially long. Hence, at a stage in the history of English when the subject relation did not exert strong pressure for the uniform treatment of all potential subjects, there would have been no reason for grammaticalizing a basic order that then had to be rearranged in most instances in performance; but as the subject relation came to play an ever-increasing role in the grammar of English, as raising rules evolved that were sensitive to it, and as the subject came to subsume more and more θ-roles of the type that had earlier been assigned to other NPs (cf. again Hawkins 1986), so the pressure presumably grew to call in all the potential subjects under this expanding generalization. EIC then backs down, permits $S$ in subject position, and agrees to rearrange sentential subjects as much as 96% of the time in performance. The general prediction made by this account, of course, is that sentential subjects in subject position will be found only in languages with a strongly grammaticalized subject relation, or else in languages where the subject position is already good for sentential subjects (and not necessarily for non-sentential subjects, as in Malagasy!). One could presumably quantify the number of subject-sensitive rules in languages in order to test this.

4.3.5 German Extraposition from NP

Consider now another rearrangement rule with offsetting benefits across different domains: Extraposition from NP. This rule postposes the $S$ from a complex NP, i.e. an NP containing a relative clause (the student that we met) or a complement clause (the fact that he was drunk), and is highly productive
in languages such as English and German in which the S is recognized on its left periphery. The effect of postposing is to improve the EIC ratios of the domains that contain the complex NP, much as Extraposition improved the ratio for the S domain when S and VP were postposed in the examples of the last section. But in the process, constituents of the complex NP are separated, and this lengthens the recognition domain for NP, making its processing more complex – on the assumption that the NP still forms a constituent and is discontinuous. I shall argue that Extraposition from NP does create a discontinuous constituent, and the data that I shall present support this.

In her discussion of extraposition rules in English, Reinhart (1983) argued that whereas an S postposed by Extraposition could be shown to be attached to VP, an S postposed by Extraposition from NP should be attached to S. Among her many arguments, she cited sentence pairs such as the following:

(4.27) a. *It \_\_\_\_[surprised heri S[that Maryi was sick]]
    b. It \_\_\_[surprised Maryi S[that shei was sick]]

(4.28) a. Nobody would \_\_\_[call heri before noon] S[who knew Maryi well]
    b. Nobody would \_\_\_[call Maryi before noon] S[who knew heri well]

The ungrammaticality of (4.27a) in the co-reference reading can be explained by saying that the pronoun her c-commands the antecedent Mary, as long as Mary and the S that contains it are dominated by VP. The corresponding (4.28a) is grammatical, however, and this can be explained on the assumption that the S in the Extraposition from NP structure is outside the domination of VP. Reinhart accordingly assumes that the attachment site is S.

This is not the only option, however. The possibility of co-reference in (4.28a) would also be predicted by an analysis in which the S remained discontinuously attached to the subject nobody, and McCawley (1982) gives a number of arguments in favor of discontinuity for this rule. For example, the extraposed S of a complex NP continues to be subject to Ross's (1967) Complex NP Constraint. If it is no longer dominated by NP, this is not predicted; if it remains under the domination of NP, it is. Moreover, Reinhart's analysis does not produce the right results when Extraposition from NP applies to non-subjects, whereas the discontinuity analysis does. Compare (4.29) and (4.30):

(4.29) a. *I \_\_\_[lent heri \_\_\_[the book that Maryi ordered] yesterday]
    b. I \_\_\_[lent Maryi \_\_\_[the book that shei ordered] yesterday]
(4.30)  a. *I \text{VP}[lent her_{i} \text{NP}[the book] yesterday \text{NP}[that Mary_{i} ordered]]
    b. I \text{VP}[lent Mary_{i} \text{NP}[the book] yesterday \text{NP}[that she_{i} ordered]]

The pronoun c-commands the antecedent in the un-extraposed (4.29a) and is predictably ungrammatical in this reading. But (4.30a), which has undergone Extrapolation from NP, has exactly the same status as (4.29a). It should not be ungrammatical, however, if the relative \( S \) is attached to the matrix \( S \), since this puts it outside the c-command domain of the VP-dominated pronoun. The discontinuity hypothesis correctly predicts that (4.29) and (4.30) should pattern identically.

For reasons such as these I shall assume the discontinuity hypothesis, and our data will be shown to support this over the S-attachment analysis. I shall also focus on German data rather than English, since the verb-final rule of German provides more structures to which Extrapolation from NP can potentially apply, and also since extensive quantitative data have been collected for this language by Christine Mooshammer, Thomas Shannon, and myself. Notice that the crucial sentences corresponding to (4.29a) and (4.30a) are also ungrammatical in German, thereby supporting the same discontinuity analysis for this language as well.\(^{15}\)

(4.29') *Ich habe \text{VP}[ih}_{r_{i}} \text{NP}[das Buch, das Maria_{i} bestellt hat] geliehen
    "I have her-Dat the book that Maria ordered has lent," i.e. I lent her_{i} the book that Maria_{i} ordered.

(4.30') *Ich habe \text{VP}[ih}_{r_{i}} \text{NP}[das Buch] geliehen \text{NP}[das Maria_{i} bestellt hat]]

The interesting tension defined by EIC for these structures is the following: the longer the \( S \) of a complex NP contained within a verb-final VP, the worse the VP recognition domain becomes; but the greater the distance between the separated NP constituents, the worse the NP domain becomes. Example (4.30') contains only one intervening non-NP-dominated constituent, the verb geliehen. The added cost for the NP in this case is minimal, while the benefit for VP is considerable, since the last IC of the VP, the verb, now stands much further to the left, and this significantly shortens the VP recognition domain.

Table 4.26 illustrates this interaction between the VP and the NP domains when Extrapolation from NP applies to a VP-dominated NP. Sections I—III are structures in which the complex NP immediately precedes the verb and is extrapolated the shortest possible distance, i.e. over one intervening word. The figures will be the same if, instead of a finite or non-finite \( V \), there is a single-word verbal dependant in final position, such as a particle. The aggregate for
the VP and NP CRDs is always lower in the basic order, and the improvement brought about by Extraposition from NP is in direct proportion to the size of \( S \).\(^{16}\)

The situation changes radically, however, when even a single constituent intervenes between the complex NP and V, as shown in sections IV–X. Section IV gives the ratios for a four-word \( S \), and for a two-word constituent, XP, intervening between NP and V. The VP now has three ICs. The aggregate for the basic order is 66.5, but that of the extraposed structure is only 55.5, i.e. there is no motivation to extrapose. The reason is that the benefit for the VP (which jumps from 33% to 60%, i.e. +27%) is offset by a deterioration in the NP score (which declines from 100% to 50%, i.e. −50%). In section I the deterioration in the NP ratio (100% to 75%, i.e. −25%) is insufficient to offset the benefit for VP (which improves from 29% to 67%, i.e. +38%) and so the aggregate for the extraposed structure is higher (71 versus 64.5). Extraposition is predicted for I, but the basic order is most optimal in IV, and it maintains this advantage when \( S \) is eight words in length (V), and even for 16 words (VI). Clearly, the prediction made by EIC in these cases is that Extraposition from NP should not apply over a distance of three intervening words.

Consider, by contrast, the case of a shorter intervening XP of just one word. With \( S \) at four words, there is still no incentive to extrapose (VII). But as \( S \) increases in size, e.g. to eight words (VIII), there is an increasing preference for the extraposed structure. But VIII is the only word-assignment type in which extraposition is preferred over both an XP and a V. Section IX illustrates the significant preference for the basic order when XP has three words and \( S = 4 \), and this preference remains significant even when \( S = 16 \) (X).

Summarizing, when a complex NP is immediately adjacent to V, Extraposition from NP is always preferred; if another constituent intervenes, the basic order is preferred, unless this constituent is just a single word length, in which case some instances of the extraposed structure are preferred. Moreover, the ratio for the VP domain always declines in the basic order in direct proportion to the increasing size of \( S \), and hence the degree of preference for the extraposed structure is also in direct proportion to the size of \( S \). This degree may systematically exceed that of the basic order, as in structures I–III; it may be less than that of the basic order (IV–VI and IX–X), improving relative to this latter as \( S \) grows in size; or the score for the extraposed structure may start lower and become progressively higher (VII–VIII). Comparing across all word-total assignments, therefore, the rela-
Testing EIC's performance predictions

Table 4.26 IC-to-word ratios for German Extraposition from NP for a VP-dominated NP

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>I</strong> S = 4</td>
<td></td>
</tr>
<tr>
<td>VP CRD: 2/7 = 29% VP CRD: 2/3 = 67% Differential = (+)6.5</td>
<td></td>
</tr>
<tr>
<td>NP CRD: 3/3 = 100% NP CRD: 3/4 = 75%</td>
<td></td>
</tr>
<tr>
<td>Aggregate = 64.5  Aggregate = 71</td>
<td></td>
</tr>
<tr>
<td><strong>II</strong> S = 8</td>
<td></td>
</tr>
<tr>
<td>VP CRD: 2/11 = 18% VP CRD: 2/3 = 67% Differential = (+)12</td>
<td></td>
</tr>
<tr>
<td>NP CRD: 3/3 = 100% NP CRD: 3/4 = 75%</td>
<td></td>
</tr>
<tr>
<td>Aggregate = 59    Aggregate = 71</td>
<td></td>
</tr>
<tr>
<td><strong>III</strong> S = 16</td>
<td></td>
</tr>
<tr>
<td>VP CRD: 2/19 = 11% VP CRD: 2/3 = 67% Differential = (+)15.5</td>
<td></td>
</tr>
<tr>
<td>NP CRD: 3/3 = 100% NP CRD: 3/4 = 75%</td>
<td></td>
</tr>
<tr>
<td>Aggregate = 55.5  Aggregate = 71</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IV</strong> S = 4 XP = 2</td>
<td></td>
</tr>
<tr>
<td>VP CRD: 3/9 = 33% VP CRD: 3/5 = 60% Differential = (-)11.5</td>
<td></td>
</tr>
<tr>
<td>NP CRD: 3/3 = 100% NP CRD: 3/6 = 50%</td>
<td></td>
</tr>
<tr>
<td>Aggregate = 66.5    Aggregate = 55</td>
<td></td>
</tr>
<tr>
<td><strong>V</strong> S = 8 XP = 2</td>
<td></td>
</tr>
<tr>
<td>VP CRD: 3/13 = 23% VP CRD: 3/5 = 60% Differential = (-)6.5</td>
<td></td>
</tr>
<tr>
<td>NP CRD: 3/3 = 100% NP CRD: 3/6 = 50%</td>
<td></td>
</tr>
<tr>
<td>Aggregate = 61.5    Aggregate = 55</td>
<td></td>
</tr>
<tr>
<td><strong>VI</strong> S = 16 XP = 2</td>
<td></td>
</tr>
<tr>
<td>VP CRD: 3/21 = 14% VP CRD: 3/5 = 60% Differential = (-)2</td>
<td></td>
</tr>
<tr>
<td>NP CRD: 3/3 = 100% NP CRD: 3/6 = 50%</td>
<td></td>
</tr>
<tr>
<td>Aggregate = 57     Aggregate = 55</td>
<td></td>
</tr>
<tr>
<td><strong>VII</strong> S = 4 XP = 1</td>
<td></td>
</tr>
<tr>
<td>VP CRD: 3/8 = 38% VP CRD: 3/4 = 75% Differential = (-)1.5</td>
<td></td>
</tr>
<tr>
<td>NP CRD: 3/3 = 100% NP CRD: 3/5 = 60%</td>
<td></td>
</tr>
<tr>
<td>Aggregate = 69     Aggregate = 67.5</td>
<td></td>
</tr>
<tr>
<td><strong>VIII</strong> S = 8 XP = 1</td>
<td></td>
</tr>
<tr>
<td>VP CRD: 3/12 = 25% VP CRD: 3/4 = 75% Differential = (+)5</td>
<td></td>
</tr>
<tr>
<td>NP CRD: 3/3 = 100% NP CRD: 3/5 = 60%</td>
<td></td>
</tr>
<tr>
<td>Aggregate = 62.5   Aggregate = 67.5</td>
<td></td>
</tr>
<tr>
<td><strong>IX</strong> S = 4 XP = 3</td>
<td></td>
</tr>
<tr>
<td>VP CRD: 3/10 = 30% VP CRD: 3/6 = 50% Differential = (-)18.5</td>
<td></td>
</tr>
<tr>
<td>NP CRD: 3/3 = 100% NP CRD: 3/7 = 43%</td>
<td></td>
</tr>
<tr>
<td>Aggregate = 65     Aggregate = 46.5</td>
<td></td>
</tr>
<tr>
<td><strong>X</strong> S = 16 XP = 3</td>
<td></td>
</tr>
<tr>
<td>VP CRD: 3/22 = 14% VP CRD: 3/6 = 50% Differential = (-)10.5</td>
<td></td>
</tr>
<tr>
<td>NP CRD: 3/3 = 100% NP CRD: 3/7 = 43%</td>
<td></td>
</tr>
<tr>
<td>Aggregate = 57     Aggregate = 46.5</td>
<td></td>
</tr>
</tbody>
</table>
tive degree of preference for Extraposition from NP should correlate directly with the size of $\bar{S}$.

Consider now the case of an NP that has been topicalized, or moved into the German Vorfeld or initial field, preceding a finite verb which will be argued in ch. 6.2.2 to construct a root $S$ (i.e. main clause) by Grandmother Node Construction. I will assume, following Emonds (1976), that the topic constituent is immediately dominated by $E$ (which stands for Expression) and is a sister to $S$. When an NP is topicalized, therefore, the structure is $E[\text{NP s[Vf...]}]$. What does EIC predict in the event that this NP is complex and potentially subject to Extraposition from NP?

Table 4.27 illustrates the major possibilities. First of all, if $S$ is very short and consists only of a single-word $V$ (section I), then the predictions are exactly those of I–III in table 4.26, since in both cases a $V$ alone intervenes between the ICs of NP. Typically, of course, an $S$ consists of much more than just a $V$, and in table 4.27 sections II–XI I add an additional constituent with variable word length. The $S$ domain remains constant across the two orders, according to the discontinuity hypothesis, and so does not need to be calculated. It does not matter, therefore, whether there is one additional XP or several; all that matters is the word length intervening between the discontinuous ICs of NP, and of course the ratio for the containing and opposed domain, which is now $E$. Sections II–IV illustrate this tension when there are two additional words in $S$ apart from $V$. The basic order is preferred when $\bar{S} = 4$ and (by a whisker) when $\bar{S} = 8$; the extraposed structure is preferred when $\bar{S} = 16$, but even here not by very much (+3). When there is a single additional word in $S$ (V–VI), the basic order is slightly preferred when $\bar{S} = 4$, and the extraposed order when $\bar{S} = 8$. Sections VII–XI illustrate the more typical cases in which the $S$ contains at least three other words: for typical lengths of $\bar{S}$ (the aggregate for extraposed and unextraposed relative clauses in the German field is 7.3 words in the German text of table 4.23), the basic order is favored, and even when $\bar{S} = 16$ (IX) there is no incentive to extrapose. When there are five other words in $S$ (X–XI), the basic order is consistently preferred, even when $\bar{S} = 16$.

Table 4.27 maintains the regularity that basic orders become progressively worse, the greater the size of $\bar{S}$, while the extraposed version is consistently better. The difference between the two structures – one a Mittelfeld extraposition, the other a Vorfeld extraposition – is that the $S$ domain typically contains many words that will necessarily intervene between the ICs of NP, whereas extraposition from a VP-dominated NP need only involve an intervening $V$ (or verbal dependant). EIC accordingly predicts that Extraposition
Table 4.27 IC-to-word ratios for German Extraposition from NP for an E-dominated NP

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>S = 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E CRD: 2/7 = 29%</td>
<td>E CRD: 2/3 = 67%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NP CRD: 3/3 = 100%</td>
<td>NP CRD: 3/4 = 75%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aggregate = 64.5</td>
<td>Aggregate = 71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>S = 4 XP = 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E CRD: 2/7 = 29%</td>
<td>E CRD: 2/3 = 67%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NP CRD: 3/3 = 100%</td>
<td>NP CRD: 3/6 = 50%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aggregate = 64.5</td>
<td>Aggregate = 58.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>S = 8 XP = 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E CRD: 2/11 = 18%</td>
<td>E CRD: 2/3 = 67%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NP CRD: 3/3 = 100%</td>
<td>NP CRD: 3/6 = 50%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aggregate = 59</td>
<td>Aggregate = 58.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>S = 16 XP = 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E CRD: 2/19 = 11%</td>
<td>E CRD: 2/3 = 67%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NP CRD: 3/3 = 100%</td>
<td>NP CRD: 3/6 = 50%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aggregate = 55.5</td>
<td>Aggregate = 58.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>S = 4 XP = 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E CRD: 2/7 = 29%</td>
<td>E CRD: 2/3 = 67%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NP CRD: 3/3 = 100%</td>
<td>NP CRD: 3/5 = 60%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aggregate = 64.5</td>
<td>Aggregate = 63.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VI</td>
<td>S = 8 XP = 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E CRD: 2/11 = 18%</td>
<td>E CRD: 2/3 = 67%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NP CRD: 3/3 = 100%</td>
<td>NP CRD: 3/5 = 60%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aggregate = 59</td>
<td>Aggregate = 63.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VII</td>
<td>S = 4 XP = 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E CRD: 2/7 = 29%</td>
<td>E CRD: 2/3 = 67%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NP CRD: 3/3 = 100%</td>
<td>NP CRD: 3/7 = 43%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aggregate = 64.5</td>
<td>Aggregate = 55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VIII</td>
<td>S = 8 XP = 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E CRD: 2/11 = 18%</td>
<td>E CRD: 2/3 = 67%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NP CRD: 3/3 = 100%</td>
<td>NP CRD: 3/7 = 43%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IX</td>
<td>S = 16 XP = 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E CRD: 2/19 = 11%</td>
<td>E CRD: 2/3 = 67%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NP CRD: 3/3 = 100%</td>
<td>NP CRD: 3/7 = 43%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aggregate = 55.5</td>
<td>Aggregate = 55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>S = 4 XP = 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E CRD: 2/7 = 29%</td>
<td>E CRD: 2/3 = 67%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NP CRD: 3/3 = 100%</td>
<td>NP CRD: 3/9 = 33%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aggregate = 64.5</td>
<td>Aggregate = 50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XI</td>
<td>S = 16 XP = 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E CRD: 2/19 = 11%</td>
<td>E CRD: 2/3 = 67%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NP CRD: 3/3 = 100%</td>
<td>NP CRD: 3/3 = 33%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aggregate = 55.5</td>
<td>Aggregate = 50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Differential = (+)6.5

Differential = (-)0.5

Differential = (+)3

Differential = (-)4.5

Differential = (-)9.5

Differential = (—)14.5

Differential = (—)14.5

Differential = (—)5.5
Rearrangements of basic word orders in performance

from NP will not, in general, be productive out of a complex NP in the Vorfeld, whereas it will be productive out of the Mittelfeld, and it should be highly productive in the event that V alone intervenes, much less so when there is an additional intervening constituent. There is one additional logical possibility. The complex NP out of which Š is extraposed could be S-dominated (either a nominative NP or an NP within an S-dominated as opposed to a VP-dominated PP). The logic of our predictions is identical here and does not need to be spelled out in detail: extraposition will correlate positively with the increasing size of Š, and negatively with the amount of S- or VP-dominated material that intervenes between the discontinuous ICs of NP.

We can now test these predictions. Table 4.28 sets out the data involving the length of Š and the relative frequency of extraposition in four text samples, three of them (all but the first one listed) collected and analyzed by Christine Mooshammer. The Š refers to relative clauses only, and the complex NP is E-, S- or VP-dominated. The EIC ratios for basic orders become progressively worse as Š increases in size: hence, the extrapositions become correspondingly better. As EIC's preferences increase, we predict more or equal numbers of extrapositions. This prediction is fulfilled: there is a perfect correlation between Š size and Extrapolation from NP.

### Table 4.28 Š Size in German Extrapolation from NP structures (relative clauses)


<table>
<thead>
<tr>
<th>Š Size</th>
<th>Unextraposed Š</th>
<th>Extrapolated Š</th>
</tr>
</thead>
<tbody>
<tr>
<td>3–5 words</td>
<td>16</td>
<td>71</td>
</tr>
<tr>
<td>6–8</td>
<td>10</td>
<td>71</td>
</tr>
<tr>
<td>9–11</td>
<td>4</td>
<td>57</td>
</tr>
<tr>
<td>12–14</td>
<td>2</td>
<td>32</td>
</tr>
<tr>
<td>15+</td>
<td>1</td>
<td>37</td>
</tr>
</tbody>
</table>

**Extrapolation ratio:**

<table>
<thead>
<tr>
<th>Š Size</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>3–5 words</td>
<td>82%</td>
</tr>
<tr>
<td>6–8</td>
<td>88%</td>
</tr>
<tr>
<td>9–11</td>
<td>93%</td>
</tr>
<tr>
<td>12–14</td>
<td>94%</td>
</tr>
<tr>
<td>15+</td>
<td>97%</td>
</tr>
</tbody>
</table>

**EIC prediction**

The extrapolation ratio out of a Complex NP will be greater or equal as the size of Š increases

Correct
With regard to the distance between the discontinuous ICs of NP, we make three related predictions: first, Extraposition from NP will not, in general, apply out of the initial field, but will apply productively out of the middle field; second, when the complex NP immediately precedes a single-word V or verbal dependant, extraposition will be optimal for all lengths of S, and so we can make an unmarked case prediction in favor of extraposition in these cases; third, since basic orders are regularly preferred when there is an XP intervening between the complex NP and V and always dispreferred when there is no XP, we expect a basic order to be productively retained only in the presence of at least one additional XP. From predictions two and three it follows that we should see evidence for complex NPs not undergoing extraposition when an XP intervenes, and the head of the NP should move into immediately preverbal position when extraposition does apply.

The first prediction is well supported. Extraposition of a relative clause out of the initial field is exceedingly rare in the data of table 4.28, and even the extensive computerized corpus of Shannon (1992a) unearthed only a handful of examples. Extrapositions out of the middle field, by contrast, are legion: the vast majority of examples in table 4.28 are of this type, as they are also in Shannon's (1992a) data. In table 4.29 I give some precise figures, taken from Shannon (1992b) and involving the extraposition of a complement clause (cf. section I). The initial-field extrapositions are given on the left, middle-field extrapositions on the right. The frequency of application for Extraposition from NP is 9% for the former, and 95% for the latter.

The second prediction is tested in section II. Of the relative clauses in this sample, 92% are extraposed over a single-word V or verbal dependant, and hence the unmarked case prediction is correct for this structural alternation.

The third prediction is tested in section III, using complement-clause data again provided by Shannon (1992b). Shannon counts the number of constituents that follow the head noun of a complex NP, both in the extraposed and in the basic order, and gives the average number of these constituents for each structural type (but not the word lengths). He distinguishes between extraposition from different case-marked NPs and from an NP within a PP. In reproducing his data here I have counted the verb or verbal dependant as one constituent, and have added to it his average for the number of additional XPs. Three pairs of basic and extraposition structures are shown in III, involving an accusative NP (X and Y), an NP within a PP (W and Z), and a nominative NP (U and V).

The prediction that the basic order will be productively retained only in the presence of V plus at least one additional XP is correct: the basic orders W
Rearrangements of basic word orders in performance

Table 4.29 Discontinuity distance in German Extraposition from NP structures

I  Extraposition from NP from the initial field versus the middle field

Data from Shannon 1992b; \( \bar{S} \) = complement clause

<table>
<thead>
<tr>
<th>Initial field</th>
<th>Middle field</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n = 189 )</td>
<td>( n = 244 )</td>
</tr>
</tbody>
</table>

| \( X_{E[NP[...N\bar{S}]S]} \) | 174 | \( W_{VP/S[NP[...N\bar{S}](XP)V]} \) | 13 |
| \( Y_{E[NP[...N]S_{NP}[\bar{S}]]} \) | 17 | \( Z_{VP/S[NP[...N](XP)V_{NP[\bar{S}]}]} \) | 231 |

Ratio of \( X/Y \): 91%  
Rearrangement to \( Y \): 9%

Ratio of \( W/Z \): 5%  
Rearrangement to \( Z \): 95%

EIC predictions

In general, the rearrangement to \( Y \) will not be productive out of the initial field.  
Correct

In general, the rearrangement to \( Z \) will be productive out of the middle field.  
Correct

II  Extraposition from NP with a one-word discontinuity

100 pages of data: first 100 pp. from Peter Handke *Die linkshändige Frau*, Suhrkamp.

\( [...NP[...N\bar{S}]V] \Rightarrow [...NP[...N]V_{NP[\bar{S}]}] \) (no other ICs follow N)

where \( V \) = a finite verb, a non-finite verb, or a verbal dependant such as a particle, not exceeding one word in length;  
\( \bar{S} \) = a relative clause

\( n = 51 \)

| \( X_{[...[...N\bar{S}]V]} \) | 4 | Mean length of \( \bar{S} \) in \( X \) = 4.75 words |
| \( Y_{[...[...N]V[S]]} \) | 47 | Mean length of \( \bar{S} \) in \( Y \) = 7.79 words |

Ratio of \( X/Y \): 8%  
Rearrangement to \( Y \): 92%

EIC predictions

Unmarked case: \( Y \) = most optimal for any length of \( \bar{S} \), i.e. \( 47/51 \) = most optimal, 92.2%  
Correct
Testing EIC's performance predictions

Table 4.29 (cont.)
III Numbers of intervening constituents in Extraposition from NP structures
Data from Shannon 1992b.

[ ... NP[ ... N $S$] (XP) $V$] $\Rightarrow$ [ ... NP[ ... N] (XP) $V_{\text{NP-}S[\}]$]

where $V$ = a finite verb, a non-finite verb, or a verbal dependant such as a particle;
$S$ = a complement clause

| $X$ [ ... NP-ACC[ ... N $S$] (XP) $V$] | 13 | 1.1 |
| $Y$ [ ... NP-ACC[ ... N] (XP) $V_{\text{NP-ACC}[S]}$] | 215 | 1.04 |

Ratio of $X/Y$: 6%
Rearrangement to $Y$: (94%)

$W$ [ ... PP[P NP[ ... N $S$]] (XP) $V$] | 31 | 2.03 |
$Z$ [ ... PP[P NP[ ... N]] (XP) $V_{\text{NP-ACC}[S]}$] | 105 | 1.18 |

Ratio of $W/Z$: 23%
Rearrangement to $Z$: (77%)

$U$ [ ... NP-NOM[ ... N $S$] (XP) $V$] | 34 | 2.29 |
$V$ [ ... NP-NOM[ ... N] (XP) $V_{\text{NP-NOM}[S]}$] | 101 | 1.18 |

Ratio of $U/V$: 25%
Rearrangement to $V$: (75%)

EIC predictions
In general, the basic order will be productively retained only in the presence of $V$ plus at least one additional XP (i.e. at least two constituents following NP):

- $X$ is not productive (6%); only 10% have an additional XP (1.1);
- $W$ is productive (23%); all have an additional XP on average (2.03);
- $U$ is productive (25%); all have at least one additional XP on average (2.29) Correct

The extraposed order will be most productive when there is no additional XP:

- $Y$: only 4% have an additional XP on average (1.04), i.e. 96% with no addit. XP;
- $Z$: only 18% have an additional XP on average (1.18), i.e. 82% with no addit. XP;
- $V$: only 18% have an additional XP on average (1.18), i.e. 82% with no addit. XP. Correct
and U are both productive, and they each have at least one additional XP; the basic order X (with accusative NPs) is not productive, and only 10% of these have an additional XP on average. Shannon's data also reveal that the extraposed structures (Y, Z, and V) are most productive without an additional XP. This is predicted by EIC, since extraposition is always most optimal under these circumstances (cf. II), whereas it will regularly not be optimal when there is an XP; hence we expect extraposition to be most frequent when it is most optimal, and it is. Derivately we also have evidence in these data for the predicted movement of the complex NP into immediately preverbal position prior to Extraposition from NP applying.

EIC thus makes some very fine-tuned, but ultimately very simple, predictions for Extraposition from NP, and these predictions are correct in our data. These data support the discontinuity hypotheses for derived constituent structures resulting from Extraposition from NP. If we assume no discontinuity of NP and attachment of the extraposed S to S, as in Reinhart's (1983) analysis, partially different predictions are made by EIC and these predictions fail to account for the data. For example, consider cases II and III in table 4.27, i.e. extrapositions from an NP in the initial field, where $\bar{S} = 4 \text{ XP} = 2$ and $\bar{S} = 8 \text{ XP} = 2$. Under the discontinuity analysis, it is correctly predicted that there will be no incentive to extrapose. But if $\bar{S}$ is attached to S, it is falsely predicted that extraposition should be productive. The structural alternation would be: $E_{[\text{NP[Det N $\bar{S}$] s}[V_f \text{ XP}]]} \Rightarrow E_{[\text{NP[Det N] s}[V_f \text{ XP $\bar{S}$]]}}$; and the ratios for S must now be calculated in addition to E and NP, since S no longer remains constant. The EIC scores are shown in (i) and (ii): in both cases the extraposed version has a higher aggregate, whereas in table 4.27 it has a lower aggregate: 17

\[
e_{[\text{NP[Det N $\bar{S}$] s}[V_f \text{ XP}]]} \Rightarrow e_{[\text{NP[Det N] s}[V_f \text{ XP $\bar{S}$]]}}
\]

(i) $\bar{S} = 4 \text{ XP} = 2$

E CRD: $2/7 = 29\%$  E CRD: $2/3 = 67\%$

S CRD: $2/2 = 100\%$  S CRD: $3/4 = 75\%$

NP CRD: $3/3 = 100\%$  NP CRD: $2/2 = 100\%$

Aggregate = 76  Aggregate = 81

Differential = (+)5
Testing EIC's performance predictions

(ii) \( \tilde{S} = 8 \) XP = 2

\[
\begin{align*}
E \ CRD: & \quad 2/9 = 22\% \\
S \ CRD: & \quad 2/2 = 100\% \\
NP \ CRD: & \quad 3/3 = 100\%
\end{align*}
\]

Aggregate = 74

Differential = (+)7

Hence this analysis falsely predicts that Extraposition from NP should be productive in these initial-field environments also. More generally, if \( \tilde{S} \) is attached to S by Extraposition from NP, and is also moved to the right periphery of (VP and) S by Extraposition, then the two sets of extraposition data should be fundamentally similar. But they are quite different. Extraposition from NP is only favored when there is a short separation between the head NP and the \( \tilde{S} \), and is disfavored out of an initial NP position in the clause. By contrast, Extraposition of a sentential subject is massively preferred (cf. section 4.3.4 and table 4.25). This difference is predicted by EIC, on the assumption that the output of Extraposition from NP is a discontinuous structure.

I have said nothing about any changes in information structure resulting from Extraposition from NP, and I believe that very little needs to be said. As in the case of Heavy NP Shift (cf. section 4.3.2), it has been proposed that Extraposition from NP is a focus construction, both in English (cf. e.g. Rochemont 1986, Rochemont and Culicover 1990) and in German (cf. Shannon 1992a, 1992b). But again the argument is not compelling. Extraposition from NP structures are more restricted in their focus options than the corresponding basic orders from which they are derived. Example (4.31) can be an appropriate response to both of the questions in (4.32):

(4.31) We sold the poster that was in the bookstore to Professor Smith.

(4.32) a. What did you sell to Professor Smith?
    b. To whom did you sell the poster that was in the bookstore?

whereas the extraposed (4.33) can only answer (4.32a):

(4.33) We sold the poster to Professor Smith that was in the bookstore.

But this does not prove that (4.33) is obligatorily and necessarily associated with a focus interpretation on the complex NP, on every occasion of use. It simply shows that you can't focus on the PP in (4.33), whereas you can in (4.31). This restriction, which mirrors the more restricted WH-extractions and question–answer options of Heavy NP Shift structures and which requires a grammatical explanation, will prevent (4.33) from being used in
Rearrangements of basic word orders in performance

209
certain contexts. What drives the occasions on which it is used, however, is ease of syntactic processing.

Shannon (1992a, 1992b) discusses some intriguing additional properties of the German Extraposition-from-NP structures in his corpus, properties which appear to support the focus theory. For example, the middle-field complex NPs that resist extraposition are very often definite rather than indefinite. They are also significantly shorter, both in their $S$ length (as table 4.29 section II shows), and also in the overall length of the whole complex NP; and shorter NPs are quite generally more definite and "mutually known" (cf. Clark and Marshall 1981) or "mutually manifest" (cf. Sperber and Wilson 1986) than longer ones (cf. section 4.4.4). Syntactic weight can explain the correlation with definiteness, therefore. Shannon also regards the stress pattern on the head of the extraposed complex NP as an indicator of focus. Both German and English have a greater than usual stress on *poster* in examples corresponding to (4.33), compared with their unextraposed counterparts. But stress can serve a number of different functions in these languages, and the function exemplified on this particular occasion reminds me of the stress pattern that one finds on the cataphoric determiner *those* before a head-noun-plus-restrictive-relative-clause in English: *those students who fail the exam will need to retake it*. The greater prominence of *those*, compared with the unstressed *the* in *the students who fail the exam*, serves to anticipate the restriction on the head noun which is still to come, by telling the hearer, in effect, to keep open the reference class defined by the NP until the relative clause has been processed, and not to close it at the head noun. German has a similar distinction: *diejenigen Studenten, die . . .* ("those students who") and *die Studenten, die . . .* ("the students who"), with a similar stress distribution. The stress on the head noun of an extraposed NP can also, it seems to me, be regarded as anticipating the relative clause (or complement clause) in the same way, by informing the hearer that more is to come. Whether this is a legitimate explanation or not, the stress pattern on these extrapositions provides no compelling evidence for a putative focus property which (as Shannon admits) is not independently definable. Once again, it is the vagueness of the concept that appears to be sustaining a unifying generalization.

Shannon's important and detailed empirical study has revealed the very patterns that EIC predicts: the limited applications of Extraposition from NP out of the initial field and its productivity in the middle field; the preferred positioning of the head immediately before the verb when extraposition applies; and so on. He argues convincingly against a grammatical restriction
that had been proposed on extrapositions from complex NP subjects in German (and Dutch): extrapositions from nominative-marked NPs are fully productive in German, as long as this NP stands in the middle field rather than the initial field (cf. table 4.29 section III). But all these facts follow simply from EIC, as do certain others which are not predicted by any informationally based theory: for example, the correlation between extraposition and the size of \( \tilde{S} \) (cf. table 4.28). One small detail which Shannon (1992b) mentions *en passant* is also interesting. Extraposition from a nominative NP is highly productive with intransitive verbs (43/44, or 98%, undergo it), much less so with transitive verbs (20/35, or 57%). Since a transitive verb has an accompanying direct object, presumably following the subject in most instances, there will be at least one intervening XP with transitives and possibly none with intransitives. Again this distribution is predicted by EIC. It does not follow from any pragmatic account.

Finally, in the context of a discussion of German it is appropriate to return to the word order generalizations proposed by Behaghel (1932), to which Shannon (1992b) appeals when giving his explanation of the data. Shannon notes that there is an inherent tension between Behaghel's *oberste Gesetz* (what belongs together mentally is placed closed together) and the *Gesetz der wachsenden Glieder* (cf. ch. 3.8.3). He points out that there is a competing motivation in the Extraposition from NP structure between keeping the ICs of NP continuous and moving a heavy constituent to the right. This is correct, but there are not two separate laws that are in opposition here. There is just a single principle, EIC, and two syntactic phrases, NP and a containing phrase, within which there are opposing EIC benefits, depending on the positioning of \( \tilde{S} \). Behaghel's two laws reduce to one, therefore. Moreover, his law about pragmatic information structure (old precedes new, cf. Behaghel 1932:4) will also be shown to be a correlate of short-before-long positioning, and hence three of his laws can be reduced to EIC.

I predict that data corresponding to tables 4.27 and 4.28 will be found in all languages in which discontinuities are permitted between daughter ICs: longer ICs will be moved out more productively than shorter ones; and the frequency of discontinuities will be in direct proportion to the distance of the separation. Kiss and Primus (1991) cite data from Hungarian that supports this: Extraposition from NP is less acceptable out of topic position, which is the leftmost position in the sentence (cf. example (4.3)), than out of positions occurring further to the right. Performance data need to be collected from Hungarian and other languages to test the predicted parallelism with German.
4.3.6 Dative–Accusative Shift in Korean

In our discussion of Japanese and Korean in sections 4.1.5 and 4.1.6 we saw clear evidence in the data from both languages for a VP node. Since the verb is clause-final, the VP could also be shown to be grammaticalized in right-peripheral position. The data also provided evidence for VP discontinuity when a direct object is preposed before an S-dominated subject or PP.

Table 4.30 Korean Dative–Accusative Shift

210 pages of data: cf. table 4.11.

$\text{vP[NP-Dat NP-Acc V]} \Rightarrow \text{NP[NP-Acc NP-Dat V]}$ (no other ICs in the VP apart from optional single-word adverbials)

| Basic Order [NP-Dat NP-Acc V] | 19 |
| Rearranged [NP-Acc NP-Dat V] | 5  |

<table>
<thead>
<tr>
<th></th>
<th>NP-Acc =</th>
<th>NP-Acc &gt; NP-Dat:</th>
<th>:3–4</th>
<th>:5+</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NP-Dat</td>
<td>1–2 words</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X [NP-Dat NP-Acc]</td>
<td>7</td>
<td>4</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Y [NP-Acc NP-Dat]</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Ratio of X/Y:

|                | 100%    | 80%   | 50%   | 0%   |
| Rearrangement to Y: | (0%)   | (20%) | (50%) | (100%) |

\[\text{NP-Dat > NP-Acc}\]

| W [NP-Dat NP-Acc] | 6 |
| Z [NP-Acc NP-Dat] | 0 |

Ratio of Z/W: 0%

EJC predictions

Unmarked case: \[18/24 = \text{most optimal (Y + W + =)}\], i.e. 75%

Marked case: Ratio of X/Y for NP-Acc > NP-Dat:

\[1–2 \geq 3–4 \geq 5+\] All correct
Grammaticalization seems to have gone one step further in Korean than in Japanese (if our data are typical, at least), since it appears to have fixed the positioning of dative and accusative NPs, both of which I assume to be VP-dominated. The data for Japanese show a 50–50 split in the distribution of [NP-Dat NP-Acc V] and [NP-Acc NP-Dat V], with all logically possible relative length differentials occurring in each order (i.e. NP-Dat > NP-Acc, NP-Acc > NP-Dat and NP-Dat = NP-Acc). Korean, by contrast, exhibits a clear skewing in favor of [NP-Dat NP-Acc V] by approximately four to one, an exclusive preference for this order when NP-Dat = NP-Acc, and a productive retention of it even when NP-Acc > NP-Dat by up to four words. EIC would predict preposing of the heavier NP-Acc in these circumstances if order were completely free. These data were collected and analyzed by Charles Kim and are set out in table 4.30. An example sentence was given in (4.11) above.

The data are not numerous, but the pattern is clearly indicative of grammaticalization, and because of the productive retention of the basic order despite weak contrary EIC preferences, the unmarked case prediction scores only 75%, which is lower than the success rate that we have come to expect. The marked case predictions are all correct, however.

4.3.7 Dative-Accusative Shift in English

It is interesting to compare the Korean data of the last section with the much-discussed Dative-Accusative Shift rule of English. In standard transformational treatments, this rule is generally considered to convert the (a) structure into the (b) structure in sentences such as (4.33) and (4.34):

(4.33) a. Mary gave a book to John.
    b. Mary gave John a book.

(4.34) a. Mary baked a cake for John.
    b. Mary baked John a cake.

In other words, it converts $\text{VP}[\text{V NP}_{\text{PP}}[\text{to/for NP}]]$ into $\text{VP}[\text{V NP}_{1} \text{NP}_{j}]$, where NP$_1$ corresponds to the PP in the basic order, and to an originally dative-marked NP in earlier stages of English (and in other languages such as German), while NP$_j$ was originally (and is elsewhere) accusative-marked. The rule has always been notoriously difficult to state, because its applicability is dependent on particular lexical items, for example *I reported the robbery to the police cannot be converted into *I reported the police the
robbery, and this gives rise to some very interesting problems of learnability (cf. Baker 1979, Pinker 1987, and Bowerman 1988 for discussion). In the present context, this restricted productivity means that, while it is always straightforward to recognize a structure that has undergone Dative–Accusative Shift, it is not so straightforward to identify all the instances of [V NP PP] that could potentially undergo it, within a given text. Large numbers of these structures will need to be converted to [V NP_i NP_j] and subjected to a grammaticality test, and while the results are in many cases clear, in some cases they are not, and I found it uncomfortable to have to rely on my own private judgments in this way. For this reason I went through the data of my English sample (cf. table 4.2) looking for [V NP_i NP_j] structures only, i.e. the outputs of Dative–Accusative Shift.

This rule is different from all the rearrangements considered hitherto in that it involves not just a re-ordering of two elements, but a category change as well: one of the NPs is contained within a PP in the basic order, and this PP is completely removed in the course of the re-ordering. Alternatively, of course, within a model in which there are no transformations of this kind, particular lexical items will be associated with two syntactic structures for the same arguments. Either way our results are intriguing, because the weight distribution between NP_i and NP_j is almost perfect from the point of view of EIC.

[VP NP_i NP_j] in English is an instance of structure type (2a) in table 3.2. Hence, we predict NP_j ≥ NP_i. There are 31 such structures in my data. Of these, 30 are as predicted (97%), with 25 involving NP_j > NP_i. There is a significant difference in the mean sizes of NP_i and NP_j: the former has a mean of 1.3 words, the latter one of 3.5. The mean of NP_i is small because a large proportion of these structures have a pronominal NP_i, i.e. Mary gave him a book, etc. A full 27 of the 31 structures are of this type, while 28 have a non-pronominal NP_j.

From these data we can conclude that if [V NP_i NP_j] structures are derived from a basic [V NP PP] structure, then in general the only instances of the basic order that will undergo the rearrangement will be those where it results in an improved shifted structure, despite the category change. In fact, the great majority of conversions are of the type vp[gave NP[a book] PP[to him]] ⇒ vp[gave NP[him] NP[a book]], and these will always result in optimal EIC ratios: the former requires four words for the recognition of three ICs, i.e. 3/4 = 75%; the latter only three, 3/3 = 100%. Since the aggregate size of the NP_j a book is actually 3.5 words, the improvement will regularly be quite considerable: from 3/5.5 on average, i.e. 55%, to 3/3 = 100%. Alternatively,
if there is no transformation as such, then the selection among competing syntactic frames for a given predicate is still being made in accordance with EIC: [V NP PP] will be most optimal when PP > NP; but when the PP is short, and especially when it dominates just P and a pronoun, [V NP_i NP_j] will have a 100% ratio for an NP_j of any length, since NP_i is just a single-word pronoun. Either way, EIC makes correct performance predictions.

Far from putting alternations such as (4.33) and (4.34) outside the scope of EIC's predictions, therefore, the category change from PP to NP can actually be explained by EIC: the removal of the preposition shortens the recognition domain for VP. In fact, these alternations need to be seen alongside a third truth-conditionally equivalent option, namely the Heavy NP Shift structure \( \text{VP}[V \text{ PP NP}] \):

(4.33) c. Mary gave to John a book that was very expensive.

(4.34) c. Mary baked for John a cake that was absolutely delicious.

These also are used when the PP is shorter than the NP (cf. table 4.22). By deleting the preposition, the VP CRD becomes even more optimal for EIC, and when the NP originally dominated by PP is just a single word, three ICs of the VP will always be recognized by a maximally optimal three words. This is what explains the predominance of pronouns in Dative–Accusative Shift: we expect structures to be most frequent in performance that are most optimal for EIC.

This account does not, as far as I can tell, offer any explanation for why the deletion of the preposition should be lexically idiosyncratic. But it does explain the existence of three relative orderings for truth-conditionally equivalent sentences, [V NP PP], [V NP_i NP_j], and [V PP NP], in each of which the aggregate length of the third IC is greater than that of the second, which is in turn greater than that of the first, i.e. the verb. The ordering [V NP PP] is always productive in performance, except (it seems) in those cases where one or the other structure is most optimal. I shall return to a consideration of Dative–Accusative Shift in ch. 5.4.2.

4.4 Comparing EIC with pragmatic predictions

In ch. 3.8.2 I pointed out that the prevailing view in the literature sees word-order alternations in languages with considerable freedom as being conditioned by pragmatic principles, and I summarized two major research traditions in this area: that of Givon (1983, 1988) and his associates; and that of
Comparing EIC with pragmatic predictions

the Prague School (cf. e.g. Firbas 1964, 1966). I expressed scepticism about the reality of these principles: the two research traditions are partially contradictory; and neither has paid any systematic attention to the interacting considerations of syntax and syntactic weight. Since there are correlations between weight and information status, it is quite conceivable that one could be in the driving seat, with the other being merely epiphenomenal. I also argued that what makes particular orders more functional or efficient is not (or not primarily) their capacity to facilitate pragmatic processing, but rather the fact that they make the processing of syntactic structure easier for the hearer.

In sections 4.3.2 and 4.3.5 I considered some proposed pragmatic properties of two rearrangement rules, Heavy NP Shift and Extraposition from NP, and I argued that the data provided no evidence for their reality. EIC's predictions, by contrast, were well supported. This suggests that we might expect similar results for free word orders in relation to the predictions of Givon and the Prague School. It is important, therefore, to juxtapose all of these predictions, so that we can see how each theory fares. EIC has been shown to make good predictions for a number of structural types and languages. What levels of success do the pragmatic theories achieve for these structures and in these languages? Which of the competing pragmatic traditions is better supported? And does EIC succeed in making them both redundant and epiphenomenal?

I have comparative data for four of the ten languages in my performance sample, and I shall now proceed to test the pragmatic theories on the same textual material that was used for EIC – or at least on significant subsets of it, for this kind of pragmatic analysis is more time-consuming than the purely syntactic analysis required for EIC. I begin with a discussion of Givon's theory and research methodology (section 4.4.1), and then test his theory, juxtaposing the pragmatic status of ICs with their respective syntactic weights, and hence with the predictions of EIC (section 4.4.2). The following two sections consider the predictions of givenness in relation to EIC, both new before given and given before new, and in the process juxtapose EIC's predictions with those of the Prague School.

4.4.1 Givon's theory of Task Urgency

Givon's (1983, 1988) theory is captured in the phrase "attend first to the most urgent task." He claims that words and phrases are ordered in such a way that either unpredictable precedes predictable information in discourse, i.e.
new before given, or important precedes unimportant information. In order to make these terms operational and testable, he provides the following methodology with which to quantify the pragmatic data of a text: degree of (un)predictability is measured by counting how far back within the last 20 clauses an entity was previously mentioned; degree of (un)importance is measured by counting how many times an entity is mentioned in the subsequent 10 clauses.\textsuperscript{18}

Givon's principle is, in essence, a hypothesis about information processing in the on-line reading of a text. In contrast to the Prague School, he considers communication to be better served when the speaker (or writer) relays new information to the hearer (or reader) as rapidly as possible, and more generally when information that is communicatively more urgent precedes that which is less so. I agree with him that this is a much more plausible functional motivation for linear ordering than that of the Prague School. Everything we know about human communication suggests that it is a highly efficient process. Sperber and Wilson (1986) have pointed to the fundamental efficiency of pragmatic inferencing in their principle of Relevance, where efficiency is measured in terms of the number of pragmatic inferences (or contextual effects) that can be derived from an utterance, divided by the processing cost of achieving them. And these authors claim that "every act of ostensive communication communicates the presumption of its own optimal relevance" (p. 158, my italics). Grammatical processing is also highly efficient in the way it enables the hearer to extract as much grammatical information as possible from the smallest amount of physical material, as I have been arguing throughout this book. Zipf's (1949) principle of least effort is ubiquitous throughout language and language change. It makes sense, therefore, that Givon should want to propose a hearer-oriented production principle of linear ordering, whereby the speaker strives to provide as much new and important information as early as possible in the left-to-right parse string.

Where I disagree with him is in the more fundamental question of the appropriate level of mental representation upon which such an efficiency metric is to be defined. Givon's metric applies, in effect, to pragmatically enriched semantic representations; EIC applies to structural representations of a syntactic nature. My point is that before you can get to any pragmatic or discourse interpretation in processing, there are numerous levels of form representation that need to be derived from the parse string, including syntactic representations, and it is these more fundamental aspects of grammatical processing that linear ordering preferences serve to make as efficient as
possible (cf. ch. 3.8.2). The empirical test of Task Urgency and EIC that I am about to undertake therefore gets to the heart of a profound issue in language processing: on what level of mental representation is the efficiency metric of linear ordering defined? Simply put, what is linear ordering good for: grammar processing or discourse?

Before we test Task Urgency, we must define exactly what this theory predicts, because matters are more complex than they appear to be at first sight. For example, given two ICs, IC_i and IC_j, that can occur in either order, and that happen to be ordered [IC_i IC_j] in a given position in a text, Givon's claim is that IC_i should be \textit{either} more unpredictable or more important than IC_j. Presumably we have to add here: or both, since there is no reason why the disjunction should be exclusive rather than inclusive. Also we need to qualify the whole claim with some rider such as: \textit{if} there is any difference between these two ICs along these dimensions, then \ldots . We do not want to claim that IC_i must be more unpredictable or more important than IC_j, because we are going to see lots of cases in which IC_i and IC_j are equally (un)predictable and equally (un)important. If these were to count as exceptions, then Givon's predictions would be in serious trouble, yet there is no reason why they should be exceptional. On the contrary, when two ICs are equal in task urgency, both orderings are predicted to be possible and productive, just as they are when syntactic weights are equal in the EIC theory. Such cases are as optimal as they can be, from the perspective of either theory, and can legitimately be added to a calculation of its overall success rate throughout a text. Yet they do not provide a critical test, obviously. For that one needs to examine the structures in which only one order is predicted; and if the number of "either-or" cases is large there is a risk of vacuity, since the amount of data on which one's theory makes a discrete prediction for one ordering and against another will be correspondingly reduced. We need to distinguish, therefore, between data that are compatible with a given theory, and those that constitute a critical test of it, and in what follows I shall concentrate on the "critical predictions" of Task Urgency and EIC.

We can formulate Givon's claim as in (4.35):

(4.35) \textit{Givon's Task Urgency theory}

Given two ICs ordered [IC_i IC_j]; if IC_i and IC_j differ in their P(redictability) and I(mportance) values, then:

- either $P_i > P_j$
- or $I_i > I_j$
- or both
In other words, either the $P$ score for $IC_i$ will be higher than that for $IC_j$, or the $I$ score will be higher, or both will be higher. The $P$ score will be higher if the entity referred to by $IC_i$ is mentioned, say, 18 clauses back, whereas the $IC_j$ entity is mentioned two clauses back, so high $P$ scores actually measure unpredictability and low $P$ scores relative predictability. If some entity is not mentioned at all in the last 20 clauses, it is assigned a zero in my coding, which means 21 to infinity, i.e. high unpredictability. For the importance factor, a high score means more mentions than a lower score within the subsequent 10 clauses. Hence, Givon predicts higher $P$ or $I$ scores for $IC_i$ than $IC_j$, if there is any difference between them.

The logic of a disjunction is compatible with several different options, and we still need to make precise what Givon would predict for a given pair of ICs. These predictions are set out in (4.36). There are three sets of possibilities, each of which comprises three major options. There are the cases where just a single order is predicted, i.e. $AB$. There are the cases where both orders are possible, either because all values for $P$ and $I$ are equal (including 0), or because one principle motivates one order, while the other motivates the other. Finally, there are counterexamples to the single-order predictions, i.e. cases where the opposite order occurs from that which Givon predicts, *BA.

\[(4.36)\] \textit{Givon's predictions} 

Given two ICs ordered $[IC_i \text{ } IC_j]$, then:

\begin{align*}
\text{Single orders predicted:} & \quad \text{either } P_i > P_j \text{ (where } I_i = I_j) \\
& \quad \text{or } I_i > I_j \text{ (where } P_i = P_j) \\
& \quad \text{or } P_i > P_j \text{ and } I_i > I_j \\
\text{Both orders predicted:} & \quad \text{either } P_i = P_j \text{ and } I_i = I_j \text{ (all values can be 0)} \\
& \quad \text{or } P_i > P_j \text{ and } I_i < I_j \\
& \quad \text{or } I_i > I_j \text{ and } P_i < P_j \\
\text{Counterexamples:} & \quad \text{either } P_i < P_j \text{ (where } I_i = I_j) \\
& \quad \text{or } I_i < I_j \text{ (where } P_i = P_j) \\
& \quad \text{or } P_i < P_j \text{ and } I_i < I_j \\
\end{align*}

The critical predictions here will be the single-order predictions, and the success or otherwise of Givon's theory will be a function of the proportion of correct ($AB$) to incorrect (*BA) predictions. For EIC the critical predictions will be those in which there is some quantifiable difference in IC-to-word
Comparing EIC with pragmatic predictions

ratios between two orders, generally when one IC exceeds another by at least one word in length.

4.4.2 Task Urgency and EIC

We can now test Givon's predictions against some of the data presented earlier in this chapter. Consider first the relative ordering of two PPs in English (cf. table 4.2). In the corresponding pragmatic analysis of these data I have tested for previous and future mentions of the immediately dominated NP within PP, whether or not the preposition itself was previously or subsequently mentioned. This is because prepositional phrases in English often correspond to NPs with particles or case affixes in other languages, and Givon's methodology is defined in terms of the predictability and importance of entities, rather than of relations and predications. When counting numbers of clauses backwards and forwards, I counted 20 finite verbs back, and 10 forwards, in order to avoid investigator analysis and subjectivity with respect to non-finite verbs and their precise clausal status in English.

The results are set out in table 4.31. I first present EIC's predictions for the relevant subset of the data in table 4.2 on which the pragmatic predictions were tested (section I). Both the unmarked and marked case predictions are well supported. In section II I consider EIC in relation to the predictions of Task Urgency. In the leftmost column, the data are broken down according to their EIC status: there are 10 instances with non-optimal orderings, i.e. with the longer PP before PP (Y); 21 in which PP = PP; 16 optimal orders in which PP > PP by one word; and so on. In the second column, I calculate EIC's success rate for the critical predictions, those in which PP! / PP. There are 38 of these orderings that are optimal for EIC and 10 that are non-optimal. The ratio of correct to incorrect critical predictions is accordingly 38/48, which is a respectable 79%. The third column sets out the number of orders for which Task Urgency (TU) makes a critical, single-order prediction, for each of the subsets of data that have been distinguished according to their EIC status. Thus, of the 10 orders that are non-optimal for EIC, TU makes a critical prediction for two of them, and an either-or prediction for the remaining eight, and so on down the column. The percentage at the foot of this column (36%) represents the proportion of the total data (n = 69) for which TU makes a critical prediction, as opposed to an either-or prediction. The final column gives the ratio of correct to incorrect for the critical predictions of TU, still divided according to their EIC status.
Table 4.31 Task Urgency and EIC in English [NP V {PP₁ PP₂}]

90 pages of data: first 90 pp. from D.H. Lawrence The Fox, Bantam Books; cf. Table 4.2 for structural template, additional EIC data and EIC calculation assumptions.

I Distribution of the data by syntactic weight

<table>
<thead>
<tr>
<th>n = 69</th>
<th>PP₁ = PP₂</th>
<th>PP₂ &gt; PP₁</th>
<th>:1 word</th>
<th>:2–3</th>
<th>:4+</th>
</tr>
</thead>
<tbody>
<tr>
<td>X [PP₁ PP₂]</td>
<td>21</td>
<td>16</td>
<td>11</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Y [PP₂ PP₁]</td>
<td>8</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratio of Y/X:</td>
<td>33%</td>
<td>15%</td>
<td>0%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

EIC predictions

Unmarked case: 59/69 = most optimal (X+ =), i.e. 85.5% Correct
Marked case: ratio of Y/X for PP₂ > PP₁ :1 ≥ 2–3 ≥ 4+ All correct

II EIC predictions in relation to Task Urgency

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-optimal: 10</td>
<td>21%</td>
<td>2/10</td>
<td>2/2 = 100%</td>
</tr>
<tr>
<td>Equals: 21</td>
<td></td>
<td>12/21</td>
<td>8/12 = 67%</td>
</tr>
<tr>
<td>Optimal (1): 16</td>
<td>79%</td>
<td>4/16</td>
<td>1/4 = 25%</td>
</tr>
<tr>
<td>(2–8): 20</td>
<td></td>
<td>5/20</td>
<td>1/5 = 20%</td>
</tr>
<tr>
<td>(9 +): 2</td>
<td></td>
<td>2/2</td>
<td>0/2 = 0%</td>
</tr>
<tr>
<td>Totals:</td>
<td>69</td>
<td>36%</td>
<td>48%</td>
</tr>
</tbody>
</table>

III Task Urgency predictions in relation to EIC

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorrect: 13</td>
<td>52%</td>
<td>9/13</td>
<td>9/9 = 100%</td>
</tr>
<tr>
<td>Either–or: 44</td>
<td></td>
<td>35/44</td>
<td>24/35 = 77%</td>
</tr>
<tr>
<td>Correct: 12</td>
<td>48%</td>
<td>4/12</td>
<td>2/4 = 50%</td>
</tr>
<tr>
<td>Totals:</td>
<td>69</td>
<td>70%</td>
<td>79%</td>
</tr>
</tbody>
</table>
Comparing EIC with pragmatic predictions

Thus, of the two critical predictions made by TU for the 10 non-optimal orderings by EIC, both are correct. Of 12 critical predictions made for the 21 instances in which $PP_1 = PP_2$ in syntactic weight, eight are correct, and so on. The ratio of correct to incorrect in each of these categories is converted to a percentage, and the percentage at the foot of the column (48%) represents the ratio of correct to incorrect for all critical predictions made by TU.

This tabulation reveals three significant patterns. First, TU only makes 25/69 critical predictions (= 36%). In other words, only approximately one third of these orderings are covered by Givon's single-order predictions, and his theory therefore claims that as many as two thirds could occur in either order. EIC, by contrast, makes critical predictions for 48/69 orders (= 70%), and so claims that less than one third should occur in either order, and that over two thirds should preferably be positioned according to this principle. Second, of the 25 critical predictions made by TU, its success rate appears to be random: 12 are correct (48%); 13 incorrect (52%). EIC's success rate for critical predictions, on the other hand, is significantly higher, 38/48 or 79%. Third, and rather intriguingly, the data reveal an inverse correlation between EIC and TU. As we go down the entries in these columns, the strength of EIC's preferences increases, from non-optimal EIC ratios, through equal ratios, through optimal ratios where the word differential is small (one word), and then progressively larger, and as it does so, the ratio of correct to incorrect predictions made by TU declines, in a highly precise manner.

This inverse correlation invites a hypothesis that seems, at first, to be very attractive. Perhaps syntactic weight and pragmatics are complementary: where EIC makes predictions, TU does not, and vice versa. If we can set them both to work on the data, within an interactive model of word-order freedom, it would appear that we would be able to make more correct predictions than with just one principle alone. I shall show below that this apparent complementarity between EIC and pragmatics becomes untenable when other languages and structures are considered, and I shall explain the inverse correlation another way.

Overall, TU is not well supported by the data of table 4.31: it makes too few critical predictions, and the success rate of these predictions is too low. EIC's predictions, by contrast, are well supported. The limited successes of TU cluster round the data for which EIC makes incorrect or weak predictions, and this remains to be explained.

Section III of table 4.31 presents the data the other way round, from the perspective of TU. The first column classifies orders according to their TU
status, giving the ratio of correct to incorrect critical predictions in the second. The number of critical EIC predictions made within each of these TU categories is presented in the third column, and the fourth column gives the EIC ratio of correct to incorrect critical predictions. Again we see the same inverse correlation as in II, though the EIC percentages are higher because the overall EIC success rate is higher.

A very similar pattern of results holds for the relative ordering of two postverbal NPs (of the type \(mNP\)) in Hungarian. The EIC predictions were tested in table 4.4. TU's predictions for the same set of data are set out in table 4.32. The data were collected by Stephen Matthews.

Table 4.32 Task Urgency and EIC in Hungarian \([V \{mNP_1 mNP_2\}]\)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-optimal:</td>
<td>10 11%</td>
<td>5/10</td>
<td>5/5 = 100%</td>
</tr>
<tr>
<td>Equals:</td>
<td>21</td>
<td>12/21</td>
<td>5/12 = 40%</td>
</tr>
<tr>
<td>Optimal (1):</td>
<td>50 89%</td>
<td>25/50</td>
<td>3/25 = 12%</td>
</tr>
<tr>
<td>(2 +):</td>
<td>35</td>
<td>26/35</td>
<td>2/26 = 8%</td>
</tr>
<tr>
<td>Totals:</td>
<td>116 89%</td>
<td>59%</td>
<td>22%</td>
</tr>
</tbody>
</table>

II  Task Urgency predictions in relation to EIC

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorrect:</td>
<td>53 78%</td>
<td>46/53</td>
<td>46/46 = 100%</td>
</tr>
<tr>
<td>Either-or:</td>
<td>48</td>
<td>39/48</td>
<td>34/39 = 87%</td>
</tr>
<tr>
<td>Correct:</td>
<td>15 22%</td>
<td>10/15</td>
<td>5/10 = 50%</td>
</tr>
<tr>
<td>Totals:</td>
<td>116 22%</td>
<td>82%</td>
<td>89%</td>
</tr>
</tbody>
</table>
Comparing EIC with pragmatic predictions

Section I sets out the data from the perspective of EIC. EIC makes critical predictions for 95/116 orders (82%), and of these a full 85/95 are correct (89%). TU makes critical predictions for 58/116 orders (59%), which is higher than the corresponding 36% figure for the English data, but significantly lower than the 82% of critical predictions made by EIC in the present case. TU's success rate of 22% is also significantly lower than its 48% score for English and than the 89% success rate of EIC. EIC's ratio of correct to incorrect critical predictions in Hungarian is therefore four times higher than that of TU, and the absolute numbers of correct predictions are almost six times higher (85 to 15). Once again, there is the same inverse

Table 4.33 Task Urgency and EIC in German NP–PP rearrangements

100 pages of data: cf. table 4.23.
For EIC's predictions, cf. table 4.23.

I EIC predictions in relation to Task Urgency

<table>
<thead>
<tr>
<th>EIC status</th>
<th>EIC ratio of correct:</th>
<th>No. of</th>
<th>TU ratio of correct:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>incorrect critical preds.</td>
<td>critical TU preds.</td>
<td>incorrect critical preds.</td>
</tr>
<tr>
<td>Non-optimal: 14</td>
<td>21%</td>
<td>7/14</td>
<td>5/7 = 71%</td>
</tr>
<tr>
<td>Equals: 21</td>
<td></td>
<td>7/21</td>
<td>4/7 = 57%</td>
</tr>
<tr>
<td>Optimal (1): 33</td>
<td>79%</td>
<td>16/33</td>
<td>8/16 = 50%</td>
</tr>
<tr>
<td>(2–6): 15</td>
<td></td>
<td>5/15</td>
<td>2/5 = 40%</td>
</tr>
<tr>
<td>(7 +): 6</td>
<td></td>
<td>2/6</td>
<td>0/2 = 0%</td>
</tr>
<tr>
<td>Totals: 89</td>
<td>79%</td>
<td>42%</td>
<td>51%</td>
</tr>
</tbody>
</table>

II Task Urgency predictions in relation to EIC

<table>
<thead>
<tr>
<th>TU status</th>
<th>TU ratio of correct:</th>
<th>No. of</th>
<th>EIC ratio of correct:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>incorrect critical preds.</td>
<td>critical EIC preds.</td>
<td>incorrect critical preds.</td>
</tr>
<tr>
<td>Incorrect: 18</td>
<td>49%</td>
<td>15/18</td>
<td>13/15 = 87%</td>
</tr>
<tr>
<td>Either–or: 52</td>
<td></td>
<td>38/52</td>
<td>31/38 = 82%</td>
</tr>
<tr>
<td>Correct: 19</td>
<td>51%</td>
<td>15/19</td>
<td>10/15 = 67%</td>
</tr>
<tr>
<td>Totals: 89</td>
<td>51%</td>
<td>76%</td>
<td>79%</td>
</tr>
</tbody>
</table>
correlation between EIC and TU, with the latter doing best when EIC does worst. Section II presents the data from the perspective of TU, with similar results.

These English and Hungarian data cast serious doubt on the validity of TU as a theory of discourse organization: the amount of data for which significant predictions are made is insufficient, and the success rate is too small. Clearly, whatever reality there is to pragmatic organizing principles, this is not the correct formulation. We must therefore try to expand both the set of pragmatic predictions made and the success rate, and we must also try to explain why it is that the few pragmatic predictions that are correct cluster round the incorrect and weak EIC predictions, and more generally why EIC and pragmatics appear to be inversely correlated.

Before we do so, consider briefly some data from a third language, German, involving the relative ordering of an accusative-marked (non-pro- nominal) NP and a PP within the middle field (cf. table 4.23). These data involve a grammaticalized basic order (of [NP PP]) and possible rearrange- ments from this basic order, rather than free ordering. Nonetheless, EIC makes correct predictions for both types of word-order alternation, so it is interesting to see how TU fares in this kind of situation, and whether gram- maticalization is reconcilable with this pragmatic principle. The data are given in table 4.33.

The pattern of results from German is fundamentally similar to the English and Hungarian data, despite the grammaticalization: EIC makes 76% critical predictions, TU only 42%; EIC's overall success rate is higher, 79% to 51%; and there is the same inverse correlation between EIC and TU, with TU's limited successes clustering round the non-optimal end of EIC's predictions. This pattern underscores the need for a revised pragmatic theory.

4.4.3 Givenness and EIC

One could try to explain away the poor showing of TU in the last section by arguing that Givon's quantified text methodology captures only a proper subset of all the pragmatic determinants of order that are compatible with task urgency, in the interests of quantifiability (as he himself readily admits). Hence, it could be argued, this method fails to capture a number of associative and inferential relationships between entities that are relevant for the predictability of an item in a text, so this is not a fair test of the role of pragmatics in linear ordering.
The premise of this argument is correct: Givon's methodology captures only those previous and subsequent mentions of an entity that are referred to explicitly. The conclusion that this invalidates the test does not necessarily follow, however. For if Givon's methodology can test only some of the predictions of his theory, we can still measure the success or otherwise of these particular predictions within different subsets of the linear orderings (these subsets being defined according to syntactic weight, for example), and it is for this reason that I am primarily interested in the ratios of correct to incorrect predictions across subsets. I assume in the process that the role of any additional pragmatic determinants of linear ordering will be roughly constant throughout the data, and that any given subset of pragmatic predictions can be selected for comparative testing on different data subsets. The overall success rate of a pragmatic theory could presumably increase (or decrease), when all possible pragmatic determinants of linear ordering are taken into account, and when entities are classified as given on the basis of associations and inferences that have been activated earlier in a text. But we can still test the relative successes of a given subset of the pragmatic predictions in the way we have done here. Givon also assumes that one can test some of his predictions in isolation from others, because that is what he does himself.

In order to try and improve the pragmatic predictions, let us consider a variant of Givon's theory, one which equates task urgency with degree of (un)predictability only, on the basis of previous mention or givenness, and which ignores (un)importance, i.e. future mention. There is, in fact, theoretical support for eliminating (un)importance as a relevant factor: both the speaker and the hearer can readily assess the (un)predictability of an entity at a given point in the discourse, but only the speaker generally knows whether an entity is going to be important in subsequent clauses that he or she is about to utter. Importance seems to involve production alone, therefore, whereas (un)predictability involves both production and comprehension. By postposing a given entity and preposing a new one, the "packaging" of communication is made more efficient for the hearer. But preposing an important entity does not add to the efficiency of the utterance in any obvious way, either for the hearer, or for the speaker for that matter. So let us relativize TU to P(predictability) marking only, as follows:

\[(4.37) \text{ New before Given } \]

Given two ICs ordered \([IC_i, IC_j]\); if \(IC_i\) and \(IC_j\) differ in their P(predictability) values, then:

\[P_i > P_j\]
Testing EIC's performance predictions

This revision predicts that less predictable items will precede more predictable ones, or equivalently that new will precede given, where these notions can be defined in terms of the distance between entities within a text, using Givon's methodology again. Example (4.37) accordingly makes the (now much simpler) predictions of (4.38):

\[(4.38) \text{ New-before-Given predictions} \]

Given two ICs ordered \[IC_i, IC_j\], then:
- Single orders predicted: \[P_i > P_j\]
- Both orders predicted: \[P_i = P_j\]
- Counterexamples: \[P_i < P_j\]

These revised predictions are tested against the English data of table 4.31 in table 4.34. New before given (NG) makes almost twice as many critical predictions as TU (70% versus 36%), but the ratio of correct to incorrect is slightly less (42% compared to 48%). Table 4.34 shows the same inverse correlation between EIC and NG that we saw between EIC and TU, with NG scoring its limited successes where EIC is incorrect or weak. Overall, NG achieves very similar results to TU, which suggests that (un)importance is not a significant contributor to this latter.

This conclusion is reinforced by an examination of the Hungarian and German data of tables 4.32 and 4.33 from the perspective of NG. For Hungarian (cf. table 4.35) the number of critical predictions made is almost identical (57% versus 59%), and the low 22% success rate of table 4.32 is now a similarly low 18%. The German data are also strikingly parallel (cf. table 4.36): 3% fewer critical predictions made (39% versus 42%); and 3% less correct (48% versus 51%). In both cases, the inverse correlation between EIC and NG remains.

These data offer little support for NG. In fact, they offer somewhat more support for the opposite of NG, namely given before new, which is what the Prague School advocates. This theory is now defined in (4.39) and its predictions are set out in (4.40):

\[(4.39) \text{ Given before New} \]

Given two ICs ordered \[IC_i, IC_j\]; if \[IC_i\] and \[IC_j\] differ in their \[P\text{redicability}\] values, then:
\[P_i < P_j\]
Table 4.34 *New–Given and EIC in English* $[NP \ V \ \{PP_1 \ PP_2\}]$

90 pages of data: cf. table 4.31.

I  **EIC predictions in relation to New–Given**

<table>
<thead>
<tr>
<th>EIC status</th>
<th>EIC ratio of correct: incorrect critical preds.</th>
<th>No. of critical NG preds.</th>
<th>NG ratio of correct: incorrect critical preds.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-optimal:</td>
<td>10 21%</td>
<td>4/10</td>
<td>4/4 = 100%</td>
</tr>
<tr>
<td>Equals:</td>
<td>21</td>
<td>14/21</td>
<td>11/14 = 79%</td>
</tr>
<tr>
<td>Optimal (1):</td>
<td>16 79%</td>
<td>14/16</td>
<td>3/14 = 21%</td>
</tr>
<tr>
<td>(2–8):</td>
<td>20</td>
<td>14/20</td>
<td>2/14 = 14%</td>
</tr>
<tr>
<td>(9 +):</td>
<td>2</td>
<td>2/2</td>
<td>0/2 = 0%</td>
</tr>
<tr>
<td>Totals:</td>
<td>69 79%</td>
<td>70%</td>
<td>42%</td>
</tr>
</tbody>
</table>

II  **New–Given predictions in relation to EIC**

<table>
<thead>
<tr>
<th>NG status</th>
<th>NG ratio of correct: incorrect critical preds.</th>
<th>No. of critical EIC preds.</th>
<th>EIC ratio of correct: incorrect critical preds.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorrect:</td>
<td>28 58%</td>
<td>25/28</td>
<td>25/25 = 100%</td>
</tr>
<tr>
<td>Either–or:</td>
<td>21</td>
<td>14/21</td>
<td>8/14 = 57%</td>
</tr>
<tr>
<td>Correct:</td>
<td>20 42%</td>
<td>9/20</td>
<td>5/9 = 56%</td>
</tr>
<tr>
<td>Totals:</td>
<td>69 42%</td>
<td>70%</td>
<td>79%</td>
</tr>
</tbody>
</table>

(4.40)  **Given-before-New predictions**

Given two ICs ordered $[IC_i \ IC_j]$, then:

- Single orders predicted: $P_i < P_j$
  - $AB$
- Both orders predicted: $P_i = P_j$
- Counterexamples: $P_i > P_j$
  - *BA*

The Prague School's predictions (GN) are set out for the English data in table 4.37. Obviously, the same number of critical predictions is made in tables 4.34 and 4.37. What differs is the success rate, which is now the mirror-image of
Testing EIC's performance predictions

Table 4.35 New–Given and EIC in Hungarian \[V \{mNP_1 , mNP_2 \}\]

112 pages of data: cf. table 4.4.

I  EIC predictions in relation to New–Given

<table>
<thead>
<tr>
<th>EIC status</th>
<th>EIC ratio of correct:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>incorrect critical</td>
</tr>
<tr>
<td></td>
<td>preds.</td>
</tr>
<tr>
<td>Non-optimal:</td>
<td>10 11%</td>
</tr>
<tr>
<td>Equals:</td>
<td>21 11/21 = 36%</td>
</tr>
<tr>
<td>Optimal (1):</td>
<td>50 25/50 = 12%</td>
</tr>
<tr>
<td>(2 +) :</td>
<td>35 27/35 = 7%</td>
</tr>
</tbody>
</table>

| Totals:        | 116 89%               |

NG status

<table>
<thead>
<tr>
<th>NG status</th>
<th>NG ratio of correct:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>incorrect critical</td>
</tr>
<tr>
<td></td>
<td>preds.</td>
</tr>
<tr>
<td>Incorrect:</td>
<td>54 82%</td>
</tr>
<tr>
<td>Either–or:</td>
<td>50 18%</td>
</tr>
<tr>
<td>Correct:</td>
<td>12 18%</td>
</tr>
</tbody>
</table>

| Totals:      | 116 89%               |

what it was for NG: 58% as opposed to 42%. In addition, the inverse correlation between EIC and NG has become a positive correlation: where EIC is non-optimal, GN scores 0% correct; for equal EIC ratios, 21%; as soon as EIC becomes optimal by as much as a one-word differential, GN scores 79%; and this figure continues to rise as EIC's preferences become stronger. GN's overall success rate (58%) is still significantly lower than EIC's (79%) and remains dangerously close to random, but the correlation requires an explanation.

The Hungarian data in table 4.38 offer much stronger support for GN over NG: 82% of GN's critical predictions are correct, compared with 18% for NG. The number of critical predictions made is still relatively small (57%
Comparing EIC with pragmatic predictions

Table 4.36 New–Given and EIC in German NP-PP rearrangements

100 pages of data: cf. table 4.23.

I EIC predictions in relation to New–Given

<table>
<thead>
<tr>
<th>EIC status</th>
<th>EIC ratio of correct: incorrect critical preds.</th>
<th>No. of critical NG preds.</th>
<th>NG ratio of correct: incorrect critical preds.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-optimal: 14</td>
<td>21%</td>
<td>7/14</td>
<td>4/7 = 57%</td>
</tr>
<tr>
<td>Equals: 21</td>
<td></td>
<td>11/21</td>
<td>6/11 = 55%</td>
</tr>
<tr>
<td>Optimal (1–6): 48</td>
<td>79%</td>
<td>14/48</td>
<td>6/14 = 43%</td>
</tr>
<tr>
<td>(7 +): 6</td>
<td></td>
<td>3/6</td>
<td>0/3 = 0%</td>
</tr>
<tr>
<td>Totals: 89</td>
<td>79%</td>
<td>39%</td>
<td>48%</td>
</tr>
</tbody>
</table>

II NG predictions in relation to EIC

<table>
<thead>
<tr>
<th>NG status</th>
<th>NG ratio of correct: incorrect critical preds.</th>
<th>No. of critical EIC preds.</th>
<th>EIC ratio of correct: incorrect critical preds.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorrect: 19</td>
<td>52%</td>
<td>14/19</td>
<td>11/14 = 79%</td>
</tr>
<tr>
<td>Either–or: 54</td>
<td></td>
<td>44/54</td>
<td>37/44 = 84%</td>
</tr>
<tr>
<td>Correct: 16</td>
<td>48%</td>
<td>10/16</td>
<td>6/10 = 60%</td>
</tr>
<tr>
<td>Totals: 89</td>
<td>48%</td>
<td>76%</td>
<td>79%</td>
</tr>
</tbody>
</table>

compared with 82% for EIC), and the 82% success rate is still less than the 89% for EIC. In absolute numbers, GN makes 54 correct critical predictions, EIC 85. Once again, the inverse correlation between them has become a positive one.

The German data of table 4.36 provide equal support for NG and GN, and hence do not decide between them, so we do not need to pursue them further here.

Consider now some data from a fourth language, Japanese, collected by Kaoru Horie. Table 4.39 examines the relative ordering of two PPs from a pragmatic perspective. The EIC predictions for the structure \([iPP_m 2PP_m] V]\) were set out in table 4.9, and table 4.39 juxtaposes EIC predictions with the
Testing EIC's performance predictions

Table 4.37 Given–New and EIC in English \{NP V \{PP_1 PP_2\}\}

90 pages of data: cf. table 4.31.

I  EIC predictions in relation to Given–New

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-optimal:</td>
<td>10 21%</td>
<td>4/10</td>
<td>0/4 = 0%</td>
</tr>
<tr>
<td>Equals:</td>
<td>21</td>
<td>14/21</td>
<td>3/14 = 21%</td>
</tr>
<tr>
<td>Optimal (1):</td>
<td>16</td>
<td>14/16</td>
<td>11/14 = 79%</td>
</tr>
<tr>
<td>(2–8):</td>
<td>20 79%</td>
<td>14/20</td>
<td>12/14 = 86%</td>
</tr>
<tr>
<td>(9 + ):</td>
<td>2</td>
<td>2/2</td>
<td>2/2 = 100%</td>
</tr>
<tr>
<td><strong>Totals:</strong></td>
<td><strong>69 79%</strong></td>
<td><strong>70%</strong></td>
<td><strong>58%</strong></td>
</tr>
</tbody>
</table>

II  Given–New predictions in relation to EIC

<table>
<thead>
<tr>
<th>GN status</th>
<th>GN ratio of correct: incorrect critical preds.</th>
<th>No. of critical EIC preds.</th>
<th>EIC ratio of correct: incorrect critical preds.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorrect:</td>
<td>20 42%</td>
<td>9/20</td>
<td>5/9 = 56%</td>
</tr>
<tr>
<td>Either–or:</td>
<td>21</td>
<td>14/21</td>
<td>8/14 = 57%</td>
</tr>
<tr>
<td>Correct:</td>
<td>28 58%</td>
<td>25/28</td>
<td>25/25 = 100%</td>
</tr>
<tr>
<td><strong>Totals:</strong></td>
<td><strong>69 58%</strong></td>
<td><strong>70%</strong></td>
<td><strong>79%</strong></td>
</tr>
</tbody>
</table>

Prague School’s given-before-new predictions. The number of critical predictions made by the two theories is now identical (54%), but the success rate of EIC is again significantly higher: 73% versus 58%. This difference is close to that between EIC and GN for two PPs in English (79% versus 58%, cf. table 4.37). There is one striking difference between the two languages, however. EIC and GN are positively correlated in English: as EIC improves its ratio of correct to incorrect critical predictions, so does GN, and vice versa. But EIC and GN are inversely correlated in Japanese: where EIC is non-optimal, GN scores 100%; where it is optimal, GN’s success rate falls to 45%; conversely where GN is incorrect, EIC scores 100%; and where it is correct, EIC scores only 50%. This kind of inverse correlation is exactly what we saw between EIC and Givon’s theory in English, i.e. between EIC and new before given
Comparing EIC with pragmatic predictions

Table 4.38 Given-New and EIC in Hungarian \[V \{mNP_1 mNP_2\}\]

112 pages of data: cf. table 4.4.

I  EIC predictions in relation to Given-New

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-optimal: 10</td>
<td>11%</td>
<td>3/10</td>
<td>0/3 = 0%</td>
</tr>
<tr>
<td>Equals: 21</td>
<td>11/21</td>
<td>7/11 = 64%</td>
<td></td>
</tr>
<tr>
<td>Optimal (1): 50</td>
<td>89%</td>
<td>25/50</td>
<td>22/25 = 88%</td>
</tr>
<tr>
<td>(2+): 35</td>
<td>27/35</td>
<td>25/27 = 93%</td>
<td></td>
</tr>
<tr>
<td>Totals: 116</td>
<td>89%</td>
<td>57%</td>
<td>82%</td>
</tr>
</tbody>
</table>

II  Given-New predictions in relation to EIC

<table>
<thead>
<tr>
<th>GN status</th>
<th>GN ratio of correct: incorrect critical preds.</th>
<th>No. of critical EIC preds.</th>
<th>EIC ratio of correct: incorrect critical preds.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorrect: 12</td>
<td>18%</td>
<td>8/12</td>
<td>5/8 = 63%</td>
</tr>
<tr>
<td>Either-or: 50</td>
<td>40/50</td>
<td>33/40 = 83%</td>
<td></td>
</tr>
<tr>
<td>Correct: 54</td>
<td>82%</td>
<td>47/54</td>
<td>47/47 = 100%</td>
</tr>
<tr>
<td>Totals: 116</td>
<td>82%</td>
<td>82%</td>
<td>89%</td>
</tr>
</tbody>
</table>

and Task Urgency, cf. tables 4.34 and 4.31. The inverse correlation became a positive correlation when the opposite theory, GN, was tested on English, but that same theory has now yielded an inverse correlation in Japanese. New before given will therefore show a positive correlation with EIC in Japanese, and this can be seen in table 4.40.

Summarizing, EIC and both GN and NG do not differ in the number of critical predictions made on this occasion, but EIC's 73% success rate is higher than the 58% of GN and much higher than the 42% of NG. The striking thing is the correlation pattern between EIC and the pragmatics: EIC is positively correlated with GN in English, but with NG in Japanese.
Table 4.39 Given-New and EIC in Japanese \( \{ \{ PP_m \} \ 2PP_m \} \ V \)

70 pages of data: first 70 pp. of data from Juugo Kuroiwa Yami No Hada, Koobunsha, Tokyo.

Cf. table 4.9 for structural template, additional EIC data and EIC calculation assumptions.

I  EIC predictions in relation to Given-New

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-optimal:</td>
<td>7 27%</td>
<td>5/7</td>
<td>5/5 = 100%</td>
</tr>
<tr>
<td>Equals:</td>
<td>22</td>
<td>10/22</td>
<td>5/10 = 50%</td>
</tr>
<tr>
<td>Optimal:</td>
<td>19 73%</td>
<td>11/19</td>
<td>5/11 = 45%</td>
</tr>
<tr>
<td>Totals:</td>
<td>48 73%</td>
<td>54%</td>
<td>58%</td>
</tr>
</tbody>
</table>

II  Given-New predictions in relation to EIC

<table>
<thead>
<tr>
<th>GN status</th>
<th>GN ratio of correct: incorrect critical preds.</th>
<th>No. of critical EIC preds.</th>
<th>EIC ratio of correct: incorrect critical preds.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorrect:</td>
<td>11 42%</td>
<td>6/11</td>
<td>6/6 = 100%</td>
</tr>
<tr>
<td>Either-or:</td>
<td>22</td>
<td>10/22</td>
<td>8/10 = 80%</td>
</tr>
<tr>
<td>Correct:</td>
<td>15 58%</td>
<td>10/15</td>
<td>5/10 = 50%</td>
</tr>
<tr>
<td>Totals:</td>
<td>48 58%</td>
<td>54%</td>
<td>73%</td>
</tr>
</tbody>
</table>

The English correlation pattern has also been shown to hold for \[ V \{ mNP_1 \ mNP_2 \} \] structures in Hungarian (cf. table 4.38), and for \{NP PP\} combinations in German (cf. the inverse correlation between EIC and NG in table 4.36, which becomes a positive one in relation to GN). Japanese is the odd man out so far, so it is important to see whether the pattern of tables 4.39 and 4.40 generalizes to other structures. It does, in fact, and tables 4.41 and 4.42 present some further data involving the relative ordering of an accusative NP0 and PP, the EIC predictions for which were set out in table 4.10. Table 4.41 compares EIC and GN. Sections I and II present the EIC predictions on the assumption that there is no VP discontinuity when the direct
Comparing EIC with pragmatic predictions

Table 4.40 New–Given and EIC in Japanese $\{1PP_m 2PP_m \} V$

70 pages of data: cf. table 4.39.

I  EIC predictions in relation to New–Given

<table>
<thead>
<tr>
<th>EIC status</th>
<th>EIC ratio of correct: incorrect critical preds.</th>
<th>No. of critical NG preds.</th>
<th>NG ratio of correct: incorrect critical preds.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-optimal</td>
<td>7 27%</td>
<td>5/7</td>
<td>0/5 = 0%</td>
</tr>
<tr>
<td>Equals</td>
<td>22</td>
<td>10/22</td>
<td>5/10 = 50%</td>
</tr>
<tr>
<td>Optimal</td>
<td>19 73%</td>
<td>11/19</td>
<td>6/11 = 55%</td>
</tr>
<tr>
<td>Totals</td>
<td>48 73%</td>
<td>54%</td>
<td>42%</td>
</tr>
</tbody>
</table>

II  New–Given predictions in relation to EIC

<table>
<thead>
<tr>
<th>NG status</th>
<th>NG ratio of correct: incorrect critical preds.</th>
<th>No. of critical EIC preds.</th>
<th>EIC ratio of correct: incorrect critical preds.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorrect</td>
<td>15 58%</td>
<td>10/15</td>
<td>5/10 = 50%</td>
</tr>
<tr>
<td>Either–or</td>
<td>22</td>
<td>10/22</td>
<td>8/10 = 80%</td>
</tr>
<tr>
<td>Correct</td>
<td>11 42%</td>
<td>6/11</td>
<td>6/6 = 100%</td>
</tr>
<tr>
<td>Totals</td>
<td>48 42%</td>
<td>54%</td>
<td>73%</td>
</tr>
</tbody>
</table>

object NPo precedes the PP_m, i.e. the critical predictions are consistently in favor of long before short ICm's (cf. table 4.10 section I). Sections III and IV assume the possibility of VP discontinuity and define EIC's predictions on those orders for which a consistent long-before-short prediction can be made, namely those in which PP_m ≥ NPo (which is the majority: 100/122, or 82%). In I and II EIC and GN make the same number of critical predictions (57%), but EIC's success rate is higher (71% versus 56%). Again we see the same inverse correlation as in table 4.39: when EIC is non-optimal, GN achieves its highest success rate of 67%; when EIC is most optimal (for 3+ word-length differentials), GN is at its worst (22%). Conversely when GN is incorrect, EIC does best (78%); and when GN is correct, EIC has its lowest score (64%). In sections III and IV EIC's success rate compared to GN's is even
higher (83% versus 58%), although the number of critical predictions made by EIC is slightly lower (48% to 55%). The inverse correlation pattern is the same as in I and II. As in all of the tables we have seen, the percentage figures for EIC's ratios of correct to incorrect critical predictions are higher, because EIC's overall success rate is higher.

Table 4.42 shows the clear positive correlation between EIC and NG. The success rate for the latter rises from 33% to 78% in all orders as EIC goes from non-optimal to optimal (3 +), cf. section I. The success rate rises from 17% to 71% in the orders in which PPₘ ≥ NPₒ, cf. section III.

Table 4.41 Given–New and EIC in Japanese [{NPₒ PPₘ} V]

70 pages of data: cf. table 4.39.
Cf. table 4.10 for structural template, additional EIC data and EIC calculation assumptions.

I EIC predictions for all orders in relation to Given–New (assuming no VP discontinuity, cf. table 4.10 section I)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-optimal:</td>
<td>20 29%</td>
<td>15/20</td>
<td>10/15 = 67%</td>
</tr>
<tr>
<td>Equals:</td>
<td>52 19/52</td>
<td>11/19 = 58%</td>
<td></td>
</tr>
<tr>
<td>Optimal (1–2):</td>
<td>34 27/34</td>
<td>16/27 = 59%</td>
<td></td>
</tr>
<tr>
<td>(3 + ):</td>
<td>16 9/16</td>
<td>2/9 = 22%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Totals: 122 71%</td>
<td>57%</td>
<td>56%</td>
</tr>
</tbody>
</table>

II Given–New predictions for all orders in relation to EIC (assuming no VP discontinuity, cf. table 4.10 section I)

<table>
<thead>
<tr>
<th>GN status</th>
<th>GN ratio of correct: incorrect critical preds.</th>
<th>No. of critical EIC preds.</th>
<th>EIC ratio of correct: incorrect critical preds.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorrect:</td>
<td>31 44%</td>
<td>23/31</td>
<td>18/23 = 78%</td>
</tr>
<tr>
<td>Either–or:</td>
<td>52 19/52</td>
<td>14/19 = 74%</td>
<td></td>
</tr>
<tr>
<td>Correct:</td>
<td>39 28/39</td>
<td>18/28 = 64%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Totals: 122 56%</td>
<td>57%</td>
<td>71%</td>
</tr>
</tbody>
</table>
### Table 4.41 (cont.)

#### III EIC predictions for $PP_m \geq NPo$ in relation to Given-New (assuming possible VP discontinuity, cf. table 4.10 section II)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-optimal:</td>
<td>8 17%</td>
<td>6/8</td>
<td>5/6 = 83%</td>
</tr>
<tr>
<td>Equals:</td>
<td>52</td>
<td>19/52</td>
<td>11/19 = 58%</td>
</tr>
<tr>
<td>Optimal (1–2):</td>
<td>28 RW</td>
<td>23/28</td>
<td>14/23 = 61%</td>
</tr>
<tr>
<td>(3 + ): 12</td>
<td></td>
<td>7/12</td>
<td>2/7 = 29%</td>
</tr>
<tr>
<td><strong>Totals:</strong></td>
<td><strong>100</strong></td>
<td><strong>55%</strong></td>
<td><strong>58%</strong></td>
</tr>
</tbody>
</table>

#### IV Given-New predictions for $PP_m \geq NPo$ in relation to EIC (assuming possible VP discontinuity, cf. table 4.10 section II)

<table>
<thead>
<tr>
<th>GN status</th>
<th>GN ratio of correct: incorrect critical preds.</th>
<th>No. of critical EIC preds.</th>
<th>EIC ratio of correct: incorrect critical preds.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorrect:</td>
<td>23 42%</td>
<td>15/23</td>
<td>14/15 = 93%</td>
</tr>
<tr>
<td>Either–or:</td>
<td>45</td>
<td>12/45</td>
<td>10/12 = 83%</td>
</tr>
<tr>
<td>Correct:</td>
<td>32 58%</td>
<td>21/32</td>
<td>16/21 = 76%</td>
</tr>
<tr>
<td><strong>Totals:</strong></td>
<td><strong>100</strong></td>
<td><strong>48%</strong></td>
<td><strong>83%</strong></td>
</tr>
</tbody>
</table>

We can summarise all of these results as follows. First, the difference between Givon's TU theory and NG alone (i.e. without the [un]importance factor) appears to be largely insignificant: compare tables 4.31–4.33 with 4.34–4.36. Secondly, with the exception of the Hungarian data, where the Prague School's rival GN theory has a distinct advantage (cf. table 4.38), both NG and GN orders are productively attested in all these languages, with GN having a slight edge overall of roughly 56%–44% (excluding Hungarian). Thirdly, EIC's success rate is consistently higher (measured in terms of ratios of correct to incorrect predictions) and the number of critical predictions that it makes is generally greater than or equal to the pragmatic predictions. Fourthly, even though the pragmatic predictions are much less successful...
Table 4.42 New–Given and EIC in Japanese [{NPo PP<sub>m</sub>} V]

70 pages of data: cf. table 4.39.
Cf. table 4.10 for structural template, additional EIC data and EIC calculation assumptions.

I  EIC predictions for all orders in relation to Given–New (assuming no VP discontinuity, cf. table 4.10 section I)

<table>
<thead>
<tr>
<th>EIC status</th>
<th>EIC ratio of correct: incorrect critical preds.</th>
<th>No. of critical NG preds.</th>
<th>NG ratio of correct: incorrect critical preds.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-optimal:</td>
<td>20</td>
<td>29%</td>
<td>15/20</td>
</tr>
<tr>
<td>Equals:</td>
<td>52</td>
<td>8/19 = 42%</td>
<td>19/52</td>
</tr>
<tr>
<td>Optimal (1–2):</td>
<td>34</td>
<td>41%</td>
<td>27/34</td>
</tr>
<tr>
<td>(3 +):</td>
<td>16</td>
<td>78%</td>
<td>9/16</td>
</tr>
<tr>
<td>Totals:</td>
<td>122</td>
<td>71%</td>
<td>57%</td>
</tr>
</tbody>
</table>

II  New–Given predictions for all orders in relation to EIC (assuming no VP discontinuity, cf. table 4.10 section I)

<table>
<thead>
<tr>
<th>NG status</th>
<th>NG ratio of correct: incorrect critical preds.</th>
<th>No. of critical EIC preds.</th>
<th>EIC ratio of correct: incorrect critical preds.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorrect:</td>
<td>39</td>
<td>56%</td>
<td>28/39</td>
</tr>
<tr>
<td>Either–or:</td>
<td>52</td>
<td>74%</td>
<td>19/52</td>
</tr>
<tr>
<td>Correct:</td>
<td>31</td>
<td>78%</td>
<td>23/31</td>
</tr>
<tr>
<td>Totals:</td>
<td>122</td>
<td>44%</td>
<td>57%</td>
</tr>
</tbody>
</table>

III  EIC predictions for PP<sub>m</sub> ≥ NPo in relation to New–Given (assuming possible VP discontinuity, cf. table 4.10 section II)

<table>
<thead>
<tr>
<th>EIC status</th>
<th>EIC ratio of correct: incorrect critical preds.</th>
<th>No. of critical NG preds.</th>
<th>NG ratio of correct: incorrect critical preds.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-optimal:</td>
<td>8</td>
<td>17%</td>
<td>6/8</td>
</tr>
<tr>
<td>Equals:</td>
<td>52</td>
<td>42%</td>
<td>19/52</td>
</tr>
<tr>
<td>Optimal (1–2):</td>
<td>28</td>
<td>39%</td>
<td>23/28</td>
</tr>
<tr>
<td>(3 +):</td>
<td>12</td>
<td>71%</td>
<td>7/12</td>
</tr>
<tr>
<td>Totals:</td>
<td>100</td>
<td>83%</td>
<td>55%</td>
</tr>
</tbody>
</table>
Table 4.42 cont’d

IV  New-Given predictions for $PP_m \geq NPo$ in relation to EIC (assuming possible VP discontinuity, cf. table 4.10 section II)

<table>
<thead>
<tr>
<th>NG status</th>
<th>NG ratio of correct: incorrect critical preds.</th>
<th>No. of critical EIC preds.</th>
<th>EIC ratio of correct: incorrect critical preds.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorrect:</td>
<td>32</td>
<td>21/32</td>
<td>16/21 = 76%</td>
</tr>
<tr>
<td>Either-or:</td>
<td>45</td>
<td>12/45</td>
<td>10/12 = 83%</td>
</tr>
<tr>
<td>Correct:</td>
<td>23</td>
<td>15/23</td>
<td>14/15 = 93%</td>
</tr>
<tr>
<td>Totals:</td>
<td>100</td>
<td>48%</td>
<td>83%</td>
</tr>
</tbody>
</table>

overall, there appears to be a significant interaction with EIC: the distribution of GN and NG orders is not random. In the English, Hungarian, and German data, GN achieves its successes within the set of data correctly predicted by EIC, and vice versa (cf. tables 4.37, 4.38, and 4.36 respectively); in Japanese, NG correlates positively with EIC in this way (cf. table 4.40 and 4.42). I shall now propose an explanation for this correlation, and draw the appropriate conclusions about the role of pragmatics in linear ordering.

4.4.4 On the role of pragmatics in linear ordering

The data of tables 4.31–4.42 lead to the following conclusion about the role of pragmatics in linear ordering: neither Givon’s theory (TU or NG) nor the Prague School’s (GN) can be considered significant generalizations about discourse organization. Clearly, they cannot both be proposed simultaneously within the context of a single theory, because they contradict each other. But whichever one is chosen as the significant organizing principle will regularly be counterexemplified by productive instances of the other, in every language type that we have considered. Hence, there is no currently valid descriptive generalization for linear ordering in pragmatic terms, a finding which reinforces the conclusion we reached about proposed pragmatic effects associated with rearrangement rules such as Heavy NP Shift (cf. section 4.3.2).
Why, then, are there significant correlations between EIC and information status? And especially, why are these correlations different for different language types?

EIC provides a simple solution. Let us assume that the data from English, Hungarian, German, and Japanese are typical of the language types that they exemplify. The correlation pattern that we have observed sets the first three of these languages against Japanese. What unites the three sets of structures examined in English, Hungarian, and German is that they involve relative orderings of NP and PP, all of which are of the type $m_{IC}$, i.e. they are constructed and recognized on their left peripheries. As a result, EIC defines a consistent short-before-long prediction for them (cf. table 3.2). The Japanese NPs and PPs, by contrast, are all of the type $IC_{m}$, for which EIC defines a long-before-short preference. The EIC correlation with the pragmatics can therefore be captured as follows: when EIC predicts short before long ICs, there is a positive correlation with given before new; when it predicts long before short, there is a positive correlation with new before given. Is this to be expected?

Most certainly it is. It follows logically because new information for the hearer will generally require more linguistic material for referent identification, and hence longer NPs and PPs, whereas predictable items can regularly be referred to by single-word pronouns and proper names, or by short(er) definite full NPs. So if EIC predicts short before long, there should be a correlation with given before new, and conversely for long before short.

Empirically we can test this to see if the NPs and PPs in our data exhibit a correlation in the direction of short–given and long–new, irrespective of relative ordering. In other words, are shorter ICs more given than longer ones? Conversely, are given items referred to by shorter ICs than new ones? Some sample data involving PPs (in English) and NPs (in German) are set out in tables 4.43 and 4.44 respectively. Table 4.43 shows that as the word length for a PP grows, so does the referential distance to a previous mention (if any), i.e. the P(predictability) score measured in terms of numbers of clauses back to a previous mention. Conversely, as the P score grows, so does the average length of the PP containing a reference to the previously mentioned item. Table 4.44 shows the same correlation for NPs. Very similar correlations have also been demonstrated for a third language, Polish, in Siewierska (1991b).

But given this correlation, which is the chicken and which is the egg, i.e. what is the primary determinant of order — information structure or syntactic weight? The answer is fairly clear. The significant generalization cannot be
Comparing EIC with pragmatic predictions

Table 4.43 Length and P(redictability) correlates for English PPs

90 pages of data: cf. table 4.31.
PPs = those occurring in the structural template of table 4.2, and in either order;
n = 138

I PP length in relation to Predictability

<table>
<thead>
<tr>
<th>PP length</th>
<th>Avge. P score (i.e. no. of clauses back to a previous mention)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 words:</td>
<td>7.65 clauses</td>
</tr>
<tr>
<td>3–4:</td>
<td>16.5</td>
</tr>
<tr>
<td>5–7:</td>
<td>17.5</td>
</tr>
<tr>
<td>8+:</td>
<td>21.0 (i.e. 21 to infinity)</td>
</tr>
</tbody>
</table>

II Predictability in relation to PP length

<table>
<thead>
<tr>
<th>P score</th>
<th>Avge. PP length</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–2 clauses:</td>
<td>2.37 words</td>
</tr>
<tr>
<td>3–15:</td>
<td>3.33</td>
</tr>
<tr>
<td>16–20:</td>
<td>4.17</td>
</tr>
<tr>
<td>21+</td>
<td>4.37</td>
</tr>
</tbody>
</table>

the pragmatics, for several reasons. First because pragmatic theories make many fewer correct predictions than EIC. Secondly, whichever pragmatic theory, Givon’s or the Prague School’s, is chosen as the correct one will have almost as many counterexamples as supporting instances, namely all the cases that support the other theory. And third, and most significantly, if we take the pragmatics as primary, we have no way of explaining why English, Hungarian, and German should pattern together against Japanese. Let us say that we propose Givon’s theory. Then, given this theory and given the length–information–status correlation of tables 4.43 and 4.44, we would predict new before given and long before short ICs – in all structures and languages. Given the Prague School’s theory, we predict given before new and short before long ICs – in all structures and languages. Looking at it from the direction of EIC and the syntax, however, we make exactly the right predictions. Given EIC and a language with ICs of the type IC_m, we predict long-before-short ordering. Given the length–information-status correlation, we then predict a positive correlation.
Testing EIC's performance predictions

Table 4.44 Length and \( P(\text{redictability}) \) correlates for German NPs

50 pages of data: first 50 pp. from P. Handke *Die linkshändige Frau*, Suhrkamp

NPs = accusative-marked full or pronominal NPs occurring in the German middle field and ordered either before or after PP in the structural template of table 4.23;

\( n = 79 \)

I \text{ NP length in relation to Predictability}

<table>
<thead>
<tr>
<th>NP length</th>
<th>Avge. P score (i.e. no. of clauses back to a previous mention)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 word:</td>
<td>3.66 clauses</td>
</tr>
<tr>
<td>2:</td>
<td>18.22</td>
</tr>
<tr>
<td>3–7:</td>
<td>18.4</td>
</tr>
<tr>
<td>8+:</td>
<td>21.0 (i.e. 21 to infinity)</td>
</tr>
</tbody>
</table>

II \text{ Predictability in relation to NP length}

<table>
<thead>
<tr>
<th>P score</th>
<th>Avge. NP length</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–2 clauses:</td>
<td>1.27 words</td>
</tr>
<tr>
<td>3–21:</td>
<td>2.4</td>
</tr>
<tr>
<td>21+:</td>
<td>3.64</td>
</tr>
</tbody>
</table>

in these languages with new-before-given ordering. For languages with ICs of the opposite type, \( mIC \), EIC predicts short before long, and the information-status correlate is now predicted to be with given before new. The number of correct predictions made by each of these pragmatic theories is not high enough for these correlations to affect the overall success rate of NG and GN in each language type, as we have seen. What is being captured in these syntactic-weight–pragmatics correlations is simply the distribution of correct to incorrect pragmatic predictions made by a given pragmatic theory in relation to EIC. In Japanese, Givon's (NG) theory does best when EIC does best; in English, the Prague School (GN) does best when EIC does best. Overall, neither pragmatic theory is well supported; but where there are correlations with EIC, the explanation for the correlations can come only from EIC.

What does this mean about the role of pragmatics in linear ordering? My data suggest a radical and controversial conclusion: pragmatics appears to
play no role whatsoever. The theories proposed add nothing to the syntactically based predictions of EIC. When I began this pragmatic investigation, I had originally hoped to find a pragmatic motivation for linear ordering in the minority structures that are non-optimal for EIC. In this way I could have appealed to some interacting contrary pressure that would have pulled orders away from the EIC ideal, especially when EIC's preferences are weak (which is where the great majority of the non-optimal orders are found). Preliminary results from English-type languages suggested that something like this was indeed going on (cf. Hawkins 1992c for discussion): EIC defines a short-before-long preference, which correlates positively with given before new. The non-optimal EIC orders could therefore be argued to be motivated by Givon's NG theory, to which I am philosophically sympathetic. However, the Japanese results rule out this interpretation, because the EIC-preferred orders now correlate positively with new before given, and the non-optimal ones with given before new, so these latter would need to be motivated by appealing to the Prague School! Once again, we can't have it both ways, and the non-optimal orders will have to be explained in syntactic processing terms: if EIC's preferences are weak, it doesn't make a lot of difference for the overall complexity of a sentence whether it is adhered to or not, and so it is precisely in these cases that we expect performance variability, and a certain number of non-optimal structures.

Clearly, this rather negative conclusion about the relevance of pragmatics must be considered tentative, until more languages have been examined from this point of view. The data that I have presented, however, have explicitly compared the major current syntactic analyses for these structures with the major current pragmatic theories, and the results suggest strongly that data that might appear to support pragmatic theories are a by-product of syntactic processing. To argue against this one will need to collect comparative data from these and other languages that point to a very different pattern of interaction between EIC and pragmatics. The kind of data cited by Givon and his associates (1983, 1988) makes no reference to syntactic analyses for the relevant structures; hence it is not clear what EIC would predict for these structures in these languages; hence their results do not necessarily establish the case for pragmatic principles of linear ordering. Therefore, the conclusion that I draw from my data, and that I would invite others to challenge, is that pragmatic principles are epiphenomenal.

The big exception to this is if there happens to be a discourse-configurational node in the grammar of the relevant language, as discussed for Hungarian and other languages in section 4.1. The selection of items to fill this node in
performance will then be driven by whatever pragmatic-semantic status it has, and not by weight, even though the grammaticalized ordering of this node relative to others is explainable by EIC (cf. Primus 1991b). In structures in which there is no grammatical discourse-configurational node, therefore, the data of this section suggest that the linear ordering of ICs in performance is not driven by pragmatics.

4.5 Temporary ambiguity preferences

In ch. 3.5.3 I formulated a prediction for preferred IC attachments in temporary ambiguities. The prediction was formulated in (3.20) and is repeated here:

(3.20) **EIC Performance Attachment Prediction**

Given: alternative grammatical attachments of an IC to different dominating phrases within a sentence S; 

Then: attachments will be preferred in performance that provide the most optimal EIC ratios for the sentence, in the unmarked case; attachments with non-optimal EIC ratios will be more or equally preferred in direct proportion to the magnitude of their EIC ratios.

This prediction, like the EIC Text Frequency Prediction (3.18) and the EIC Performance Ordering Prediction (3.17), comes in two parts, an unmarked case prediction and a marked case prediction. In the unmarked case it is expected that IC attachment preferences will be those that are best for EIC when the parser has a choice between two grammatical assignments. Similarly, the choices that are made between two alternative orderings of ICs were predicted to be optimal for EIC in general. In the marked case, if any preferred IC attachments are for some reason not optimal for EIC, then it should still be the case that attachment preferences are in proportion to the size of their EIC ratios, with the next highest ratio being most preferred, etc. Similarly, for non-optimal linear orderings, the most frequent ones should be those that are only slightly dispreferred by EIC, the next most frequent will be slightly less preferred, and so on.

The unmarked case prediction will, as we shall see, predict most of the Late Closure effects observed by Frazier (1979a). Frazier's other preferences will be argued to follow from complexity differences between alternative Structural Domains (cf. ch. 2.2.1). The precise nature of these structural predictions is complicated, however, by some still unresolved disagreements.
in the psycholinguistic literature over the nature of temporary ambiguity resolution. One major issue involves the timing of the interaction between different knowledge sources, i.e. between structural preferences on the one hand, and semantic considerations (cf. Tannenhaus and Carlson 1989), and even real-world plausibility information (cf. Pearlmutter and MacDonald 1992), on the other. For example, it has been shown in Tannenhaus and Carlson (1989) and Trueswell, Tannenhaus, and Guernsey (1992) that the temporary ambiguity in the sequence *The defendant examined* (between a main clause reading *The defendant examined the document*, and a reduced relative, *The defendant examined by the lawyer was nervous*) disappears in the sequence *The evidence examined* (which has the reduced relative interpretation only). The reason is that the main- clause interpretation requires an agentive subject of *examined*, and *the evidence*, being inanimate, cannot be an agent. But this means that semantic information of this sort is immediately available to restrict a structural assignment that is possible in other cases. Pearlmutter and MacDonald (1992) show that even pragmatic information can have this effect.

The conclusion I draw from this is that the availability of a structural choice, and hence of a temporary ambiguity and possible garden path, can depend on non-syntactic factors. EIC’s prediction (3.20) is accordingly relativized to those sentence types in which there is a genuine choice. This is captured by the phrase “Given alternative grammatical attachments of an IC . . . .” If there are no alternatives, then EIC has nothing to choose from. Fortunately, most of the sentence types that have been used to motivate Late Closure, Minimal Attachment, etc., are ones in which two alternative readings are semantically and pragmatically plausible and in which the effects of structural complexity alone have been measured. I leave it to the psycholinguists to sort out precisely when non-syntactic factors can eliminate the availability of a choice and to develop an appropriate interactive theory of parsing.

There is another unresolved theoretical issue of potential relevance to EIC’s predictions. In the model of Frazier (1979a, 1985) temporary ambiguity resolution proceeds in accordance with her First Analysis Constraint, given in (4.41):

\[(4.41) \text{ Frazier's First Analysis Constraint} \]

At choice points the processor adopts the first syntactic analysis available (specifically the analysis that follows from Minimal Attachment, Late Closure, and the Most Recent Filler Strategy)
Testing EIC's performance predictions

rather than pursuing multiple syntactic analyses or delaying analysis of the input.

If this is how ambiguity resolution works, then the only predictions that EIC can make involve the preferred, first analyses assigned by the parser. Are the parser's attachments in accordance with (3.20), or are they not? By contrast, in a model such as that of Gorrell (1987), which allows for parallel activation and retention of alternative analyses, with preferences and strengths assigned to these alternatives, or in any model that allows for multiple syntactic analyses in the temporarily ambiguous portions of a sentence, it would be possible to test EIC in a more ambitious way and predict that, all things being equal, the activation level for any analysis should be in accordance with its EIC score. Thus, if two alternatives have equal or roughly equal EIC scores, they should be equally accessible on-line. If one is less preferred than another, the strength of activation for the weaker reading should be in accordance with its EIC score and with the degree of difference between the ratios for the stronger and for the weaker reading.

Since I shall be relying on the experimental tests of ambiguity resolution conducted by Frazier (1979a, 1985), the only prediction I can exemplify will be in terms of the single, preferred readings for sentences, i.e. the first analyses according to her theory. My basic claim is that EIC provides an explanation for an important subset of her experimental findings, those that are subsumed under the principle of Late Closure, defined as follows:

(4.42) Late Closure
Incorporate a new item into the phrase or clause currently being processed, rather than to an earlier (or subsequent) phrase or clause.

The effect of Late Closure can be illustrated with the English temporary ambiguity of (3.19), repeated here as (4.43):

(4.43) a. John \( \text{VP}_1[\text{read NP[the book } \text{§[that he had } \text{VP}_2[\text{selected PP[in the library]}]]]} \]  
   b. John \( \text{VP}_1[\text{read NP[the book } \text{§[that he had } \text{VP}_2[\text{selected}]] PP[\text{in the library}]}] \)

Frazier has shown experimentally that there is a preference for assigning the PP in the library to the lower VP\( _2 \) as an argument of selected (4.43a), rather than to the higher VP\( _1 \) as an argument of read. This can be seen particularly clearly in garden-path sentences of this type like (4.44) in which low attachment of the PP takes place first in accordance with Late Closure, whereupon
the PP is subsequently assigned to the higher VP when it is recognized that this is grammatically required by the subcategorization frame for *put*:

(4.44) a. John _VP_1[put _NP_[the book _S_[that he had _VP_2[selected _PP_[in the library]]]]]

   b. John _VP_1[put _NP_[the book _S_[that he had _VP_2[selected]]] _PP_[in the library]]

Late Closure stipulates a preference for low attachment. EIC can explain it. The EIC ratios for the sentence as a whole are improved when attachment is low. Consider the IC-to-word ratios of the two structures in (4.44), focusing only on the two phrases whose ratios are affected by the different assignments, VP₁ and VP₂:

(4.44')

<table>
<thead>
<tr>
<th>Structure</th>
<th>VP₁ CRD</th>
<th>VP₂ CRD</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. John _VP_1[put <em>NP</em>[the book <em>S</em>[that he had _VP_2[selected <em>PP</em>[in the library]]]]]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VP₁ CRD:</td>
<td>2/2</td>
<td>2/2</td>
</tr>
<tr>
<td>VP₂ CRD:</td>
<td>1/1</td>
<td>1/1</td>
</tr>
<tr>
<td>EIC ratio:</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>b. John _VP_1[put <em>NP</em>[the book <em>S</em>[that he had _VP_2[selected]]] <em>PP</em>[in the library]]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VP₁ CRD:</td>
<td>3/8</td>
<td></td>
</tr>
<tr>
<td>VP₂ CRD:</td>
<td>1/1</td>
<td></td>
</tr>
<tr>
<td>EIC ratio:</td>
<td>37.5%</td>
<td>100%</td>
</tr>
</tbody>
</table>

If the PP in the library is attached to VP₂, both VP₁ and VP₂ have optimal ratios, as in (4.44'a). If it is attached to the matrix VP₁, however, the result is a considerably longer CRD with a much poorer IC-to-word ratio of 37.5%, as in (4.44'b). VP₂ remains optimal. Aggregating over these scores, EIC predicts a preference for low attachment, a Late Closure effect. Late Closure is accordingly explained by the theory of Structural Domains in conjunction with the complexity metric of ch. 2.2.1 that underlies EIC, and EIC in turn predicts the preference for (4.44a) over (4.44b).

Another garden-path sentence that receives a similar explanation is (4.45):

(4.45) a. Mary _VP_1[worded _NP_[the letter she had been _VP_2[writing carefully]]]

   b. Mary _VP_1[worded _NP_[the letter she had been _VP_2[writing]]] carefully

Frazier's experiments have shown that the adverb *carefully* is first assigned to the lower VP₂ dominating *writing*, cf. (4.45a). A high attachment to the matrix VP₁ would produce a very long CRD and a low EIC ratio, cf.
Testing EIC's performance predictions

(4.45b) which parallels (4.44\textsuperscript{1}b). The processing of the subcategorization requirements of the matrix verb \textit{worded} subsequently forces high attachment, however, and results in a garden path.

A further sentence type that Frazier discusses in connection with Late Closure is (4.46), where the ambiguous portion is italicized:

\begin{equation}
\text{(4.46) While Mary was reading the book} \text{ fell down.}
\end{equation}

This sentence is temporarily ambiguous between a reading in which the NP \textit{the book} is the object of \textit{reading}, and another in which it is the subject of \textit{fell down}. Frazier presents evidence for an on-line preference for (4.46\textsuperscript{1}a) over (4.46\textsuperscript{1}b):  

\begin{equation}
\text{(4.46\textsuperscript{1}) a.}_\textsf{ s}\{\text{While Mary was reading the book} \ldots \}
\end{equation}

\begin{equation}
\text{b.}_\textsf{ s}\{\text{While Mary was reading the book} \ldots \}
\end{equation}

and she attributes this preference to Late Closure.

There is another way of accounting for this, however. Frazier assumes that both attachment sites, the matrix S and the subordinate \textit{S}, have been constructed at the time \textit{the book} is being parsed, and hence that the First Analysis Constraint will choose between them on the basis of Late Closure. But according to the parsing principles of this book (cf. ch. 3.2 and ch. 6.2), it is questionable whether the matrix S is available as well as the \textit{S} at this point. The complementizer \textit{while} constructs \textit{S} by Mother Node Construction, but the construction of the \textit{S} also requires some dominated matrix category that uniquely determines it, such as an auxiliary verb or an appropriate inflectional affix on the verb or a nominative-marked NP. \textit{S} cannot construct a matrix S, in general, since \textit{S} may be dominated by a number of different higher nodes. Nor can a sentence-initial constituent lead automatically to the postulation of a matrix S, as I argue in ch. 6.2.2, because S is not the only possible highest constituent in a sentence. The first constituent encountered may be immediately dominated by E (cf. Emonds 1976) rather than by S. And more generally the parsing principles that I am assuming, which are supported by performance and grammatical evidence involving word order, are based crucially on the notion of unique determination: an abstract syntactic category is only postulated when there is some material in the parse string that guarantees its existence.\textsuperscript{19} Applied to the sentence (4.46), this means that only one higher node is available at the time \textit{the book} is encountered, as shown in (4.46\textsuperscript{1}b). When the finite verb \textit{fell} is parsed, the verb will construct VP by Mother Node Construction, and the morphology
of finiteness will construct (the matrix) S by Grandmother Node Construction, whereupon the book will be reassigned to S by Immediate Constituent Attachment, since S requires a subject and this S cannot be the subject. S may be attached to S or to a higher E (cf. ch. 6.2.2).

\[(4.46') \text{ While Mary was reading the book . . .}\]

If this is correct, the preference for low attachment of the book (which results in the garden-path effect) would be explained by the fact that there is only one possible attachment site, and not by Late Closure.

Some support for this alternative comes from two quarters. First, EIC, which correctly explains and predicts the Late Closure preferences in the other cases, would actually define a slight preference for high attachment, i.e. for \[(4.46'^1)\text{ b} \] if both of the options of \[(4.46'^1)\] were available. If the six words of \[(4.46'^1)\] are all dominated by a single constituent, as in \[(4.46'^1)\text{ a} \], then the IC-to-word ratio for that portion of the CRD for S will be \[1/6 = 17\% \]. If they are distributed between two ICs, S and NP (or S and some second sentential IC, both of which are dominated by E), then the IC-to-word ratio will be much higher, e.g. \[2/6 = 33\% \]. Hence, EIC prefers to keep left-branching structures short in an English-type language, and this results in the preference for high attachment if the NP could be attached low or high. The garden-path effect therefore suggests that S is the only attachment site that is available on-line.

The second piece of support comes from temporary ambiguities and garden paths that are predicted by Frazier's Minimal Attachment principle, defined in \[(4.47)\]:

\[(4.47) \text{ Minimal Attachment}\]

Postulate the fewest nodes consistent with the grammar.

Some of the examples she discusses here are of the same type as \[(4.46)\], only more so. It is highly doubtful whether there is any psychological reality to the existence of a choice, because our parsing principles make only one attachment site available on-line. Consider \[(4.48)\]:

\[(4.48) \text{ Bill knew the answer was correct.}\]

The ambiguity is between \[(4.48'^1)\text{ a} \] and \[(4.48'^1)\text{ b} \]:

\[(4.48'^1) \text{ a. } s[Bill \text{ VP[knew NP[the answer . . . }\]

\[(4.48'^1) \text{ b. } s[Bill \text{ VP[knew } s[IS[NP[the answer . . . }\]
But I would argue that the only attachment site that is available to the parser at the time the attachment is made is VP, as in (4.48a). \( \tilde{S} \) and the lower S are not constructable in the absence of a complementizer. If the complementizer that is present, as in (4.49), it will force the construction of \( \tilde{S} \) by Mother Node Construction and of S by Sister Node Construction (cf. ch. 6.2.1), and this will remove the ambiguity.

(4.49) Bill knew that the answer was correct.

(4.49') \( s[Bill \, \text{vpfc} new \, g[that \, s] \ldots \)

Without the complementizer there will be no MNCC for \( \tilde{S} \), there will be no construction of the subordinate S, and since the verb know is compatible with several subcategorization frames it will not construct \( \tilde{S} \) or S either. Our parsing principles can therefore explain the garden-path effect here without having to assume an on-line structural choice resolved by Minimal Attachment. This supports a similar explanation for the apparent Late Closure effect in (4.46).

Other examples of Minimal Attachment cited by Frazier offer stronger evidence for the existence of a choice. For example, (4.50) which parallels The defendant examined... discussed above, seems to involve a genuine choice between a matrix VP analysis and a reduced relative clause, resulting from the morphological ambiguity of the verb expected:

(4.50) The candidate expected to win lost.

Her garden-path sentences seem to be of two kinds, therefore: those in which a structurally simpler analysis is first assigned because this is the only one compatible with independently motivated parsing principles; and those in which there is a genuine choice, and in which the First Analysis Constraint operates to assign the structurally simpler interpretation and to eliminate the alternative until it is called back on the basis of subsequent input.

Moreover, all of her structural preferences appear to be explainable by, and subsumable under, a single generalization. The preferred analyses involve less complex Structural Domains. The EIC-preferred attachments involve smaller and more efficient CRDs, which are simply Structural Domains for constituent recognition. The Minimal Attachments involve postulating fewer dominating nodes, and hence the SDs for the NP the answer in (4.48) and for the verb expected in (4.50) include fewer nodes if \( \tilde{S} \) and S are not constructed over them. Frazier's Most Recent Filler Strategy involves smaller Movement Domains, as these were defined in (2.27) of ch. 2.3. In all cases, therefore, we
have a single generalization, not three separate statements. We also have an explanation in terms of a highly general simplicity criterion defined on nodes within a tree.

Notice finally that Frazier's structural preferences for temporary ambiguity resolution have also been found to be operative in a language of the opposite type from English, namely Japanese. In ch. 3.3 and 3.4 I discussed the Japanese sentence (3.9b), repeated here as (4.51):

(4.51) \[ S_2[S_1[\text{Kinoo John-ga kekkonsi-ta}] \text{ to} \ N_P[\text{Mary-ga}] \text{ vp[it-ta]]} \]

"yesterday John got-married that Mary said," i.e.

Mary said that John got married yesterday.

The adverb \textit{kinoo} is preferably attached to the first verb encountered \textit{kek-konsi-ta}, rather than to \textit{it-ta}. This is predicted by Minimal Attachment, and in our terms by the fact that the matrix \( S_2 \) is not available as an attachment site at the time that the first verb is encountered in the bottom-up parse. Hence attachment to \( S_1 \). This reduces the SD for \textit{kinoo} at the time that it is attached, by limiting the dominating nodes to \( S_1 \) (subsequently the \( S \) and \( S_2 \) will have to be included when they are constructed, but at that point the parsing of \textit{kinoo} will have been completed). It also improves the IC-to-word ratios for the sentence as a whole. If \textit{kinoo} is attached high to \( S_2 \), this will give a very long CRD for \( S_2 \). If it is attached low to \( S_1 \), the ratios for both \( S_1 \) and \( S_2 \) can be optimal, in conjunction with a bottom-up parse. Hence the SDs of this sentence are systematically improved by attaching \textit{kinoo} to \( S_1 \). Empirical support for this preference has been given by Ueda (1984, cited in Frazier and Rayner 1988) and by Inoue (1991).
5 Testing EIC's grammatical predictions

In ch. 3.6 I formulated three sets of predictions made by EIC for grammatical conventions involving linear order: predictions for basic orders (3.6.1); for grammatical rearrangements of basic orders (3.6.2); and for implicational hierarchies of linear orderings (3.6.3). These predictions are derived from the assumption that there is a precise correlation between the data of language performance and the grammatical conventions of the world's languages. The explanation that has been proposed for this correlation is processing complexity. In the area of linear ordering, the metric of complexity is that of EIC. More generally, I argued in ch. 3 that the notion of a Constituent Recognition Domain (CRD), upon which the linear ordering metric is defined, is just one instance of a Structural Domain (SD) of structurally related nodes, whose relative size and complexity determines the applicability of numerous grammatical operations, such as relativization and extraction, pronoun retention, and so on. Whenever there are differences in structural complexity between different points in a syntactic tree, or between alternative structural renderings of one and the same semantic proposition, I claimed that we should see a “performance-grammar correlation” (cf. ch. 2.3): grammars will conventionalize the easier structures before they conventionalize the more difficult ones (hence the cut-off points on various hierarchies), assign more surface coding to arguments in complex structural positions than to those in simpler positions (in order to make semantic and syntactic interpretation easier in performance), and grammaticalize those linear orderings that result in faster and more efficient structural recognition in performance. This is the basic claim that is to be put to the test in this chapter. Is there a precise correspondence between the performance data from several language types that we have just examined and grammaticalized ordering conventions? Do the unmarked and marked case predictions of EIC hold for grammars as well? I shall argue in what follows that they do.

I begin by testing EIC's predictions for basic orders, where the grammaticalization appears to be sensitive to the aggregate weights of different
syntactic categories and has positioned them accordingly. In order to have a clear test of EIC, we must therefore concentrate on categories whose relative syntactic weights are not in doubt (cf. the discussion in ch. 3.6.1). This will not be an issue in our predictions for binary branching structures, since these will generally involve a non-phrasal constructing category and a phrasal node, but it does become critical for multiple branching structures. I repeat EIC's Basic Order Prediction for convenience:

(3.21)  **EIC Basic Order Prediction**

The basic orders assigned to the ICs of phrasal categories by grammatical rules or principles will be those that have the most optimal ratios for ICs of aggregate length, in the unmarked case; any basic orders whose EIC ratios are not optimal will be more or equally frequent across languages in direct proportion to their EIC ratios.

It will be apparent that (3.21) makes essentially the same frequency prediction for grammars that the EIC Performance Ordering Prediction (3.17) and the Text Frequency Prediction (3.18) made for performance: the most optimal orders will be found in the unmarked case; any attested non-optimal orders should still have relatively good EIC ratios, more precisely they should be more or equally frequent across languages in proportion to these ratios.

### 5.1 Binary branching structures

#### 5.1.1 EIC's predictions

Consider a plausible tree structure for the phrase \((X) \text{ has played with friends of John} \) in a language such as English:

(5.1)

```
S
  /\  /
Aux VP
  /\  /
has V PP
  /\  /
played P NP
  /\  /
with N PossP
  /\  /
friends Poss NP
  /\  /
of Name
  /
John
```
Each dominating node contains two ICs, one a category that is a mother-node-constructing category (MNCC) in our terms, the other a phrasal IC that is not unique to its mother. Thus, *played* is of category V, which constructs VP, and PP is its non-constructing sister (attached by Immediate Constituent Attachment (3.4), cf. ch. 3.2). All the phrasal nodes are on the right of their MNCC sisters, and so (5.1) is a right-branching structure.

It is also a highly optimal structure from the perspective of EIC. Each MNCC constructs its mother on the basis of a single word, and the very next word constructs the second IC of the CRD, all the way down the tree. Thus, the CRD for S consists of just two words, S(hasAux, playedVP), that of VP consists of just VP(playedV, withPP), PP consists of PP(withP, friendsNP), and so on. The IC-to-word ratios here are consistently optimal, 2/2 = 100%. In terms of IC-to-non-IC ratios, they are also maximally optimal: each CRD has a 2/3 = 67% ratio, which is the highest it can be.

In terms of the typology of structural types enumerated in table 3.2, all the phrases of (5.1) are of the type [CmIC]. I pointed out in ch. 3.6.1 that the mirror-image of this can also be optimal, i.e. [ICmC]. Applied to (5.1), this would result in the reverse ordering for all these binary branching phrases, as shown in (5.2):

(5.2)

Example (5.2) is a left-branching structure, and the ordering of elements is exactly that found in languages such as Japanese. The tree is parsed bottom-up, and each CRD consists of just two words, as before: PossP consists of PossP(JohnNP, ofPoss), NP of NP(ofPossP, friendsN), PP of PP(friendsNP, withP),
and so on. The IC-to-word ratios are consistently $2/2 = 100\%$, and IC-to-non-IC ratios are $67\%$.

It has been known for some time that the great majority of the world's languages pattern either like (5.1) or like (5.2), and hence that there are strong correlations between the positioning of an MNCC in one phrasal node and its positioning in another. Greenberg (1966) was the first to formulate a set of implicational universals between these MNCC orders; these were reformulated in W. P. Lehmann (1978) and Vennemann (1974); they were reformulated again in Hawkins (1983) using an expanded Greenbergian language sample; and most recently they have been tested and reformulated yet again, using a much expanded database, in Dryer (1992). Various grammatical generalizations have been appealed to when describing the grammatical facts. Greenberg used traditional category labels and formulated a separate implicational statement for each pair of ordered categories that were strongly correlated in his data (e.g. if (S)OV, then postpositions (in PP), etc.). Vennemann attempted a more abstract generalization in terms of "operator" and "operand" categories, the latter being very similar to the "heads of phrases" of other grammatical models, and proposed a consistency principle for all operator-operand pairs (called the Natural Serialization Principle). The head-of-phrase principle figured prominently in the grammatical generalizations of Hawkins (1983), both in accounting for implicational patterns, and in the statement of distributional regularities, for which a Jackendoff (1977)-style format of X-bar rules was argued to be particularly appropriate. More recently, ordering conventions have been described not just in terms of the head or non-head status of the categories in examples (5.1)–(5.2), but in terms of the direction in which case and $\theta$-roles are assigned by a head (cf. Travis 1984, 1989; Koopman 1984).

I argued in ch. 3.8.1 that grammatical generalizations in these terms may be descriptively useful in certain grammars, but they are not explanatory. They have numerous counterexamples and allow for many word-order combinations that are non-occurring across languages (cf. e.g. section 5.2). They often fail to discriminate between those combinations that are highly frequent across languages, and those that are non-productive or non-existent. And adding an element of direction to an intrinsically structure-defining process, such as a constituency-defining PS rule, or case assignment from a head, is a technical device with very little explanatory motivation. EIC, by contrast, follows from a general theory of structural complexity, and is supported by data from language performance.
In what follows I shall examine the most up-to-date cross-linguistic evidence we have for the word-order patterns of (5.1)–(5.2), in relation to EIC's predictions. There are, of course, not just two logical possibilities for each pair of sisters, but four: the optimal ones, \([C_m IC]\) and \([IC_m C]\); and two non-optimal ones, \([C IC_m]\) and \([m IC C]\). EIC predicts, first of all, that the two optimal orders should be the ones that grammars generate in their rule outputs in the unmarked case. In other words, the combined total of the grammars generating these outputs should account for at least the same kind of proportion, within the set of all grammars, as did the combined total of all orders that were most optimal for EIC within the set of all attested instances of a given structure in performance. If there is a correlation between performance and grammar, whereby grammars conventionalize rule outputs in accordance with their frequency of occurrence in performance, that is what we should expect.

I have not formulated an explicit prediction for the frequency of the two optimal orders relative to each other. All things being equal, we expect a random distribution in response to equally good EIC ratios, and indeed we shall see in the data below that the ratio between them is very often close to 50–50. I have no guarantee, however, that all other things will be equal, especially when there are more than two optimal orders in the multiple branching structures of the next section.

With regard to the marked case predictions, the EIC ratio for the non-optimal \([C IC_m]\) turns out to be higher than that for the corresponding non-optimal \([m IC C]\). Within the set of minority languages that are predicted to be infrequent by the unmarked-case prediction, we therefore expect a preference for \([C IC_m]\) over \([m IC C]\).

To see why EIC prefers \([C IC_m]\) to \([m IC C]\), compare the CRD for VP in (5.3) with that in (5.4):

\[
\begin{align*}
(5.3) & \quad \text{VP} \\
& \quad V \\
& \quad \quad \text{played} \\
& \quad \quad \quad \text{NP} \\
& \quad \quad \quad \quad \text{Det} \\
& \quad \quad \quad \quad \quad \text{the} \\
& \quad \quad \quad \quad \quad \text{friends} \\
& \quad \quad \quad \quad \quad \text{with} \\
& \quad \quad \quad \quad \quad \text{P} \\
& \quad \quad \quad \quad \quad \text{P} \\
& \quad \quad \quad \quad \quad \text{PP} \\
& \quad \quad \quad \quad \quad \text{V} \\
(5.4) & \quad \text{VP} \\
& \quad PP \\
& \quad \quad \text{P} \\
& \quad \quad \quad \text{N} \\
& \quad \quad \quad \quad \text{Det} \\
& \quad \quad \quad \quad \quad \text{the} \\
& \quad \quad \quad \quad \quad \text{friends} \\
& \quad \quad \quad \quad \quad \text{with} \\
& \quad \quad \quad \quad \quad \text{NP} \\
& \quad \quad \quad \quad \quad \text{V} \\
\end{align*}
\]
The CRD for each comprises all the words and non-ICs dominated by VP, which gives much lower ratios than in (5.1) and (5.2): 2/4 = 50% for IC-to-words; 2/8 = 25% for IC-to-non-ICs. If we make the calculation left to right, however (cf. (3.15)), (5.3) is much preferred because the first IC of the domain, V, is constructed early by a single word, whereas the first IC of (5.4) is a whole phrase, PP. Example (5.3) has the L-to-R IC-to-word and IC-to-non-IC ratios for the VP shown in (5.3'), and (5.4) has those in (5.4'):

\[(5.3') \quad \text{L-to-R IC-to-word:} \quad \frac{1}{1} \quad \frac{2}{4} \quad 100\% \quad 50\% \quad \text{Aggregate} = 75\% \]

\[\quad \text{L-to-R IC-to-non-IC:} \quad \frac{1}{1} \quad \frac{2}{7} \quad 100\% \quad 29\% \quad \text{Aggregate} = 65\% \]

\[(5.4') \quad \text{L-to-R IC-to-word:} \quad \frac{1}{3} \quad \frac{2}{4} \quad 33\% \quad 50\% \quad \text{Aggregate} = 42\% \]

\[\quad \text{L-to-R IC-to-non-IC:} \quad \frac{1}{7} \quad \frac{2}{8} \quad 14\% \quad 25\% \quad \text{Aggregate} = 20\% \]

5.1.2 Testing EIC's predictions

Table 5.1 sets out data for the cross-categorial positioning of a verb and direct object within VP and an adposition (i.e. preposition or postposition) and NP within PP. The data are drawn from two samples: my own expansion of Greenberg's sample in Hawkins (1983) and abbreviated ES (for Expanded Sample); and Matthew Dryer's database, reported in Dryer (1992) and other publications. Dryer's quantities refer to what he calls "genera" rather than individual languages, i.e. groups of languages with a (presumed) time depth corresponding to the subfamilies of Indo-European and no greater than 4000 years (cf. Dryer 1989a). For the data of table 5.1 he kindly supplied me (in June 1991) with individual language quantities as well. The languages listed are exclusively those that have free-standing adpositions, i.e. separate words, rather than affixes. The data of ES do not make this distinction. Also, the great majority of ES's entries refer to individual languages, but there are a number of entries for whole language families, when the word-order properties of the members are similar.

The data of table 5.1 have been compiled from the statistics available on the correlation between verb position in relation to a direct object (VO or OV) and adposition order within PP ([P NP] or [NP P]). This compilation therefore makes the assumption that a PP will occupy the same position relative to V as a direct object NP, and then quantifies the distribution of
the constructing category P within PP. This assumption is generally valid, but not exceptionlessly so. Dryer (1992: 92) gives figures showing that 122/132 of his genera (92.4%) have identical positioning for PP and a DO. For languages with VO (SVO, VSO, and VOS) the correlation is almost perfect, 59/60 (98.3%), i.e. VO co-occurs with V-PP. For languages with OV, however, the correlation is 63/72 (87.5%). What this means is that there are some OV languages with V-PP orders; hence the order type labeled B in table 5.1, [[NP P] V], should have slightly fewer members, and either A [V [P NP]] or C [V [NP P]] should have slightly more (unfortunately Dryer 1992 does not give the precise adposition order within these OV and V-PP structures). If A has slightly more, this will put the two optimal orders, A and B, closer to 50–50. If C has slightly more, this will not affect EIC's marked case prediction, which predicts that C will have more or equal languages compared with the other non-optimal type D, [[P NP] V] – which it does anyway. The overall number of mismatches (7.6% of Dryer's genera) is not large enough to make a significant difference to the unmarked case prediction, even if most were to go to the non-optimal type C.

In all three samples of table 5.1, EIC's unmarked and marked case predictions are well supported. The most reliable sample is probably Dryer's set of 389 languages, all of which involve non-affixed adpositions only, and here the success rate for the unmarked case prediction is 94%, while structure C outnumbers D by three to one in the marked orders, as predicted by EIC ratios measured left to right. The ratio of optimal to non-optimal EIC orders in these samples is strikingly similar to what we saw in performance in the last chapter. Clearly, if these figures for grammars are typical, there will indeed be a strong correlation between performance and grammar.

Notice that both my own and Dryer's samples make the assumption that adposition and verb have been grammaticalized in the relevant order in all the languages of these samples. For these binary branching structures, this assumption is fairly safe. The position of the adposition within PP is among the most strongly grammaticalized of any orders (cf. Hawkins 1983). In the vast majority of languages, only one adposition order occurs and the other is completely ungrammatical. Even when there is some "doubling," as in German which has both prepositions (in das Haus "into the house") and postpositions (dem Haus gegenüber "the house opposite"), one order is often clearly basic and productive (in German, prepositions). Verb position is more variable, but again many languages permit only one order of V and O, and when both occur grammatical arguments can often motivate one as basic and the other as derived. Performance data can also provide evidence
Table 5.1 Grammaticalized $VP\{VP\{NP\}\}$ orders

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>136</td>
<td>40.5%</td>
<td>70</td>
<td>35.7%</td>
<td>161</td>
<td>41.4%</td>
</tr>
<tr>
<td>B</td>
<td>162</td>
<td>48.2%</td>
<td>107</td>
<td>54.6%</td>
<td>204</td>
<td>52.4%</td>
</tr>
<tr>
<td>C</td>
<td>26</td>
<td>7.7%</td>
<td>12</td>
<td>6.1%</td>
<td>18</td>
<td>4.6%</td>
</tr>
<tr>
<td>D</td>
<td>12</td>
<td>3.6%</td>
<td>7</td>
<td>3.6%</td>
<td>6</td>
<td>1.5%</td>
</tr>
</tbody>
</table>

Totals: 336 196 389

EIC predictions

Unmarked case: A + B = most optimal

- ES: 298/336 = most optimal, 88.7% Correct
- Dryer's genera: 177/196, i.e. 90.3% Correct
- Dryer's languages: 365/389, i.e. 93.8% Correct

Marked case: C ≥ D

- ES: 7.7% > 3.6% Correct
- Dryer's genera: 6.1% > 3.6% Correct
- Dryer's languages: 4.6% > 1.5% Correct
for grammaticalization, as we showed in the last chapter, by retaining a basic order despite weak (or equal) contrary preferences defined by EIC; and the general conclusion we were able to draw from these data was that there is more grammaticalization of linear ordering in the world’s languages than has been assumed hitherto, rather than less. Any errors that have been made in this regard when classifying languages have probably been in the direction of assuming free order when there is some grammaticalization, rather than the reverse. This means that the reliability of these samples as reflections of basic-order grammaticalizations is probably quite high, especially in Dryer’s (1992) sample, which has done most to refine the original Greenbergian sample and has incorporated the most recent grammatical analyses.

There is another consideration that is relevant for binary branching structures: one of the daughters is always an MNCC, and the positioning of constructing categories appears to be highly regular and consistent cross-categorially, even in those languages that appear to be inconsistent. Thus, I shall argue in ch. 6.3.3 that when one considers the totality of devices that are available for constructing higher phrasal nodes (i.e. beyond MNC), an apparently mixed language such as German can be shown to be largely consistent, with as much left-peripheral higher node construction as English. And the sequence of historical changes that led from a German-type Old English syntax to Modern English, specifically the fronting of the verb and various morphological and syntactic changes involving case marking on NPs, can be motivated in terms of the replacement of one left-peripheral constructing category (for VP) by another. Across all languages, therefore, there is considerable grammaticalization of positioning for constructing categories, and in a binary branching structure this means that the position of the unique non-constructing sister IC is also fixed as well. In a multiple branching structure, consisting of at least two non-constructing ICs, the constructing category will generally be fixed at or near one periphery, whereas the non-constructing ICs will exhibit more variability. This is what we saw in English Heavy NP Shift structures (cf. ch. 4.3.2): NP and PP could be rearranged in response to EIC, but V is fixed on the left periphery of VP, even when NP = one word and both [V NP PP] and [NP V PP] would be equally optimal. The reason for this more consistent positioning of constructing categories, I submit, is precisely that these nodes construct the ICs of their respective domains. The order [NP V PP] could be fine for the VP itself, but it would extend the recognition domain for whatever higher phrase (most often S) contained VP as an IC, since the V would now occur later in the parse string. Only if constructing categories are consistently left-peripheral in
English-type languages, and consistently right-peripheral in Japanese-type languages, will the relevant containing CRDs be most optimal, and it is for this reason that their positioning is regularly conventionalized.

Consider now the grammaticalized data for another pair of binary branching ICs in (5.1)–(5.2), \( PP\{P \{NP\{N PossP\}\}\} \), i.e. an NP consisting of a head noun and a genitive or possessive phrase immediately embedded within a PP. The optimal orders are \( PP[P NP[N PossP]] \) and \( PP[NP[PossP N P]] \), i.e. A and B in table 5.2. A full 93% of grammars in my Expanded Sample exemplify one or the other of these types, which satisfies EIC’s unmarked case prediction. The preference for structure C with the single-word P to the left of the multi-word NP is only very slight on this occasion, but since the marked-case prediction requires only that \( C \geq D \), this prediction is satisfied as well.

The structures of (5.1)–(5.2) also contain an auxiliary verb as the highest constructing category in the tree. Dryer (1992) gives data on the ordering of different types of auxiliaries in relation to verb and direct-object position. Table 5.3 reproduces his figures for auxiliaries designating tense or aspect. Of these genera, 90% are most optimal for EIC, in accordance with the unmarked case prediction. The marked case prediction is not quite correct on this occasion, but the margin of error is so small as to be insignificant. Table 5.4 gives the figures for languages containing a negative auxiliary, for which both the unmarked and marked case predictions are correct.

Table 5.5 considers the positioning of a head noun and possessive phrase within an NP that is a direct object of V rather than a sister of P, as in table 5.2, i.e. phrases corresponding to the English met friends of John. The optimal

<table>
<thead>
<tr>
<th>Order</th>
<th>ES (1983)</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A ( [P [N PossP]] )</td>
<td>134</td>
<td>39.9%</td>
</tr>
<tr>
<td>B ( [[PossP N] P] )</td>
<td>177</td>
<td>52.7%</td>
</tr>
<tr>
<td>C ( [P [PossP N]] )</td>
<td>14</td>
<td>4.2%</td>
</tr>
<tr>
<td>D ( [[N PossP] P] )</td>
<td>11</td>
<td>3.3%</td>
</tr>
<tr>
<td>Total</td>
<td>336</td>
<td></td>
</tr>
</tbody>
</table>

**EIC predictions**

Unmarked case: \( A + B \) = most optimal \( 311/336 = 92.6\% \) Correct

Marked case: \( C \geq D \) \( 4.2\% > 3.3\% \) Correct
Testing EIC's grammatical predictions

Table 5.3 Grammaticalized $S\{Aux_{T/A} VP\{V NP\}\}$ orders

<table>
<thead>
<tr>
<th>Order</th>
<th>Dryer's 1992 genera</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A [Aux [V NP]]</td>
<td>28</td>
<td>39.4%</td>
</tr>
<tr>
<td>B [[NP V] Aux]</td>
<td>36</td>
<td>50.7%</td>
</tr>
<tr>
<td>C [Aux [NP V]]</td>
<td>3</td>
<td>4.2%</td>
</tr>
<tr>
<td>D [[V NP] Aux]</td>
<td>4</td>
<td>5.6%</td>
</tr>
<tr>
<td><strong>Total:</strong></td>
<td><strong>71</strong></td>
<td></td>
</tr>
</tbody>
</table>

**EIC predictions**

Unmarked case: A + B = most optimal  64/71 = most optimal, 90.1%

Correct

Marked case:  C ≥ D  4.2% < 5.6%

Close

Abbreviation

$Aux_{T/A}$ = auxiliary verb designating tense or aspect

Table 5.4 Grammaticalized $S\{Aux_{Neg} VP\{V NP\}\}$ orders

<table>
<thead>
<tr>
<th>Order</th>
<th>Dryer's 1992 genera</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A [Aux [V NP]]</td>
<td>13</td>
<td>52%</td>
</tr>
<tr>
<td>B [[NP V] Aux]</td>
<td>8</td>
<td>32%</td>
</tr>
<tr>
<td>C [Aux [NP V]]</td>
<td>3</td>
<td>12%</td>
</tr>
<tr>
<td>D [[V NP] Aux]</td>
<td>1</td>
<td>4%</td>
</tr>
<tr>
<td><strong>Total:</strong></td>
<td><strong>25</strong></td>
<td></td>
</tr>
</tbody>
</table>

**EIC predictions**

Unmarked case: A + B = most optimal  21/25 = most optimal, 84%

Correct

Marked case:  C ≥ D  12% > 4%

Correct

Abbreviation

$Aux_{Neg}$ = negative auxiliary verb
Table 5.5 Grammaticalized $V_P\{V\ N_P\{N \ PossP\}\}$ orders

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A $[V\ [N\ PossP]]$</td>
<td>124</td>
<td>36.9%</td>
<td>63</td>
<td>29.0%</td>
</tr>
<tr>
<td>B $[[PossP\ N]\ V]$</td>
<td>153</td>
<td>45.5%</td>
<td>112</td>
<td>51.6%</td>
</tr>
<tr>
<td>C $[V\ [PossP\ N]]$</td>
<td>38</td>
<td>11.3%</td>
<td>30</td>
<td>13.8%</td>
</tr>
<tr>
<td>D $[[N\ PossP]\ V]$</td>
<td>21</td>
<td>6.3%</td>
<td>12</td>
<td>5.5%</td>
</tr>
</tbody>
</table>

Totals: 336 217

**EIC predictions**

Unmarked case: $A + B$ = most optimal

- **ES**: 277/336 = most optimal, 82.4%, Correct
- **Dryer's genera**: 175/217, i.e. 80.6%, Correct

Marked case: $C \geq D$

- **ES**: 11.3% > 6.3%, Correct
- **Dryer's genera**: 13.8% > 5.5%, Correct
orders are attested in 82% and 81% of the two samples, and among the non-optimal orders there is a clear marked-case preference for the EIC-preferred structure C, which corresponds to the English alternative *met John's friends*.

Finally, in table 5.6 I reproduce Dryer’s data for the positioning of a higher verb such as “want” and a VP complement consisting of a verb and direct object. Of his genera, 84% are most optimal, and there is a clear skewing in favor of structure C among the non-optimal orders.

Tables 5.1–5.6 present the available cross-linguistic data for six different pairs of binary branching structures, each of which is of the form \{C ICi{C ICj}\}, i.e. one is immediately embedded in the other. The phrases corresponding to ICi are either NP, PP, or VP; those corresponding to ICj are either NP or PossP. In all of these cases EIC’s predictions are well supported and the proposed correlation between performance and grammar is confirmed. In no case is either of the phrasal nodes, ICi or ICj, sentential, however. For example, if a sentential complement were to occur in the position of a direct object as a sister of V, then ICi would be an S, and ICj would be S: vP{V s{Comp S}}. If a relative clause were to occur in the position of PossP in table 5.5, then ICj would be an S: vP{V NP{N S}}. Under these circumstances the testing of EIC’s predictions for the grammar becomes more complicated, because these sentential nodes are regularly subject to rearrangement, by rules such as Extrapolation, Extrapolation from NP, and Heavy NP Shift, often in the majority of structures and performance instances to which they can potentially apply, as we saw in ch. 4.3. This means that we cannot simply examine basic orders alone, as we have done for nodes of smaller aggregate

<table>
<thead>
<tr>
<th>Order</th>
<th>Dryer’s 1992 genera</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A [want [V NP]]</td>
<td>42</td>
<td>49.4%</td>
</tr>
<tr>
<td>B [[NP V] want]</td>
<td>29</td>
<td>34.1%</td>
</tr>
<tr>
<td>C [want [NP V]]</td>
<td>10</td>
<td>11.8%</td>
</tr>
<tr>
<td>D [[V NP] want]</td>
<td>4</td>
<td>4.7%</td>
</tr>
<tr>
<td>Total:</td>
<td>85</td>
<td></td>
</tr>
</tbody>
</table>

**EIC predictions**

Unmarked case: A + B = most optimal 71/85 = most optimal, 83.5% Correct

Marked case: C ≥ D 11.8% > 4.7% Correct
length, like NP, PossP, and PP, whose rearrangement is less frequent. We must now consider both the basic orders and the effects of rearrangement, for these unusually large and complex categories.

I shall be considering EIC's rearrangement predictions in greater detail in section 5.4. Notice in this context what EIC predicts for the binary branching structures with sentential nodes that I have just defined. In (5.5) I set out the four logically possible orderings for VP{V S{Comp S}}:

(5.5)  
A  [V [Comp S]]  
B  [[S Comp] V]  
C  [V [S Comp]]  
D  [[Comp S] V]

The structure in question is that of [knows [that she will win]] in English. A and B (the English type and the Japanese type respectively) are again the optimal orders for EIC. But because of the inherent complexity of S, the EIC ratios for the non-optimal C and D will be worse than they are for all the corresponding C and D structures in tables 5.1–5.6, where only an NP or a PossP intervened between the MNCC for the higher phrase and that for the lower phrase, i.e. between C_1 and C_j in \{C_i 1\leq i\leq j C_j\}. EIC accordingly predicts that the proportion of languages with A + B should now be higher than it is in the other cases, and that C + D should be exceedingly rare.

The available data seem, at first, to confirm this in a quite striking manner. Consider the English-type languages with an initial V in VP. I pointed out in Hawkins (1990) that verb-initial languages are exceptionlessly Comp-initial in S, i.e. structure C is unattested, a fact which Dryer (1992) has corroborated in his data. Verb-final languages, on the other hand, exemplify both final complementizers as in Japanese, and initial complementizers as in Persian. The precise productivity of the latter has not yet been established in the available samples, but if the clearly related category of what Dryer (1992) calls "adverbial subordinators" (i.e. words corresponding to when and although in English) is any indication, the proportion of the Persian type to the Japanese type may be as high as 17/55 (i.e. 31%). This makes it appear as if structure D is productive, in contrast to C; but appearances are deceptive, because grammars that would potentially generate D invariably seem to have an extraposition rule converting D into A (or alternatively D is not generated as a basic order to begin with). This is true for Persian and for German. It is also true for the finite S structures of Yaqui and Turkish (cf. Dryer 1980). Moreover, in all the languages mentioned, Extraposition is obligatory in this environment, with the result that these languages exhibit a "left–right
Testing EIC's grammatical predictions

asymmetry” (cf. section 5.6): a rightward skewing for sentential direct objects, even in languages that are SOV for non-sentential objects (i.e. when the CRD for VP is not strained). Languages with basic SVO order, by contrast, can maintain this order with optimal EIC ratios in both cases.

When the effects of grammatical rearrangements are taken into account, therefore, the available evidence suggests that A and B are indeed the only two productive orders, as originally predicted. A may be a basic order only, or the result of a transformation; B is a basic order only. What still needs to be explained is the precise basis for the left–right asymmetry: if languages with basic SOV (for non-sentential objects) can exhibit an SVO alternant when the object is sentential and of the type [Comp $S$], why can’t SVO languages have an SOV alternant for sentential objects of the type [$S$ Comp]? In section 5.6.1 I shall suggest an answer to this.

Consider finally the four logically possible orderings for $v_p\{V_{NP}\{N \tilde{S}\}\}$, i.e. relative-clause constructions in relation to a higher verb.

(5.6) A $[V [N \tilde{S}]]$
B $[[\tilde{S} N] V]$
C $[V [\tilde{S} N]]$
D $[[N \tilde{S}] V]$

Verb-initial languages display an almost exceptionless preference for A over C, 60/61 in Dryer’s (1992) data (98.4%). Verb-final languages are again mixed, with postnominal relatives actually outnumbering their prenominal counterparts by 37/63 (i.e. 58.7%). But again this does not mean that structure D is fully productive, because of the existence of Heavy NP Shift and Extraposition from NP rules in these languages, which quite regularly transform D into A or into $[NP[N V_{NP}[\tilde{S}]]]$. This latter (Extraposition from NP) was a conversion that we examined in detail for German in ch. 4.3.5, where we found that in this particular environment the great majority of NPs were predicted to undergo it, and did undergo it, thereby removing structures of type D from the parse string. The difference between (5.6D) and (5.5D), however, is that a relative clause is ultimately NP-dominated and hence subject to positioning rules for NP, whereas an $\tilde{S}$ under the immediate domination of VP can be subject to different positioning rules. Thus, one could imagine a grammar making a discrimination between NP and $\tilde{S}$, positioning the former before the verb (SOV) and the latter after (SVO) (which would dispense with the need for an extraposition rule), whereas a preverbal NP is automatically going to contain all the daughter categories that NP can dominate, including relative clauses, and hence we would not expect NPs with
relative clauses to be ordered differently from those without by basic-order rules. A rearrangement is predicted in these cases, and Extraposition from NP is particularly attractive from the perspective of EIC, since it maintains the original NP positioning while removing the major source of the complexity (S). We can expect, therefore, that structures of type (5.6D) will be more productive than those of type (5.5D), and this is the case (cf. section 5.6).

Again the absence of type (5.6C) in otherwise verb-initial languages needs to be explained, and I shall return to this in section 5.6.1. Notice also that within the relative clause itself, the category that constructs the S, which may be a complementizer or subordinating affix of some kind (cf. sections 5.2.1 and 6.3.2), is positioned in accordance with EIC's unmarked and marked case predictions. Abbreviating this category as C, we find productive instances of languages with the type A structure [N [C S]] (English and Persian) and with type B [[S C] N] (Lahu and Basque). There is a handful of type C languages with [N [S C]] (e.g. Dyirbal), but no languages at all of type D [[C S] N], i.e. with an initial complementizer or other relative clause subordination indicator within a prenominal relative clause. As a result, the type A structure is extremely frequent since it comprises all head-initial languages as well as a number of otherwise head-final languages. Type B is well-attested, but only among strongly head-final languages, and many languages that could have been of this type have gone over to the A structure, or to C (maintaining their final MNCC). D is unattested, and hence the distribution of attested languages now reverts to the familiar pattern: A + B are productive; C is a minor type; and D does not occur at all. In section 5.6.1 I return to this distribution and offer an explanation for the complete absence of D.

Summarizing, there is a productive left–right asymmetry involving head-noun positioning with relative clauses, whereby both prenominal and post-nominal relative clauses are found in languages that are otherwise head-final, whereas head-initial languages are entirely postnominal. There is also an asymmetry within the relative clause itself: a right-peripheral C can occur with both final and initial head nouns, but a left-peripheral C can occur only with initial head nouns.

5.1.3 Frazier's head adjacency explanation

Frazier (1979b, 1985) has also proposed a processing explanation for the ordering of binary branching sisters, and her alternative motivates the
Testing EIC's grammatical predictions

adjacency of heads of phrases down trees such as (5.1)–(5.2). Heads of phrases are regularly MNCCs in our terms (cf. ch. 6.1). Frazier assumes the two-stage parsing model of Frazier and Fodor (1978) in which the first stage has a limited viewing window of five or six adjacent words, and she reasons that on-line syntactic and semantic processing could not take place in structures such as (5.7)–(5.8), since the number of words that would have to be parsed in order to identify phrasal packages forming coherent semantic units would regularly exceed this word limit:

(5.7) $\text{VP[NP[Head Noun Rel. Clause] Verb]}$

\begin{align*}
Y & \text{ Complement } X \\
of Y
\end{align*}

(5.8) $\text{PP[NP[Head Noun Rel. Clause] Postposition]}$

\begin{align*}
Y & \text{ Complement } X \\
of Y
\end{align*}

In these configurations, $X$, the head of one phrase, and $Y$, the head of the other, are separated by a sister (here a complement) of $Y$.

Frazier's explanation makes the assumption of absolute unprocessability, as this was discussed in ch. 1.3. She is, in essence, proposing an absolute limit on the human processing architecture, and structures that go beyond this by regularly requiring access to more than five or six adjacent words in the parser's first stage should be unattested.

But structures such as (5.7)–(5.8) are not unattested. The ordering of (5.7) occurs in all the structures of German that have not undergone Extraposition from NP (cf. ch. 4.3.5), for example (5.9) taken from my German data:

(5.9) Wie soll ich also $\text{VP[NP}[deinem Vater $\text{g[der wochenlang im Ausland war]] erklären}]$ daß . . .

"How shall I therefore to-your father who for-weeks abroad was explain that . . .," i.e. How shall I therefore explain to your father who was abroad for weeks that . . .

The ordering of (5.8) is found in Lakhota (example from C. Lehmann 1984: 82):

(5.10) $\text{PP[NP[xé wä $\text{g[Tamalpais éciya-pi] wä] él]}}$

"mountain INDEF Tamalpais they-called INDEF on," i.e. on a mountain that they called Tamalpais
Both of these structures are clearly straining processing resources. The great majority of structures corresponding to (5.7)/(5.9) undergo Extraposition from NP, as I showed in table 4.29 (section II), thereby making the head noun adjacent to the verb. And Lehmann (1984) points out that structure (5.8)/(5.10) is avoided in many languages in which it could potentially occur, by extraposing the Š to the right of the postposition (cf. German in this regard *dem Haus gegenüber, in dem er wohnt* "the house opposite in which he lives", i.e. opposite the house in which he lives). Nonetheless, (5.9) is clearly grammatical in German, and (5.10) is clearly grammatical in Lakhota. The Lakhota structure is also attested in languages in which the postposition is a suffix and is attached to the last word of the postnominal relative clause, as in Galla (cf. Lehmann 1984: 255).

Thus we are dealing here with a dispreference for structures (5.7)–(5.8), not with an absolute prohibition, and this dispreference is exactly what EIC predicts: the CRD for VP and for PP proceeds all the way from the head noun to the verb and postposition respectively, and this results in very low EIC ratios when there is a long intervening constituent such as Š. But since the length of the CRD can evidently exceed Frazier's five-to-six-word limit in some languages, then either this proposed limit on the first stage of her parser needs to be extended, or else structures of this sort are processed in both the first and second stages of the parser's operations. Either way, the proposed absolute limit on language structure is not correct.

There is a second set of problems for Frazier's explanation. Even though structures (5.7)–(5.8) exist, they are clearly dispreferred, and this dispreference might be taken as evidence for the fact that a five-to-six-word viewing window is a real constraint on processing in some form or another within a suitably reformulated model. It is not clear to me how inbuilt restrictions such as this could make predictions for preferences rather than absolute prohibitions, but let us assume for the sake of argument that there might still be some reality to a viewing window of, say, five to six words. The performance evidence involving Extraposition from NP in German, however, presents problems even for this weaker account. For I showed in ch. 4.3.5 that if the VP in structures such as (5.7) contains an additional constituent XP intervening between the complex NP and the verb, then EIC predicts that extraposition should not in general apply, and that the longer VP structure with the relative clause intervening between the head noun and the verb should be maintained. And in table 4.29 I demonstrated that the unextraposed NPs are exactly those in which there is at least one additional XP following NP and preceding V. Hence, there is no preference for a viewing
window of less than five-to-six words in these cases: longer viewing windows for VP are actually preferred over shorter ones, and the longer viewing windows will regularly exceed the five-to-six-word limit! Frazier's theory, of course, makes the opposite prediction: the longer the phrasal package is, the more likely it is that extraposition should apply, in order to avoid exceeding the proposed limit. By defining structural complexity in terms of the efficiency ratios for all the phrasal domains of a sentence, as EIC does, it can be shown that the potential benefits for VP resulting from Extrapolation from NP can be more than offset by the deterioration in NP scores when discontinuous material intervenes between the head noun and S. A theory that simply proposes an absolute word limit on processing is insufficiently sensitive to structure, and does not capture the fact that longer viewing windows for phrasal packages such as VP may often be preferred over shorter ones. The preferences will even exceed any proposed absolute limit in terms of fixed numbers of words, given appropriate assignments of words to constituents in structures of this type. All of these considerations argue against absolute limits on viewing windows when processing phrasal packages that form coherent syntactic and semantic units. There are relative degrees of structural complexity and of preference, and these seem to be of the type that EIC defines.

There is another fact which is unaccounted for in Frazier's theory. The mirror-image structures to (5.7) and (5.8), combining verb-initial VPs and prepositions with prenominal relatives, are exceedingly rare:

\[(5.7') \text{ Vp}[\text{Verb NP}[\text{Rel. Clause Head Noun}]]\]
\[\text{X Complement Y}\]
\[\text{of Y}\]

\[(5.8') \text{ PP}[\text{Preposition NP}[\text{Rel. Clause Head Noun}]]\]
\[\text{X Complement Y}\]
\[\text{of Y}\]

Mandarin Chinese is the only language known to me that provides solid evidence for these structures, and their almost total absence appears at first sight to support Frazier's theory. But both (5.7) and (5.8) are well attested, as we have seen, which argues against Frazier's theory. And the discrepancy between (5.7)--(5.8) and (5.7')--(5.8') poses the question of why there should be a difference between them. As far as Frazier is concerned, both options should be non-existent. An alternative proposal will be made in section 5.6.1, which motivates the unproductivity of (5.7')--(5.8').
5.2 Multiple branching structures

In this section I shall consider two sets of predictions made by EIC for multiple branching structures, one involving daughters of the NP, the other involving the VP. The phrases to be tested will consist of an MNCC (N and V respectively) and two sister ICs. Depending on the internal structure of these latter (i.e. on whether they are of the type mIC or ICm, cf. table 3.2), EIC defines different ordering preferences in performance, which were tested in ch. 4. If there has been grammaticalization of a basic order, EIC predicts optimal ratios for categories of aggregate length in that order. Because of the possibility of cross-linguistic differences in length aggregates (discussed in ch. 3.6.1), the testing of multiple branching predictions needs to focus on those categories whose relative length differences are clearest.

5.2.1 \{N Adj Š\}

Consider an illustrative prediction for the NP domain involving the relative ordering of a head noun, a single-word adjective, and a relative clause (Š). This pair of sister ICs to N is particularly appropriate in this context: if the length of one item (the adjective) is kept constant, there will be a significant length difference between Adj and Š, whatever the precise aggregate for the latter is in a given language. Also, adjectives and relative clauses perform a very similar semantic function as (restrictive or appositive) modifiers of the head noun, and in grammatical models with rich phrase-internal constituent structure they can occupy the same (bar-)level in the hierarchic structure of the NP (cf. e.g. Jackendoff 1977). All of these considerations serve to reduce confounding factors involving constituent structure and language-particular differences. In what follows I shall assume a flat structure for at least that portion of the NP domain which contains these three categories.

Consider first what EIC predicts when Adj and Š are both on the same side of the head noun, as shown in table 5.7.

The orders with perfect IC-to-word ratios are (a) and (c), where only three words are required for the recognition of three ICs in the CRD for NP. In (b) and (d), on the other hand, the Š is center-embedded between the two single-word ICs, and as a result all six words need to be scanned in order to recognize the three ICs of NP, making a ratio of 3/6 = 50%. It is significant that in languages in which the head noun is initial, the basic order of single-word adjectives and relative clauses is always that of (a) and never that of (b), according to C. Lehmann (1984: 201). Noun-final languages permit more
Table 5.7 Grammaticalized $NP\{N \ Adj \ S\{C \ S\}\}$ orders with Adj and $S$ on the same side of $N$

Assume: $N = 1$ word, $Adj = 1$, $C = 1$, $S = 3$, $S = 4$

N, C = constructing categories (for NP and $S$ respectively)

<table>
<thead>
<tr>
<th>Structure</th>
<th>Example</th>
<th>IC-to-word ratio (NP CRD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) $[N \ Adj \ [C \ S]]$</td>
<td>$NP[movies \ good \ [that \ Bill \ will \ see]]$</td>
<td>$3/3 = 100%$</td>
</tr>
<tr>
<td>(b) $[N \ [C \ S] \ Adj]$</td>
<td>$NP[movies \ [that \ Bill \ will \ see] \ good]$</td>
<td>$3/6 = 50%$</td>
</tr>
<tr>
<td>(c) $[[S \ C] \ Adj \ N]$</td>
<td>$NP[S[Bill \ see \ will \ that] \ good \ movies]$</td>
<td>$3/3 = 100%$</td>
</tr>
<tr>
<td>(d) $[Adj \ [S \ C] \ N]$</td>
<td>$NP[good \ S[Bill \ see \ will \ that] \ movies]$</td>
<td>$3/6 = 50%$</td>
</tr>
</tbody>
</table>

variation, but again the predicted (c) is cross-linguistically preferred. There appears to be no language with (d) as the basic or unmarked order (cf. ibid.: 201–203), although it does occur as an alternant in Lahu and Chinese, for example. The basic order in noun-final languages such as Tamil and Japanese, and even in Lahu and Chinese, is therefore (c).

This preliminary test provides strong support for EIC’s unmarked case prediction. But there are more logically possible orderings than the four in table 5.7: Adj and $S$ are frequently not on the same side of $N$ (as in English: *good movies that Bill will see*); and in all positions in which $S$ occurs, the category that constructs it ($C$) could potentially occur on the left or on the right periphery. The advantage of the (a) order is that $C$ is the leftmost daughter of $S$. $C$ is most typically a complementizer or relative pronoun in languages with postnominal relative clauses (cf. ibid.: 85–109), and in the great majority of them it is positioned to the left. Postnominal relatives with a right-peripheral $C$ are also attested, however (cf. ibid.: 73–80), as in Dyirbal, in which $C$ is a subordinating morpheme -ŋu that is attached to the verb in lieu of the tense suffix (example from ibid.: 73):

(5.11) *Yibi ŋnalŋa-ŋu djilwal-ŋu*

"woman child-ERG kick- SUBDN," i.e.

the woman whom the child kicked

Prenominal relative clauses are typologically more consistent, preserving the right-peripheral construction of the structure (c) in table 5.7, even when different devices serve as the constructing category for $S$. For example, the
Dravidian languages add a subordinating suffix -a to tense-marked verb stems in final position, as in the Telugu example (5.12) (from ibid.: 50):

(5.12) mūru nāku ic-cin-a pustukamu
"you me give-PAST-SUBDN book-NOM," i.e.
the book that you gave me

Turkish possesses non-finite verbal forms that are uniquely indicative of subordinate clause status, occurring again in final position, e.g. those with an -en suffix as in example (5.13) (from ibid.: 52):

(5.13) mekteb-e gid-en adam
"school-DAT go-suBDN man," i.e.
a man who is going to school

Chinese uses a nominalizing particle de immediately following the relative clause and immediately preceding the head noun.

Notice also that category C is sometimes absent in languages with either postnominal or prenominal relative clauses, whereupon subordination is indicated positionally. Thus, in many verb-initial languages the juxtaposition of a head noun plus initial verb of the relative clause is sufficient to render C dispensable (as in Jacaltec, cf. ibid.: 80–85); and in Japanese, the juxtaposition of a final verb plus head noun has the same effect (ibid.: 70–72). The S of the relative clause is therefore still recognized as subordinate on its left periphery in these verb-initial languages (on the basis of the immediately preceding head noun), and is still recognized as subordinate on the right periphery in Japanese (by an immediately following head noun occupying the immediately postverbal position that C would have filled if present). I shall return to consider all these alternations in more detail in ch. 6.3.2.

Consider now all the logically possible orders of NP{N Adj s{C S}}. There is a total of 12, shown in table 5.8 and labeled A–L. Four of these, A–D are optimal for EIC: in each, three adjacent words suffice for the recognition of three ICs. The remaining eight all require six words in the viewing window for NP, and so have an IC-to-word ratio of 3/6 = 50%. We can achieve finer discriminations within these eight, however, by calculating our ratios left to right (cf. (3.15) in ch. 3.4). When we do so, the four optimal orders retain their 100% ratios, and three further groupings are established among the non-optimal orders: there are two orders with an 83% aggregate (E and F); four with a 63% aggregate (G–J); and two with 38% (K and L). The orders with 83% have higher ratios because the two single-word ICs are both leftmost in the CRD, and provide more constituency information sooner in
Testing EIC's grammatical predictions

the left-to-right parse string. The 63% orders have one single-word IC to the left and one to the right of S. The 38% orders are worst because both single-word ICs are rightmost in the CRD, and so two thirds of the constituency information for NP is delayed until the last two words of the parse string.

EIC predicts that the optimal orders A–D will be those that are grammaticalized across languages in the unmarked case. It also predicts that E + F will be more or equally attested compared with all other non-optimal orders, and that G–J will be more or equally frequent than K + L. These predictions are strongly confirmed. All the optimal orders are extensively attested. All head-initial languages (roughly 50% of the world’s total) are of either type A or type B, and the majority of head-final languages exemplify one or the

Table 5.8 Grammaticalized \( N \text{ Adj } S \{C \ S\} \) orders

Assume: \( N = 1 \) word, \( \text{Adj} = 1 \), \( C = 1 \), \( S = 3 \), \( \bar{S} = 4 \)

\( N, C = \) constructing categories (for \( \text{NP} \) and \( \bar{S} \) respectively)

<table>
<thead>
<tr>
<th>Structure</th>
<th>L-to-R IC-to-word ratio (aggregates)</th>
<th>Attested language numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>A ( [N \text{ Adj } [C \ S]] )</td>
<td>1/1 2/2 3/3</td>
<td>100% 100% 100% Extensive, e.g. Romance, Arabic, Ewe</td>
</tr>
<tr>
<td>B [\text{Adj N } [C \ S] ]</td>
<td>100%</td>
<td>Extensive, e.g. Germanic, Greek, Finnish</td>
</tr>
<tr>
<td>C [ [S C] \text{ N Adj} ]</td>
<td>100%</td>
<td>Extensive, e.g. Basque, Burmese</td>
</tr>
<tr>
<td>D [ [S C] \text{ Adj N} ]</td>
<td>100%</td>
<td>Extensive, e.g. Telugu, Turkish, Lahu</td>
</tr>
<tr>
<td>E [N \text{ Adj } [S C] ]</td>
<td>1/1 2/2 3/6</td>
<td>100% 100% 50% Attested, e.g. Lushei, Dyirbal</td>
</tr>
<tr>
<td>F [\text{Adj N } [S C] ]</td>
<td>83%</td>
<td>Attested, e.g. Yaqui, Hurrian</td>
</tr>
<tr>
<td>G [N {C S} \text{ Adj} ]</td>
<td>1/1 2/5 3/6</td>
<td>100% 40% 50% 63% None</td>
</tr>
<tr>
<td>H [\text{Adj } [C S] \text{ N} ]</td>
<td>63%</td>
<td>None</td>
</tr>
<tr>
<td>I [N {S C} \text{ Adj} ]</td>
<td>63%</td>
<td>None</td>
</tr>
<tr>
<td>J [\text{Adj } [S C] \text{ N} ]</td>
<td>63%</td>
<td>Marked variant in Lahu, Chinese</td>
</tr>
<tr>
<td>K [ [C S] \text{ N Adj} ]</td>
<td>1/4 2/5 3/6</td>
<td>25% 40% 50% 38% None</td>
</tr>
<tr>
<td>L [ [C S] \text{ Adj N} ]</td>
<td>38%</td>
<td>None</td>
</tr>
</tbody>
</table>

EIC predictions

Unmarked case: \( A + B + C + D = \) most optimal; all head-initial languages = A or B; plus majority of head-final = A or B or C or D. Correct
other of A–D. The only other orders that are attested as basic orders at all are E and F. This supports the marked-case prediction, and although I do not have figures on the precise distribution of E + F compared with A–D orders in head-final languages, the majority appear to be of the latter kind, and this majority coupled with all the head-initial languages is sufficient to satisfy the unmarked case prediction as well. There is even a marked-case discrimination between the 63% and the 38% orders, since one of the former occurs as a productive marked variant in some languages, whereas none of the latter do.

Summarizing, the 100% orders are all extensively attested; the 83% orders are found in a minority of head-final languages; at least one of the 63% language types occurs as a marked variant; while the 38% languages are not attested at all. The higher the EIC score for an NP ordering, the more languages there are that make use of it, and the great majority of the world’s languages appear to have grammaticalized one of the maximally optimal orders.

These facts provide strong support for EIC. They provide no support, however, for current grammatical generalizations in terms of head-of-phrase positioning across categories (cf. Zwicky 1985), or directionality of case and θ-role assignment (cf. Travis 1984, 1989; Koopman 1984). A head noun does not assign case or θ-roles to a modifying Adj or S, so this generalization cannot even begin to take effect here. And while there is an almost exceptionless preference for postnominal relative clauses in head-initial languages, head-final languages are found with both prenominal and postnominal relatives, and adjectives occur productively in prenominal and postnominal position in both language types. Consistent head ordering is not a correct cross-linguistic generalization, therefore, nor even a statistical one (apart from the correlation between [N S] and head-initial); cf. Dryer (1988, 1992) for a similar conclusion.

Notice finally that whereas the positioning of a relative clause in relation to its head noun is typically strongly grammaticalized across languages, adjectives exhibit much more variation. There is more “doubling” (cf. Hawkins 1983), resulting in both postnominal and prenominal orders in the Romance languages, in Dyirbal, and in many other cases. Often a basic order can still be discerned: the prenominal adjectives of Romance are more limited and less frequent than their postnominal counterparts, and are generally associated with particular interpretations. On other occasions, frequency alone suggests a basic order, as in Dyirbal (cf. Dixon 1972: 291). The Romance languages therefore combine structure A in table 5.8 with a marked variant B. Dyirbal combines structure E with a marked variant F. EIC’s Basic Order Prediction
(3.21) is not set up to make predictions for these marked variants within languages, unless they can be subsumed under the predictions for grammatical rearrangements (cf. ch. 3.6.2 and section 5.4 below). But our clear expectation is that these marked orders will still have relatively good EIC ratios. Moreover, if a language does not have a single basic order, but permits (for example) both A and B equally, then A and B might be regarded as equally basic, and in this case (3.21) as currently formulated would make the usual prediction for both orders within a single language, predicting optimality in the unmarked case. Alternatively (and more plausibly) the grammar could be regarded as allowing free ordering in these cases, and the actual occurrence of different orderings in performance would then be subsumed under EIC's predictions for performance (cf. ch. 3.5.1 and ch. 4). Either way, EIC predicts that the kinds of alternations that are noted in grammatical descriptions of particular languages should have optimal ratios in the unmarked case, both within and across languages. This leads us to expect an alternation between A and B in languages such as Romance, but not between A and G, for example. And while E and F may both occur in Dyirbal, we do not expect E in combination with I in this language. Moreover, on those occasions in performance when structure J does occur as a marked variant of D in Lahu and Chinese, EIC predicts that the S will not be long and complex. One consequence (and advantage) of having a single theory for both performance and grammar (cf. ch. 3.8) is that very similar predictions are made and are testable, whatever the precise nature and degree of the grammatical response.¹

5.2.2 The positioning of NP and PP within a VP

Several of the multiple branching structures that were examined in ch. 4 involved an NP and a PP in combination with a verb. For a number of languages the structural template used defined NP as an accusative or direct object NP only. There are grammatical arguments in the languages in question for a VP constituent, and I assumed that the accusative NP was necessarily VP-dominated, while the PP could be either VP- or S-dominated.

With these assumptions we can now make predictions for the grammaticalized ordering of NP and PP within a VP, in those languages that have a VP and a basic order within it. Notice first that we have seen clear evidence from several languages for the greater length of PP compared with NP, and hence the aggregate for PP exceeds that for NP. This is reasonable because the PP consists of an NP plus one additional element, the adposition, but this does
not actually guarantee that PPs will generally be longer than NPs in performance (NPs within a PP might just happen to be shorter than those dominated by VP or S), though it makes it very likely. Empirically, the facts are as follows. The mean length of direct-object NPs within our English sample is 2.5 words; the mean length of (VP- or S-dominated) PPs is 4.6 words, i.e. over two words longer (cf. table 4.22). There are 414 structures in these data in which either PP exceeds NP in length, or NP exceeds PP (excluding the cases where NP = PP). Of these, 349 (or 84%) are in the direction of the mean, with PP exceeding NP. The Japanese data of table 4.10 include 153 structures in which PP\textsubscript{m} > NP\textsubscript{o} or NP\textsubscript{o} > PP\textsubscript{m}. In 100 of these (65%), the PP is longer. In Korean (cf. table 4.11), the proportion of longer PPs is 56/74, or 76%. In German (cf. table 4.23), the proportion is almost identical to English: 51/68 (or 86%). In all these cases, the ratio of longer PPs to shorter direct object NPs is always significantly higher than the reverse: by two to one in Japanese, three to one in Korean, and over five to one in English and German.

Consider first English-type languages, in which MNCCs are regularly left-peripheral. Since V constructs VP, our unmarked case prediction for the VP\{V\textsubscript{m}NP\textsubscript{m}PP\} structure is that V will be the leftmost of these constituents. This will provide immediate recognition of the VP within a higher domain such as S. It will also be optimal for EIC ratios within the VP itself: V is typically a single-word IC, and by positioning it to the left of NP and PP, both of which have aggregates longer than one and are recognized on their left peripheries, the best possible ratios are consistently guaranteed. If V is non-initial within an English-type language, however, EIC makes a rather interesting prediction. From the perspective of the VP alone, any IC or ICs that precede V should not exceed the one-word aggregate of V. This means that grammaticalized one-word ICs, such as pronouns, will be possible in this position. The Bantu languages and Romance languages exemplify exactly this possibility with their preverbal clitic pronouns. Even if there is a plurality of them in any one VP (as in the French je \textit{VP[le lui donne]} “I it to-him give,” i.e. I [will] give it to him), the IC-to-word ratio can be perfect (3/3 = 100% in the French example), since each pronoun is a separate IC. Alternatively EIC is also compatible with whole phrases in preverbal position, on condition that the phrase adds only one word to the CRD for VP. This can happen under only one circumstance: the relevant phrase must be constructed on its right periphery, even though phrases in English-type languages are generally of the opposite type, m\textsubscript{IC}. We will see a striking example of this in the English VP in section 5.3.2.
From the perspective of a higher node containing VP, such as S, any preverbal constituent will be particularly advantageous if it can also construct the VP in lieu of the verb, by Grandmother Node Construction (cf. ch. 6.2.2). Non-subject clitic pronouns are typically case-marked, and do seem, in general, to be capable of indicating the onset of VP within the parse string, though I have not investigated the matter in detail (cf. further ch. 6.3.3). Pre-verbal phrases with right-peripheral MNCCs have a marked status within a predominantly left-constructing language, and this leads me to expect that such structures will be less common than pre-verbal clitics. This also requires further investigation.

What about the relative ordering of mNP and mPP within the CRD for VP? EIC's basic order predictions here are crystal clear: the preferred grammaticalized order is \([V \text{ mNP} \text{ mPP}]\), with the direct object adjacent to the verb, on account of the greater length of PP. This is the order that is strongly grammaticalized in English (cf. table 4.22). It is also my impression that a PP can only intervene between V and a direct object in VO languages under conditions that are analogous to those that motivate Heavy NP Shift in English, and hence that \([V \text{ mNP} \text{ mPP}]\) is generally the grammaticalized order in these languages. The existence of a VP and of a basic order for these ICs within it still needs to be established for many of the languages in current typological surveys, and until this kind of grammatical fine-tuning is added to the data of these surveys, it will not be possible to give precise figures. But the impressionistic evidence is very strongly supportive.

Summarizing, EIC predicts \([V \text{ mNP} \text{ mPP}]\) orders in English-type languages. The NP can precede V if it is a pronominal clitic, and/or if it can construct the VP in left-peripheral position (cf. ch. 6.2.2 and 6.3.3). The PP will be rightmost, and \([V \text{ mPP}\text{ mNP}]\) will occur only under conditions of Heavy NP Shift, but \([\text{ mPP V m NP}]\) should not occur at all: the minimum word content of PP (2) exceeds that of V, and it does not appear that a PP is capable of constructing VP within a higher domain.

Consider now Japanese-type languages, in which MNCCs are consistently right-peripheral. The unmarked case prediction for \(V\{V\text{ NP mPP}\}\) is that V will be rightmost within its CRD, thereby shortening head-final containing CRDs (in conjunction with a bottom-up parse) and the CRD for VP itself. Within the VP EIC makes a clear long-before-short prediction, and hence if any order is grammaticalized, it must be \([\text{PP m NP m V}]\), with the direct object adjacent to V just as it is in English. Both Japanese and Korean show clear evidence for the grammaticalization of this sequence in the performance data of tables 4.10 and 4.11, since this order is productively tolerated even when
NP_m > PP_m, and a rearrangement to \([\text{NP}_m \text{PP}_m \text{V}]\) takes place only when the NP is significantly longer, just as Heavy NP Shift in English requires a significantly longer \(m\)PP for the mirror-image movement to the right.

The grammaticalization of VP-internal orderings in verb-final languages is complicated, however, by the possibly conflicting demands imposed by containing domains, in which early recognition of the VP can be an advantage, and by the predictions for the VP domain itself. If case-marked NPs can render a VP predictable, then positioning a short case-marked NP at the onset of VP can be advantageous for a matrix S, even though the VP alone would prefer a longer IC such as PP_m to the left of this NP. In contrast to English-type languages in which the benefits for VP and S are largely harmonic, in Japanese these two CRDs can pull in different directions. In this kind of a situation it seems reasonable to predict that, if there is going to be a grammaticalized basic order at all, priority must be given to the most immediately containing node, and hence any grammaticalized orders will follow the predictions for VP, with alternative S-motivated orderings arising through a rearrangement of the basic order. This can be motivated by at least two considerations. First, the containing nodes for VP are not always constant. If VP is embedded within another VP, then EIC's predictions for the higher VP will have much less impact on internal orderings within the lower VP, since there will be a consistent bottom-up parse. The lower VP will prefer long before short ICs, and the higher VP will prefer construction of the lower VP as late as possible, either by V or by a case-marked NP. V will always be final anyway, and a PP is generally longer than a case-marked NP, so there will be no reason for reversing the preferred \([\text{PP}_m \text{NP}_m \text{V}]\) order. Only in those instances in which the case-marked NP is longer than the PP will there be a conflict, with the lower domain preferring \([\text{NP}_m \text{PP}_m \text{V}]\) and the higher one \([\text{PP}_m \text{NP}_m \text{V}]\). When S is the containing node, on the other hand, there will always be an advantage for the S domain to prepose the case-marked NP, regardless of its length relative to PP, if the construction of S has already been initiated by a ga-marked subject. But this brings us to the second relevant consideration: there is frequent deletion of arguments in Japanese, including subjects, and hence the S domain may consist only of a VP, and may be constructed simultaneously with this latter. Leaving aside the mechanics of how the S might actually be recognized under these conditions (cf. ch. 6.2), it is clear that an S that dominates VP only will not impose any contrary demands on VP-internal orders.

For these reasons it is logical to propose a principle of the Priority of the Most Immediately Containing CRD, when predicting grammaticalized basic
orders. Even when the total number of rearrangements approaches a majority, as in the Extrapolation from NP structures of German (cf. ch. 4.3.5), there is still evidence for an S positioned within the NP in the basic order, rather than outside it, and in exactly the (rightmost) order relative to all other NP daughters that would be predicted by NP-internal considerations alone. Within Japanese, the grammaticalized basic order of [PP NP V] that is visible in the data of table 4.10 also supports the priority principle, because the grammaticalization follows the predictions for VP. The very fact that long before short was supported in all our Japanese performance data suggests that the priority principle operates in performance as well, as we would expect given the proposed correlation between performance and grammar.

Some very interesting additional support for the priority principle comes from Turkish, a language in which NPs do not look fundamentally like PPs, as they do in Japanese and Korean, and in which the case-marking of NPs is very different in surface syntax from the free-standing postposition of a PP. In ch. 4.1.4 and 4.1.7 I considered EIC's predictions for two types of NPs, NP_m in which there is a clear left-branching relative clause or possessive phrase, and mNP, which refers to all other NPs and which are assumed to be recognizable on their left peripheries by some NP-dominated constituent (cf. further ch. 6.3.4). Charlotte Reinholtz has examined her Turkish data (cf. table 4.6) with reference to the relative positioning of a non-subject NP of both types and a PP. Her data confirm the relative length differences between NP and PP that we have seen in the other languages: in 24/32 structures (or 75%) PP > NP, while only 8/32 had NP > PP. When the NP is of the Japanese type, i.e. NP_m, there is clear evidence for a grammaticalization of VP[PP NP V], just as in Japanese: 12/15 relevant structures (80%) have the order [PP_m NP_m V]. Within these 12, the PP_m is longer than, equal to, or shorter than NP_m, another sign of grammaticalization. In the three instances of the order [NP_m PP_m V], however, NP_m is shorter than or equal to PP_m. This suggests that the NP is being preposed because of its brevity. If so, the typically shorter mNP structures should exhibit even more departures from the basic order, and they do. The basic order [PP_m, NP V] is now retained in only 12/22 instances (55%), and of the 10 structures with [mNP PP_m V], every single one consists of a single-word case-marked mNP.

These data indicate that there is a VP-motivated basic order [PP NP V] for Turkish, just as there is for Japanese and Korean. The S-motivated departures from this assert themselves with especially short NPs, where the benefits for S are greatest since the VP can now be recognized as soon as possible, and significantly sooner than if a longer PP precedes the NP. In none of these
languages, however, does the grammaticalized basic order follow the preferences dictated by S; and the overall quantity of S-motivated rearrangements in Turkish, namely 13/37 (or 37%), still leaves a clear majority of basic orders intact in language performance. Clearly, the VP has priority in determining the relative ordering of its own ICs, and this both supports our proposed priority principle, and makes life a lot easier when trying to formulate and test EIC's predictions!

In all the languages and structural types considered so far, \([V_{m}NP_{m}PP]\), \([PP_{m}NP_{m}V]\), and \([PP_{m}mNP_{m}V]\), there has been one fundamental regularity predicted by EIC, despite the differences: a VP-dominated direct object or other NP should be adjacent to the verb in the basic order, and not separated by a PP. This is reminiscent of Tomlin's (1986) principle of Verb–Object Bonding, for which he presents considerable grammatical evidence from different types of languages. One piece of evidence for regarding EIC as the ultimate explanation for the facts that motivate Verb–Object Bonding, however, comes from a rather crucial test case in which the predictions diverge. There is one structural type in which EIC predicts that a direct object will not be adjacent to the verb and will be separated from it by a VP-dominated PP. That type is exemplified by German, in which the verb is final within VP, and NP and PP are fundamentally English-like in being recognized on their left peripheries, i.e. \(mNP\) and \(mPP\), by a determiner or noun, or by a preposition respectively. The type is therefore \(VP[\{mNP_{m}PP\}V]\).

Relevant data from German were considered in ch. 4.3.3 and table 4.23. The grammaticalized order is clearly \(mNP_{m}PP_{m}V\), with the shorter NP before the longer PP, which provides better left-to-right IC-to-word ratios than the reverse \(mPP_{m}NP_{m}V\). The German combination of \(mIC\) sisters with final verb position is, of course, cross-linguistically highly marked, as the data of table 5.1 showed: at most 3.6% of the world’s languages combine verb-finality with prepositions. But within this marked type, EIC ratios are improved if NP precedes PP within a multi-branching VP, and indeed 69/89 structures (or 78%) in table 4.23 exemplified this order. Grammaticalization was further supported by showing that \(NP PP V\) was productive when NP was shorter than, equal to, or longer than PP, and that the rearrangement to \(PP NP V\) was favored when NP was significantly longer than PP, but did not occur when PP was significantly longer than NP.

German is clearly not a verb–object bonding language. EIC appears to predict correctly, therefore, which language types will adhere to Tomlin's principle and which will not. It also makes another interesting prediction
for such languages. If the NP is pronominal, it should always stand to the left of a PP, which necessarily consists of at least two words. There will never be a motivation for postposing a pronoun to the right of PP, because a pronoun never exceeds PP in length. Hence, pronouns should always precede PPs within the VP, given that the NP dominating the pronoun stands in a grammaticalized basic order before the PP (cf. ch. 6.3.3). Moreover, EIC predicts the possibility of having separate serialization rules for pronouns in this and other languages, whereby a pronoun precedes all other VP-dominated constituents with word-length aggregates equal to or in excess of one word. In English-type languages this will result in pronouns either immediately preceding the verb (as in Romance and Bantu), or immediately following it (as in English, cf. section 5.4.2), with gradual diachronic cliticization of the pronoun onto the verb. But in German we predict that these pronouns will be in the leftmost position of the VP, at the opposite extreme from V, thereby precluding cliticization!

And indeed it is well known that non-subject pronouns in German do precede both PPs and full NPs and occur in left-peripheral position, regardless of the case of the pronoun and of the full NP. (5.14) is the only grammatical ordering for the pronoun es:

(5.14) Ich habe vP[es PP[an die Frau] geschickt]

"I have it to the woman sent," i.e.
I have sent it to the woman.

(5.15) *Ich habe vP[PP[an die Frau] es geschickt]

and (5.16) appears to be the basic order for the phonologically heavier pronoun ihm, with rearrangement to (5.17) being possible only under conditions of strong stress:

(5.16) Ich habe vP[ihm NP[das Buch] gegeben]

"I have to-him the book given," i.e.
I have given him the book.

(5.17) *Ich habe vP[NP[das Buch] ihm gegeben]

Without stress on ihm, (5.17) is ungrammatical.

The cross-linguistic data of this section have been less quantitative than in previous sections, and the testing of EIC's predictions for basic orders within the VP has had to be more anecdotal than I would like. This is an area in which more grammatically fine-tuned typological data are needed. Nonetheless, EIC makes correct predictions for the languages of our
performance sample, in which the grammaticalized orders are clearly visible, and these predictions can perhaps help to guide typological surveys in the future.

Also in need of further investigation are the languages that are marked from the point of view of EIC and that appear to position the verb medially within the VP. Thus, Chinese has many \( \text{VP}[ \text{PP \ V \ NP} ] \) orders, but is apparently quite unique in this regard: the correlation of VO with V–PP is otherwise 100% in Dryer's (1992) sample, cf. section 5.1.2. The mirror-image order \( \text{VP}[ \text{NP \ V \ PP} ] \) is more productive, but still infrequent: 9/72 (or 13%) of Dryer's genera combine OV with a postverbal rather than a preverbal PP.

The preferred correlations between the binary branching ICs \{V O\} and \{V PP\} have already been motivated and tested in section 5.1.2 and table 5.1. The unmarked preference is for O and PP to occur on the same side of V, and our primary concern in this section has been to predict their relative order when they do so. The greater frequency of \[V \text{ PP}_m\] versus \[m \text{ PP V}\] in the marked case was motivated in terms of higher left-to-right IC-to-word ratios, and this could result in some \[V \text{ PP}\] orders even in languages that are otherwise \[NP \ V\], i.e. OV. This will be especially favored if the PP is prepositional, as in German, since the preposition provides immediate recognition of this IC after the verb. But even if the PP is postpositional, the early placement of the verb results in relatively good left-to-right ratios, despite the added length of the CRD for VP compared with the optimal \[PP_m \ V\] order. These considerations motivate the peripherality of V relative to both O and PP in the unmarked case, and a marked case preference for V–PP in OV languages over PP–V in VO languages, and hence for \[NP \ V \ PP\] over \[PP \ V \ NP\] when these two complements of the verb are combined within a single multi-branching VP.

The preferred peripherality of the verb is further motivated by the requirements of whatever CRD immediately contains VP. The VP is quite a heavy constituent, compared with other ICs. Like an \( \text{S}\), it therefore prefers to be either the rightmost IC of its domain, with a left-peripheral MNCC (the verb) comparable to a left-peripheral complementizer in \( \text{S}\); or else the leftmost IC of its domain, with a right-peripheral MNCC comparable to a right-peripheral complementizer. In this way the verb can construct VP either as early as possible (in an English-type language), or as late as possible (in the Japanese type), and the maximum number of nodes dominated by VP are thereby excluded from the CRD that immediately contains VP, so that EIC ratios are maximized.
5.3   A look at English

Of the ten languages that were examined in ch. 4, English is probably the one with the most strongly grammaticalized word order. Accordingly it has many basic orders of relevance to our Basic Order Prediction (3.21), and it is therefore instructive to look at this one language in more detail, in order to see whether these basic orders are consistent with EIC. In the first two subsections we shall examine the distribution of single-word and multi-word ICs within the VP (5.3.1) and the NP (5.3.2). A performance approach to English grammar reveals an intriguing regularity, which has not, to my knowledge, been explicitly pointed out before. In the third subsection we shall document a cross-categorial generalization in the positioning of complements of the head, which has been noticed before, but for which EIC now provides an explanation (5.3.3).

5.3.1   Single-word and multi-word ICs in the VP

In the last subsection I made a prediction for English-type languages with left-peripheral MNCCs: if any constituent precedes the verb within VP, it should add only one word to the CRD for VP. Since the verb is also a one-word item, this will guarantee the highest possible IC-to-word ratios for the VP. Either the IC must dominate a single word and nothing else, e.g. a pronoun; this we might refer to as "unique-word domination." Alternatively, if the IC is phrasal, it must be of the type ICₘ, constructed and recognized on the right periphery by the last word of its domain. Either way the IC will add only a single word to the CRD for VP, and I shall refer to both these types as "single-word ICs" in this context.

English does not make use of the preverbal option for pronouns, preferring instead to position them immediately after the verb (cf. p. 286 below). Since almost all its phrases are of the type ICₘ, with left-peripheral construction, the possible candidates for preverbal positioning are severely limited. A PP is impossible, because it is prepositional:

(5.18)   a. *He VP[within days recovered his form]

           b. He VP[recovered his form within days]

An NP (or quantifier phrase) is impossible, because it is constructed by the left-peripheral determiner (or quantifier):
(5.19) a. *He vp[two times recovered his composure]
    b. He vp[recovered his composure two times]

However, adverbs and adverb phrases are productive in this position. Adverb phrases (henceforth AdvPs) frequently dominate only a single-word adverbial, and preverbal placement of these is, predictably, completely grammatical:

(5.20) a. He vp[rapidly recovered his form]
    b. He vp[recovered his form rapidly]

Notice how the ungrammatical (5.19a) becomes grammatical when the synonymous adverb twice replaces the multi-word phrase two times:

(5.21) a. He vp[twice recovered his composure]
    b. He vp[recovered his composure twice]

And even when an adverb is modified by another adverb, making a multi-word AdvP, the result is still grammatical:

(5.22) a. He vp[quite rapidly recovered his form]
    b. He vp[recovered his form quite rapidly]

This can be explained by observing that rapidly is the head of the phrase quite rapidly: the adverbial modifier quite is added to the left of the head, or in our terms, to the left of the category that constructs the higher AdvP. Hence, this AdvP is of the type IC\textsubscript{m}, and adds only one word to the VP CRD that contains it in (5.22a). This is shown in the tree diagram (5.22'):
The CRD for VP in (5.22) consists of a highly optimal three words that provide recognition of the three ICs: \( vp(\text{rapidly}_{\text{AdvP}}, \text{recovered}_v, \text{his}_{\text{NP}}) \), making a ratio of 100%.

EIC now makes a further prediction. AdvPs may also contain right-branching daughter complements, such as PP. Jackendoff (1977) points out that complement categories are not as productive within the AdvP as they are in other phrases (cf. section 5.3.3). Nonetheless, they do exist in English, and when they occur they follow the adverb head with the result that AdvP is no longer right-headed or IC\text{m}. Such phrases would add more than a single word to the CRD for VP if they were to occur preverbally. Hence EIC makes the unmarked case prediction that AdvPs with PP complements will occur only postverbally, just like the phrases in (5.18) and (5.19). This prediction is correct. Compare the sentences of (5.23):

(5.23)  
a. *He \( vp[\text{quite rapidly for an old man recovered his form}] \) 
b. He \( vp[\text{recovered his form quite rapidly for an old man}] \) 
c. He \( vp[\text{quite rapidly recovered his form for an old man}] \) 
d. Quite rapidly for an old man he \( vp[\text{recovered his form}] \)

The preverbal AdvP in (5.23a) is ungrammatical. On the other hand, (5.23b) has a good EIC ratio, because the AdvP is now heavier than both V and NP, and even though it is not recognized by the leftmost word, it is recognized by the second word dominated, i.e. much closer to the left than to the right periphery. Also grammatical is (5.23c), in which the PP has been postposed out of the AdvP. I leave open whether this is a discontinuous structure comparable to Extraposition from NP (cf. ch. 4.3.5), or whether the PP has been immediately attached to the VP, or whether it is some kind of "afterthought." The important point is that only the IC\text{m} \text{quite rapidly} remains in leftmost position, parallel to the structure (5.22a) in which there is no PP complement at all. Finally, the whole AdvP is topicalized and fronted in (5.23d), which supports the analysis of \text{quite rapidly for an old man} as a single constituent.\textsuperscript{2} Parallel grammaticality facts can be seen in (5.24):

(5.24)  
a. *He \( vp[\text{rather gracefully for a viola player played the violin}] \) 
b. He \( vp[\text{played the violin rather gracefully for a viola player}] \) 
c. He \( vp[\text{rather gracefully played the violin for a viola player}] \) 
d. Rather gracefully for a viola player he \( vp[\text{played the violin}] \)
The ungrammaticality of (5.23a) and (5.24a) is completely unexpected from a narrowly grammatical point of view. Not only is this a descriptive complication within the grammar of English; it goes against a fundamental structural principle of some generality, and the fact that it does so requires an explanation. The problem is this. If English allowed only single-word adverbials to the left of V, as in (5.20a) and (5.21a), we would be able to write the appropriate ordering convention in terms of the category Adv alone, allowing this category (optionally) in preverbal position, while whole adverb phrases (with or without modifiers and complements) could occur postverbally, cf. (5.20b), (5.21b), (5.22b), (5.23b), and (5.24b). But because adverbs can modify other adverbs in preverbal position, as in (5.22a), we have to assume a dominating AdvP in this position as well, as shown in (5.22\textsuperscript{1}). This AdvP can be independently shown to contain a PP complement (optionally), as illustrated in (5.23) and (5.24). Yet the presence of this complement is ungrammatical in one of the independently motivated positions for AdvP, and grammatical in others.

In other words, we can motivate the structure $\text{VP}[\text{AdvP V \ldots}]$, and we can motivate $\text{AdvP}[\text{Adv PP}]$, but when we combine them the result is ungrammatical: $\text{*VP}[\text{AdvP}[\text{Adv PP} \text{ V \ldots}]]$. The structural principle that is being violated here is transitivity of domination. If a node A dominates B, and B dominates C, then A dominates C, and any rule (of positioning, rearrangement, or deletion) that applies to A should affect whatever it dominates. For example, we would not expect a rule that was independently motivated to delete an AdvP to delete just the adverb head along with modifying adverbs, and to leave the PP complement intact. Yet exactly this is happening in the rule positioning AdvP before V: there is independent motivation for positioning the higher node A, but this does not extend to the dominated node C in some environments. What could explain this exception to an otherwise extremely general and independently motivated principle? EIC provides the counter motivation that is needed: the grammatical complication is good for performance, since it keeps the VP optimal in processing, by limiting preverbal AdvPs to single-word ICs of the type IC\textsubscript{m}. We shall see a fundamentally similar state of affairs involving AdjPs within NP in the next subsection, for which the same explanation will be offered, thereby supporting its relevance in the present context.

So far we have established that if an IC is preverbal within VP, then it is necessarily a single-word IC within the CRD for VP. From this it follows that if an IC is possibly multi-word, it must be postverbal. The complement categories NP, PP, VP, and S are all possibly multi-word categories, and in
some cases necessarily so (e.g. PP, Š), and these are all positioned to the right of V (cf. section 5.3.3 for a discussion of their relative order). Obviously, if some category is necessarily a multi-word IC, then it is also possibly multi-word, so we can unify this class in terms of the possibility of dominating more than one word, which contrasts with the categories that necessarily add just a single word to the CRD for VP. Notice that AdvPs are possibly multi-word ICs in postverbal position, but necessarily single-word ICs before the verb (since complement PPs are excluded in that position).

If the single-word/multi-word distinction is to have descriptive significance, it must have generality. There is at least one other category in the English VP that is necessarily single-word: pronouns. The prediction made by EIC is that all necessarily single-word ICs will precede all possibly multi-word ICs in the unmarked case, since that will make CRDs as short as possible on aggregate and provide the best IC-to-word ratios. There may not be a grammaticalization in all cases; but if there is, the predictions are clear.

The prediction is correct for AdvPm and for V itself, as we have seen. For pronouns, there is no separate basic-order rule in English positioning them before the verb (as there is in Romance, where the pronoun rule contrasts with the positioning of NPs and PPs that are generally or necessarily multi-word). But there is a clear "conspiracy" throughout the rearrangement rules of English to generate surface structures in which pronouns are in immediate postverbal position. Thus, if a pronoun starts out as a direct object in the possible Dative–Accusative Shift structure, I gave it to the man, it cannot be moved to the right of the multi-word NP, *I gave the man it. If there is another pronoun within the PP, however, the rule applies grammatically, I gave it to him ⇒ I gave him it, because the result is two postverbal pronouns with an optimal ratio of 3/3 = 100%. I gave it him with deletion of the preposition also seems to be possible here, and this is again optimal (cf. ch. 5.4.2). If the direct object is a multiword NP and the PP contains a pronoun, I gave the book to him, then Dative–Accusative Shift always results in significantly improved ratios (cf. I gave him the book), and this is exactly the condition in which the rule applies most productively in performance (cf. ch. 4.3.7). In all these structures the single-word pronouns precede multi-word NPs, and they also precede any other PP, VP, or Š complements within the VP. A pronoun cannot undergo Heavy NP Shift to the right of PP, for example, *I sent to my friend him.

The grammaticalized surface structures of the English VP are therefore in accordance with our generalization: the necessarily single-word categories in the CRD, AdvPm, V, and Pro precede all possibly multi-word ICs. Consider
now particles in English. These are also typically one-word items, but as Jackendoff (1977: 79) points out, they can be multi-word, e.g. I'll look the answer right up, sir. As a result, they are predicted to be postverbal, and they are. But since their aggregate is very close to one (it is only very rarely that they are modified like this in performance), they should precede all the postverbal categories that are necessarily multi-word, such as PP and S, and they do. The only category that can precede Part is, of course, NP, and we saw in ch. 4.3.1 that [V NP Part] is indeed the grammaticalized order. NP is only a possibly multi-word category, since it can also dominate a single word, but in contrast to particles and particle phrases, the aggregate for NP is much higher than one (namely 2.5 words in table 4.22), so it is predicted that [V NP Part] will regularly be converted into [V Part NP] in performance, and this is what we saw. The basic order is productively retained only when the NP is a single word such as a pronoun, looked it up. When the NP was non-pronominal, a full 75% of these structures were converted to [V Part NP], rising to 80% for NP ≥ 2 words. The anomaly of the basic order here, first discussed in ch. 3.5.2, can now be put in proper perspective: the aggregate for Part is lower than that for NP, yet it follows this latter in the basic order; it precedes all the necessarily multi-word ICs of VP, however, so no rearrangements are needed in these cases.

Summarizing, we have the following descriptive generalizations governing linear ordering in the English VP:

\[(5.25)\] a. If an IC is preverbal within VP, then it is necessarily single-word in the CRD for VP.

b. If an IC is possibly multi-word in the CRD for VP, then it is postverbal.

c. If an IC is necessarily single-word in the CRD for VP, then it precedes all possibly multi-word ICs.

These generalizations are clearly motivated by EIC and support its predictions, even to the point of creating a stipulated exception in the grammar of English to an otherwise general structural principle.

5.3.2 Single-word and multi-word ICs in the NP

It is no accident that the English NP exhibits a pattern that is fundamentally similar to (5.25): single-word ICs are clustered to the left, and multi-word ICs to the right, as shown in (5.26):
(5.26) \( \text{NP}[\text{these three interesting students} \ \ \text{PP}[\text{from Germany}]] \)
\( \text{AdjP}[\text{full of enthusiasm}] \)
\( \text{VP}[\text{to be taught}] \)
\( \text{s}[\text{that I shall teach}] \)

The determiner \textit{these}, the numeral \textit{three}, the adjective \textit{interesting} and the noun \textit{students} are all single-word categories in this example, whereas \textit{from Germany} and \textit{that I shall teach}, etc, are all multi-word. In contrast to the VP, however, the NP does not have a productive set of categories that necessarily dominate a single word only. The head noun does, but apart from this all the prenominal categories in (5.26) can dominate more than one word, even though they typically do dominate just a single word in performance. Thus, numerals can be multi-word, \textit{three thousand}, prenominal adjectives can be modified by adverbs, \textit{very interesting students}, and so on.

The corresponding descriptive regularity in the NP is therefore formulated in terms of possible single-word domination: if an IC is prenominal, then it is possibly single-word in the CRD for NP; if an IC is necessarily multi-word, it must be postnominal; and ICs that are possibly single-word precede all necessarily multi-word ICs. ICs that can possibly dominate a single word (and that frequently do so in performance) will have lower aggregate lengths than those that necessarily dominate more than one word; so positioning the former before the latter will result in the most optimal EIC ratios, just as it does in the VP. Also if an IC is necessarily single-word, like the noun itself, it will \textit{a fortiori} join the class of possibly single-word ICs, just as the necessarily multi-word ICs in the VP patterned with the possibly multi-word ICs. The grammaticalization of orderings in (5.26) is therefore fully optimal for all the different postnominal possibilities. The set of ICs in the NP (or the set of most immediately dominated daughter categories, if we follow the working assumptions about constituent structure made in (3.12) in ch. 3.4) will be \( \text{NP}[\text{Det Num Adj N PP}] \), when there is a postnominal PP. Since the PP is recognized on its left periphery, the EIC ratio is \( 5/5 = 100\% \).

We now have an explanation for a traditional puzzle in English grammar: single-word adjectives stand before the noun, \textit{a yellow book}, adjective phrases headed by \textit{yellow} stand after, \textit{a book yellow with age} (cf. C. Smith 1969). Compare also \textit{an interesting book} and \textit{a book interesting to read}. These multi-word phrases are ungrammatical in front of the noun, \(*\text{a yellow with age book}*, \(*\text{an interesting to read book}*)\), even if Williams' (1981) head-final filter is adhered to, cf. \(*\text{a with age yellow book}*, \(*\text{a to read interesting book}*)\).
Notice that we cannot state this ordering distinction in terms of the difference between an Adj and AdjP, because (as Jackendoff 1977 observes) the prenominal adjective can be modified by an adverb, *a very yellow book* etc., and the addition of this adverb constitutes an AdjP. Hence, AdjP needs to be assumed both before the noun and after it, just as AdvP needs to be assumed before and after the verb within a VP, and in both cases right-branching complements of AdjP and AdvP, such as PP, are not allowed before N and V respectively, but are allowed after them.

These facts can be subsumed under our performance generalization. An AdjP without a PP complement is a possibly single-word IC in the CRD for NP; hence it can precede N. With the addition of a complement, an AdjP such as *yellow with age* is necessarily multi-word, so it follows. When an adverb alone is added, *very yellow*, this AdjP is still a possibly single-word IC within the CRD for NP. If AdjP is the first IC of the NP, i.e. with no preceding determiner, cf. *very yellow books*, it adds just a single word to the NP CRD. The AdjP here is of the type AdjP_m and is parallel to the modified AdvP_m before a V in (5.22a): *yellow* constructs the AdjP and there is an optimal IC-to-word ratio of 2/2 for the NP. If a determiner precedes, cf. *a very yellow book*, the AdjP will be a multi-word IC. This will lower the IC-to-word ratio for NP compared with *very yellow books*, but since adverb modifiers are short items, it will not generally lower it by very much. Moreover, a postposed AdjP (*a book very yellow*) would have the same IC-to-word ratio as the preposed AdjP (3/4 = 75%), so there is no reason for preferring one AdjP ordering over the other in these cases. Ordering *very yellow* before rather than after *books*, on the other hand, improves the ratio for NP: *very yellow books* has a 100% ratio; *books very yellow* would have 67% (2/3).

When there is a succession of adverb modifiers, as in *quite utterly divine concerts*, each of the left-branching modifier phrases can have a perfect IC-to-word ratio in conjunction with a bottom-up parse, as shown in (5.27):

```
(5.27)
```

```
NP
  /\  
AdjP N
|   |
|   |
AdvP_2 Adj concerts
|   |
|   |
AdvP_1 Adv divine
|   |
|   |
Adv utterly
|   |
|   |
Adv quite
```
AdvP₁ has a 1/1 ratio (= 100%), AdvP₂ is 2/2 (100%), and so are AdjP and NP. If there is a string of adjectives, as in *nice old yellow books*, I assume that we are dealing with a string of four ICs (i.e. NP[AdjP AdjP AdjP N]), and with a perfect 4/4 EIC ratio (cf. n. 4 ch. 5 again).

One apparent problem remains. If left-branching modifiers of AdjP are grammatical in prenominal position in English, why are left-branching complements not possible in this position also, i.e. *a with age yellow book?* There are, it seems to me, several reasons for the absence of this structure in English, all of them derived from EIC.

First, English is a left-MNCC language, so its complement phrases are of the type mIC. They are therefore predicted to occur to the right, within their containing phrases, not to the left. Secondly, there is no precedent elsewhere in the grammar of English for a complement to stand to the left of its head, as there is in the marked case language German. Hence, whereas in German there is a structural precedent for complements to the left of Adj within an AdjP, there is no such precedent in English. German has complex (right-headed) AdjPs of this sort, which are a historical expansion of its AAdvP pattern (cf. Weber 1971), whereas English does not. Thirdly, there is no reason for English to have *[a [with age yellow] book]*, since *[a book [yellow with age]]* does exist and is an optimal 3/3 (100%) order for EIC. And finally, whereas left-branching AdjPs can be tolerated after a determiner (*a very yellow book*) since they add very little to the IC-to-word ratio for NP, complement phrases in this position, being necessarily multi-word, would make the NP ratios worse.

For all these reasons English has limited AdjP daughters in prenominal position, and this creates the same structural anomaly that we observed in the AdvP data before a verb: a rule that is independently motivated to apply to AdjP does not apply when AdjP has a set of daughter categories that it is independently motivated to have. A language-particular stipulation simply has to be written into the grammar of English in these cases, blocking the relevant structures. The stipulation is quite ad hoc, but it follows from a performance principle that shapes all grammars, for it precludes an inefficient structure for processing. It is not generally appreciated that particular grammars are full of these kinds of stipulations and conditions on rules, whose significance is precisely that they follow from principles of a performance nature, interacting with other general determinants of structure and grammar (cf. ch. 1.5).

There have been some attempts in the grammatical literature to try and "regularize" such exceptions and stipulations by bringing them into the
grammatical fold and subsuming them under current generalizations. Thus, Abney (1987) proposes that a prenominal adjective is the head of an AdjP, with the immediately following noun its NP complement, i.e. $\text{AdjP}[^\text{yellow} \text{NP}[\text{book}]]$. He observes that postadjectival complements such as PP are normally possible, whereas according to the standard analysis NP is not. If $\text{book}$ is an NP within an AdjP, this fills the gap. But this is an act of grammatical desperation, since as Radford (1993) points out, $\text{yellow book}$ does not have the distribution of an AdjP, and ad hoc mechanisms are required in order to make it pattern like one. Since something has to be stipulated in this area, better to recognize the source of the problem (processing efficiency) and filter the undesirable structures, rather than propose some revised grammatical principles with little supporting evidence for what is not ultimately a limitation coming from the theory of grammar itself.

For language typology, this discussion of adjectives and adjective phrases results in the following prediction: multi-word AdjPs containing complements in English-type languages should prefer postnominal position; but single-word adjectives (or adjectives modified only by adverbials) can have good EIC ratios either in prenominal or in postnominal position, hence they are predicted in both positions. I therefore predict a parallelism between single-word pronouns in VP, and single-word adjectives in NP. The former may occur either before or after the verb, while full NPs follow; the latter may occur before or after the noun, while multi-word AdjPs follow. The available evidence suggests that this is correct. Certainly, single-word adjectives are productive in both orders in English-type languages: 40/95 (or 42%) of Dryer's (1992) VO genera have AdjN, while 55/95 (58%) have NAdj. Precise data for different types of AdjPs are not available in these languages, but it is significant that the semantically very similar relative clause constructions are almost exclusively postnominal in (98% of) VO genera (cf. section 5.1.2), and impressionistically the combination of a single-word adjective before noun with a noun before a multi-word AdjP is quite commonplace, whereas the reverse is not.

We can summarize our performance generalization for the English NP as follows:

$$(5.28) \quad \begin{align*}
\text{a. If an IC is prenominal within NP, then it is possibly single-word in the CRD for NP.} \\
\text{b. If an IC is necessarily multi-word in the CRD for NP, then it is postnominal.}
\end{align*}$$
c. If an IC is possibly single-word in the CRD for NP, then it precedes all necessarily multi-word ICs.

Syntactic categories can therefore be ranked on the following scale: they can be necessarily single-word, possibly single-word (and possibly multi-word), and necessarily multi-word. The grammaticalized ordering rules in the NP are sensitive to the distinction between ICs that are necessarily multi-word and those that are possibly or necessarily single-word. These latter form a single class and are positioned before the necessarily multi-word ICs. The VP contains more ICs that are necessarily single-word, and that may explain why this class is now positioned before the combined class of possibly and necessarily multi-word ICs. The middle category on this scale, possibly single-word and possibly multi-word, combines with either the first or the last category, giving possibly single-word versus necessarily multi-word, and necessarily single-word versus possibly multi-word respectively, and this results in the descriptive generalizations of (5.28) for the NP and (5.25) for the VP. In both cases single-word ICs precede multi-word ICs.

Consider finally a category that appears at first sight to be problematic for the generalization in (5.28): genitives. As is well known, English has two structural possibilities here: a prenominal possessive phrase, the king's daughter; and a postnominal PP that occurs typically with the preposition of, the daughter of the king. Since genitives of both types can be analyzed as consisting of a head category, of or Poss (i.e. -s), plus an NP, this amount of structural complexity seems unexpected in prenominal position. In fact, the distribution of the two is entirely consistent with (5.28). The prenominal PossP can dominate just a single word, and frequently does so: John's daughter. The PP genitive is of necessity multi-word, and it is exclusively postnominal. Hence, genitives are exactly as predicted. Moreover, the prenominal genitive is a left-branching IC, PossP_m, just like prenominal adjectives, AdjP_m (cf. (5.27)). It is processed bottom-up, and each successive CRD within it can be optimally efficient, as in (5.29) which parallels (5.27):
However, despite the conformity of PossPs with the performance general-
ization in (5.28), there are reasons for expecting that this option will not be as
productive as the postnominal counterpart in English-type languages. First of
all, if we replace John in (5.29) with the man who lives next door, the CRD for
PossP₁ becomes very inefficient and complex, cf. the man who lives next door's
(friend's daughter). That is to say, these phrases differ from left-branching
AdjPs and AdvPs as in (5.27), which are always optimal. And secondly, this
greater length of NP compared with AdvP means that NPs with prenominal
PossPs become very inefficient when they are embedded in a higher PP or VP
to the right of P and V respectively, since the CRDs for PP and VP are then
increased unnecessarily compared with the postnominal genitives. And
indeed we saw in tables 5.2 and 5.5 that prenominal genitives are typologi-
cally rare in these languages in comparison with postnominal genitives.

5.3.3 A cross-categorial regularity in the ordering of complements

Consider now the relative ordering of complements of a head in English.
Within the verb phrase we have seen clear evidence for the basic order [V
NP PP], with the NP preceding PP. The NP has an aggregate length of 2.5
words, and the PP of 4.6 words (cf. table 4.22). Since V dominates a single
word, the grammaticalized serialization is consistently short before long, and
this guarantees the most optimal IC-to-word ratios for the VP in an English-
type language, in accordance with the Basic Order Prediction. If there is an
additional Ș complement, as in (5.30), the basic order defined by Jackendoff
(1977: 81) is [V NP PP Ș], i.e. with the Ș as the last complement in the string:
I vppnformed Npfthe boy PP[with great reluctance] §[that he must leave]

I do not have the precise figures for VP-dominated S complements, but it is safe to say that an S has a greater aggregate word length than either NP or PP.

These verb-phrase data lead to the following hierarchy of categories in terms of length aggregates:

(5.31)  S > PP > NP > V (">" = "has greater aggregate length than")

When this hierarchy is linearized by grammatical rules in a language such as English, it is predicted to do so in the opposite order, i.e. [V NP PP S], and this appears to be the basic order.

It is not the only grammatically permissible order, of course. NP and PP can be inverted by Heavy NP Shift, as we have seen (ch. 4.3.2). PP and S can be inverted in response to similar considerations of relative length, e.g. I informed the boy [that he must leave] [with the reluctance of somebody who was carrying out an unwelcome obligation]. I predict that such rearrangements in performance will follow EIC's Text Frequency Prediction (3.18). As with Heavy NP Shift, it is not obvious that the grammar of English contains an actual rule of rearrangement in this case. If we formulate a "Heavy PP Shift," we will need a similar rule for just about every complement category; yet heaviness is not an intrinsically grammatical notion (cf. ch. 3.5.2).

The rearrangement of [V NP PP S] to [V NP S PP] may also result in a "Late Closure" effect, whereby the PP is attached low within the S rather than within the matrix VP, thereby improving the EIC score for the sentence as a whole (cf. ch. 4.5). This may have some consequences for the performance data, but since I view EIC as the ultimate explanation for both linear ordering and Late Closure effects, I do not expect considerations of temporary ambiguity avoidance to interfere significantly with EIC's primary predictions in terms of relative sequencing. Ordering predictions are made quite generally for all phrases and their ICs, whether there is temporary ambiguity on particular occasions or not, and these predictions are well supported. On the other hand, the avoidance of what would otherwise be a regular garden path does appear to motivate the selection of one EIC-preferred order over another (cf. section 5.6.1; cf. also Frazier's 1985 Impermissible Ambiguity Constraint, summarized in ch. 1.2). And it is ultimately an empirical matter to test whether EIC's linear ordering predictions are affected when a temporary ambiguity and/or garden path would result in one or more of the
alternative orderings. The data of ch. 4, which are extensive enough to cover the major logically possible interactions between weight and ambiguity, suggest that linear ordering is fine-tuned to weight alone. For example, the relative ordering of two PPs in English was exclusively of the type \([PP_1 PP_2]\) in table 4.2 when \(PP_2 > PP_1\) by five or more words, and the possible ambiguity of the \([PP_1 PP_2]\) or of the \([PP_2 PP_1]\) sequence produced no exceptions to this.

The other phrasal categories of English obey the same linearization regularity in the ordering of complements as the VP, even though not all complement categories are grammatically permitted in all phrases. Whichever ones are possible occur in the same relative order.

The noun phrase permits both PP and \(\bar{S}\) complements of N. The basic order according to Jackendoff (1977) is \([N PP \bar{S}]\), as in (5.32):

\[
(5.32) \quad NP[students_{PP}[from West Germany]_{\bar{S}}[that I will teach]]
\]

The reverse ordering \([N \bar{S} PP]\) is again grammatical, and I predict that it will be favored in performance when \(PP > \bar{S}\), as in students [that I will teach] [from the major industrialized countries of the world]. Normally, of course, \(\bar{S} > PP\), as reflected in the grammaticalized basic order.

Prepositional phrases may contain an additional PP complement in addition to NP:

\[
(5.33) \quad PP[across_{NP}[the street]_{PP}[from Bill's house]]
\]

The ordering is always NP before PP, and the reverse is not now a grammatically possible rearrangement: *across from Bill's house the street.

Adjective phrases pattern like noun phrases in permitting both PP and \(\bar{S}\) complements. Jackendoff defines \([Adj PP \bar{S}]\) as the basic order, as in (5.34):

\[
(5.34) \quad Adj_P[unfortunate_{PP}[for our hero]_{\bar{S}}[that the city burned]]
\]

The reverse \([Adj \bar{S} PP]\) is grammatical, predictably under performance conditions that mirror the conversion from \([N PP \bar{S}]\) to \([N \bar{S} PP]\), i.e. when \(PP > \bar{S}\), as in it was unfortunate [that the city burned] [for the courageous hero of our story].

We can therefore expand our hierarchy of (5.31) as follows:

\[
(5.31') \quad V \quad \bar{S} > PP > NP > N_P \quad ("\succ" \text{ = "has greater aggregate length than"}) \quad Adj
\]
The head categories, V, N, P, and Adj, each consist of just one word. The relative ranking of the complements has been taken over from the verb-phrase data. I do not have precise figures for these complements when they occur in other phrasal categories, but it would be surprising if their relative ranking were different, and impressionistically it is not. On each occasion when there is a basic word order defined on a head plus any two complements, the linearization is always the mirror image of the aggregate length ranking in (5.31), i.e. short before long, and this is exactly what EIC’s Basic Order Prediction leads us to expect.

What remains to be established in this area is the answer to the following question: how reliable are Jackendoff’s (1977) claims about grammaticalized basic orderings among the complements of these phrases? For example, is the ordering of PP before S really grammaticalized within the VP (cf. (5.30)), the NP (cf. (5.32)), and the AdjP (cf. (5.34)), or is this simply the most frequent pattern in performance, given that S is typically longer than PP? The preference for [X PP S] when S > PP, and for [X S PP] when PP > S, assuming that my own acceptability intuitions are reliable here, provides no argument for grammaticalization over free ordering. Performance data can potentially provide an argument, however, along the lines that were illustrated for [V NP PP] versus [V PP NP]. In ch. 4.3 I showed that the effects of grammaticalization are visible in performance data in those subsets of the word-total assignments to ICs for which EIC makes weak or equal contrary preferences, and these subsets can therefore discriminate between free word orders and grammaticalized basic orders. What needs to be established in the present case is whether the order [X PP S] remains productive in a significant subset of the cases for which EIC would predict [X S PP], whereas the converse fails: i.e. does the latter occur only when EIC predicts it in performance, and does the former occur both when it is predicted in performance and, on occasion, when it is not? Until such data are collected, it remains a moot point whether [X PP S] is indeed the grammaticalized basic order of English.

5.4 Grammatical rearrangements

In ch. 3.6.2 I formulated some testable predictions made by EIC for grammatical rearrangement transformations. These predictions were based on the assumption that rearrangements would apply only to those subsets of the grammaticalized basic orders that were not optimal for EIC (typically a minority on account of the Basic Order Prediction (3.21)), and they would
make them just as optimal for EIC as the basic order itself. This led to the predictions of (3.23), which are repeated here:

(3.23) **EIC Grammatical Rearrangement Prediction**

If the grammar encodes a rearrangement rule in response to EIC, this rule will apply only to those grammatical categories and only in that (leftward or rightward) direction that will potentially increase the EIC ratio for a sentence; if there is a rule for rearranging a category with a non-optimal but relatively good EIC ratio, there will be a rule (possibly the same rule) for all categories with a worse ratio, and any conditions of application on the latter (e.g. obligatoriness) will result in more or equal applications; no rules will apply to those (sub)categories or in a direction that necessarily results in no EIC increase in every application.

The following general prediction results from (3.23): a heavy category with left-peripheral recognition, e.g. §[Comp S] or NP[N S], will be reordered to (or will remain to) the right of its sister ICs; a heavy category with right-peripheral recognition, e.g. §[S Comp] or NP[S N], will preferably be reordered to (or will remain to) the left. Very light categories, on the other hand, such as single-word particles or pronouns, will gravitate to the left of their CRDs by rearrangement rules. Only these reorderings for these different types of categories will result in regular EIC improvements. Prediction (3.23) also makes some more fine-tuned predictions for the conditions of application on rearrangement rules, and it discriminates between categories that will, and those that will not, undergo a particular rule.

In section 5.4.1 I shall test sample predictions involving heavy categories, and in 5.4.2 I shall do the same for the light categories.

### 5.4.1 Heavy categories

The prediction of (3.23) is that a heavy NP in an English-type language, consisting of (e.g.) NP[Det N S] or NP[N S], will be postposed to the right; heavy NPs in Japanese-type languages with e.g. NP[S Det N] or NP[S N], will preferably be preposed. This appears to be a correct cross-linguistic generalization. Certainly it is supported by all the relevant performance data from the language sample in ch. 4. What is not clear, however, even for English, is whether this rearrangement is brought about by a conventionalized grammatical rule. In ch. 3.5.2 I discussed the possibility that English Heavy NP Shift
Testing EIC's grammatical predictions

might be a performance rearrangement rather than a transformational rule. In free-word-order languages, heavy NPs will be positioned by the same performance mechanisms that position all other NPs and all other categories, and there will again be no grammaticalized rule specifically for heavy NPs. Until these issues are resolved, Heavy NP Shift cannot serve as a test case for EIC's predictions for grammatical rearrangement rules. Instead, only the performance predictions can be tested, and these have been shown to be well supported (cf. e.g. ch. 4.3.2 for English, 4.1.5 for Japanese).

By contrast, the repositioning of an S by extraposition appears to involve more grammaticalization, as discussed in ch. 3.5.2, and so provides us with a more reliable test case. This process has also received a lot of attention in different languages. In what follows I shall focus on the cases where S occupies the position of a grammatical subject or object, resulting in a sentential subject and sentential object respectively. In many grammatical analyses, these Ss will be immediately dominated by NP, and the heaviness of S will then make these NPs ICs of non-aggregate length (cf. the Basic Order Prediction (3.21)), whose basic orders will often be dispreferred by EIC.

Consider the predictions for languages with complementizer-initial Ss, set out in (5.35):

(5.35)  \[\tilde{S} \text{ Postposing in Comp-initial Languages}\]

A Comp-initial language in which \(\tilde{S}\) is rearranged in response to EIC will always postpose it and never prepose (\(S = \text{sentential subject, } s = \text{simple NP subject, } O = \text{sentential object, } o = \text{simple NP object}\))

\begin{align*}
\text{a. } & V s[\text{Comp } S] \text{ NP } \Rightarrow V \text{ NP } s[\text{Comp } S], \text{ NOT } s[\text{Comp } S] \text{ V NP} \\
& vOs \Rightarrow vsO, \text{ not Ovs} \\
& vSo \Rightarrow voS, \text{ not Svo} \\
\text{b. } & s[\text{Comp } S] \text{ V NP } \Rightarrow V \text{ NP } s[\text{Comp } S] \\
& Svo \Rightarrow voS \\
\text{c. } & \text{NP } s[\text{Comp } S] \text{ V } \Rightarrow \text{NP V } s[\text{Comp } S], \text{ NOT } s[\text{Comp } S] \text{ NP V} \\
& sOv \Rightarrow svO, \text{ not Osv} \\
\text{d. } & s[\text{Comp } S] \text{ NP V } \Rightarrow \text{NP V } s[\text{Comp } S] \\
& Sov \Rightarrow ovS
\end{align*}

These predictions hold whether the extrapoosed \(\tilde{S}\) is attached to VP (as argued for English by Reinhart 1983) or to S, and indeed whether the relevant language has a VP or not. The EIC improvements brought
about by postposing a lengthy IC with left-peripheral recognition, i.e. mIC, have already been amply illustrated in ch. 3.7 and table 3.2, and do not need to be further justified here. For illustrative EIC calculations for \[\text{[Comp S]}\] structures in English, cf. table 4.24.

These predictions appear to be overwhelmingly supported. Verb-initial languages regularly have postposing rules operating as in (5.35a). For example, all VOS languages (all of which are Comp-initial, cf. sections 5.1.2 and 5.6) obligatorily postpose a sentential direct object to the right of a subject (according to Dryer 1980: 132), as in the following Malagasy example (from Keenan 1976):

   "thinks that looks-for the child Rasoa Rabe," i.e.
   Rabe thinks that Rasoa is looking for the child.

   b. Mihevitra Rabe ʂ[fa mitady ny zaza Rasoa].

VSO languages also regularly extrapose a center-embedded sentential subject to the right. SVO languages, such as English, regularly extrapose a left-peripheral sentential subject, as in (5.35b), often obligatorily. Relevant data from the following SVO languages are given in Dryer (1980: 136–138): Kinyarwanda, Thai, Tuscarora, and Colloquial Egyptian Arabic.

SOV languages with initial complementizers, such as Persian and German, pattern just like these verb-initial and SVO languages (cf. (5.35c) and (d)). Sentential subjects postpose (obligatorily in Persian, optionally in German), and so do sentential objects, obligatorily in both languages. Example (5.37) is from Persian (repeated from ch. 1.4, cf. (1.8)):

(5.37) a. *An zan ʂ[ke an mard sangi partab kard] mi danat.
   "the woman that the man rock threw CONT knows," i.e.
   The woman knows that the man threw a rock.

   b. An zan mi danat ʂ[ke an mard sangi partab kard].

Similar examples from other complementizer-initial SOV languages, including Yaqui, Turkish, and Latin, are provided by Dryer (1980: 130–132).

Complementizer-final S structures result in different EIC ratios and predictions. In fact, for reasons discussed in section 5.6.1, final complementizers occur only in head-final (SOV or postpositional) languages. Any predictions formulated for verb-initial and SVO complementizer-final languages are quite academic, therefore, and can be dispensed with. In (5.38) we formulate the prediction for SOV languages:
(5.38) S Preposing in Comp-final Languages
A Comp-final language in which S is rearranged in response to EIC will preferably prepose it, but may also less preferably (and in fewer languages) postpose it, in accordance with EIC ratio increases:
NP $\bar{s}[S \text{ Comp}] V \Rightarrow$ (preferably) $\bar{s}[S \text{ Comp}] NP V$
$\Rightarrow$ (less preferably) $NP V \bar{s}[S \text{ Comp}]
\text{sOv} \Rightarrow$ (preferably) OsV
$\Rightarrow$ (less preferably) svO

The rationale for this prediction can be seen by reconsidering the sOv structure from Japanese presented earlier as (1.6a) and (3.9a) and renumbered here (5.39):

(5.39) Mary-ga $\bar{s}[\text{kinoo John-ga kekkonsi-ta to}] \text{-it-ta}.$
"Mary su yesterday John su got married that said," i.e.
Mary said that John got married yesterday.

For ease of exposition, let us number nodes from the top of the tree down rather than from left to right (cf. n. 7 ch. 3), i.e. the highest clause will be labeled $S_1$, the immediately dominated subordinate clause $S_2$, and so on. This will simplify comparison between several variants of a common structure. Also, let us count particles as belonging to the word to which they are morphologically and/or syntactically attached, as in our discussion of this sentence in ch. 3.4, rather than as separate words, as we did in the performance calculations involving Japanese in ch. 4.1.5. This will again simplify the comparison and enable us to focus more clearly on essential differences. The IC-to-word ratios for the matrix $S_1$ and VP1 in the center-embedded (5.39) are accordingly 33.3% and 100% respectively, the aggregate for which is 66.7%:

(5.39') $s_1[NP[\text{Mary-ga}] \text{ VP1}[\bar{s}[S_2[\text{kinoo John-ga kekkonsi-ta to}] \text{-it-ta}]]$

$S_1$ CRD: \[\frac{2}{6} = 33.3\%\]
$VP_1$ CRD: \[\frac{2}{2} = 100\%\]
Aggreg. = 66.7%

As we pointed out in ch. 3.4, a preposed $\bar{s}[S \text{ Comp}]$ results in a much-improved aggregate. There are two possible structural analyses: one in which the fronted S is no longer a VP constituent and is immediately attached to $S_1$, and the discontinuous analysis in which it remains attached to the VP. The
EIC ratios for S-attachment and discontinuity are given in (5.40a) and (5.40b) respectively:

\[(5.40)\]

<table>
<thead>
<tr>
<th></th>
<th>CRD:</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3/3 = 100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1/1 = 100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aggreg.</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[a. \text{SitVPit}^\text{st}^\text{noo} \]

\[\text{S1} \text{VP1} \text{S2} \text{VP} \text{NP} \text{John-ga kekkonsi-ta} \text{ta} \text{to} \text{Mary-ga} \text{it-ta} \text{S} \text{VP} \text{X} \text{CRD:} \text{I I} \text{VP} \text{CRD:} \text{I I} \text{Aggreg. = 100%} \]

\[b. \text{SitVPit}^\text{st}^\text{noo} \]

\[\text{S1} \text{VP1} \text{S2} \text{VP} \text{NP} \text{Mary-ga} \text{it-ta} \text{S} \text{VP} \text{X} \text{CRD:} \text{I I} \text{VP} \text{CRD:} \text{I I} \text{Aggreg. = 83.4%} \]

Both analyses produce a significantly improved EIC aggregate compared with (5.39').

However, postposing of $\text{S}[\text{S Comp}]$ to yield svO can also bring about a slightly improved aggregate, whether the $S$ is attached to S or to VP. In (5.41) this is illustrated using Japanese morphemes: the EIC aggregate increases from 66.7% to 75% in (5.41a) (S-attachment); and from 66.7% to 70% in (5.41b) (VP-attachment).

\[(5.41)\]

<table>
<thead>
<tr>
<th></th>
<th>CRD:</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3/6 = 50%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1/1 = 100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aggreg.</td>
<td>75%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[a. \text{SitVPit}^\text{ary-ga} \]

\[\text{S1} \text{NP} \text{VP1} \text{it-ta} \text{S2} \text{VP} \text{NP} \text{kinoo John-ga kekkonsi-ta} \text{ta} \text{to} \text{NP} \text{Mary-ga} \text{it-ta} \text{S} \text{VP} \text{X} \text{CRD:} \text{I I} \text{VP} \text{CRD:} \text{I I} \text{Aggreg. = 75%} \]

\[b. \text{SitVPit}^\text{ary-ga} \]

\[\text{S1} \text{NP} \text{VP1} \text{it-ta} \text{S2} \text{VP} \text{NP} \text{kinoo John-ga kekkonsi-ta} \text{ta} \text{to} \text{NP} \text{Mary-ga} \text{it-ta} \text{S} \text{VP} \text{X} \text{CRD:} \text{I I} \text{VP} \text{CRD:} \text{I I} \text{Aggreg. = 70%} \]

Alternatively, if these verb-final languages lack a VP, then the EIC ratio for a center-embedded $S$ in the structure $\text{S1[NP S V]}$ as in (5.39) will be $3/6 = 50\%$. Postposing the $\tilde{S}$ as in (5.41a) will produce the same $3/6 = 50\%$ ratio, but results in an improvement when ratios are measured left to right (cf. (3.15)),
because the two shorter ICs, NP and V, now precede S. Preposing S produces the optimal $3/3 = 100\%$ ratio of (5.40a). Either way, with or without the VP assumption, EIC defines a clear preference for rearrangements to Osv, but it does not rule out svO altogether in Comp-final languages, and it is no accident, as I see it, that postposing of $S[S \text{ Comp}]$ is attested, for example in Lakhota. An example (taken from Dryer 1980: 132) is given in (5.42):

(5.42) Tohá slolyáya he §[wakpála ektá oihpaya ki]?

“When you-know qu creek into fall that,” i.e.

When did you find out that he fell in the creek?

Preposing, as in Japanese, is predicted to be more frequent.

Let us consider now some more fine-tuned predictions for the operation of Extrapolation within a single language. German presents us with some interesting distinctions in the conditions of application on this rule, for which (3.23) provides an explanation. The reader is referred to Hawkins (1986: 144–147) for a summary and tabulation of the relevant research literature on German. The problem posed by these data is the following. For certain types of clausal embeddings, and in some environments, Extrapolation is an optional rule, i.e. there are grammatical pairs of sentences in which Extrapolation has and has not applied. But for other types of clausal embeddings and/or in other environments, Extrapolation appears to be obligatory: only the extraposed version is grammatical. This distribution of obligatory and optional applications needs to be explained.

I pointed out above that the extrapolation of an $S$ in direct-object position is obligatory in German, just as it is in Persian (cf. (5.37)). This is illustrated in (5.43):

(5.43) a. *Er hatte §[daß er nicht lange leben würde] gewußt.

“He had that he not long live would known,” i.e.

He had known that he did not have long to live.

b. Er hatte gewußt §[daß er nicht lange leben würde].

If the embedding is a non-finite $VP$, however, as opposed to a finite $S$, then the extrapolation is optional. Both (5.44a) and (b) are grammatical:

(5.44) a. Er hatte $VP[\text{den Preis zu gewinnen}]$ gehofft.

“He had the prize to win hoped,” i.e.

He had hoped to win the prize.

b. Er hatte gehofft $VP[\text{den Preis zu gewinnen}].$
Whereas (5.43a) must be converted to (5.43b), (5.44a) may or may not be converted to (5.44b).

An examination of the EIC scores for these two pairs of sentences makes it clear why the grammar discriminates between them. S embeddings are on aggregate significantly longer than VPs. In (5.43) the S is six words in length, while the VP has been assigned four words in (5.44). The aggregate word length for German infinitival VPs in the text sample of table 4.23 is 3.4 words. I do not have precise figures for the S aggregate when S is functioning as a sentence complement, but it is likely to be at least two words longer. The figures for S within a complex NP structure, shown in table 4.28, add up to over six words on aggregate. When EIC ratios are calculated for (5.43) and (5.44), the score for the center-embedded S within the matrix S and VP domains is significantly worse than that for the center-embedded VP, 27% versus 50% (cf. table 5.9 below). This is partly the result of the aggregate word-length difference, partly also because an S is constructed by a left-peripheral Comp whereas a VP is constructed at or near the right periphery of this node, most plausibly by the infinitival complementizer zu, and this improves the ratio for the matrix VP when VP is center-embedded. The prediction of (3.23) is that if there is a rule for rearranging the center-embedded VP with its relatively good 50% score, there will also be a rule, possibly the same rule, for the center-embedded S whose EIC score is much worse. This is of course the case. Further, (3.23) predicts that any conditions of application on the extraposition of S will result in more or equal numbers of applications compared with VP. Since the extraposition of S is obligatory while that of VP is optional, this is also fulfilled.

The justification for these EIC scores is given in table 5.9. The table sets out the assumptions that have been made in the calculations, e.g. S and VP remain attached to VP when they are extraposed. It gives the EIC ratios for (5.43a) and (5.43b) in sections Ia and Ib respectively. Not only is the aggregate of the S and VP CRDs very low in Ia (27%), but the extraposed S has a perfect 100% ratio in Ib. In sections II–IV the table gives EIC scores for pairs such as (5.44a) and (5.44b) involving VPs of different lengths, six words, four words, and two words respectively. The purpose of this is to show that even when a center-embedded VP has the same number of words as the S in our calculations, six words, its EIC score is still significantly higher than that of the center-embedded S, 46% versus 27% (cf. IIa). Hence, its grammaticality is still motivated vis à vis the ungrammaticality of (5.43a) (i.e. Ia), as is the optionality of the conversion to IIB compared to the obligatoriness of the conversion to Ib. By recognizing the VP constituent near its right periphery,
the score for $VP_1$ is consistently high in the center-embedded IIa–IVa, namely 67%, and the cost of the center embedding is measured principally in the $S_1$ domain. When $VP$ is extraposed, $S_1$ is consistently improved to 100%, but the $VP_1$ domain generally declines (cf. IIb and IIIb), except when the $VP$ consists of just two words and the constructing category $zu$ is now left-peripheral (cf. IVb), whereupon $VP_1$ is also improved to 100%. In II and III we therefore see the same tension between $S_1$ and $VP_1$ that we saw between (certain instances of) these nodes in English Extraposition (cf. ch. 4.3.4) and between a higher NP and a VP in the German Extraposition from NP (cf. ch. 4.3.5). The benefits for the $S_1$ outweigh the losses for $VP_1$ in the German $VP$ Extraposition structures, and as a result the application of this rule is significantly favored over its non-application, even though the EIC scores for IIb and IIIb are not as optimal as they are for Ib (and for IVb).

From the point of view of our performance predictions in ch. 4, the declining EIC scores for the center-embedded $VP$s down the grammatical IVa (59%), IIIa (50%), and IIa (46%) lead us to expect that the frequencies of these center embeddings in text samples relative to their extraposed counterparts will also be greater or equal. Data clearly supporting this prediction are given in Hawkins (1992c: 203).

The $S$ and $VP$ complements of table 5.9 occur in sentences containing two-place matrix verbs, wissen “to know” and hoffen “to hope.” With three-place verbs, the pattern of grammaticality data has been argued to be rather different (cf. Ebert 1973: 170 and Hawkins 1986: 145–147). $S$ complements are extraposed obligatorily as usual, as in (5.45b). But the extrapolation of some $VP$ complements now appears to be obligatory as well, as shown in (5.46b).

(5.45)  
\begin{enumerate}
  \item a. *Er hat die Frau $S[\text{daß sie den Mann töten möchte}]$ gebeten.
  \hfill (5.45)  
  \item b. Er hat die Frau gebeten $S[\text{daß sie den Mann töten möchte}].$
\end{enumerate}

(5.46)  
\begin{enumerate}
  \item a. ?Er hat die Frau $VP[\text{den Mann zu töten}]$ gebeten.
  \hfill (5.46)  
  \item b. Er hat die Frau gebeten $VP[\text{den Mann zu töten}].$
\end{enumerate}

Ebert (1973) states that Extraposition is “virtually obligatory” in structures such as (5.46), and he points out that this rule is only truly optional when the
Table 5.9 Extrapolation of $\bar{S}$ and $\bar{VP}$ in German (two-place verbs)

\[ s_i[V_f \ VPI[\bar{S}/\bar{VP} \ V]] \Rightarrow s_i[V_f \ VPI[V S/\bar{VP}]] \quad (V_f = \text{finite verb}) \]

Assume: (i) $S_i$ is constructed by $V_f$ (cf. ch. 6.3.1);  
(ii) $VP_i$ is constructed by $V$;  
(iii) $S$ is constructed by Comp, cf. $s_i[\text{Comp} S]$;  
(iv) $VP$ is constructed by infinitival complementizer $zu$, cf. $\bar{VP}[, ... , zu V]$;  
(v) $\bar{S}$ and $\bar{VP}$ remain attached to $VP_i$ by Extrapolation.

<table>
<thead>
<tr>
<th>I</th>
<th>$\bar{S} = 6$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>*Er $s_i[\text{hatte } \bar{VP}_i[\text{daß er nicht lange leben würde}] \ \text{gewußt}]]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$S_i$ CRD: [2/8 = 25%]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$VP_i$ CRD: [2/7 = 29%]</td>
<td>Aggreg. = 27%</td>
</tr>
<tr>
<td>b.</td>
<td>Er $s_i[\text{hatte } \bar{VP}_i[\text{gewußt } s[\text{daß er nicht lange leben würde}]})$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$S_i$ CRD: [2/2 = 100%]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$VP_i$ CRD: [2/2 = 100%]</td>
<td>Aggreg. = 100%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II</th>
<th>$\bar{VP} = 6$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>Er $s_i[\text{hatte } \bar{VP}_i[\text{den sehr wertvollen Preis zu gewinnen}] \ \text{hofft}]]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$S_i$ CRD: [2/8 = 25%]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$VP_i$ CRD: [2/3 = 67%]</td>
<td>Aggreg. = 46%</td>
</tr>
<tr>
<td>b.</td>
<td>Er $s_i[\text{hatte } \bar{VP}_i[\text{höfft } \bar{VP}_i[\text{den sehr wertvollen Preis zu gewinnen}]]]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$S_i$ CRD: [2/2 = 100%]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$VP_i$ CRD: [2/6 = 33%]</td>
<td>Aggreg. = 67%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III</th>
<th>$\bar{VP} = 4$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>Er $s_i[\text{hatte } \bar{VP}_i[\text{den Preis zu gewinnen}] \ \text{höfft}]]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$S_i$ CRD: [2/6 = 33%]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$VP_i$ CRD: [2/3 = 67%]</td>
<td>Aggreg. = 50%</td>
</tr>
<tr>
<td>b.</td>
<td>Er $s_i[\text{hatte } \bar{VP}_i[\text{höfft } \bar{VP}_i[\text{den Preis zu gewinnen}]]]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$S_i$ CRD: [2/2 = 100%]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$VP_i$ CRD: [2/4 = 50%]</td>
<td>Aggreg. = 75%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IV</th>
<th>$\bar{VP} = 2$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>Er $s_i[\text{hatte } \bar{VP}_i[\text{zu gewinnen}] \ \text{höfft}]]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$S_i$ CRD: [2/4 = 50%]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$VP_i$ CRD: [2/3 = 67%]</td>
<td>Aggreg. = 59%</td>
</tr>
<tr>
<td>b.</td>
<td>Er $s_i[\text{hatte } \bar{VP}_i[\text{höfft } \bar{VP}_i[\text{zu gewinnen}]]]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$S_i$ CRD: [2/2 = 100%]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$VP_i$ CRD: [2/2 = 100%]</td>
<td>Aggreg. = 100%</td>
</tr>
</tbody>
</table>
Testing EIC's grammatical predictions

embedded infinitival appears without any surface arguments, i.e. both (5.47a) and (5.47b) are fully grammatical:

(5.47) a. Er hat die Frau zu arbeiten gebeten.
    "He has the woman to work requested," i.e.
    He requested the woman to work.

b. Er hat die Frau gebeten zu arbeiten.

There is some uncertainty over the precise status of (5.46a) in German, but it is clear that there is a strong preference for (5.46b), and this contrasts with the complete grammaticality of both (5.44a) and (5.44b) in the corresponding two-place structure. EIC can explain the difference. Table 5.10 gives the relevant IC-to-word ratios, starting with the embedded S of (5.45) (section I). The aggregate for the SI and VP1 CRDs in the center-embedded Ia is the same very low 27% that it was for Ia in table 5.9. But because we are now dealing with an extra NP argument within VP1 (die Frau "the woman") which constitutes the first IC of this domain, it is no longer the case that VP1 can be recognized much more efficiently when a center-embedded S is replaced by a VPI. Die Frau will always be the first IC of the domain and gebeten "requested" the last, and the size of the EIC score will simply reflect the length of whatever complement category intervenes. By contrast, the right-peripheral recognition of VPI in the two-place structures generally gave a better ratio to the center embeddings than to their extraposed counterparts within VP1. In the three-place structures the SI scores will be a direct consequence of the length of a center-embedded S or VPI, and so will the scores for VP1. As a result, a six-word VPI will be just as bad as a six-word S when both are center-embedded (see IIa), namely 27%. A four-word VPI as in (5.46a) (see IIIa) will only be slightly better, at 34%. And only the two-word center-embedded VPI of (5.47a) (see IVa), will have a half-way reasonable score of 47%.

The figures of table 5.10 show that the infinitival center embeddings are almost as bad as the S center embeddings, and this explains why native-speaker judgments treat them alike and regard VPI extraposition as "virtually obligatory" with three-place predicates, except for the shortest VPI of all, consisting of no surface arguments (cf. (5.47) and section IV), whose 47% score is close to the ratios for the grammatical IIa and IIIa in table 5.9. That is, whereas the center-embedded VPI complements in table 5.9 are always significantly better than a center-embedded S, the two sets of complements pattern alike in table 5.10, except for VPI = 2. (This point
Table 5.10  *Extraposition of S and VP in German (three-place verbs)*

\( s_1[V_f V_P[[NP S/VP V]] \Rightarrow s_1[V_f V_P[[NP V S/VP]]] \)

Assumptions: cf. table 5.9

<table>
<thead>
<tr>
<th></th>
<th>( S = 6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>a. *Er s_1[hat ( V_P )[die Frau [daß sie den Mann töten möchte] gebeten]]</td>
</tr>
<tr>
<td></td>
<td>( S_1 ) CRD: [2/10 = 20%]</td>
</tr>
<tr>
<td></td>
<td>( V_P ) CRD: [3/9 = 33%]</td>
</tr>
<tr>
<td></td>
<td>b. Er s_1[hat ( V_P )[die Frau gebeten [daß sie den Mann töten möchte]]</td>
</tr>
<tr>
<td></td>
<td>( S_1 ) CRD: [2/4 = 50%]</td>
</tr>
<tr>
<td></td>
<td>( V_P ) CRD: [3/4 = 75%]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II</th>
<th>( VP = 6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a. *Er s_1[hat ( V_P )[die Frau ( \bar{V} )[den sehr reichen Mann zu töten] gebeten]]</td>
</tr>
<tr>
<td></td>
<td>( S_1 ) CRD: [2/10 = 20%]</td>
</tr>
<tr>
<td></td>
<td>( V_P ) CRD: [3/9 = 33%]</td>
</tr>
<tr>
<td></td>
<td>b. Er s_1[hat ( V_P )[die Frau gebeten ( \bar{V} )[den sehr reichen Mann zu töten] ]</td>
</tr>
<tr>
<td></td>
<td>( S_1 ) CRD: [2/4 = 50%]</td>
</tr>
<tr>
<td></td>
<td>( V_P ) CRD: [3/8 = 38%]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III</th>
<th>( VP = 4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a. *Er s_1[hat ( V_P )[die Frau ( \bar{V} )[den Mann zu töten] gebeten]]</td>
</tr>
<tr>
<td></td>
<td>( S_1 ) CRD: [2/8 = 25%]</td>
</tr>
<tr>
<td></td>
<td>( V_P ) CRD: [3/7 = 43%]</td>
</tr>
<tr>
<td></td>
<td>b. Er s_1[hat ( V_P )[die Frau gebeten ( \bar{V} )[den Mann zu töten] ]</td>
</tr>
<tr>
<td></td>
<td>( S_1 ) CRD: [2/4 = 50%]</td>
</tr>
<tr>
<td></td>
<td>( V_P ) CRD: [3/6 = 50%]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IV</th>
<th>( VP = 2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a. Er s_1[hat ( V_P )[die Frau ( \bar{V} )[zu arbeiten] gebeten]]</td>
</tr>
<tr>
<td></td>
<td>( S_1 ) CRD: [2/6 = 33%]</td>
</tr>
<tr>
<td></td>
<td>( V_P ) CRD: [3/5 = 60%]</td>
</tr>
<tr>
<td></td>
<td>b. Er s_1[hat ( V_P )[die Frau gebeten ( \bar{V} )[zu arbeiten]]</td>
</tr>
<tr>
<td></td>
<td>( S_1 ) CRD: [2/4 = 50%]</td>
</tr>
<tr>
<td></td>
<td>( V_P ) CRD: [3/4 = 75%]</td>
</tr>
</tbody>
</table>
remains valid if, following the assumptions of ch. 6.3.3, the VP can be recognized earlier in the S1 domain on the basis of case marking on NPs. The EIC ratio for S1 will then be optimal and constant throughout all the sentences of table 5.10, and the cost of the center embedding will be measured in the VP1 domain alone.)

It is doubtful, however, whether the conditions on the rule of Extraposition actually differ in the two environments. A rule can be either obligatory or optional, but it cannot be "virtually obligatory." Nor is it clear that the different types of VPs can be distinguished in a grammar based on their different sister ICs and/or based on whether they dominate daughter ICs corresponding to surface arguments, as in (5.46) versus (5.47). Hence, it is not clear that we could actually write a grammatical rule that would make Extraposition obligatory in (5.46) but optional in (5.47), even if we wanted to (but cf. section 5.5). The most plausible resolution of this issue is to say that Extraposition is obligatory for a center-embedded S, but optional for a center-embedded VP. Center-embedded VPs are therefore grammatical and will be associated with different degrees of acceptability in performance (and frequency of use, etc.) as a function of their EIC scores. Center-embedded VPs with three-place verbs will be much less acceptable than those with two-place verbs, since their scores approach the low levels associated with S, to which the grammar of German has actually made a conventionalized response. This obligatoriness of S extraposition versus the optionality of VP extraposition is what EIC explains.

5.4.2 Light categories
Consider now EIC's rearrangement predictions for light categories, such as pronouns and particles. Following (3.23), we would expect them to gravitate to the left of their CRDs, if they depart from the basic grammaticalized order assigned to them. This accordingly predicts that there will be no postposing rule of "Light NP Shift" in English-type languages, converting e.g. I gave it to the boy who came yesterday into *I gave to the boy who came yesterday it, or I introduced her to the mayor of Colchester into *I introduced to the mayor of Colchester her, or I bought this at the hardware store into *I bought at the hardware store this. On the other hand, in the structure VP[V NP Part ...] a single-word particle is predicted to move to the left of a multi-word NP, if there is any grammaticalized rearrangement, and this is what Particle Movement brings about.

Therefore (3.23) predicts the possibility of some rules and the impossibility of others. It also provides an explanation for some subtle conditions of
application on the relevant transformational rules, which have not been adequately explained hitherto. These conditions encode what are technically exceptions to, or limitations on the productivity of, independently motivated processes (cf. ch. 2.1). In a similar way, we could explain some conditions of application on the rearrangement of heavy categories in the last section.

The Particle Movement example has already been extensively discussed. The rearrangement in (5.48) improves the EIC ratio for the VP:

(5.48)  

a. \[I_{\text{vp}}[\text{looked}_{\text{np}}[\text{the long number}] \text{ up}]]\]  

VP CRD: \[\frac{3}{5} = 60\%\]

b. \[I_{\text{vp}}[\text{looked up}_{\text{np}}[\text{the long number}]]\]  

VP CRD: \[\frac{3}{3} = 100\%\]

A corresponding rearrangement when the NP is itself a single word does not:

(5.49)  

a. \[I_{\text{vp}}[\text{looked it up}]\]  

VP CRD: \[\frac{3}{3} = 100\%\]

b. *\[I_{\text{vp}}[\text{looked up it}]\]  

VP CRD: \[\frac{3}{3} = 100\%\]

Moreover, when the particle occurs within a particle phrase, preposing is not possible, as pointed out by Sag (1987), cf. *I looked [right up] [the number] versus I looked [the number] [right up]. Only a single-word particle will systematically improve EIC ratios (and achieve 100% ratios) when preposed before all NPs. If NP = one word, preposing a particle phrase would produce a lower score. Grammaticalization has therefore limited these rearrangements to those in which the outputs are always optimal.

The ungrammaticality of (5.49b) can be explained by appealing to this same functional motivation for Particle Movement: there is no point in rearranging the grammaticalized basic order (5.49a), because this structure is already optimal and cannot possibly be improved. The prediction of (3.23) is that rearrangement rules will not apply to any (sub)categorics or in a direction that necessarily results in no EIC increase in every application. Hence Particle Movement in Emonds's (1976) formulation is associated
Testing EIC's grammatical predictions

with a condition on its application stating that the rule blocks when the NP is a pronoun.

EIC's predictions therefore rule out two distinct kinds of rearrangements when formulated in this way: those in which there is always a decrease, as in the Light NP Shift structures exemplified above, \([V\ Pro\ PP] \Rightarrow [V\ PP\ Pro]\); and those in which there is never any improvement, as in (5.49). If we were merely to prohibit a decrease, we would have no explanation for (5.49a) versus (5.49b). The more general prediction blocking rearrangements that result in no EIC increase in every application is supported both by this empirical distinction in grammaticality, and by the absence of any functional motivation for rearranging an already optimal order that has been conventionalized. Naturally, the motivation for blocking a regular decrease will be even stronger, and it may be that the motivation against structures like (5.49b) will be overridden on other occasions by contrary considerations favoring their generation (e.g. the generality of the relevant rule, the desired absence of exceptions, etc.). A possible example will be discussed below, involving Dative Movement. But from the perspective of EIC alone, the stronger formulation is supported.

This formulation also explains the distinction between pronouns and another subcategory of NP that regularly dominates only a single word: proper names. Notice that names do undergo Particle Movement. Look John up when you go to New York and look up John when you go to New York are both grammatical, even though John is a single word. But names can dominate more than one word (John Smith, John Fitzgerald Kennedy), and hence in the terminology of section 5.3, whereas pronouns are necessarily single-word, names are possibly multi-word, just like other NPs. As a result, there will be some instances in which the EIC score for a VP will be improved when it contains a proper name, namely whenever this latter consists of at least two words ([look John Smith up] \Rightarrow [look up John Smith], etc.), and so the grammar needs to allow for this possibility. The grammar could in principle block movement with the subcategory Name, just as it does with pronouns, but there is no longer a reason to do so. Names can therefore undergo Particle Movement, whatever their terminal content, and the performance prediction we make for them is the same that we make for all other types of NPs: the longer they are, the more frequent the application of Particle Movement (cf. ch. 4.3.1). Postposed single-word names will be the least frequent, but they will still be grammatical, for the reason given.
There is a parallel set of grammaticality distinctions in the various verb-second inversion rules of English discussed in Emonds (1976). An example is Subject–Simple Verb Inversion, illustrated in (5.50):

(5.50)  
   a. The little boy ran down the hill.  
   b. Down the hill ran the little boy.

This rule blocks when the subject is a pronoun:

(5.51)  
   a. He ran down the hill.  
   b. *Down the hill ran he.

To see that this pronoun prohibition is not an isolated phenomenon, compare (5.52)–(5.54):

(5.52)  
   a. He sat among the guests.  
   b. *Among the guests sat he.  
      (Inversion after Locative Phrase Preposing)

(5.53)  
   a. He was no less corrupt.  
   b. *No less corrupt was he.  
      (Comparative Substitution)

(5.54)  
   a. He is speaking to the President now.  
   b. *Speaking to the President now is he.  
      (Participle Preposing)

Let us assume some topicalization of a constituent to the left of S in these structures. The inversion will then consist of rearranging the subject NP and V within S. If the subject is a single-word pronoun, there will be no EIC increase in every application, and the result will be ungrammatical: down the hill s[he ran] ≠ *down the hill s[ran he]. But if the subject is multi-word, there will be an increase: down the hill s[NP[the little boy] ran] ⇒ down the hill s[ran NP[the little boy]]. The former has a ratio of 2/4 = 50% for the S CRD; the latter has 2/2 = 100%. EIC can account for these ungrammaticalities, therefore. It also predicts that these inversions should be grammatical with single-word proper names. And they are: cf. down the hill ran John, among the guests sat John, etc.

Consider now the rule of Dative Movement. In ch. 4.3.7 I presented some performance data from English showing the systematic EIC improvements brought about by the application of this rule. A typical example from our data is (5.55b), in which the VP has an optimal 3/3 = 100% ratio, compared with the 3/4 = 75% ratio in the input (5.55a):
(5.55)  

a. \( I \ VP [gave \ NP [\text{the book}] PP [\text{to him}]] \)  

b. \( I \ VP [gave \ him \ NP [\text{the book}]] \)

A full 27/31 (87%) of the shifted structures in the data had a pronoun in the immediate postverbal position, which guarantees a ratio of 100% whatever the length of the other NP, and I argued that it was no accident that Dative Movement is favored over the alternative [V PP NP] (i.e. [gave PP [to him] NP [the book . . .]]) in these cases. By removing the PP and P nodes in (5.55b), a fully optimal structure results. Clearly, Dative Movement subserves the EIC principle in language performance, and even the structural changes that it brings about can be explained by EIC. It should come as no surprise, therefore, that it blocks when a pronoun would be moved to the right as in (5.56). Example (5.56a) has an optimal ratio; (5.56b) results in a decrease and is ungrammatical.

(5.56)  

a. \( I \ VP [gave \ it \ PP [\text{to the woman}]] \)  

VP CRD: \[
\begin{array}{c}
3/3 = 100\%
\end{array}
\]

b. \( *I \ VP [gave \ NP [\text{the woman}] \ it] \)  

VP CRD: \[
\begin{array}{c}
3/4 = 75\%
\end{array}
\]

However, if the output can retain the optimality of the input structure, as in (5.57), the result does appear to be more or less grammatical, in contrast to the other rearrangements we have seen so far (e.g. (5.49b)):

(5.57)  

a. \( I \ VP [gave \ it \ PP [\text{to him}]] \)  

VP CRD: \[
\begin{array}{c}
3/3 = 100\%
\end{array}
\]

b. \( I \ VP [gave \ him \ it] \)  

VP CRD: \[
\begin{array}{c}
3/3 = 100\%
\end{array}
\]
What could explain the difference? One possibility is to say that the overall structure of the VP is simpler in (5.7b) than in (5.7a), on account of the removal of the PP and P nodes, and that this motivates the conversion, even though the CRD for the VP is not being improved. Some support for this comes from the fact that there is yet a third structure that needs to be considered here:

\[(5.57)\quad \text{c. } I_{VP} [\text{gave it him}]\]

\[
\begin{array}{c}
\text{VP CRD: } \\
3/3 = 100\%
\end{array}
\]

This structure has not undergone Dative Movement, but it has simplified the VP in the same way that (5.7b) has, by removing nodes, and it too is grammatical, without there being any EIC improvement as such. This suggests that complexity differences in Structural Domains, involving on this occasion the total set of nodes dominated by VP, can motivate the elimination of dominated nodes—presumably as long as recoverability is guaranteed.

More generally, we might suggest the following refinement in the predictions of (3.23). An EIC-motivated rearrangement will not permit regular decreases in EIC scores, as in (5.56); and it will not apply to (sub)categories that can never increase EIC scores, when all else is equal; but if equally good EIC scores are accompanied by a structural simplification, as in (5.57), then rearrangements may be permitted that retain the EIC scores of the input, since they are now functionally motivated. Many more cases of this sort will need to be examined, especially in other languages, to see whether this is a viable refinement, and indeed to see whether these predictions for light categories hold at all.

I shall end this section by pointing to a relevant grammaticality distinction in German, which is predicted by (3.23), and which has not, to my knowledge, been discussed in the German linguistics literature. German has a Pronoun Fronting rule that operates in subordinate clauses, as shown in (5.58b):

\[(5.58)\quad \text{a. } \text{Ich weiß, daß die Frau ihn liebt.} \]

"I know that the woman him loves," i.e.

I know that the woman loves him.

b. Ich weiß, daß ihn die Frau liebt.
Testing EIC's grammatical predictions

This rule fronts a non-subject pronoun such as *ihn* ("him," accusative) to the left of subjects such as *die Frau* ("the woman.") The rule blocks when the subject is itself a pronoun, however:

(5.59) a. Ich weiß, daß sie ihn liebt.
    "I know that she him loves," i.e.
    I know that she loves him.

b. *Ich weiß, daß ihn sie liebt.

A single-word pronoun can move leftwards over the full NP in (5.58b), but it cannot move over another single-word pronoun.

To see why not, consider the EIC scores for these sentences. The analysis is now more complicated because we are dealing with two higher nodes, S and VP. I also need to anticipate my discussion of VP recognition in German (cf. ch. 6.3.3) by pointing out that these non-subject pronouns can construct VP by Grandmother Node Construction (cf. ch. 6.2.2), and for this reason they are typically placed on the leading leftmost edge of the VP in German. If the pronoun is fronted to the left of the subject, I assume that it continues to construct VP, which then results in a discontinuous VP, though this is not crucial to my predictions. EIC ratios are calculated left to right in order to reveal the precise benefits of rearranging the single-word pronoun, if any.

(5.58') a. ... daß _S[NP[die Frau] VP[ihn liebt]]_

| S CRD: |                 |
|       | 1/2 2/3         |
|       | 50% 67%         |

S Aggreg. = 59%

| VP CRD: |                 |
|         | 1/1 2/2         |
|         | 100% 100%       |

VP Aggreg. = 100%

Average: 79.5

b. ... daß _S[VP[ihn] NP[die Frau] VP[liebt]]_

| S CRD: |                 |
|       | 1/1 2/2         |
|       | 100% 100%       |

S Aggreg. = 100%

| VP CRD: |                 |
|         | 1/1 2/4         |
|         | 100% 50%        |

VP Aggreg. = 75%

Average: 87.5
A full NP subject in (5.58a) delays access to the VP within S. By preposing the non-subject pronoun in (5.58'b), the two ICs of S, NP and VP, will be constructed by the first two words of the parse string, resulting in a 100% ratio. Even though this VP is now discontinuous, the benefits for S outweigh the disadvantages for VP, and the average for the two CRDs rises from 79.5 to 87.5. Without the discontinuity assumption, the average for (5.58'b) would be higher (96): there would be three ICs of S (aggregate 92%); and one IC of VP (100%). Either way, Pronoun Fronting is functionally motivated.

For the corresponding (5.59'), however, the basic (a) order is already 100%, and so there can be no improvement associated with the rearrangement. In fact, if the VP is discontinuous, as shown in (5.59'b), Pronoun Fronting results in a decrease from 100 to 92. Without the discontinuity, both the input and the output would stand at 100, and there would be no increase. Either way, Pronoun Fronting is predicted not to occur, and it does not.

5.5 Word-order hierarchies
In ch. 3.6.3 I formulated a prediction for grammatical hierarchies involving word order, repeated here:

(3.24) EIC Grammatical Hierarchy Prediction
Given: an IC position P center-embedded in a CRD for a phrasal node D (i.e. there is non-null material to both left and right of P within D);
Testing EIC's grammatical predictions

a set of alternative categories \( \{C\} \) that could occur in position \( P \) according to the independently motivated and unordered PS rules of the grammar;

Then: if a category \( C_i \) from \( \{C\} \) produces low EIC ratios for \( D \) and is grammatical, then all other categories from \( \{C\} \) that produce improved ratios will also be grammatical.

We can now proceed to test this.

Consider the Prepositional Noun-Modifier Hierarchy of Hawkins (1983) (henceforth PrNMH), given in (5.60):

\[(5.60) \quad PrNMH\]

If a language is prepositional, then if RelN then GenN, if GenN then AdjN, and if AdjN then DemN.

This hierarchy defines a sequence in which various noun modifiers are preposed before the noun in prepositional languages. The sequence, together with exemplifying languages, is shown in (5.61):

\[(5.61) \quad \begin{align*}
\text{a. Prep: } & \emptyset & [\text{NDem, NAdj, NGen, NRel}] \text{ e.g. Arabic, Thai} \\
\text{b. Prep: } & \text{DemN} & [\text{NAdj, NGen, NRel}] \text{ e.g. Masai, Spanish} \\
\text{c. Prep: } & \text{DemN, AdjN} & [\text{NGen, NRel}] \text{ e.g. Greek, Maya} \\
\text{d. Prep: } & \text{DemN, AdjN, GenN} & [\text{NRel}] \text{ e.g. Maung} \\
\text{e. Prep: } & \text{DemN, AdjN, GenN, RelN} & \emptyset \text{ e.g. Amharic}
\end{align*}\]

If a category low on PrNMH precedes the noun, so do all the higher categories.

In structural terms this hierarchy is defined on categories that can precede \( N \) within an NP that is contained in a PP, as shown in (5.62):

\[(5.62) \quad PP[P_{\text{NP}}[\underline{N} \ldots ]]\]

\begin{align*}
\text{a. } & \emptyset \\
\text{b. } & \text{Dem} \\
\text{c. } & \text{Adj} \\
\text{d. } & \text{PossP} \\
\text{e. } & \text{S}
\end{align*}

These categories will be center-embedded in the CRD for PP, as defined in (3.24), if they intervene between the category that constructs the first IC of PP, P, and whatever category constructs the second, NP. According to our theory, the NP can be constructed by \( N \) and the determiner Dem. Both \( PP[P_{\text{NP}}[N \ldots ]] \) and \( PP[P_{\text{NP}}[\text{Dem} \ldots ]] \) can have \( 2/2 = 100\% \) ratios, therefore, and so Dem does not introduce any center embedding when it precedes the noun and it can be just as optimal for EIC as when it follows. The frequency distribution between these two options across languages is almost exactly
50–50. The other categories of (5.62) do introduce center embedding and less-than-optimal ratios. A single-word Adj will delay access to NP by a small amount (unless morphological agreement affixes, for example, are capable of constructing NP by additional construction principles, cf. ch. 6.2.3; i.e. the category Adj is not uniform from a parsing perspective and this will play an additional role in determining its ranking on PrNMH; in the present context I shall consider length alone). PossP consists of an NP plus Poss, and this will add more words to the CRD for PP when it precedes N. Š is the most complex category of all and will obviously add the most.

Table 5.11 *EIC and the PrNMH*

<table>
<thead>
<tr>
<th>Structure</th>
<th>IC-to-word ratio (PP)</th>
<th>N in sample</th>
<th>Percentage preposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. (PP[PNP[N \text{Dem}]])</td>
<td>100%</td>
<td>29</td>
<td>48%</td>
</tr>
<tr>
<td>2. (PP[PNP[\text{Adj} N]])</td>
<td>67%</td>
<td>17</td>
<td>28%</td>
</tr>
<tr>
<td>3. (PP[PNP[\text{PossP} N]])</td>
<td>50%</td>
<td>8</td>
<td>13%</td>
</tr>
<tr>
<td>4. (PP[PNP[\bar{S} N]])</td>
<td>33%</td>
<td>1</td>
<td>2%</td>
</tr>
</tbody>
</table>

Total n in sample = 61 languages (from Hawkins 1983)

Percentage postposed

<table>
<thead>
<tr>
<th>Structure</th>
<th>Percentage postposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1'. (PP[P \text{NP}[N \text{Dem}]])</td>
<td>52%</td>
</tr>
<tr>
<td>2'. (PP[P \text{NP}[N \text{Adj}]])</td>
<td>72%</td>
</tr>
<tr>
<td>3'. (PP[P \text{NP}[N \text{PossP}]])</td>
<td>87%</td>
</tr>
<tr>
<td>4'. (PP[P \text{NP}[N \bar{S}]])</td>
<td>98%</td>
</tr>
</tbody>
</table>
The EIC scores for the PP therefore decline down the hierarchy of (5.62), as shown in table 5.11. I have assigned an aggregate of one word to Adj, two to PossP, and four to S. Dem, N, and P are all assumed to be (single-word) MNCCs. With these assumptions the IC-to-word ratios decline sharply down (5.62b)–(5.62e), see 1'–4' respectively in table 5.11. Meanwhile the corresponding structures with postposed noun modifiers all have optimal ratios, see 1'–4'. As a result, the disparity between preposing and postposing becomes all the greater, and as it does so the number of exemplifying languages in our sample with the preposed order declines. This is predicted by EIC's Basic Order Prediction of (3.21) (cf. ch. 3.6.1): “any basic orders whose EIC ratios are not optimal will be more or equally frequent across languages in direct proportion to their EIC ratios,” a prediction that mirrors the corresponding marked case prediction for performance (cf. (3.18) in ch. 3.5.1 and the exemplifying data of ch. 4).

The implicational structuring of PrNMH is predicted by (3.24). A center-embedded S has the lowest EIC score of 33%. If this is grammatically tolerated, then the other center-embedded categories, i.e. PossP and Adj, will be tolerated as well, with 50% and 67% respectively. They are. Similarly, if the PossP is tolerated, so is the Adj. This is also correct, at least with significantly more than chance frequency.

The remaining implication, if AdjN then DemN, is explained as follows. The Basic Order Prediction (3.21) predicts optimal EIC ratios for the PP in the unmarked case. There are two candidate structures, as we have seen: [P [N ... ]] and [P [Dem ... ]]. Between them they account for just about all the prepositional languages in our sample, i.e. almost without exception prepositional languages have a leftmost N or Dem within the NP, permitting recognition of the ICs of PP within an optimally efficient viewing window of two words.

From this it follows that if the Adj precedes the N, then Dem, as the other available MNCC for NP, should precede both Adj and N. That is, a prepositional language with [Adj N] should have the order [Dem Adj N] when all three categories co-occur, not *[Adj Dem N] nor *[Adj N Dem]. As a result, we get the implicational universal “if [P [Adj N]] then [P [Dem N]]” when we consider binary branching structures only: the preposing of Adj before N implies the preposing of Dem. Empirically, this is again a statistically significant generalization (cf. Dryer 1993).

Conversely, if N precedes Adj in a prepositional language, our theory predicts that Dem can precede or follow N, since both can construct NP: [Dem N Adj], [N Dem Adj], and [N Adj Dem] are all possible, though there is a preference for the first and third over the second (cf. Hawkins 1983: 119).
We therefore have the following center-embedding hierarchy for PrepP:

\[(5.63)\quad \text{PP: Adj} > \text{PossP} > \bar{S} \quad (" > \) = "more center-embeddable than")\]

This hierarchy defines the grammaticalization sequence in which categories can be preposed before N in an NP contained in a PrepP. The structures that are worst for processing, according to EIC, will be grammaticalized only if all easier structures are also grammaticalized within the same domain. It was this same pattern that we observed when explaining Keenan and Comrie's Accessibility Hierarchy in terms of increasingly complex Structural Domains (specifically Relativization Domains) in ch. 2.3. A grammar that resorts to a complex SD of a certain type (Relativization Domain, CRD for PP, etc.) must already have grammaticalized all potentially available structures of this same type that are of intermediate and lesser difficulty. In other words, there are gradual and cumulative increases in complexity in the grammatical variation data within common SDs across languages. We do not find a very complex structure co-occurring with nothing but optimal structures elsewhere in its SD, e.g. RelN in a CRD for PrepP co-occurring with noun-initial order for every other modifier, or relativization on a genitive when the only other relativizable position is a subject. Rather, if complexity is allowed at all, it is allowed first to a small degree, then to a greater degree, and only then to a large degree, and it is this that explains the structuring of a great many universal hierarchies. They are grammaticalizations of different degrees of processing difficulty resulting from structural complexity (recall the discussion in ch. 1.4), or from other forms of complexity.\(^{12}\)

The general logic of this explanation for (5.63) is supported by other word-order hierarchies. There are many languages in which an \(\bar{S}\) center-embedded in the CRD for S is ungrammatical, while an NP is grammatical. English is an example. Recall the data of (2.1a, b) repeated here as (5.64):

\[(5.64)\quad \begin{align*}
\text{a. } & \quad \text{\(S\)}[\text{Did } \text{s}[\text{that John failed his exam}] \text{vp[surprise Mary]]}\] \\
\text{b. } & \quad \text{\(S\)}[\text{Did } \text{NP[this fact]} \text{vp[surprise Mary]]}\]
\end{align*}\]

The CRDs for S are indicated, making it clear that \(\bar{S}\) and NP are center-embedded in the sense of (3.24). This gives us a hierarchy for S:

\[(5.65)\quad \text{S: } \text{NP} > \bar{S} \quad (" > \) = "more center-embeddable than")\]
If a language permits the center embedding of $S$, it will permit the center embedding of NP. Japanese and Korean permit both (cf. example (1.6a) in ch. 1.3); English permits a center-embedded NP only.

The German VP structures of table 5.10 suggest an extension of this hierarchy (cf. section 5.4.1). A center-embedded NP within VP, e.g. $vp[NP NP V]$, is completely grammatical; a center-embedded $S$ is not, *$vp[NP S V]$.

A center-embedded infinitival $VP$ was shown to be almost as bad for EIC as the center-embedded $S$, unless the former consists only of the verb plus zu. There is uncertainty over the precise grammaticality status of ? $vp[NP VP V]$ in German, which I commented on. In this context, the EIC scores coupled with this uncertainty are sufficient to position $VP$ between NP and $S$:

\[(5.66) \quad VP: \quad NP > VP > S \quad (" > " \quad "more center-embeddable than")\]

And this hierarchy defines a clear sequence in which languages can grammaticalize these options: if the $S$ is grammatical, so are the others; if the $VP$ is ungrammatical, so is the $S$; etc.

The hierarchies of (5.63), (5.65), and (5.66) can be seen to conform to a very simple pattern: the aggregate length of the categories increases down each of them. As a result, the distance separating the first from the last IC of the CRD increases, and IC-to-word ratios decline. There are presumably more such hierarchies awaiting discovery in all phrasal categories, once the relevant daughter categories have been investigated in more languages.

Consider finally the hierarchy of (5.63) in languages with PPs that contain postpositions rather than prepositions. The relative order of the ICs, NP and P, is now reversed, and if Adj, PossP, and $S$ are center-embedded, they will follow the MNCC for NP and precede P, the MNCC for PP, i.e. $[[N Adj] P]$, $[[N PossP] P]$, etc. The prediction of (3.24) is that the center embedding of a lower category requires the center embedding of all higher categories. $[[N PossP] P]$ requires the co-occurrence of $[[N Adj] P]$, therefore. That is, if NGen then NAdj in postpositional languages (equivalently: if AdjN then GenN). This generalization holds exceptionlessly for the Expanded Sample of Hawkins (1983: 67–68). The corresponding prediction for $[[N S] P]$ does not hold, however. There are many languages with this basic order co-occurring with both prenominal adjectives and genitives (e.g. Brahui, Quechua, Yaqui, cf. Hawkins 1983: 83, 85), and although the postposing of the relative clause $S$ to the right of P (cf. $PP[NP[N] P NP[S]]$) is required in many of these, creating what is most plausibly a discontinuous NP structure of the kind we saw for Extrapolation from NP in German (ch. 4.3.5), it can be retained in many (cf. section 5.1.3). As a result, we have a systematic set of exceptions to
(5.63) involving the sentential category, $\tilde{S}$. The very productivity of these counterexamples suggests that we are dealing with some additional principle of generality involving the processing of sentential nodes. In the next section I shall return to these data in the context of a broader discussion of left–right asymmetries.

5.6 Left–right asymmetries

In Hawkins (1988b) and Hawkins and Cutler (1988) I discussed some “left–right asymmetries” in syntax and morphology. As the name suggests, these asymmetries involve a skewing in the distribution of categories throughout the parse string, for example light categories to the left, heavy $\tilde{S}$s with initial complementizers to the right. This skewing has been grammaticalized in many rules and principles, in response, I would argue, to EIC. The asymmetries are ultimately consequences of the fact that language is produced and comprehended in an item-by-item manner from left to right, i.e. in a temporal sequence. Such asymmetries were explicitly predicted not to exist by Miller and Chomsky (1963: 472), not surprisingly given the assumption of pure acceptability in Chomsky (1965); see ch. 1.3 for discussion. If grammars are viewed as being independent of processing, except perhaps at a very general and abstract level in the evolution of the innate grammar, then one would not expect to find sensitive reflections of on-line processing mechanisms in the grammatical details of individual languages and language types. But if one holds, as I do, that much of syntax and just about all the rules of linear ordering are direct responses to processing ease and are grammaticalizations of efficiency principles such as EIC, then one does expect to see the temporal sequence of speech reflected in the grammar. In this section we shall review the left–right asymmetries considered up to this point, fill in some explanatory gaps, and consider some more examples.

Left–right asymmetries can be observed and described at different levels of generality. We see them in the positioning rules for individual categories within individual structures. We see them in the positioning of whole sets of categories of a certain type throughout an entire phrase or phrases. And we see them in certain discrepancies between grammars of one type and those of another, e.g. between head-initial and head-final languages.

At the lowest level of generality, we have seen many examples of asymmetrical positioning for individual categories in individual structures. Pronouns are skewed to the left of the VP by basic order rules or are rearranged
Testing EIC's grammatical predictions

leftwards, in English and many other languages (cf. sections 5.2.2, 5.3.1, and 5.4.2). This skewing provides earlier access to the daughter ICs of VP in the left-to-right temporal sequence of the VP than if a longer IC were first to add several words to the CRD for VP. If a long IC is constructed and recognized on its right periphery, it will also be skewed to the left, where it adds at most one word to the CRD for VP and precedes shorter ICs including pronouns. This explains the basic order \( \text{vp[pp[NP P] NP V]} \) in many verb-final languages (cf. 5.2.2). If the NP preceded PP, the CRD would be longer, and temporal access to VP structure would be delayed.

A relative clause of the type \( \text{§[C S]} \) is skewed to the right of a multi-branching NP (cf. 5.2.1) where it permits shorter ICs to be processed first (e.g. N and Adj), providing faster access to the internal structure of NP. Relative clauses of the opposite type, \( \text{§[S C]} \), are skewed to the left, where they add at most one word to the CRD for NP (namely C, the category that constructs the relative clause, cf. ch. 6.3.2). Categories such as N and Adj can then follow \( \text{§[S C]} \) and achieve optimal EIC ratios (cf. C and D in table 5.8). Conversely N and Adj can both precede in a minority of languages and achieve relatively good scores (cf. E and F in table 5.8), because two of the three ICs of NP can still be recognized early in the CRD for NP. Both orderings of \( \text{§[S C]} \) are therefore advantageous for a temporally sequenced parse string, with preposting clearly preferred. The positioning of sentence complements within VP or S exhibits an identical skewing (cf. 5.4.1). \( \text{§[Comp S]} \) is ordered or rearranged to the right, \( \text{§[S Comp]} \) preferably to the left, where it adds only one word to the higher CRD. In some languages the latter is positioned to the right.

At a higher level of generality, we can state a generalization about all single-word and multi-word ICs within a phrasal category. This was attempted for English in sections 5.3.1 and 5.3.2, and the result was a clear leftward skewing for the former and a rightward skewing for the latter: if an IC is possibly single-word within NP, then it precedes all necessarily multi-word ICs \((= (5.28c))\); if an IC is necessarily single-word within VP, then it precedes all possibly multi-word ICs \((= (5.25c))\). The result of this skewing is faster access to NP and VP structure respectively. We can also capture a generalization across phrasal categories (VP, NP, PP, and AdjP) for the ordering of complement categories of different aggregate lengths: a shorter \( m \) IC precedes a longer \( m \) IC, cf. (5.31'). For example, \( \text{vp[V NP pp[P NP]]} \) is the preferred relative order of NP and PP in the VP and elsewhere in English-type languages. The result is the fastest possible temporal access to as much of the internal structure of these phrases as possible.
5.6.1 Typological asymmetries

From a comparative linguistic perspective we see left-right asymmetries across different language types. The asymmetries seem to involve the sentential nodes, $\bar{S}$ and $S$, and examples were given in section 5.1.2. Verb- (and head-) initial languages are almost exclusively Comp-initial in $\bar{S}$, i.e. $\bar{S}[\text{Comp S}]$. Verb- (and head-) final languages exhibit both Comp-final positioning, $\bar{S}[\text{S Comp}]$, as in Japanese and Korean, and Comp-initial positioning as in Persian and German. This cross-categorial asymmetry results in a skewing in favor of $\bar{S}[\text{Comp S}]$ throughout the world's languages, given that the distribution of VO to OV is otherwise roughly equal (cf. Hawkins and Cutler 1988, Hawkins and Gilligan 1988). It also results in a further asymmetry in the relative positioning of $V$ and $\bar{S}$. Since there are more $\bar{S}[\text{Comp S}]$ languages than $\bar{S}[\text{S Comp}]$ languages, and since the former will prefer postposing in the containing $S$ and the latter preposing, there will be more $[V \bar{S}]$ structures and languages than $[\bar{S} V]$ structures and languages.

In Hawkins (1990) I offered an explanation for these typological left-right asymmetries in terms of a second principle interacting with EIC and involving temporary ambiguity and garden-path avoidance. The explanation was based on Frazier's (1979a, 1985) Minimal Attachment principle, and on an idea that had been first proposed by Antinucci et al. (1979), and then experimentally validated by Clancy et al. (1986). The logic of the explanation was as follows. If an $\bar{S}[\text{S Comp}]$ (or an $\bar{NP}[\bar{S} N]$) is contained within a verb-final clause, there will be regular opportunities for misanalyzing main and subordinate clause arguments as arguments of the same verb. For example, in $[\bar{NP} \bar{S}[\text{Comp V}]]$ and $[\bar{NP} \bar{NP}[\bar{S} N] V]$ structures, the matrix NP will often be interpreted on-line as an argument of the subordinate verb within the left-branching sentential node, as in the following Japanese example (taken from Clancy et al. 1986):

\[(5.67)\] Zoo-ga \(\bar{NP}[\bar{S}[\text{kirin-o taoshi-ta}] \text{shika-o]} \text{nade-ta}.\]

"Elephant-su giraffe-OBJ knocked down deer-OBJ patted," i.e.

The elephant patted the deer that knocked down the giraffe.

By Minimal Attachment this sentence is first interpreted as "the elephant knocked down the giraffe," and this interpretation is then subsequently revised to the translation given. In order to avoid these kinds of garden paths, a number of head-final languages bring forward the relevant MNCC for $\bar{S}$ and NP, thereby indicating unambiguously the onset of subordinate
clause material, i.e. \([\text{NP}_\text{S}[\text{Comp S}] \text{ V}]\) and \([\text{NP}_\text{NP}[\text{N S}] \text{ V}]\). This then results in the cross-categorial and ordering asymmetries.

I still think that this explanation is on the right lines. It is not completely adequate, however, as Dryer (1991) has shown. The problem is that it is not general enough. With left-branching relative clauses, the deletion of the NP relativized on means that there are always missing argument positions in the subordinate \(\tilde{S}\) that can potentially be filled by matrix NPs that have already been encountered, as in (5.67). But when the \(\tilde{S}\) is a complete subordinate clause with a complementizer or adverbial subordinator (cf. section 5.1.2), the potentially ambiguous structures are significantly fewer, and the ambiguities can always be avoided by \(\tilde{S}\)-sensitive rules preposing, or even postposing, the \(\tilde{S}\), so that the arguments of each verb are gathered up unambiguously as that verb is encountered. Yet the pattern of left-right asymmetries is the same in both cases.

In this context I want to propose a more general solution. The crucial parsing parameter that underlies all these asymmetries seems to reduce to whether or not the grammar provides immediate recognition of subordinate-versus main-clause status. Recall the English garden path discussed by Bever (1970) that results when a complementizer is deleted in sentential subject position (cf. (1.5), renumbered here as (5.68)):

\[
(5.68) \quad \begin{align*}
a. \text{That Mary is sick surprised Bill.} \\
\text{b. *Mary is sick surprised Bill.}
\end{align*}
\]

Example (5.68b) is ungrammatical because if the complementizer is deleted, the initial material in the parse string will always be understood as a main clause, and there will always be a structural misanalysis and garden path, in violation of Frazier's (1985) Impermissible Ambiguity Constraint. English is therefore a language in which the grammar provides "immediate matrix disambiguation." The parser makes an immediate decision about the main- or subordinate-clause status of a clause: if there is a complementizer, the clause is subordinate; otherwise, it is the matrix, or "root" clause (cf. Emonds 1976). If the complementizer is deleted in environments such as I believe John is an idiot, then John will first be assigned to the matrix, and the task of constructing the subordinate clause will fall on the finite verb is, which constructs S and whose positioning after the matrix will indicate subordination. John will then be reassigned to the subordinate clause (cf. ch. 4.5).

English is an immediate matrix disambiguation, or IMD language. More generally, all head-initial languages can be predicted to be IMD languages,
because all phrases are constructed by left-peripheral MNCCs in these languages, and subordination indicators such as complementizers and adverbial subordinators are simply a type of MNCC. If the MNCC is present, the clause is unambiguously subordinate; if it is not, there will be no unique constructor of subordinate-clause status, and a matrix interpretation will result, as in (5.68b). Moreover, given the significance of the main clause/subordinate clause distinction, we do not expect exceptions to the predicted, leftward positioning of MNCCs for sentential nodes, even though there are minority exceptions in the case of smaller and less significant nodes (e.g. PossP before rather than after N within NP in certain prepositional languages).

A head-final language, on the other hand, will not be able to signal immediate matrix disambiguation via a complementizer if the complementizer stands on the right periphery of its clause. In the processing of Japanese the parser has to wait for evidence about subordination, and it cannot immediately assign matrix or subordinate status to each new S encountered, as we have pointed out repeatedly. The result is considerable on-line ambiguity; cf. Inoue (1991) for numerous examples. (Better terms here would be “vagueness” or “indeterminacy,” given the huge number of structural options that are potentially available in a bottom-up parse, most of which the hearer is psychologically unaware of; “ambiguity” normally refers to a small number of choices that the hearer is psychologically aware of and from which a selection must be made.)

But because the main clause/subordinate clause distinction is a significant one, it can be argued that there will be a parsing preference for immediate matrix disambiguation in all languages, and we should expect to find this preference reflected even in head-final languages. One option available to them is to invert the order of Comp and S within ـ، and the order of N and S within NP, and thereby indicate immediately whether a clause is subordinate. If an otherwise head-final language chooses this option for sentential nodes, the associated parser can provide immediate matrix disambiguation, just as it does for English. These head-final languages (Persian, German, Yaqui, etc.) will therefore be inconsistent with respect to these nodes, though they can still achieve good EIC scores by an appropriate positioning or rearrangement of the relevant categories.

So head-final languages are compatible with a choice. All things being equal, EIC prefers final complementizers, and final head nouns within NP, both of which will not permit the parser to provide immediate matrix disambiguation, because there is indeterminacy in the parse string. But an
indeterminacy can be resolved, and all things being equal the parser also prefers early decision-making to a delay. So some of these head-final languages do resolve the indeterminacy immediately, and position their MNCCs for subordinate clauses (complementizers and adverbial subordinators) to the left.\textsuperscript{13} By contrast, head-initial languages will automatically be compatible with immediate matrix disambiguation, and there will be no motivation in such languages for introducing a mirror-image inconsistency. Hence we have a left–right typological asymmetry whereby \texttt{s}{\texttt{Comp S}} and \texttt{NP}{\texttt{N S}} structures are found productively in both head-initial and head-final languages, whereas \texttt{s}{\texttt{S Comp}} and \texttt{NP}{\texttt{S N}} are found only in head-final languages. Correspondingly, immediate matrix recognition can be provided by parsers accompanying grammars of both types, whereas its absence is characteristic only of head-final languages.

We can now see the Japanese garden path of (5.67) in its proper context. This sentence fails to give a clear on-line indication of subordinate-clause onset; and in addition leads the parser to assign arguments from different clauses to one and the same verb. The latter problem entails the former, but the former does not entail the latter. A grammar can simply fail to give immediate information about subordination, and this alone, we now claim, is a sufficient motivation for the productive existence of inconsistent head-final languages such as Persian, German, and Yaqui, and for the left–right asymmetry in sentential nodes. This explains the otherwise unpredicted productivity of structures of the type (5.5D) [[Comp S] V] and (5.6D) [[N S] V] in section 5.1.2, and the absence of [[S Comp] and [S N]] structures in head-initial languages.\textsuperscript{14} We can also explain a quantitative difference between (5.5D) and (5.6D). If, as in (5.67), prenominal relative clauses involve additional Minimal Attachment problems over and above the simple failure to provide immediate matrix disambiguation, then we would expect more grammars to convert [S N] into [N S] than convert [S Comp] into [Comp S]. This is the case. I mentioned in section 5.1.2 that 31\% of OV languages have clause-initial adverbial subordinators in Dryer’s (1992) sample. For relative clauses the proportion is almost double: 37/63 = 59\% of OV languages have [N S] in his sample.

We also have an explanation now for the otherwise surprising absence of a parallelism between the internal structure of relative clause Šs and that of sentence complements. Corresponding to the productive (5.5D) [[Comp S] V], there is no productive (5.69D) with the constructing category C to the left of the relative clause (I repeat and juxtapose (5.5), and also (5.6), for convenience):
Cross-linguistic frequencies for (5.69) pattern like they do for the other binary branching structures of section 5.1: (5.69A) and (5.69B) are productive; (5.69C) is less so; and (5.69D) is on this occasion not attested at all. Why are there no C-initial relative clauses? The answer I would propose is that the (5.69D) structure is pre-empted by (5.69A) and (5.69C) in a head-final language (cf. (5.6D)). The motivation for the left-right asymmetries involving (5.5D) and (5.6D) is to indicate subordinate-clause status immediately. A relative clause has two potential means of doing this: bring forward the head noun ((5.69A) and (C)); or bring forward the constructor of S (5.69D). Either way the problematic structure (5.69B) is avoided. The problem with the (D) solution is that it constructs the lower node, i.e. the relative clause, first. If the constructing category, C, is (as it often is) a general subordinating conjunction, then there will still be structural ambiguity, this time with sentence complements. An initial head noun, on the other hand, will construct the highest node, NP, and thereby indicate the subordinate status of the clause as soon as material from this clause is processed, and avoid a structural ambiguity with other types of subordinate clauses. It also has yet another advantage: the head noun precedes the empty position relativized on and provides a lexically filled phrase for the on-line interpretation of the empty node; the reverse ordering results in a processing delay. There is independent evidence from languages with head-internal relative clauses in the syntax in favor of a leftward skewing for the NP that functions semantically as the head, for example in Yavapai (cf. C. Lehmann 1984: 119–121). In addition, parsing a full NP before its gap is an easier operation for the parser than the reverse (cf. Janet Fodor 1984).
For all of these reasons, then, (5.69D) is functionally unmotivated: the parallelism with (5.5D) fails because (5.69D) is pre-empted by (5.69A) and (C). As a result the binary branching predictions of EIC go through in the usual way (A + B most productive, C significantly less, D even less so). There is still a typological asymmetry, however. Head-initial languages are exclusively of type (5.69A), whereas head-final languages are found with (5.69A), (B), and (C), the net result of which is to make postnominal relative clauses (much) more frequent across all languages than prenominal relatives. The difference in this case involves the total absence of the (D) structure (5.69D), and its productive presence in the other typological left–right asymmetries.

Consider at this point two further examples of left–right asymmetries that have been much discussed in the literature and that can now be placed within a more general explanatory framework: the relative positioning of subject and object NPs, and the positioning of topic and focus nodes.

5.6.2 Subject, object, and verb position

It was Greenberg (1966) who pointed out that the dominant order of subject and object across languages is "almost always one in which the subject precedes the object," while the verb can occur in all three positions, SOV, SVO, and VSO. In the context of the present discussion, this means that there is a leftward skewing for subjects, and a rightward skewing for objects, within the clause. In order to see why, we need to be clear about the structure of transitive sentences, and about the impact of different orderings on the relevant CRDs. In particular, we need to consider the consequences of having, versus not having, a VP. In languages that have a VP, the direct object will typically be contained within this node, whereas the subject will be immediately dominated by S. Within the CRD for S we will therefore be dealing with the relative order of NP and VP, rather than with subject and object NPs directly. In languages without a VP, there will be three ICs of S: the subject NP; the object NP; and V.

Let us begin with the predictions for languages with a VP. The major prediction that we make (by the Basic Order Prediction (3.21) in ch. 3.6.1) is that \( s[\text{NP VP}] \) will be grammaticalized in the unmarked case, in preference to \( s[\text{VP NP}] \). The VP is a more complex node than the subject NP. It can include a direct object, an indirect object, various oblique NPs, adverbial phrases, etc., in addition to the verb. The subject, by contrast, is just a single NP. In head-initial languages like English, with ICs of the type \( mIC \), positioning the shorter subject NP before the VP results in significantly improved EIC
Left–right asymmetries 329

ratios compared with the reverse. Hence, \( mS \ [\text{V}_m \text{O}] \) is expected, and \([\text{V}_m \text{O}] \)
mS is not. But equally, in head-final languages like Japanese with ICs of the
type ICm, it turns out that there is still a preference for positioning NP before
VP in the S domain, even though heavy constituents are typically arranged or
rearranged to the left in this language (cf. ch. 4.1.5). That is, \( S_m \ [\text{O}_m \text{V}] \) is
expected, and \([\text{O}_m \text{V}] \) S_m is not. One reason for this is the fact that this node
appears to be constructable by categories other than V that are unique to VP
(such as accusative case marking, cf. ch. 6.2.2) and that occur in advance of
the verb. The right-peripheral recognition of the subject \( S_m \), in conjunction
with the construction of VP by (say) \( O_m \), results in a shorter recognition
domain for the sentence node when \( S_m \) precedes \( O_m \) (and VP) than it does
when VP precedes \( S_m \) (cf. table 5.13 below).

Tomlin (1986) gives quantitative data of relevance to these predictions.
Within his sample of 402 languages, there are 168 with SVO (42%) and
180 with SOV (45%); by contrast there are just 12 with VOS (3%), and
five with OVS (1.2%). (The remaining languages, 37 in all [9%], have
VSO, cf. (5.70)). The combined distribution of SVO and SOV to VOS and
OVS is therefore 348/17, or 95.3% to 4.7%. If all of these languages have a
VP, then the unmarked case prediction in favor of [NP VP] is clearly fulfilled.
They all have an adjacent V and O, and so the existence of a VP node is
certainly possible. But even if some (or many) do not, these data are still
significant, since there are at most 17 that can be exceptional with the order
[VP NP], compared with 348 potential [NP VP] structures, which makes it
highly probable that the unmarked case prediction is also fulfilled within the
subset that does have a VP.

Consider now all three attested language types with V before O: SVO, VSO,
and VOS. Their relative frequencies in Tomlin’s sample are given in (5.70):

(5.70)  

<table>
<thead>
<tr>
<th>Language Type</th>
<th>Number of Languages</th>
<th>Percentage of Sample</th>
<th>Percentage of VO Languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVO</td>
<td>168</td>
<td>42%</td>
<td>77%</td>
</tr>
<tr>
<td>VSO</td>
<td>37</td>
<td>9%</td>
<td>17%</td>
</tr>
<tr>
<td>VOS</td>
<td>12</td>
<td>3%</td>
<td>6%</td>
</tr>
</tbody>
</table>

There are two frequency patterns of note here. First, SVO is favored in the
great majority of these languages relative to both the other options, by a ratio
of 77% to 23%, i.e. over three to one. Secondly, frequencies decline as the
subject moves to the right. Within the minority types, VSO and VOS, the
former is also preferred by a ratio of three to one.

EIC can explain these two patterns. The Basic Order Prediction predicts
SVO in the unmarked case, while VSO turns out to have better ratios than
VOS, often significantly so. VSO can be quite an efficient order, therefore,
though not as good as SVO, and this explains its relatively frequent, but still not preferred, status.

In order to illustrate EIC's predictions, we need to make assessments of the aggregate lengths of S, V, and O, and we must consider these orders both with and without a VP. I shall assume that VO languages, being predominantly head-initial, have NPs of the type $m_NP$ rather than $NP_m$. As for the aggregate lengths, I shall assign one word to V, as usual. For subject and object we need to examine the performance data from languages that are head-initial. The evidence that I have suggests that the object is longer than the subject in these languages. For example, in the Polish data of table 4.19 the mean length of S in the order SVO is 2.5 words, while that of O is 3.7, i.e. 1.2 words more. SVO accounts for 69% of all orders in Siewierska's (1991b) data. In the VSO order (accounting for a further 7% of the data), the O is 5.5 words longer than S. Only in the VOS order (11% of the data) is the S longer than O — by 3.7 words. In the remaining (13%) of her data, O and S are roughly equal in size. The aggregate for O is therefore longer in Polish than for S. The Greek data of table 4.20 pattern the same way, with O significantly longer than S. I do not have aggregated data for English, but impressionistically the O is longer than the S. Indirect support for this comes from Givon's (1979: 51–53) text counts from written English, in which 91% of his subjects were definite, compared with only 56% of direct objects. Indefinite NPs are typically longer than definite ones, since they require more material to identify the unknown referent for the hearer (cf. section 4.4.4). This in turn suggests that O is longer than S in English as well.

I shall therefore assign one more word to O than to S in VO languages, e.g. two to S and three to O. With these assignments, EIC's scores for the three VO orders of (5.70) are given in table 5.12, assuming a VP (and VP discontinuity in VSO languages). The order with the highest aggregated ratio for the S and VP domains is SVO (84%), followed by VSO (75%), and then VOS (70%). EIC therefore predicts SVO in the unmarked case, and more VSO than VOS languages in the marked case, which matches the data of Tomlin's sample. Notice that VSO has a better ratio than VOS, even though there is discontinuity between the ICs of VP, V, and O. This discontinuity extends the VP domain and results in a low ratio for VP. But the intervening subject produces a perfect ratio for the S (i.e. clause) domain: the V constructs VP, the first IC of S; and the subject can then immediately construct the second IC, i.e. $2/2 = 100\%$. By contrast, VOS has a very low ratio for the S domain because of the long initial VP (40%), and even though the VP is optimal, the aggregated score is lower than that for VSO.
The predictions of table 5.12 do not depend crucially on these particular word-total assignments. If \( m_S = 1 \) word and \( m_O = 2 \), for example, SVO would have an aggregate of 100%, VSO would be 84%, and VOS 75%, preserving the same relative ranking. Even if we were to assign the same number of words to \( m_S \) and \( m_O \), e.g. two, we would still define an unmarked preference for SVO (84%), but we would not be able to discriminate between VSO and VOS in this case, both of which would have 75%.

Consider now the effect of having a flat structure, without a VP. Let us assign \( m_S = 2 \) and \( m_O = 3 \), as in table 5.12. There is only one domain to consider in this case, the S domain, consisting of three ICs, \( m_S \), \( m_O \), and V. The ratio for SVO is \( \frac{3}{4} = 75\% \). This is lower than the 84% with the VP assumption. I do not think it is far-fetched to suggest that this higher score provides a motivation for having a VP in an SVO language. If words are grouped into phrases in this way, the parser can achieve a more efficient recognition of the sentence and its structure, even though there is more structure to process! This seeming contradiction follows from our general theory, though we have not made any explicit predictions for constituent

Table 5.12 *EIC and VO languages*

Assume: \( m_S = 2 \) words, \( m_O = 3 \), \( V = 1 \)

- VP dominates V and \( m_O \) (even when discontinuous)
- V or O constructs VP

Total n in sample = 402 languages (217 VO languages), from Tomlin 1986

<table>
<thead>
<tr>
<th>Structure</th>
<th>Agg. IC-to-word ratio</th>
<th>N in sample</th>
<th>Percentage of VO languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ( s[m_S , \text{VP},[V , m_O]] )</td>
<td>S CRD: 2/3 = 67% ( \frac{2}{3} = 67% )</td>
<td>168</td>
<td>77%</td>
</tr>
<tr>
<td></td>
<td>VP CRD: 2/2 = 100% ( \frac{2}{2} = 100% )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. ( s[\text{VP},[V , m_S , \text{VP},[m_O]] )</td>
<td>S CRD: 2/2 = 100% ( \frac{2}{2} = 100% )</td>
<td>37</td>
<td>17%</td>
</tr>
<tr>
<td></td>
<td>VP CRD: 2/4 = 50% ( \frac{2}{4} = 50% )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. ( s[\text{VP},[V , m_O] , m_S] )</td>
<td>S CRD: 2/5 = 40% ( \frac{2}{5} = 40% )</td>
<td>12</td>
<td>6%</td>
</tr>
<tr>
<td></td>
<td>VP CRD: 2/2 = 100% ( \frac{2}{2} = 100% )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
structure as such, preferring to take that as given and define linear ordering preferences relative to it.

VSO languages have a 75% aggregate in table 5.12. Interestingly, they have an identical 75% (3/4) without a VP. This leads us to predict variation in these languages. And indeed there are those such as the Celtic languages that exhibit clear evidence for a VP (cf. e.g. Anderson and Chung 1977; McCloskey 1983; Sadler 1988), while at the same time VSO languages are strongly correlated with head-marking patterns in Nichols’s (1986) sense, and one of the characteristics of head-marking languages, according to Nichols, is flat syntax, i.e. without a VP. With a flat structure, therefore, SVO and VSO have identical ratios of 75%. Since a VP is readily available for the former, however, and since head marking is only weakly correlated with SVO, SVO languages can receive higher EIC scores than VSO, and so our predictions remain.\(^\text{15}\)

VOS languages have an aggregate of 70% in table 5.12. Without a VP, the ratio for the clause would be 3/5 = 60%. This lower score again motivates having a VP, just as it does in SVO languages, and VOS languages are predicted to be less frequent than both SVO and VSO whether they have a VP or not (cf. table 5.14 below).

Let us consider now languages of the opposite type, with O before V. There are three logical possibilities: SOV, OSV, and OVS. Their relative frequencies in Tomlin’s (1986) sample are:

(5.71) \begin{align*}
\text{SOV} & \quad 180 \text{ languages} \quad 45\% \text{ of sample} \quad 97\% \text{ of OV languages} \\
\text{OVS} & \quad 5 \text{ languages} \quad 1.24\% \text{ of sample} \quad 3\% \text{ of OV languages} \\
\text{OSV} & \quad 0 \text{ languages} \quad 0\% \text{ of sample} \quad 0\% \text{ of OV languages}
\end{align*}

One order is again greatly favored compared with the other two (SOV), and there is a preference for one of the two marked orders (OVS). But in contrast to the VO languages, the number of languages with marked orders is extremely small, no more than 3% of the total for OV, and the declining frequency does not match the increasing rightward positioning of the subject.

The best studied OVS language is Hixkaryana, a Carib language of northern Brazil (cf. Derbyshire 1979). Some other OVS languages are briefly discussed in Derbyshire and Pullum (1981) and Pullum (1981). These authors enumerate some languages whose basic order is probably OSV, but most of these are rejected as basic orders in later work by Polinskaja (1989). Even if there are some languages with basic OSV, therefore, the relative ranking of Tomlin’s sample appears to be valid, as is the fairly massive preference for SOV over OVS.
Languages with O before V are predominantly of the type IC$_m$, as in Japanese, except when there are left–right asymmetries for the sentential nodes (cf. section 5.6.1). As we have seen, to possess structures of the IC$_m$ type has profound consequences for parsing and for grammar, including bottom-up processing (cf. ch. 3.3), the preposing of heavy constituents (cf. chs. 4.1.5, 4.1.6, and 5.4.1), etc. It also leads to a hypothesis about the aggregate weights for different constituents in these languages, on the basis of which we can make our EIC calculations and run our basic order predictions.

We have seen clear evidence in the head-initial languages Polish and Greek for direct objects being longer than subjects. We have also seen evidence for longer constituents positioned to the right of shorter constituents in the basic orders of the NP (cf. ch. 5.2.1) and of the VP (cf. ch. 5.2.2) in all head-initial languages. This consistent skewing of heavy material to the right makes sense in these languages, given that ICs are recognized on the left. It explains the positioning of the subject NP before the longer and more complex (and left-recognized) VP. Moreover, since the subject is recognized on its left periphery, SVO sentences will have much better ratios if subjects are short, at the same time that it does not matter how long direct objects are. The direct object is going to be recognized immediately after the verb anyway, so as long as its own internal structure is efficient, a sentence can have good EIC ratios for direct objects of any length. But with each additional word added to a left-recognized subject, the sentence domain deteriorates, and this results in a real performance constraint on subjects in these languages which explains the observed length differential between subjects and objects: object length is unconstrained, so the performance aggregate for objects ends up being longer than for subjects.

For an SOV language of the type IC$_m$, it is not subjects that are constrained, but objects. The subject in Japanese is not recognized as such until the subject-constructing particle -$ga$ is recognized on the right periphery of the NP (cf. ch. 4.1.5). Hence, only at this point is the sentence domain initiated, and the subject can be as long as it likes prior to this. But once the sentence domain is initiated, it is advantageous to achieve VP recognition as fast as possible, just as it is in an SVO language. The VP will be constructed by a V, or by any other daughter that is unique to VP, such as an accusative-marked direct object NP (cf. ch. 6.2.2). Either way, there is a strong constraint in an SOV sentence to minimize the distance between the subject -$ga$ particle, and the verb or direct object particle -$o$, and this means that there should be a performance preference for short direct objects. As a
Testing EIC's grammatical predictions

consequence, the aggregate for objects should now be shorter than for subjects. Our theory therefore leads to a performance prediction regarding the distribution of syntactic weight in basic transitive sentences: SVO languages should have longer objects; SOV languages longer subjects. This, in turn, predicts an altogether different allocation of pragmatic NP types to subject and object positions in VO and OV languages, in accordance with the syntactic weight–pragmatic correlations of ch. 4.4.4. SVO languages should exhibit good given-before-new scores, correlating with short-before-long constituents; SOV languages should exhibit more new-before-given scores, correlating with long before short. Hence, subjects should be more given and definite in SVO languages, less so in SOV languages.

Perhaps remarkably, this different weight distribution for SOV languages turns out to be supported by the Japanese data of table 4.8 section II. Subject NPs with -ga are longer than or equal to object NPs with -o in 79% of transitive clauses (67/85), and the ratio of longer subjects to longer objects is almost two to one (63% to 37%). The figures for Korean are in the same direction as those for Japanese, though less strikingly so. Subject NPs are longer than or equal to object NPs in 68% of transitive clauses in table 4.11 (124/182), and the ratio of longer subjects to longer objects is 53% to 47%. In both languages it is clearly not the case that the direct object is significantly longer than the subject.

These considerations motivate assigning one more word to subjects than to objects in our EIC calculations for OV languages, e.g. three to subjects and two to objects. In fact, nothing hinges on this crucially, and the same predictions will be made if the number of words assigned to each is equal. Therefore the motivation for the reversed assignment of word quantities to S and O is primarily one of performance accuracy. These assignments can also be motivated for OVS and OSV languages, since all these languages seem to have a productive SOV variant as well. If the preposed object leads to a preference for longer objects, then the competing structures in these languages will lead to alternative length preferences that balance each other out, and equal word assignments will be more appropriate. EIC's relative frequency predictions are identical if the word-total assignments are equal, as we have mentioned.

EIC's predictions are set out in table 5.13, assuming a VP. $S_m [O_m V]$ has the highest aggregated score of 84%. The VP has a perfect ratio of 2/2 = 100%, because of the immediate adjacency of the category that constructs the direct object (-o in Japanese) and the verb. The S domain proceeds from the final word or particle of the subject ($S_m$) to the first word or particle that unambiguously constructs VP. If $O_m$ can do this (as is assumed here), the two
Table 5.13 EIC and OV languages

Assume: $S_m = 3$ words, $O_m = 2$, $V = 1$

1. VP dominates $V$ and $O_m$ (even when discontinuous)
2. $V$ or $O$ constructs VP

Total $n$ in sample = 402 languages (185 OV languages), from Tomlin 1986

<table>
<thead>
<tr>
<th>Structure</th>
<th>Agg. IC-to-word ratio</th>
<th>N in sample</th>
<th>Percentage of OV languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s[S_m V P [O_m V]]$</td>
<td>S CRD: 2/3 = 67%</td>
<td>180</td>
<td>97%</td>
</tr>
<tr>
<td></td>
<td>VP CRD: 2/2 = 100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s[V P [O_m V] S_m]$</td>
<td>S CRD: 2/5 = 40%</td>
<td>5</td>
<td>3%</td>
</tr>
<tr>
<td></td>
<td>VP CRD: 2/2 = 100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s[V P] S_m V P [V]$</td>
<td>S CRD: 2/4 = 50%</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>VP CRD: 2/5 = 40%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ICs of the matrix, NP and VP, will be recognized in a three-word viewing window, making a ratio of $2/3 = 67\%$. If only the verb can construct VP, the ratio will be $2/4 = 50\%$. Because of the extra advantages afforded by early VP recognition, it is clearly preferable for these grammars to have unique constructors of VP positioned in advance of the verb.16

The VP ratio for $[O_m V] S_m$ remains 100\%, but the S domain is now longer. It is initiated by the last word or particle of $O_m$, which constructs VP, and proceeds to the last word or particle of $S_m$, which unambiguously constructs the subject NP. The result is five words for the recognition of two ICs, i.e. $2/5 = 40\%$. The aggregate for both domains is therefore 70\%, which is significantly lower than the 84\% for SOV. The aggregate for $[O_m] S_m [V]$ is an extremely low 45\%. The VP domain is now discontinuous, requiring five words for the recognition of two ICs, $2/5 = 40\%$. And the S domain proceeds from the first constructor of VP, i.e. the last word or particle of $O_m$, to the last word or particle of $S_m$, making four words for the construction of two ICs, i.e. $2/4 = 50\%$. Both domains have low scores, therefore, and it is surely
no accident that this extremely low aggregate is matched by no attested languages whatsoever in Tomlin's sample.\textsuperscript{17}

This same relative ranking is defined by EIC even when $S_m$ and $O_m$ are assigned equal word totals of e.g. two. The aggregates are 84\%, 75\%, and 59\% for SOV, OVS, and OSV respectively.

Consider now the EIC scores without the VP assumption, and with the word-total assignments of table 5.13. For [$S_m \ O_m \ V$] three ICs can be recognized in a four-word viewing window, making a $3/4 = 75\%$ ratio. Since this is lower than the 84\% in table 5.13, SOV languages have a motive for having a VP. For [$O_m \ V \ S_m$] a five-word viewing window is required, i.e. $3/6 = 60\%$. This is also lower than the 70\% in table 5.13, so these languages are also better off with a VP. [$O_m \ S_m \ V$] has the same five-word viewing window and 60\% ratio. This is significantly higher than the 45\% score on the VP assumption. Hence, OSV languages have no motivation for having a VP, and if this is a basic order at all, it is predicted to occur in languages that do not have a VP. Since OVS languages can readily incorporate a VP, given the adjacency of $O$ and $V$, and since this results in the higher EIC score of 70\%, our relative frequency prediction of OVS $>$ OSV remains.

We have motivated the unmarked preference for SOV over OVS and OSV, and the preference for OVS over OSV within the marked orders. It now remains to collapse the data of tables 5.12 and 5.13 and to see whether EIC makes the right predictions for all languages (VO and OV) considered together. Tomlin (1986) established the following four-way split between the six logically possible orders, based on their relative frequencies in his 402-language sample:

\begin{align*}
(5.72) \quad & SVO \quad > \quad VSO \quad > \quad VOS \quad > \quad OSV \\
\quad \text{SOV} & \quad \text{OVS} \\
\quad 348 \text{ languages} & \quad 37 \text{ languages} \quad 17 \text{ languages} \quad 0 \text{ languages} \\
\quad (87\%) & \quad (9\%) \quad (4\%) \quad (0\%)
\end{align*}

The difference between SVO (168) and SOV (180) is not statistically significant in his sample, so he collapses them. The difference between VOS (12) and OVS (5) is significant at the 0.05 level in a chi-square test, but he argues that the small number of these languages in his sample coupled with uncertainties about the correctness of their basic order classification are sufficient to rule out this significance. The result is four groups of basic orders, accounting for 87\%, 9\%, 4\%, and 0\% of the world's languages respectively.
Tomlin then goes on to argue that this four-way split can be accounted for by three interacting principles: the Theme First principle, Verb-Object Bonding, and the Animated First principle. The most frequent orders (SVO and SOV) adhere to all three, VSO adheres to two, VOS and OVS each to one, and OSV to none. We have already reduced the effects of Verb-Object Bonding (where it is correct) to EIC (see section 5.2.2). Theme First is reminiscent of Given before New, which we argued to be derivative of EIC's predictions for short-before-long ordering in languages with ICs of the type mIC (cf. ch. 4.4.3). This is not a correct predictor for all language types, however (cf. ch. 4.4.4), and any statistical effects here in certain languages were argued to be epiphenomenal. I have not investigated the animacy status of NPs, but I suspect that this is epiphenomenal in the same way, even though there are languages that have grammaticalized an animacy-based ordering principle, e.g. the Bantu language Sesotho (cf. Morolong and Hyman 1977; see ch. 7.2 for further discussion). At any rate, any theory that can account for the data of (5.72) in terms of one principle will clearly be preferred to a theory that requires three distinct principles, and that makes the rather questionable assumption, when deriving its predictions for language frequencies, that each carries equal weight. Does EIC explain the ranking of (5.72), therefore? Not only does it do so: it also explains why SVO and SOV are significantly more frequent than all four remaining orders together (by 87% to 13%).

Our EIC calculations have been made both for those structures that contain a VP and for those that do not. Table 5.14 summarizes the EIC scores under both analyses. I have put a box around the highest possible ratio that a given ordering can achieve: in most cases this is the ordering with a VP. Now, whether we aggregate across the two scores or, more reasonably, whether we select the highest possible score for each order, the EIC ratios of table 5.14 correlate exactly with Tomlin's four-way ranking in (5.72). SVO and SOV both achieve the highest possible score of 84% (and 75% without the VP). VSO languages score a maximum of 75% in either analysis. VOS and OVS also pattern identically, with a 70% maximum (and 60% without the VP). And finally OSV scores a maximum of 60% without the VP (and 45% with it). EIC therefore explains Tomlin's ranking on the basis of just one principle. It also predicts (in accordance with the Basic Order Prediction of (3.21)) that the most optimal order or orders, here SVO and SOV, should be grammaticalized in the unmarked case, at the expense of all other orders. Their combined 87% figure versus 13% for all the rest satisfies this prediction, and is reminiscent of the overall level of success achieved by the
Testing EIC's grammatical predictions

Table 5.14 EIC and all six orders

Assume: in VO languages $m_S = 2$ words, $m_O = 3, V = 1$

in OV languages $S_m = 3$ words, $O_m = 2, V = 1$

cf. tables 5.12 and 5.13 for further assumptions

<table>
<thead>
<tr>
<th></th>
<th>With a VP</th>
<th>No VP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_S V m_O$</td>
<td>84%</td>
<td>75%</td>
</tr>
<tr>
<td>$V m_S m_O$</td>
<td>75%</td>
<td>75%</td>
</tr>
<tr>
<td>$V m_O m_S$</td>
<td>70%</td>
<td>60%</td>
</tr>
<tr>
<td>$S_m O_m V$</td>
<td>84%</td>
<td>75%</td>
</tr>
<tr>
<td>$O_m V S_m$</td>
<td>70%</td>
<td>60%</td>
</tr>
<tr>
<td>$O_m S_m V$</td>
<td>45%</td>
<td>60%</td>
</tr>
</tbody>
</table>

EIC Ranking:
Best scores: 84% 75% 70% 60%
SVO > VSO > VOS > OSV

% of languages: 87% 9% 4% 0%

unmarked case predictions for language performance (cf. (3.17) and (3.18)) that were tested on several languages in ch. 4.

We do not, therefore, need to appeal in this context to a collection of causative factors most of which are probably epiphenomenal. Tomlin's data are explained by EIC alone, and more generally by a theory of syntactic parsing. The preferred versus non-preferred orderings of subject, object, and verb are motivated by the efficiency with which they provide on-line information about word groupings or constituent structure. The massively preferred orders will be SVO and SOV, in both of which the subject precedes the object. This follows ultimately from the preference for [NP VP] in both language types, and/or from the relative weight distributions of subjects and objects in these languages: objects being longer in SVO (whether there is a VP or not); and subjects being longer in SOV (again with or without a VP). The only other productively attested type, VSO, also has relatively good scores because of the recognition of the subject immediately after the construction of VP, and/or because objects are again longer than subjects in these head-initial
languages. The three most preferred word orders all have subject before object, therefore. But this left–right asymmetry follows from general principles and does not require any separate descriptive statements for subjects and objects as such, as in Tomlin's approach. Similarly, the preference for VOS and OVS over OSV, in all of which the object precedes the subject, does not reduce to any special principle(s) involving subjects and objects, but simply follows from the general theory, namely EIC.

5.6.3 Topic and focus positions

There are many languages in which there appears to be a grammaticalized structural node or position with certain characteristic semantic and/or pragmatic properties that are most typically referred to as "topic" and "focus" nodes. Hungarian and Finnish are examples and were discussed in ch. 4.1.2 and 4.1.9. Japanese topic phrases (with -wa) were discussed in ch. 4.1.5, cf. also ch. 6.2.2. The precise allocation of semantic-pragmatic properties to these nodes exhibits a certain variation across languages and has often eluded an exact analysis. Nonetheless the existence of grammaticalized semantic-pragmatic properties going beyond the purely syntactic, and associated with these nodes and positions across languages, can scarcely be in doubt, and we have seen some evidence for their reality in our performance data from languages such as Hungarian, in which the selection of entities to fill these positions in language use is driven by considerations other than syntactic weight (cf. ch. 4.1.2).

But while the semantic-pragmatic motivation for the existence of these "discourse-configurational nodes" appears to be real, it does not follow, as is often assumed, that their surface positioning in relation to the other, syntactic nodes of a sentence should also be explained in semantic-pragmatic terms. For example, the predominant leftward positioning of topic phrases has been explained in terms of given-before-new ordering (cf. Gundel 1988 for discussion), and cognitively in terms of the computational priority of presenting an entity prior to the sentence that is a predication about that entity (cf. Sasse 1987). We dismissed the relevance of Given before New as a universal principle of ordering in ch. 4.4.4. Sasse's cognitive principle is more plausible, but it is not obvious how a cognitive approach could generalize to explain focus positions as well. More generally, these discourse-configurational nodes co-occur in a tree structure with word groupings and nodes that are primarily syntactic in nature and whose relative ordering has been shown repeatedly in this book to be driven by efficient
constituent structure recognition. If this is what determines linear ordering in the normal case, why should discourse-configurational nodes be any different? Our theory therefore leads to a very different analysis of the positioning (but not of course of the pragmatics and/or semantics) of these nodes. What determines their order is their syntactic weight, the manner of their parsing, and the nature of the respective sister nodes that make up their constituent recognition domains.

This hypothesized explanation has been developed in detail by Primus (1991b) in relation to a variety of languages possessing these nodes. She argues that the proposed pragmatic and cognitive theories do not provide a good explanation for the positioning of discourse-configurational nodes, especially focus nodes, whereas EIC does provide a simple and independently motivated explanation with good predictive power. In the interests of space I shall not reproduce Primus's arguments in detail, but shall merely refer the reader to her paper and summarize the basic points.

For topic position Primus distinguishes between different kinds of topic constructions (e.g. hanging topics versus dislocations), and between the obligatoriness or optionality of moving a topic phrase into this position in different languages. Characteristic of all these construction types and variants is a clear left–right asymmetry in which the topic phrase occurs preferably to the left of the clause that functions as the predication on the topic, i.e. [TopP PredP]. According to Primus, this is because the PredP is typically significantly longer than the TopP. Positioning the shorter IC before the longer one within this domain produces better EIC scores than the reverse, and hence an additional pragmatic or cognitive explanation is simply not needed: the relative ordering follows from the same principle that explains all other linear orders.

Notice that the [TopP PredP] asymmetry is reminiscent of the [NP VP] preference that we documented in the last subsection and that resulted in a preference for subjects being positioned before objects. Again this is not because topic and subject share givenness (in head-final languages they appear not to): it is because of EIC. Notice also that [TopP PredP] ordering will produce optimal ratios for this domain if PredP can be recognized early. This explains, for example, the frequent verb- or auxiliary-second languages in which the verb or auxiliary acts as a constructor of the clause or PredP on the left periphery of this latter, immediately following the first-position TopP (cf. further ch. 6.2.2). It also explains the fact noted by Gundel (1988) that languages with topic markers such as -wa in Japanese always place these on the right periphery of TopP, i.e. as postpositions. This shortens the distance
from the construction of TopP to the construction of PredP (by a verb/auxiliary or by a nominative-marked NP, cf. again ch. 6.2.2). TopP is recognized on its right periphery, PredP at or near its left.

The situation with focus nodes is more complex, but again Primus points to a clear left–right asymmetry. She distinguishes between wide focus, where the focus is a complex (and heavy) constituent such as VP, S, or S̅, and narrow focus on, for example, an NP or PP. She then observes that focus constituents can occur in different positions, in response to their weights and the EIC ratios of their CRDs. For example, narrow focus nodes exhibit a clear skewing to the left in languages such as English. Compare it-clefts (which dislocate a focused constituent to the left, e.g. Harry in It is Harry that I am going to take to the movies) with wh-clefts (which position the focused constituent to the right, e.g. What I did was take Harry to the movies), in which the VP is focused. Primus points out, citing data from Prince (1978) and Erdmann (1988), that the preposed focused constituent in the it-cleft is significantly shorter than the co-occurring sister constituent that functions as the background of the focus. Conversely, in the wh-cleft the focused constituent is longer than the preceding material that serves as the background of the focus. This is shown in (5.73) (taken from Prince 1978: 886), which gives the average number of words occurring in the positions indicated:

\[(5.73) \quad \begin{array}{c|c|c}
    & oral & written \\
\hline
it is Focus & 2.5 & 5.7 \\
what ... is Focus & 4.6 & 11.4 \\
\end{array} \]

\[WH\text{-cMis clearly favor heavy focus constituents, at the right periphery of the clause; whereas it-clefts favor light focus constituents immediately following the copula be. Similar weight sensitivities are shown to hold for narrow and wide focus phenomena in other languages as well, leading Primus to conclude that there can be no consistent pragmatic or cognitive explanation for focus placement, whereas there is a clear weight sensitivity of exactly the type that EIC predicts. Both topic and focus nodes therefore provide further evidence for EIC in Primus’s analysis, to which the reader is referred for further details. \]
6 Grammaticalized node construction

In ch. 3.2 I defined two parsing principles, Mother Node Construction (MNC) and Immediate Constituent Attachment (ICA). The former is a construction principle: it defines a method for recognizing abstract phrasal groupings on the basis of terminal elements within a parse string. Without such a principle, constituent structure would not be recognizable on-line, there would be no such thing as a Constituent Recognition Domain (CRD), and so on. Clearly, the construction of abstract syntactic nodes, and the correct assignment of words to their appropriate dominating nodes, is a vital prerequisite for the everyday use of human language.

This chapter will consider the theory of node construction in more detail. MNC is an extremely general and all-pervasive principle which has been largely sufficient to illustrate the wide range of EIC-derived word order predictions presented in chs. 4 and 5. But the construction of a higher node by an immediate daughter is not the only way of adding non-terminal structure to the terminal categories of the parse string, and I have already alluded to the existence of some additional construction principles in previous chapters. The purpose of this chapter is to define these principles and to consider the predictions they make for variation in both category selection and ordering across languages. A further goal is to consider how grammars have responded to the parser's need to recognize higher constituent structure in a rapid and efficient manner. The head of phrase generalization, so prevalent in most current grammatical models, is essentially a grammatization, as I see it, of higher node construction on the basis of lower terminal categories, and the proposed defining properties of heads begin to make sense when viewed from this parsing perspective. There is a reason for the existence and co-occurrence of these properties, and by searching for this reason, in addition to examining the relevant descriptive generalizations across languages, we begin to understand why grammars are the way they are, and we have an additional rationale for deciding between competing formulations of the same grammatical insight.
The head of phrase generalization

In the first section of this chapter I shall consider the head of phrase generalization and discuss the relationship between it and MNC. The second section defines additional construction principles, and the third illustrates category and ordering alternations across languages that can now be explained by the existence of alternative construction principles for one and the same higher abstract node.

I begin by repeating, for convenience, the definition of MNC, as well as the parsing axioms from which MNC was argued to follow (in ch. 3.2):

(3.3) *Mother Node Construction* (MNC)
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.

(3.5) *Axiom of MNCC Existence*
For each phrasal mother node M there will be at least one daughter category C that can construct M on each occasion of use.

(3.6) *Axiom of MNCC Uniqueness*
Each MNCC will consistently construct a unique M on each occasion of use.

6.1 The head of phrase generalization

Many current grammatical models assume a "head of phrase" generalization within their phrase-structure rules or principles. These include Government-Binding Theory (cf. Chomsky 1981, 1982, 1986; cf. also Jackendoff 1977); Generalized Phrase-Structure Grammar (cf. Gazdar et al. 1985), and Head-driven Phrase-Structure Grammar (cf. Sag and Pollard 1989). Within these models, a head category such as a noun will be dominated by one or more phrasal projections of the head, such as N¹, including ultimately the highest, or maximal, projection, e.g. N¹¹. Accompanying specifiers or complements of the head are attached to the appropriate projection. Theories differ with regard to the set of head categories that they postulate, the number of bar levels or phrasal projections that they assume, and the precise internal constituency of the daughters. Despite the differences, it is argued that there are some profound linguistic generalizations associated with the notion of "head," and that it structures the syntax of languages to a considerable extent (cf. Zwicky 1985; Hudson 1987; Corbett et al. 1993).
The theory of heads is also the central idea of X-bar theory, according to Kornai and Pullum (1990). X-bar theory was first introduced by Chomsky (1970) using the following rules for English:

(6.1)  a. $X'' \rightarrow (\text{Spec } X') X'$
       b. $X' \rightarrow X \text{ Comp}$

Within these rules, $X$ stands for the head categories N, V, Adj, and Prep; Comp for the complements of the head categories which enter into the strict subcategorization of the latter (cf. Chomsky 1965: ch. 2), such as NP, PP, VP, and S; and $(\text{Spec } X')$ for the various specifiers of these head categories plus their complements (e.g. determiners for noun phrases, auxiliaries for verb phrases, adverbial modifiers of adjective phrases). Since Chomsky (1986) the set of heads has been expanded from these lexical categories to the functional categories Complementizer, Infl, and Det, which project to Complementizer Phrase (CP), InflP, and DetP respectively (cf. Abney 1987 for discussion of DetP).

6.1.1 Heads are MNCCs

From the perspective of language processing, any grammatical model that incorporates the head of phrase generalization, whether in the form of an X-bar theory or in any other way, is defining a set of syntactic relations that make it very easy for the hearer to recognize abstract syntactic structure. Within each pair of rewrite rules such as (6.1), the identity of the dominating category on the left-hand side is always determined by (and determines) the identity of $X$ (or $X'$) on the right. Thus, an NP (or $N''$) on the left must ultimately be rewritten as ... $N$ ..., where $N$ is the lexical head, and a rule such as $N'' \rightarrow (\text{Spec } P') P'$ would be impossible in this theory. When parsing terminal elements within a parse string, therefore, the recognition of a head category immediately enables the hearer to infer the appropriate mother node by activating the relevant PS rules for his/her language. Any additional daughters can then be attached by Immediate Constituent Attachment (3.4) (cf. ch. 3.2). In other words, heads of phrases are mother-node-constructing categories (MNCCs) in our sense, and the head of phrase generalization can be seen as a grammaticalization of the parsing principle of MNC.

Let us pursue this idea further. To what extent do the grammatical details of the head of phrase generalization follow from this proposed explanation?

The parsing axioms of MNCC Existence and Uniqueness define what are plausibly necessary conditions for the successful recognition of abstract
syntactic structure on-line. The former asserts that there has to be some MNCC with which to construct every M on every occasion of use – otherwise how would M be recognized? The latter states that each MNCC must consistently construct one and the same mother node. An N cannot construct an NP one minute and a PP the next, for example.

Grammars that employ the head of phrase generalization define a number of properties that appear to be motivated by these axioms. One key property of heads is their obligatoriness (cf. Hudson 1987: 111): the head is the category that has to be present, whereas non-heads may be omitted or deleted. This follows straightforwardly from the axiom of MNCC Existence. If there is no daughter MNCC for M on a particular occasion of use, then M will not be recognizable and constructable, and the axiom will not be fulfilled. In a theory in which heads are the MNCCs for higher constituents, obligatoriness follows from the need to recognize what constituent is actually being parsed on each occasion of use.

A second property is distributional equivalence (cf. Hudson 1987): the head is the category whose distribution is similar to that of the mother. This property is closely related to obligatoriness and is also a consequence of the axiom of MNCC Existence. MNCC and M share the same distribution precisely because the MNCC constructs M and may be the only daughter of M that is present, being obligatory. The distributional equivalence of an MNCC and its M therefore follows from the construction potential of the former. The distributional equivalence of heads and mothers follows, in turn, from the fact that heads are MNCCs.

Thirdly, heads stand in a relation of uniqueness to their mothers or maximal projections. The category V, for example, cannot be the head of both VP and of AdjP. This is explained by the parsing axiom of MNCC Uniqueness: an MNCC cannot construct more than one mother. If it did, there would be systematic ambiguity with respect to which mother node was intended on a particular occasion of use. Making the wrong selection would result in structural misanalyses, garden paths, etc., all of which cause processing difficulty, and would defeat the whole purpose of having MNCCs at all, which is to render higher structure recognizable and constructable. Hence, it is impossible for any head to project to a plurality of distinct mothers, since heads are MNCCs. It follows that if a language is going to make use of a given phrasal category, e.g. AdjP (cf. Thompson 1988), then this language must also possess a separate form class of categories distinct from others, i.e. adjectives, that can uniquely construct the higher phrasal node. The consequences of MNCC Existence and Uniqueness for PS rules are quite profound, therefore.
Each phrasal node on the left will be rewritten so as to contain at least one uniquely determining MNCC on the right. In the theory of heads, the uniquely determining MNCC is the head of phrase.

Fourthly, heads can be both lexical categories (N, V, Adj, etc) and (for Chomsky 1986) functional categories as well. This extension of the head of phrase generalization to include the functional categories is consistent with, and plausibly motivated by, the axioms of MNCC Existence and Uniqueness. Any daughter can be an MNCC, as long as it guarantees unique recognition and construction of the mother. As a result, both lexical categories and closed-class grammatical items such as inflected auxiliaries, complementizers, and determiners can serve in this capacity. Some phrasal nodes will be constructed by the one (e.g. VP by V), and others by the other (e.g. InflP by Infl), but in both cases the same fundamental generalization is being captured by PS rules that incorporate the two types of heads, and this generalization is ultimately a parsing one: all heads are MNCCs.

A fifth general property of heads is so fundamental that it is often not even explicitly discussed: projection from a daughter. Heads are always daughters of the phrasal nodes onto which they project the relevant categorial status — never, say, sisters or nieces. Yet one could imagine a set of syntactic rules in which such additional projection possibilities were allowed. Heads can subcategorize for, and govern, their sisters (cf. Hudson 1987, and section 6.1.3 below), so why can’t they determine the actual categorial status of a sister? Imagine that a given phrasal node co-occurring with a verb were to be assigned the category NP, within the VP projected from V, i.e. $VP[V\ NP]$, and that some other phrasal node quite non-distinct from this NP in terms of the daughter categories it could dominate but occurring in conjunction with an adjective in an AdjP were assigned the category PP, $AdjP[Adj\ PP]$, i.e. solely on the basis of the different sister head categories. In such a case, the head would project both to a mother node, assigning it a category that is a more complex variant of its own (cf. the discussion of category constancy in section 6.1.2), and to a sister node, assigning it a category that was uniquely dependent upon the choice of head. Such cases of sister projection seem to be non-occurring, however, although they are far from absurd and one could readily imagine that a grammar could make use of this possibility. Nonetheless, the existence of distinct sisters is not a sufficient basis for distinguishing between phrasal categories: at least partially distinct daughters are always required, one of which will construct the mother node.

The requirement of projection from a daughter follows from the axiom of MNCC Existence, which requires that each $M$ must be constructed by a
daughter category C. This rules out phrasal construction by a sister alone, as in the example discussed; and this is not a trivial claim. It makes it impossible for a head to determine the actual categorial status of a sister, even though it may assign case to it, and may select one type of sister rather than another within subcategorization frames. Moreover, in section 6.2.1 a parsing principle of Sister Node Construction will be argued to exist for structures such as $\bar{s}[\text{Comp } S]$ in English, in which Comp constructs both a mother $\bar{S}$ and an S as right sister. But the axiom requires that no phrasal node can be constructed only on the basis of a unique sister.

The ultimate functional motivation for the axiom of MNCC Existence, I would argue, is that categories that are to be treated as distinct by the syntax, such as NP and PP, must be recognizably distinct in performance. By containing a daughter category that is an MNCC unique to M, higher constituent structure can always be readily and rapidly inferred. By contrast, making the identity of M entirely dependent upon a sister would not be generally efficient, for several reasons. First, if the determining categories, V and Adj in the examples above, were to follow NP/PP, a decision about the categorial status of this phrase would have to be delayed until V or Adj were encountered, i.e. after NP/PP had been completely scanned. Decisions about the attachment site for a phrase must often be delayed in parsing, but a phrase's categorial identity should always be rapidly recognizable, and on-line indeterminacy and delay in this regard can be avoided if mothers are systematically constructed by daughters. Second, making a category dependent upon its sisters for recognition imposes a strong adjacency requirement on these ICs, regardless of their ordering. Any kind of distance between them will lessen the construction potential of the determining sister. Such adjacency may not always be desirable, however. V and a direct-object NP are most frequently adjacent in English, but when the NP is longer than any additional VP-dominated constituents, EIC prefers a postposed NP, which destroys the adjacency. Third, sister construction imposes a strong predictability requirement on adjacent ICs, again regardless of ordering. Such predictability is exceptional, however. Even a PP in English, whose immediate daughter constituents are limited to P and NP (and preceding adverbial specifiers), does not allow us to infer an accompanying NP after P, because of stacked PPs such as $\text{PP}[\text{from } \text{PP}[\text{under } \text{NP}[\text{the bridge}]]]$. Fourth, regular sister construction would introduce considerable complexity into grammatical rules, since all rules that referred to a sister-dependent category would need to be made context-sensitive. Any time reference was made to NP in the example above, the sister V would have to be included in the rule.
specification, and this would both add an extra symbol and reduce the
generality of the rule, making it inapplicable to NPs in other environments.

For all of these reasons, mothers are systematically constructed by daugh-
ters, and Sister Node Construction is limited to a handful of structures in
which sisters are regularly adjacent and predictable, such as $g[\text{Comp } S]$. Even
here, the $S$ must also be constructed by a daughter MNCC, on account of the
axiom of MNCC Existence.

There are some additional properties of heads that have been proposed in
the published literature, over and above the five I have just summarized. I
shall return to these in section 6.1.3. In the next subsection I shall continue
the comparison of heads and MNCCs by pointing out some differences
between them. In particular, we will see that whereas all head categories in
a grammar are MNCCs in parsing, since heads construct their mother nodes,
ot all MNCCs need be heads.

6.1.2 MNCCs that are not heads

The axioms of MNCC Existence and Uniqueness explicitly allow for the
possibility that there may be more than one daughter category that can
serve to construct a given $M$. For example, both a noun and a determiner
are unique to NP within Standard Theory PS rules (cf. Chomsky 1965) and
both could, in principle, construct NP within a parser referring to these rules.
In a language like English in which the determiner comes first in the parse
string, the task of constructing NP will always fall on Det, when it is present,
since it is the first uniquely determining category that the parser encounters.
In this way the parser does not need to wait for N in order to construct NP,
and the EIC ratios of higher containing phrases such as VP can be optimal
when NP is the last IC of the domain.

This functional equivalence of Det and N as MNCCs for NP makes a
typological prediction that was tested in ch. 5.5. Since each can perform
the same parsing job, we expect variation, with either Det or N occupying
the position that is independently predicted for the MNCC for NP in the
relevant language. Consider languages that are left-mother-node-construct-
ing (left-MNC), i.e. languages such as English that have VO order in the VP
and prepositions in the PP. EIC predicts that the head noun and Det will be
equally optimal in the leftmost position of NP, since each can construct the
NP immediately following V or P. Noun-initial order is found in Yoruba,
and Det-initial in English. I observed in ch. 5.5 that $PP[P \text{ NP}[N \ldots ]]$ and
The head of phrase generalization

The head of phrase generalization is almost equally frequent, and that there is only a tiny minority of prepositional languages that does not have one or the other as its leftmost daughter of NP. The productivity of this alternation between Det and N is explained by our parsing axioms in conjunction with EIC.

For most theories of heads, however, there cannot be more than one head per phrase, i.e. there is a relation of bi-uniqueness between mothers and heads: each head projects to only one mother, and each mother is a projection of just one head. If the determiner is to be a head in addition to N, there must be two distinct dominating phrases, DetP and NP, the first a projection of Det and the second a projection of N (cf. Abney 1987).

It is important to stress that bi-uniqueness is not a necessary consequence of our parsing axioms, though it is compatible with them. The axioms insist only on the uniqueness of a mother to a given constructing category, and do not require that this uniqueness should be reversible. As a result, there can be a plurality of MNCCs for a given M. Our parser is compatible with a grammar in which one of these MNCCs is a head of phrase, such as N, and another is not, e.g. Det. It is compatible with a grammar in which some phrasal nodes are projections of heads, while others are not, e.g. only lexical categories may be heads, and neither the complementizer nor S need be a head of S. The complementizer simply has to render the identity of its immediately containing phrase predictable, since it is an MNCC. Our parser is even compatible with a model of grammar that makes no reference to heads at all.

Although the head of phrase generalization can therefore be seen as a grammaticalization of the parser's need for mother node construction, it is not the only way of reflecting and realizing this parsing constraint on grammars. MNCCs do not need to be heads for syntax to be parseable. And whereas all heads are necessarily MNCCs and their defining properties appear to be largely motivated by their construction potential, the converse fails: not all MNCCs need be heads.

Bi-uniqueness makes it possible for most theories of heads to impose an additional requirement on head–mother relations: category constancy. The mother must be of the same general category type as the head. If a plurality of heads were allowed, this requirement would be unworkable, unless the different heads were all of the same category, since at most one head would be able to project its categorial status onto the mother. The principle of MNC does not insist on category constancy for this reason. It requires only that the phrasal mother be uniquely recognizable on the basis of the relevant MNCC, which is very different. Both Det and
Grammaticalized node construction

N can construct NP on this account. An Aux, or Infl, can construct S, without us having to assume that the mother is an AuxP or InflP, and Comp can construct a subordinate $\bar{S}$.

As I see it, both bi-uniqueness and category constancy are unnecessary, and rather over-enthusiastic grammatical formalizations of the construction intuition that ultimately motivates them. Category constancy is particularly unfortunate, since it leads to some very counterintuitive dominating node labels, especially for the functional head categories, and this detracts from the correct MNC intuition that underlies them. There are powerful syntactic and semantic arguments for saying that a VP is a type of V (or a more complex version of V), and that an NP is a type of N, etc. But a complementizer such as *that* in English exists primarily in order to provide a clear boundary for the onset and recognition of a subordinate clause; and to assert (even implicitly) that this clause is really a complementizer phrase (CompP) and hence a more complex type of complementizer, by analogy with VP and V, and NP and N, seems mistaken. So is the assumption of an InflP. Just because a finite verb or auxiliary can construct an S does not warrant labeling that S an InflP. It is not — it's a clause. More generally, category construction does not entail category constancy.

Once this is recognized, we can dispense with category constancy as a requirement on constructing categories in a grammar, and with bi-uniqueness. We can also entertain the more radical proposal that we don't need heads at all, either in the theory of grammar or in (at least some) particular grammars. If the ultimate explanation for this aspect of syntactic structure is a parsing one, and if there are other rule types that are equally parsable, then we need to look more critically at whether the head of phrase generalization is the best formalization of the grammar's response to MNC. It is no good appealing to the inherent simplicity and generality of rule schemas such as (6.1), because these schemas are now derivative of processing, and a major criterion of evaluation becomes the goodness of fit between proposed parsing principles and different grammatical formalizations. Another criterion involves the capturing of significant generalizations: do grammars make reference to heads in their grammatical regularities?

Notice that a phrase-structure grammar that dispenses with heads can still be highly constrained by a processing module that incorporates our parsing axioms. Each phrasal node will have to be rewritten so as to contain at least one MNCC that is unique to it. Thus, a rule such as $\text{VP} \rightarrow \text{NP} + \text{PP}$ would be an impossible rule since there would be no MNCC for VP; $\text{VP} \rightarrow \text{Adj} + \text{PP}$ would be impossible alongside $\text{AdjP} \rightarrow \text{Adj} + \text{PP}$, on account of the
non-uniqueness of Adj as an MNCC; and VP \( \rightarrow V + PP \) and AdjP \( \rightarrow V + PP \) would be impossible for the same reason.

My major interest in this context is not in choosing between one version of the heads theory and another, or between a theory of phrase structure that uses heads and one that does not. I consider these to be largely descriptive issues. Instead I am concerned with the following more explanatory question: to what extent can fundamental features of phrase-structure organization, in any of their current formalizations, be said to follow from the need to recognize that structure on-line using parsing principles of the kind I have proposed? If the parser can motivate such features in a simple and convincing manner, then syntax can become less stipulative and mysterious than it is at present. The five properties of heads discussed in section 6.1.1 seem to be natural consequences of the principle of MNC and of the parsing axioms that motivate it: obligatoriness; distributional equivalence; uniqueness; lexical and functional head projection; and projection from a daughter. Two further properties that have been proposed for heads are more questionable and do not follow from MNC and the axioms: bi-uniqueness and category constancy. I turn next to some additional properties that have been proposed for heads, which I shall analyze from a processing perspective. Are these properties correct, and can we give a processing explanation for them? These properties have been most extensively discussed within the context of a head-of-phrase grammar, but again I am not committed to any one particular formalization, and we can regard these properties (where they are correct) as properties that any grammatical model is going to have to formalize in some notationally equivalent way.

### 6.1.3 Some other proposed properties of heads

One major property of heads, proposed by both Zwicky (1985) and Hudson (1987), is that heads serve as the morphosyntactic locus for inflections, i.e. the head is the constituent on which any inflections relevant to the mother are located. Aux, V, and N all carry inflections in English, and these categories are all heads (of S/AuxP/InflP, VP, and NP respectively). Languages with richer inflectional morphology illustrate additional head-marking possibilities: Adj bears the inflections relevant to AdjP in German; and in Abkhaz postpositions exhibit inflectional agreement with their sister NPs within the PP (cf. Hewitt 1979).

Why should heads be the morphosyntactic locus for inflections? The answer is surely related to the obligatoriness of head categories, which
follows from the axiom of MNCC Existence. It would be counterproductive to position these inflections on a daughter that was regularly optional and omissible, since the grammatical information that they carry would then be lost. By positioning these affixes on the obligatory constituent, their regular retention and presence is guaranteed.

What is crucial to this functional explanation is not that inflection-bearing categories are grammatical heads: from a narrowly grammatical perspective this morphosyntactic correlate is a complete mystery. What motivates it is the fact of higher node construction. It is because MNCCs construct their mother nodes that they are the appropriate bearers of inflection assigned to these mothers. MNCCs are indispensable to the recognition of their mothers. This is why they are obligatory, and why MNCC and M are distributionally equivalent. It is also why MNCCs are appropriate morphosyntactic loci, whereas non-MNCCs are not.

These proposed properties of heads are really the properties of constructing categories, or derivative consequences of these properties. And by viewing MNC and the parsing axioms as the explanatory primitives, we can actually explain the properties of heads. We can also make sense of some disagreements in the heads literature. Take Det and N. There is considerable confusion over the proper analysis of these two categories. Is each a head, of DetP and NP respectively, as in the theory of Abney (1987)? Or is N alone the head, with Det a specifier, i.e. a non-head, as argued by Ernst (1991)?

The controversy arises, as I see it, because both Det and N can be MNCCs for the same mother node. Within a theory of heads, it therefore seems reasonable to propose that both can be heads, and this can be supported by showing that each can stand alone as the obligatory constituent in some languages, that each can be a morphosyntactic locus e.g. in German. What is not satisfactory about this is that it fails to capture the fact that a constituent headed by N and one headed by Det are actually the same constituent, i.e. NP in traditional analyses. As a result, rules that refer to this constituent must now apply to a disjunction of categories, DetP or NP; and there are further grammatical complications, because an NP within a DetP should always behave like a maximal projection if N and Det are both heads – yet it doesn’t, as Ernst (1991: 195) shows.¹

Our parsing theory of MNC, by contrast, allows for just such a plurality of MNCCs for one and the same M, without having to assume that all (or any) of them are heads. It predicts ordering variation between them, e.g. Det-initial or N-initial NPs in languages with VO and prepositions. It predicts that at least one of these MNCCs must be present to construct M on every
occasion of use, i.e. either N alone or Det alone. And it makes a plurality of
categories available as morphosyntactic loci: both Det and N were inflection-
ally marked in Old English and remain so in Modern German. Our theory
does not insist that all MNCCs must be inflectionally marked within the NP.
It predicts that if there is to be inflectional marking, it will fall on an MNCC
and not on a non-MNCC. If there is a plurality of MNCCs, the inflections
may be marked on just one, or on all of them, or may be distributed between
them, presumably with consequences for deletion and omission possibilities
in specific instances depending on how the distribution is made.

It seems to me that this variation predicted by MNC, in ordering, dele-
tions, and inflectional positioning, is essentially correct. But neither Abney’s
(1987) nor Ernst’s (1991) theory can quite do it justice. If N alone is the head,
then there are counterexamples to the proposed head properties of nouns
whenever Det behaves like a head (e.g. when N is deleted), and of course
the head-like properties of Det are unaccounted for. If both Det and N are
heads, then there are counterexamples to the proposed head properties of
each, whenever a given property such as obligatoriness applies only to the
other; this approach fails to capture the fact that each can be the head of what
is distributionally and grammatically the same phrase; and the dominated
phrase (NP) does not always behave like a maximal projection. It is consid-
erations such as these that make me suspicious of the correctness and ulti-
mate utility of the head of phrase generalization. Where it is correct, its
properties follow from MNC. Where it is incorrect, MNC is still sup-
ported, and the head of phrase generalization gets in the way. The true
explanatory primitive appears to be MNC, therefore.

Two other general properties of heads, defined by Hudson (1987), are the
following: the head is the SUBCATEGORIZAND within its phrase, i.e. it subcate-
gorizes for its sisters; the head is also the GOVERNOR within its phrase, i.e. it
determines the morphosyntactic form of its sisters, by assigning case to them.
Subcategorization and government are, therefore, generally considered to
involve a relationship of co-constituency between ICs, and the category that
does the subcategorizing and governing is claimed to be the head of phrase.

Within Government-Binding theory, a crucial and necessary part of the
definition of subcategorization and government is the c-command relation
(cf. Reinhart 1983). For example, if A governs B (e.g. a verb governs a direct-
object NP), then A c-commands B. B is then said to be in the syntactic
domain of A, as defined by c-command.\(^2\) The trouble with this is that it
does not quite succeed in capturing the domains in which government and
subcategorization apply, in a natural and intuitive way. If verbs are contained
within a VP, and if case assignment and subcategorization are lexically specified properties of verbs (as they surely are), then such processes are predicted within the VP domain, e.g. for direct objects, but not outside it. Subjects present a problem, therefore. Instead of saying that the relevant verb assigns case, etc., to the subject, c-command forces us to say either that a sister Aux/Infl governs a nominative NP, or that the whole VP does. But this has a number of disadvantages: it imposes an artificial distinction between subject and object government; it results in a non-lexical category, Aux/Infl, assigning the case that is intrinsically associated with the lexical verb within VP; alternatively the whole VP must govern a nominative NP, yet VP is not a lexical category. For languages such as English, with uniform nominative marking for subjects and highly constrained and predictable syntactic positioning for these subjects, nominative case can be assigned, in effect, syntactically on the basis of configurational properties, and these disadvantages can be circumvented. They become intolerable, however, in languages in which VP-external properties of subjects are variable and intimately controlled by individual VP-dominated lexical verbs. Productive examples are ergative languages, in which lexical verbs that subcategorize for a direct object (i.e. transitive verbs), assign one case to the VP-external grammatical subject, the ergative, while those that do not (intransitive verbs) assign another case, the absolutive, to the VP-external subject (cf. Comrie 1978). Alternative grammatical models, such as GPSG (cf. Gazdar et al. 1985) circumvent this problem by allowing the verb to project its subcategorization and case-assignment properties to S as well as to VP, through the Head Feature Convention; and the S then defines the domain within which these properties are operative. This forces an analysis in which both S and VP are projections of one and the same head category, V, i.e. they are represented as the same phrasal category with different bar levels (cf. also Jackendoff 1977). This analysis also relies crucially on the theory of heads itself. I am not persuaded by either assumption.

Instead, I want to argue that the reason why a lexical, VP-dominated verb can assign structural properties outside of VP, and specifically to an S-dominated NP subject, is because V can construct S as well as VP. If V is finite, it will construct S by Grandmother Node Construction (cf. section 6.2.2); if it exhibits agreement with the subject, it will construct S by Agreement Projection (cf. section 6.2.3). In other words, we need to extend the concept of a c-command domain in this context, and the linguistically significant generalization that I propose is the notion of a Construction Domain, defined as follows:
The head of phrase generalization

(6.2) Construction Domain of a Node

The Construction Domain of a category X consists of nodes dominated by the highest phrasal constituent constructed by X.

A verb can therefore subcategorize for, and assign case to, NPs that are in its Construction Domain, in accordance with the following definitions:

(6.3) Subcategorization

If a category A subcategorizes for a category B, then B must be in the Construction Domain of A.

(6.4) Case Assignment

If a category A assigns case to a category B, then B must be in the Construction Domain of A.

Since S is the highest phrasal constituent constructed by V (when finite), it follows that V can subcategorize for, and assign case to, NPs in S, as well as in VP.

Subcategorization and government or case assignment do not offer strong support for a head-sister relationship, therefore: the relevant ICs are not always sisters; and the assigning category is not necessarily a head within the largest phrase in which these grammatical properties are assigned. Instead, these properties are more naturally explainable in terms of the capacity of the verb to define a Construction Domain within which the syntactic, semantic, and morphological co-occurrence requirements of that verb can hold. Hence, if a category is going to define syntactic relations of dependency on other elements within a certain phrase, this category must actually construct the higher phrase. The functional motivation for this is quite intuitive. It makes sense that syntactic dependencies should be limited so as to apply within, rather than outside, the highest phrase actually constructed by the category that defines the dependancies, because then the phrase to which the dependant is to be attached will inevitably be constructed on every occasion of use. If dependants could be attached outside Construction Domains, there would be no way to guarantee that the desired containing category would actually be present on all occasions, without additional stipulations and co-occurrence requirements being made. But if the syntactic dependencies of a category are limited to Construction Domains, then the higher structure of a tree within which these dependencies hold, and the attachment sites for the dependants, are necessarily and automatically present.
Grammaticalized node construction

Notice also that if the construction of phrasal nodes were limited to just MNC, there would be no difference between the Construction Domain of a node and the c-command domain of that node. It is because phrasal node construction can extend beyond the first branching constituent, by principles such as Grandmother Node Construction and Agreement Projection, that Construction Domains can grow beyond c-command domains, in just the way that seems empirically right for subject-verb relations.

The status of subcategorization (and government) within the NP also fits with this account of VP and S. Zwicky (1985) and Hudson (1987) both argue that Det is the subcategorizand, because of lexical differences between determiners such as each (which requires a singular count noun), many (which requires a plural), and much (which requires a mass noun). But one could argue equally plausibly that N is the subcategorizand here. Just as all transitive verbs are subcategorized for a direct-object NP, so all singular count nouns can subcategorize for certain classes of determiners only (a and each in English), etc. It might be objected that it is simpler to state such subcategorizations on the determiner, since the number of relevant subcategorizing determiners is smaller than the number of subcategorizing nouns, and hence fewer statements would be needed in the lexicon overall. But the force of this simplicity argument is weakened when we recall that every single transitive verb will be listed in the lexicon as subcategorizing for a direct-object NP. The most plausible resolution of the NP issue, therefore, is to say that subcategorization can work in both directions, from Det to N and from N to Det. It does so, moreover, because the Construction Domains of Det and N are identical in our account: each constructs NP, hence each can in principle subcategorize for other ICs within its domain. Any priority for Det in this regard is simply a matter of economy, and does not follow from any principled considerations involving its putative head of phrase status.

Properties such as subcategorization are therefore properties of constructing categories, holding within a Construction Domain; they are not just head of phrase properties.

Yet another property of heads proposed by Hudson (1987) is a semantic one: the head is the semantic functor category, i.e. the meaning of the head has the status of a functor in relation to an argument category. Here I would argue that semantic functors are not always heads; nor are they correlated with MNCCs. Moreover, a processing approach actually explains why there should be no necessary correlation between the syntax and the semantics in this area.
Discussions of function and argument status are not helped by the unclarity in most of the logical literature over which categories are semantic arguments and which are functors. Zwicky (1985: 4) comments on this as follows: “With a certain amount of formal ingenuity, a Montague-style semantics that treats Det as a functor on the argument N can be redone as a system treating N as a functor on the argument Det.” One serious attempt to come up with independent criteria for function and argument categories is Keenan (1979), and if we follow his reasoning it is clear that N is the argument within Adj–N and Relative Clause–N pairs, with Adj and Rel the functors, while V is the functor and NP the argument within VP. Hence, if V is both the functor and the head within VP, then whether one regards N or Det as the head of its phrasal category, Adj and Rel will be functors and not heads, and there will be no general correlation between functors and heads.

Nor is there any reason why there should be. Heads are MNCCs, and there is no a priori reason why there should be a correlation between semantic functors and MNCCs. MNCCs exist in order to give the listener on-line cues about the groupings of words into syntactic constituents, but nothing requires that these categories should simultaneously be of one semantic type versus another. Which daughters are arguments and which are functors will be determined quite independently by a semantic theory that maps the categories of the syntax onto their associated logical structures. Therefore, considerations of processing can actually motivate the lack of any correlation between heads and their associated semantic categories.

6.1.4 The status of heads

The head of phrase generalization is an attempt to unify a broad range of grammatical phenomena. The proposed properties of heads include: obligatoryness; distributional equivalence; uniqueness; lexical and functional head projection; projection from a daughter; the head as morphosyntactic locus; as subcategorizand; and as governor. I have argued that all these properties follow, either directly or indirectly, from the parsing axioms of MNCC Existence and Uniqueness. Three additional correlates of heads are often proposed (bi-uniqueness, category constancy, and semantic functors), but these properties do not follow from the axioms, and I have argued that their motivation and justification are correspondingly moot.

There are different versions of the theory of heads, depending for example on whether both lexical and functional categories are included in the set of heads (cf. Corbett et al. 1993). Despite the differences, all these versions are,
as I see it, attempts at formalizing the grammatical consequences of our parsing principle of MNC. Heads are MNCCs, and they have the properties they do because they are grammaticalized construction principles. This broad range of properties begins to make sense when viewed from a processing perspective: even the disagreements between different theories can now be better understood. But the whole thing is simply mysterious from a grammar-internal perspective, and the best we can do is to say that the head of phrase generalization is a candidate for an ultimately innate principle of grammar, since it does not follow from anything else in the grammar.

But the properties of heads do follow from something else. They follow from the primitives of a theory of language processing, many of which are also plausibly innate and hard-wired, so as to guarantee that syntactic structure can be rapidly recognized in real time, along the lines of Jerry Fodor's (1983) modularity thesis. Thus, the grammatical properties of heads follow from independent principles of performance that are needed anyway.

What remains to be shown grammatically is which version of the heads theory is the best formalization of the ultimate MNC explanation, and whether the head of phrase generalization is even needed at all. Heads are a subset of the set of MNCCs – in many theories a proper subset – and the linguistically significant generalizations in this area are consequences of MNC itself, not of any intrinsic and ultimately stipulated properties of heads. Hence, heads are in principle dispensable. The PS rules of all individual grammars will be heavily constrained by the axioms of MNCC Existence and Uniqueness, requiring the constructability of a mother by at least one daughter, regardless of the precise grammatical format used. Constituent structures can now be much less complex, and flatter than those built around a head of phrase theory such as (6.1). I have also suggested that the syntactic domains for various dependencies can be defined in terms of Construction Domains, a concept that provides a simple and functionally motivated solution to problems involving the subcategorization and government of VP-external subjects by lexical verbs within VP.

6.1.5 Extended Mother Node Construction

Despite the scepticism that I have expressed regarding the ultimate utility of the head of phrase generalization, it remains the case that heads are a subset of MNCCs and that most grammatical models do make use of them. We need to take account of the fact that the phrase-structure generalizations to which
MNC refers in these models of grammar will be structured in terms of heads, by extending MNC as follows:

(6.5) **Extended Mother Node Construction (EMNC)**

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; if C is a head of phrase, construct the set of uniquely determined projections of C over C, up to the maximal projection M.

That is, if C is a head, it can construct a set of uniquely determined dominating nodes, up to the maximal projection or mother node M. Non-mother-node-constructing daughters of M must then be attached to the appropriate projection of C by ICA (cf. ch. 3.2).

### 6.2 Some additional construction principles

This section will define and illustrate some other construction principles that parsers can make use of, in addition to (Extended) Mother Node Construction. These new principles will be less general than MNC in the range of structures to which they apply, and they will be more relevant for some language types than others. Nonetheless, a number of properties of grammars appear to be direct responses to, and grammaticalizations of, these construction principles, in much the same way as the head of phrase generalization is a grammaticalization of MNC. In particular, these new principles will enable us to explain a number of alternations in category selection, and in ordering, across and within languages. The new principles will be presented in the following order: Sister Node Construction (section 6.2.1); Grandmother Node Construction (section 6.2.2); Agreement Projection (section 6.2.3); and Attachment Promotion and Construction (section 6.2.4). Section 6.2.5 redefines our parsing axioms for phrasal node construction in the light of these principles. Some sample predictions that they make for grammars are considered in section 6.3.

#### 6.2.1 Sister Node Construction

This principle is defined as follows:

(6.6) **Sister Node Construction (SNC)**

In the left-to-right parsing of a sentence, if any word of syntactic category C uniquely determines an adjacent right sister node N, in
accordance with the PS rules of the grammar, then N is immediately constructed as an adjacent right sister.

For example, a complementizer will construct an S as adjacent right sister in English frames such as $S_{[\text{Comp } S]}$. As a result, words such as unstressed *that* in *that John is sick*, or "adverbial subordinators" (cf. Dryer 1992) such as *because* in *because John is sick*, will render a following S predictable and can construct it on-line.

SNC is necessarily asymmetric. A category C can construct a sister category to its right, i.e. one that has not yet appeared in the on-line parse and that is predictable on the basis of C. But it cannot construct a category to its left, i.e. one that has already been processed. This follows from the axioms of MNCC Existence and Uniqueness: a phrase must always dominate at least one daughter MNCC and cannot be exclusively dependent upon a sister for category recognition (recall section 6.1.1). It follows that any phrase that is the left sister of a category C that could potentially construct it, must already have been constructed and recognized by the time C is encountered. Hence, C cannot literally construct it, and the relevant phrase will simply be attached to the mother of C by Immediate Constituent Attachment (ICA); cf. (3.4) in ch. 3.2. But a phrase standing to the right of a potentially constructing sister C can be constructed by C, since C is encountered first in the left-to-right parse. A left-peripheral Comp can construct an $S$ as mother and an $S$ as sister, therefore, before the MNCC of $S$ itself is reached. All ICs of $S$ are then attached to this pre-existing $S$, and MNC and ICA are, in effect, collapsed. The left-to-right nature of parsing, coupled with the principle of construction by a daughter, combine to make a following sister constructable, but not a preceding one.

From the point of view of EIC, an SNC structure appears to be very efficient: one word, e.g. *because*, constructs not only the mother $S$, but a sister $S$ as well. Hence, two daughter ICs, Comp and $S$, are recognized on the basis of a single word, giving an IC-to-word ratio of 2/1 for that recognition domain, or 200%! One wonders why such structures are not more frequent.

The answer is that sister construction is not a generally efficient parsing strategy, for the reasons discussed in section 6.1.1. It imposes a strong requirement of adjacency, predictability, and context-sensitivity on the constructing and constructed categories, and the constraints and grammatical complications that these entail will more than offset the added efficiency in IC-to-word ratios. As a result, construction by a daughter category is
Some additional construction principles

systematically required. This rules out any sister node construction to the
left, and construction to the right is limited to those structures in which
sisters are regularly adjacent and predictable, such as $s[\text{Comp S}]$. Even here
there is a certain redundancy, since the sister $S$ must be constructable by a
daughter MNCC as well, in accordance with the axiom of MNCC
Existence. Presumably the added efficiency of SNC offsets the redundancy
when adjacency and predictability are independently motivated, as they are
with a complementizer and a right-adjacent sister $S$. Another plausible
environment in which $S$ is constructed by SNC is when it occurs to the
right of a Topic Phrase in languages such as Japanese, cf. (6.13) below.

6.2.2 Grandmother Node Construction

A construction principle which is more productive than SNC is Grandmother
Node Construction. Sometimes a category may determine not just a mother
node $M$, but also the mother of that mother, i.e. a grandmother $G$. This
principle can be defined as follows:

(6.7) Grandmother Node Construction (GNC)

In the left-to-right parsing of a sentence, if any word of syntactic
category $C$ uniquely determines a grandmother node $G$ directly domi-
nating its mother node $M$, in accordance with the PS rules of the
grammar, then $G$ is immediately constructed over $M$.

Consider finite verbs in English, such as *sings* and *saw* in *sings in the
garden* and *saw Bill at school*. These verbs, like their non-finite counterparts
(to) *sing in the garden* and (to) *see Bill at school*, will construct VP by
MNC. But their finiteness provides additional structural information,
namely the same information that is provided on other occasions by a
tense-marked Aux such as *will* (cf. *will sing in the garden*) or by a modal
such as *may* (cf. *may sing in the garden*), both of which I assume to be
dominated by an AuxP. All these finite verbal elements signal the construc-
tion of $S$. *Sings in the garden* is a VP immediately dominated by $S$, i.e. $s[...$
$vp[sings in the garden]]; *will sing in the garden* has the structure $s[...$
$auxp[will] vp[sing in the garden]]. I will therefore say that a $V_f$ such as
*sings* constructs VP by MNC, and $S$ by GNC. An Aux such as *will con-
structs AuxP by MNC, and $S$ by GNC. A $V$ per se, whether finite or non-
finite, constructs VP alone. The morphology of finiteness provides the
additional information about the attachment of VP to a grandmother $S$. 
Inflectional morphology is, in general, a productive indicator of grandmother nodes, and we shall see several examples of its function in this capacity throughout this chapter.

What about a preceding subject such as *John* in *John sings in the garden*? *John*, as a proper name, determines a dominating NP node uniquely, and will construct NP by MNC. But is an initial NP also sufficient to construct S by GNC? I believe not. An initial NP in English may either be a subject, or some fronted initial constituent, as in *John, Bill likes*, which I assume (following Emonds 1976) to be dominated not by S but by a distinct category, which he labels E for Expression. We therefore have the following structures for these sentence types in English:

(6.8) $s_{NP}[John] \, v_p[\text{sings in the garden}]$

(6.9) $e_{NP}[John], \, s_{NP}[Bill] \, v_p[\text{likes}]$

An initial *John* does not allow us to infer the grandmother node dominating NP, since both S and E are possible. Hence, NP alone is constructed. In (6.8) the finite *sings* (or the Aux in *John will sing in the garden*) then constructs S by GNC, and *John* is immediately attached to this S as an NP subject by ICA, as soon as the VP and S (or AuxP and S) have been constructed. I would argue that in (6.9) the comma intonation plays a crucial role in constructing the mother node E. $NP[John]$ is then immediately attached to this E by ICA. The comma may also render the following S predictable by SNC. Alternatively, $s[Bill \, likes]$ will be processed just like the S in (6.8).

It follows from these assumptions that a subject in English is not recognizable as such until a finite V or Aux is encountered to which the initial NP can be attached (unless the comma renders S predictable in (6.9) by SNC, in which case the next NP encountered will be S-attached and hence immediately recognizable as a subject). But for all basic sentence types such as (6.8), the processing of S-dominated constituents, and the recognition of subjecthood, is crucially dependent upon the presence of constructing categories for S, namely $V_f$ and Aux. There is one class of subject NPs, however, that does permit an S to be immediately constructed on-line, namely nominative-marked pronouns, *I, he, she*, etc. These pronouns are uniquely subjects, necessarily S-attached, and do not occur as fronted constituents under E:

(6.8') $s_{NP}[he] \, v_p[\text{sings in the garden}]$

(6.9') $e_{NP}[him], \, s_{NP}[he] \, v_p[\text{likes}]$
We can therefore say that *he* in English constructs NP by MNC, and S by GNC. More generally, the MNCCs for NP include not just Det and N (discussed in section 6.1.3), but also Pro, and Name as well. Since *he* is the first item encountered in the left-to-right parse in (6.8'), it will also construct S over the subject NP, and the VP will be attached to this S by an application of GNC whose effect will be the same as ICA, given that S has already been constructed. *Sings* therefore constructs the same S as *he*, and this S is, in effect, constructed twice, by two alternative grandmother-node-constructing categories (GNCCs). Once again it is the morphology of the pronoun that signals the construction of a grandmother node over the mother NP.

We can summarize these construction principles for English as follows:

\[
\begin{align*}
(6.10) & \quad V: \quad \text{constructs VP (MNC)} \\
V_f: & \quad \text{constructs VP (MNC); S (GNC)} \\
\text{Aux:} & \quad \text{constructs AuxP (MNC); S (GNC)} \\
\{ \text{Pro, Name} \} & : \quad \text{construct NP (MNC)} \\
\{ \text{Det, N} \} & \quad \text{constructs NP (MNC); S (GNC)} \\
\text{[+ Nom]} & \quad \text{[+ Nom]} \\
\text{IC-1 Comma:} & \quad \text{Comma constructs E; attach IC-1 to E (ICA)}
\end{align*}
\]

That is, all lexical verbs construct VP by MNC. A finite lexical verb, *V_f*, will also construct S by GNC. An Aux constructs both AuxP and S. Four NP-dominated categories can construct NP, including all pronouns, while a nominative-marked pronoun will add the feature [+Nom] to this NP, thereby constructing a feature-marked NP by MNC, as well as S by GNC. Comma intonation constructs E by an appropriate extension of node construction in parsing that allows for construction on the basis of prosodic features as well as syntactic categories.

Consider now some facts from Japanese. Japanese is a “topic-prominent” language (cf. Thompson 1978), with sentence-initial topic phrases (cf. ch. 5.6.3) marked by the postposed particle *wa*, as in (6.11) (taken from Kuno 1973b):

\[
(6.11) \quad \text{Mary wa sono mura ni kita.} \\
\quad \text{“Mary TOP the village to came,” i.e.} \\
\quad \text{Mary came to the village.}
\]
There are some clear pragmatic, semantic, and syntactic properties associated with this topic position (cf. Kuno 1973b), sufficient for us to say that there is a syntactic Topic Phrase in Japanese, constructed by \( \text{wa} \), with the internal constituent structure of \( \text{TopP[NP[Mary] wa]} \). For example, Topic Phrases do not occur in subordinate clauses, but are limited to main or root clauses only. An embedded \( \text{wa} \), as in (6.12a), is ungrammatical and must be replaced by the subject marker \( \text{ga} \), as in (6.12b), cf. Kuno 1973b and also Gundel 1988:

\[
\begin{align*}
\text{(6.12) a. } & \quad \text{*TopP[NP[Mary wa kita mura] wa] doko desu ka} \\
& \quad \text{"Mary TOP came village TOP where is QU", i.e. Where is the village that Mary came to?}
\end{align*}
\]

\[
\begin{align*}
\text{b. } & \quad \text{TopP[NP[Mary ga kita mura] wa] doko desu ka?}
\end{align*}
\]

\( \text{Wa} \) therefore provides two pieces of syntactic information on-line: construct a Topic Phrase, by the usual MNC; but also, construct the highest S, or root clause, \( \text{S}_r \), as the mother of TopP, by GNC. Gundel (1988) speculates on why Japanese should have such a topic marker, and more generally on why these markers are typically found in SOV languages only. The answer she proposes is intimately linked to their GNC function in our terms and corroborates the significance of providing immediate matrix disambiguation, as this was discussed in ch. 5.6.1. In a consistent SOV language with left-branching subordinate clauses and right-peripheral complementizers, the first clausal elements will be assigned simply to S, and a decision about main or subordinate status will have to wait until disambiguating cues are encountered, in the form of a complementizer, or perhaps morphological material on the verb, or just the presence versus absence of syntactic material following the verb. In ch. 5.6.1 I argued that the desirability of having immediate matrix disambiguation explains the existence of left-peripheral complementizers in a number of SOV languages. The -\( \text{wa} \) marker provides another solution, by indicating immediately that the relevant clause is a main clause. Hence both TopP and \( \text{S}_r \) are constructed immediately, and the interpretational properties of these sentences (aboutness in Reinhart’s 1981 terms, the assertion versus presupposition status of main versus subordinate clauses, etc.) can be assigned straight away. Some psycholinguistic support for these hypothesized advantages of immediate disambiguation in clause type comes from the experiments on Japanese conducted by Abe \textit{et al.} (1988), who found that sentences in which a sentence-initial subject had the topic marker -\( \text{wa} \) were easier to comprehend than corresponding sentences with -\( \text{ga} \). In our terms, this would be because of the more rapid and efficient processing of
the syntactic, semantic, and pragmatic information association with the TopP and $S_r$ syntactic nodes, resulting from immediate matrix disambiguation.

The MNC and GNC functions of -wa raise the question of whether the other case particles of Japanese also have GNC as well as MNC functions in parsing. I argued in ch. 4.1.5 that they do. Let us say that $ga$, $o$, and $ni$ construct feature-marked NPs by MNC, $NP_{+Nom}$, $NP_{+Acc}$, and $NP_{+Dat}$ respectively, with all the remaining material in these NPs forming a single constituent as left sister of the particle and dominated by a non-feature-marked NP node, e.g. $NP_{+Acc}[NP_{+Dat}[sono hon] o] \text{ "this book".}$ Now, assuming a division of the basic Japanese sentence structure into $NP_{+Nom}$ and VP (following the grammatical arguments and performance data summarized in ch. 4.1.5), this $NP_{+Nom}$ becomes a constructor of $S$, by GNC, parallel to the construction of $S$ by $Pro$ in English. But equally, $NP_{+Acc}$ and $NP_{+Dat}$ become constructors of VP, again by GNC, since they determine VP uniquely. We therefore have the following construction principles for the NP particles of Japanese:

$$
\begin{align*}
\text{wa:} & \text{ constructs TopP (MNC); } S_r \text{ (GNC); right S (SNC)} \\
\text{ga:} & \text{ constructs } NP_{+Nom} \text{ (MNC); } S \text{ (GNC)} \\
\text{o:} & \text{ constructs } NP_{+Acc} \text{ (MNC); VP (GNC)} \\
\text{ni:} & \text{ constructs } NP_{+Dat} \text{ (MNC); VP (GNC)}
\end{align*}
$$

The GNC function of case marking was also illustrated for Turkish in chs. 4.1.7 and 5.2.2, where it was argued that the VP-internal preposing of short case-marked NPs was motivated by their ability to construct a VP, thereby shortening the CRD for the containing $S$.

The fact that $S$ and VP can be constructed by surface cases within the NP, as well as by auxiliary and verbal elements, is highly significant. Because the subject is typically the first NP of its clause in the great majority of languages (cf. ch. 5.6.2), it will initiate the CRD for $S$. The VP can then be constructed rapidly within this CRD, either by $V$ or by an appropriate case-marked NP. An SVO language typically constructs its VP by $V$, the left-peripheral daughter of VP, unless a clitic pronoun precedes that is capable of constructing VP (cf. ch. 5.2.2). But if $V$ were the only means of constructing VP, then many SOV languages and structures would be doomed to having a necessarily inefficient CRD for $S$. The existence of alternative devices that can construct VP well in advance of the verb provides a solution to this problem (cf. also ch. 5.6.2).

In contrast to the explicit Topic Phrase of Japanese, a number of languages (even those with basic SOV and head-final properties) permit a single constituent in sentence-initial position, often with no clear pragmatic/semantic
correlates, immediately followed by some verbal element in second position. Languages with such "verb-second" constructions include Warlpiri, Luiseño, and German. The verbal element is always finite in these cases: an Aux in Warlpiri and Luiseño (cf. Steele 1975), or any finite lexical or auxiliary verb in German. These Vf or Aux constituents are quite plausibly directly attached to S and can construct it by MNC, while the initial constituent is most likely a sister of this S, both being dominated by a higher sentential node such as Sr. In German, for example, a finite verb in second position is a unique indicator of main-clause status, so that Vf will construct Sr by GNC, in addition to S by MNC, and the initial and functionally diverse constituent will be attached to Sr by ICA. Precise details will vary from language to language. The main point is that it may not be the topic marker itself that constructs the Sr node, as in Japanese, but some verbal element. The initial constituent of a German sentence does not uniquely determine its dominating node. As with the English sentences (6.8) and (6.9), there is the possibility of domination by E as well as by the highest S (Sr). A [-Nom] case-marked NP will not construct a VP in initial position, but may do so in postverbal position. The second-position verb seems to be the crucial higher node constructor, therefore. It constructs (by GNC) the Sr node to which the immediately preceding and typically initial constituent is to be attached, as well as its own mother (by MNC), S; cf. further section 6.3.1.

It should be clear from this brief discussion that GNC makes available some additional ways of constructing phrasal nodes on-line. In particular, it provides alternative phrasal-node-constructing categories (PNCCs) for one and the same phrasal node. I have already suggested that several categories can be MNCCs for the same M: e.g. Det, N, Pro, and Name can all construct NP. Now, with the addition of GNC, we can show that Aux and Vf are alternative PNCCs for S in a language such as English, that \[NP\text{[-Nom]}\] and V can be PNCCs for VP in SOV languages like Japanese and Turkish, and that Sr can be constructed by a topic marker within TopP, or by Vf or Aux occurring in second position after a preceding IC-1. These functionally equivalent alternations enable us to explain a number of variation facts across and within languages, involving the ordering, grammaticalization, and complementary distribution of different PNCCs (cf. section 6.3).

6.2.3 Agreement Projection

Many languages exhibit morphological agreement. C. Lehmann (n.d.) draws a distinction between two kinds, which he labels "internal" and "external"
Some additional construction principles

respectively. An example of the former is adjective and determiner agreement with a noun, in languages such as Latin (6.14):

(6.14) illarum bonarum feminarum
   “that-GEN-PL-FEM good-GEN-PL-FEM woman-GEN-PL-FEM,” i.e.
   of those good women

External agreement is exemplified by Subject–Verb Agreement in English (I sing/he sings), and by the various “head-marking” agreement features characteristic of languages such as Abkhaz (cf. Nichols 1986), e.g. (sara) s-q’ɔ+n+t”, “(I) obl-1sg-from,” i.e. from me.

A general feature of both kinds of agreement is that the agreeing category, B, stands in a very constrained syntactic relation to the category it agrees with, A. Either, in the simplest case such as (6.14), the agreeing adjectives and determiners are plausibly sister ICs of the noun they agree with; or, as in Subject–Verb Agreement in English, the verb is an immediate daughter of VP, and VP is the sister of the NP that the verb agrees with. Similarly, a relative pronoun that agrees with a head noun in Latin, German, and English will be positioned in the Comp position of ɔ[Comp ɔ], and will be an immediate daughter of the ɔ that is a sister to the head noun, ɔ[N ɔ].

This kind of local syntactic relation is referred to as “bijacency” in Culicover and Wilkins (1984), and is defined as follows:

(6.15) **Bijacency**
A node B is bijacent to a node A iff:
(a) B is a sister of A; or
(b) B is immediately dominated by a sister of A.

Agreement appears to operate under a condition of bijacency, therefore, as defined in (6.16):\(^5\)

(6.16) **Agreement**
If a category B agrees with a category A, then B must be bijacent to A.\(^6\)

This bijacency property is functionally motivated, I believe, and can be seen as a consequence of the major functions that agreement performs in language processing. The first function is to tie together the ICs of a single phrase and thereby facilitate the recognition of sisterhood and shared constituency. The determiner and adjective are sisters of the noun in (6.14). The VP immediately dominating an agreeing verb is a sister of the subject NP in Subject–Verb Agreement languages. The relative clause is a sister of the head
noun, and so on. The clear morphological signaling of co-constituency then makes it possible for ICs to be discontinuous from one another, as in Latin *magno cum periculo* "great with danger," i.e. with great danger, in which adjective and noun are separated by the preposition *cum* (cf. Vincent 1987). The role of morphological marking in facilitating non-adjacency can be seen particularly clearly in a language like Warlpiri (cf. Blake 1987: 78). If noun and adjective are adjacent, ergative case marking is attached at the rightmost periphery of the NP, as in (6.17a); if noun and adjective are non-adjacent, as in (6.17b), each must carry the ergative marker, thereby exhibiting surface agreement through case marking, which is a prerequisite for discontinuity in this language (example from Blake 1987):

(6.17)  
a. Tyarntu wiri-ngki-tyu yarlki-rnu.  
"dog big-ERG-me bite-PAST," i.e.  
The big dog bit me.

b. Tyarntu-ngku-tyu yarlku-rnu wiri-ngki.  
"dog-ERG-me bite-PAST big-ERG," i.e.  
The big dog bit me.

There is a second function of agreement, which further motivates its existence and the domain of its application. Agreement can expand the construction potential of a node in parsing. The nominal agreement features on *illarum* and *bonarum* in Latin indicate unambiguously that these are NP-dominated constituents. For *illarum*, the determiner, an NP would be constructable whether there was agreement or not (by MNC). But for *bonarum* it would not be. Hence, adjective-initial NPs, such as *bonarum feminarum*, can be recognized as NPs immediately on the left periphery, without waiting for the noun. Agreement can therefore serve to construct a mother or a grandmother node that is categorically distinct from the word that carries the agreement features. Whether the higher node constructed is a mother or a grandmother will depend on the precise constituent structure involved, e.g. on whether Adj is immediately dominated by AdjP or by NP, and V by VP or by S.

We therefore have a parsing principle of Agreement Projection, defined in (6.18):

(6.18) *Agreement Projection (AgP)*  
In the left-to-right parsing of a sentence, if any word of syntactic category B exhibits agreement features with a syntactic category A,
then the mother node of A is immediately constructed over (the mother node of) B.

The adjective *bonarum* constructs an NP, either over Adj or over AdjP depending on the correct analysis, by projecting from the nominal agreement features to the noun agreed with, and by constructing the mother of this noun over the adjective (phrase). If a verb agrees with a subject, the mother of the subject, S, is constructed over the mother of the verb, VP, if there is one – otherwise S is constructed directly over V. AgP expresses the parsing intuition that B and A are recognized as sisters of the same constituent, or that B is the immediate daughter of the sister of A, whenever B agrees with A. Agreement on B therefore constructs either a mother node for B or a grandmother for B, plausibly because MNC and GNC are also independently motivated and productive principles of parsing. This then constrains the syntactic domains within which agreement relations can hold, and results in the bijacency limitation.

The usefulness of agreement and of AgP in parsing can be seen in several structural contexts. First, if B precedes A, as in the Latin Adj-N construction, then AgP will construct NP, in advance of MNC. This will be advantageous for EIC in all containing CRDs whose ratios are improved by having left-peripheral constructing categories. The CRD for NP itself will be unaffected.

Second, if Adj and N are non-adjacent, as in the Warlpiri example (6.17b), and if each is dominated by an (ultimately co-indexed) NP, instead of being discontinuous daughters of a single NP, then both the adjective and the noun will be able to construct a mother node NP, the former by AgP, the latter by MNC. I leave it open whether repeated co-indexed NPs or crossed branching is the best analysis in any given case. The point is that the former requires an additional construction principle such as AgP in order to construct NP.

Third, AgP makes it possible for the syntactic category agreed with, A, to be deleted altogether, without losing important information about containing structure on-line. *Bonarum* can stand alone in Latin, and is recognizably NP-dominated when it does so. It is known to agree with a noun, whether that noun is physically present or not. VPs with agreement markers on V can also frequently occur without the accompanying subject NP that they agree with. These are the so-called “Pro-drop” languages (cf. Gilligan 1987), and their agreement features are sufficient to construct S by AgP, whether or not the subject is overtly present. In “head-marking” languages such as Abkhaz, the category A is regularly omissible, as in the example cited at the beginning of this section where the PP may consist of an agreement-marked postposition.
only, and the preceding NP agreed with is deleted. In all of these examples, higher node information is either not threatened by the deletion, as in the Abkhaz PP which is constructed by P as usual; or if it is threatened, it can be supplied by AgP, as in the _NP[Adj] and _s[VP] structures. This is not to say that all grammars that could regularly omit category A will do so. There are languages that could be Pro-drop, but are not, such as Modern German. There are also some restrictions on the _NP[Adj] structure in this language that are not found in the corresponding _NP[Det] or _NP[N] constructed by MNC. But my point is that AgP regularly facilitates this kind of deletion and makes it available to grammars as an option.

In Modern English, on the other hand, where there is no longer adjective agreement, we find that additional MNCCs are required to support adjectives when they stand alone within NPs, either an accompanying pronoun _one e.g. _tall ones and _handsome ones, or an accompanying determiner e.g. _the tall and _the handsome, or both, _the tall _ones and _the handsome ones. The determiner and _one can both construct NP, while the uninflected adjective cannot. The residual third-person singular agreement on the verb is also insufficient to activate AgP on a regular basis, and so English does not have the Pro-drop option on the basis of agreement.

Notice that I referred to agreement in (6.18) using a conditional “if any . . . category B exhibits agreement features with a . . . category A,” without discussing which categories will agree with which. The bijacency condition (6.15b) imposes a partial restriction: if B is immediately dominated by the sister of A, then B must be the agreeing category and A the category agreed with, rather than the other way round. But, for sisters (cf. (6.15a)), why doesn’t the noun agree with the adjective, rather than vice versa?

This question has been addressed by Keenan (1979), who provides a semantic solution. The categories that agree are function categories, applying to Ns or NPs as arguments. He argues that Adj is a function on N, Det on N, _S on N, VP on NP, P on NP, and so on, and it is consistently the former that agrees with the latter. He then goes on to propose (by the Meaning-Form Dependency Principle) that the reason why the form of these function categories varies with the choice of arguments is because their precise interpretation varies with these arguments as well. Compare, for example, the different interpretations of the adjective _flat in combination with nouns such as _beer, _road, and _tyre.

The most productive case of agreement across languages appears to be Subject-Verb Agreement. Even languages with little or no agreement elsewhere in their grammars, such as English, may exhibit Subject-Verb
Agreement, however residually. And among Nichols's (1986) "dependent-marking" languages, one "head-marking" feature that frequently co-occurs with dependent marking is Subject–Verb Agreement. This fact needs to be explained. In addition, why should agreement with a subject be significantly more common than agreement with a direct or indirect object (as in Basque and Abkhaz)?\[^7\] AgP can contribute to a solution.

One major effect of subject agreement on a verb is to extend the Construction Domain of that verb, as this term was defined in (6.2). V will construct VP by MNC, but a V with Subject–Verb Agreement will also construct S, the mother of the subject, over VP by AgP. Such agreement therefore functions like finiteness on a lexical verb in English, which constructs S over VP by GNC. Finiteness is unique to S, and this explains its productivity as a constructor of S, and its absence from clauses that are dominated only by VP and that lack a subject NP. But equally, agreement with a subject can construct S as well, and since subjects are the most productive and frequent argument types in all clauses, S-construction through agreement with a subject is a grammaticalization option that is readily available to very many languages.

In order to explain why subject agreement is much more common than direct- and indirect-object agreement, therefore, we can begin by observing that subjects are more common than direct or indirect objects. But this can't be the whole story, because we still need a reason for why any kind of agreement system should be grammaticalized. Part of the reason is plausibly the clear marking of co-constituency: the VP is a sister of the subject NP, V is a sister of the direct object NP, and so on. But this does not distinguish between subject and object agreement. We must therefore look to the second function of agreement: its construction potential.

Notice the following asymmetry. If the verb agrees with a VP-dominated non-subject NP, then since V alone constructs VP (by MNC), and since the mother of the non-subjects is also VP, agreement adds nothing to the construction potential of V. But if the verb agrees with an S-dominated subject, then agreement will extend the Construction Domain of V beyond the VP constructed by MNC alone. Hence, there is greater utility in having a grammaticalized subject agreement system than in having object agreement.

Moreover, the need for S-construction by V follows from the fact that V imposes subcategorization and case requirements on arguments outside the VP (regularly, in fact), as discussed in section 6.1.3; yet V alone can only construct VP. Additional parsing devices are therefore needed that are unique to S and that can construct it, thereby extending the Construction Domain of
the verb so that it can define subcategorization etc. outside of the VP. If V is 
finite, it will construct S by GNC; if it exhibits agreement with the subject, it 
will construct S by AgP. These are the two major morphological devices that 
languages make use of for this purpose, and the processing approach of this 
book can make sense of their role and grammaticalization. First of all, it 
explains why there should be a finiteness category sensitive to the presence or 
absence of a subject, such that the verb takes one morphological form in 
clauses that contain a subject, and another in clauses that do not. 
Finiteness is like subject agreement in this regard, and the reason for their 
similarity is that both are constructors of S. Secondly, it explains why subject 
agreement should be more frequent than object agreement. The former 
extends the Construction Domain of V beyond the VP and indicates co-
constituency between this VP and the subject NP; the latter indicates co-
constituency only (within the VP).

Our approach also makes a prediction. Any verb that defines syntactic or 
semantic co-occurrence requirements on NPs outside of VP in any given 
sentence, i.e. outside the highest node constructed by V alone on the basis 
of MNC, must have distinctive morphological marking that can extend the 
Construction Domain of V to S. In English, sentences with overt subjects 
have co-occurring verbs with finite morphological marking, and also residual 
subject–verb agreement. By contrast, sentences with non-overt subjects (i.e. 
those whose subjects have been deleted or raised) will not need to construct a 
higher S, and these have non-finite verbs and exhibit no agreement with the 
subject. If S is to be constructed over VP, therefore, we predict the need for 
some additional morphological marking that is unique to S. The precise 
options selected by different languages can vary, but the need for S-construc-
tion in these cases cannot.

Notice finally that just as category construction does not entail category 
constancy for MNCCs like Det, as argued in section 6.1.2, so it does not 
entail it either when the relevant constructing principle is GNC or AgP. In 
other words, just because a finite Infl or agreement can construct S does not 
warrant setting up an Infl Phrase or an Agreement Phrase, as in current 
Government-Binding theory. To do so introduces an unnecessarily complex 
syntactic structure, with numerous nodes and constituency groupings that 
have very little surface justification. It incorporates the correct insight that 
the relevant dominating category is, say, a projection from the inflectional 
element, but derives from this (at least implicitly) the counterintuitive con-
sequence that this category is somehow an inflectional phrase. This misses the 
same generalization that the NP/DetP dichotomy misses: we are dealing here
with different ways of constructing the same higher constituent or word grouping. There is no good reason why there should be a bi-uniqueness relation between the constructing category or affix and the phrase type constructed. Uniqueness is sufficient for processing, and I would argue that it is processing that shapes the syntax in this area and that explains the grammaticalization of the various PNCCs we have seen.

6.2.4 Attachment Promotion and Construction

We have seen that the Construction Domain for a category X can be extended from the nodes dominated by the mother of X to nodes dominated by a higher constituent constructed by X, using principles such as GNC and AgP. A Construction Domain is therefore defined in (6.2) as a set of nodes dominated by the highest phrasal constituent constructed by X. These nodes constitute a Structural Domain for X in C, as this term was defined in ch. 2.2.1 (cf. (2.11)), i.e. a set of nodes that structurally integrate X within a larger constituent C, here the highest constituent constructed by X.

There are a number of phenomena across languages, involving the immediate attachment of a category Y to a constructing category X, that are rather remarkable, but which begin to make sense when viewed from a processing perspective. It seems that the immediate attachment of Y to X enables Y to behave syntactically and semantically as if it were attached to the highest node constructed by X. If Y is a non-constructing category, it can behave as if it is a structural sister to the highest node constructed by X. If it is a constructing category, and if the highest node constructed by X dominates the highest node constructed by Y, then by attaching Y to X the Construction Domain for Y can actually be extended to that for X, so that Y has the same Construction Domain as X. I shall refer to the former cases as Attachment Promotion, and to the latter as Attachment Construction. In both it is the regular predictability of higher node construction when X is being parsed that makes the higher constituent structure immediately available on-line. A node Y attached to X can then be interpreted as attached to X itself, or to a projection of X, and the Construction Domain for Y can also grow to subsume that of X.

Let us define Attachment Promotion as follows:

(6.19) **Attachment Promotion** (AtP)

In the left-to-right parsing of a sentence, if a category Y, Y a non-constructing category, is immediately attached to a sister X, X a
PNCC, then the sisterhood relation between Y and X can be extended to the highest node constructed by X.

In other words, Y can, in effect, c-command the whole Construction Domain for X, by being immediately attached to X itself, and a node within this Construction Domain can function as if it is c-commanded by Y, even though it may c-command Y, even asymmetrically, in the actual surface structure. If we think in terms of the Principle of Structural Determination (2.34) defined in ch. 2.4, we should therefore expect to find that some nodes in the Construction Domain for X can be syntactically and/or semantically dependent on Y, even though they do not appear to have Y in their SDs.

For example, there is a so-called focus position immediately prior to the verb in a number of SOV languages (cf. A.H.-O. Kim 1988). Turkish is an example: cf. the analysis of Erguvanli (1984). In fact, this position provides a means of extending syntactic dependencies and logical scope relations in accordance with AtP. Assume that the verb in Turkish constructs VP by MNC, and S by GNC or AgP on the basis of the appropriate morphological affixes. Then, by attaching a category directly to the verb, that category can function as a sister of S, the highest category constructed by V. Hence, one would expect to be able to find antecedent NPs in preverbal position, whose anaphors were contained within S, but not necessarily within VP. Murvet Enc has confirmed (in a 1990 lecture at the University of Southern California) that there are indeed anaphors in Turkish that can asymmetrically c-command their antecedents in focus position. Consider here a related example involving a WH-word and the whole sentence over which it takes its scope. A WH-word occurs regularly in the immediately preverbal position in Turkish, often obligatorily, as in example (6.20a) from Erguvanli (1984: 48):

(6.20) a. Para-yi kim cāl-di?
   “money-ACC who steal-PAST?,” i.e.
   Who stole the money?

   b. *Kim para-yi cāl-di?

Kim “who” is the grammatical subject in this example, and by being positioned immediately prior to the verb it takes the whole sentence in its scope, including the accusative NP which is now separated from the verb by the WH-word. The meaning of (6.20a) is: “Who is such that that person stole the money?” with the whole Construction Domain of V in the scope of the question word.
Some additional construction principles

Example (6.20) suggests that there are two basic ways in which languages can grammaticalize structural positions for the surface expression of logical scope relations. Either the category with wide scope can be moved into a structural position where it directly c-commands the categories within its scope; or, alternatively, it can be placed immediately adjacent to a PNCC, where it can be interpreted as a sister of the highest projection of this latter, thereby c-commanding everything in its Construction Domain in the extended constituent structure derived by AtP. The first solution is the one adopted by English. The WH-word is moved directly into the position of immediate sisterhood with the S within its scope, i.e. ɛ[WH S]. The second solution is the Turkish one, where WH moves directly to V, possibly forming a single constituent with it, i.e. ɣ[WH V]. Kim (1988) points out that this solution is one that is regularly adopted in SOV languages. The final possibility, of course, is to leave the WH-word in its basic word order, as dictated by the normal constituency and ordering rules for the relevant language, and to convert this structure into the appropriate logical representation in which c-command relations are expressed. The resulting mapping from surface structure to logical structure is rather non-transparent and complex, and it is presumably for this reason that grammars regularly help the parser by grammaticalizing structural positions in the syntax for scope-bearing elements, in the two ways that we have seen.

More generally, I would suggest that the set of c-command relations that determine syntactic and semantic dependencies in surface syntax can be extended on the basis of AtP to include Construction Domains and sister categories that are immediately adjacent to the relevant PNCCs, as defined in (6.19) and as exemplified by the preverbal position of many SOV languages. From the perspective of the Principle of Structural Determination (2.34), the set of such dependencies is correspondingly extended and a number of reversals of logical scope or antecedent–anaphor relations are accordingly predicted from what would be predicted by surface structure alone, without the principle of AtP. Considerations of space preclude a more detailed discussion in this context, and it remains to explore the full extent of AtP's operation in different languages and language types, the precise categories that can be attached, and the precise manner of their structural attachment. My only goal here is to show that there is a clear functional motivation for extending the sisterhood relation in syntax beyond categories that are actually adjacent, in the event that one of them constructs a higher phrasal node.

Similarly, in the event that both adjacent nodes Y and X are constructing categories, it seems that the construction potential of the one can amalgamate
with that of the other, so that the lower Construction Domain is, in effect, extended to that of the higher domain. Let us define this possibility as follows:

(6.21) **Attachment Construction (AtC)**

In the left-to-right parsing of a sentence, if a category Y, Y a PNCC, is immediately attached to a sister X, X a PNCC, and the highest node constructed by X dominates the highest node constructed by Y, then the Construction Domain of Y is extended to that of X.

German and Dutch provide an example of this in their clause-final verb-cluster constructions (cf. e.g. Jacobs 1992 for German, and Steedman 1984 for Dutch). Let us assume that the subcategorization and case-assignment requirements of verbs are defined to hold within Construction Domains, cf. (6.3) and (6.4). Then if the main, finite verb can form a cluster, and quite plausibly a constituent, with subordinate non-finite verbs, and if the (preceding) arguments of these different verbs can be rearranged within the clause, then subcategorization and case-assignment requirements may no longer be satisfied within Construction Domains so defined. Consider the German examples of (6.22) (from Jacobs 1992):

(6.22) a. weil s[ein kluges Kind]VP[das ohne weiteres v[verstehen kann]]

because a clever child can understand that without difficulty

b. weil s[das ein kluges Kind]VP[ohne weiteres v[verstehen kann]]

There are numerous co-occurrence requirements that must be satisfied in a complex embedding structure such as this. Regardless of the particular syntactic analysis that one adopts, it is clear that the verb _verstehen_ must assign accusative case to _das_. If the positioning of _das_ is as shown in (6.22a), then _das_ will not technically be in the Construction Domain for _verstehen_. _Kann_ alone, being finite and exhibiting agreement with the subject, can potentially construct both VP and S. But in the compound _v[verstehen kann]_, it is the higher V that constructs the mother VP, and the morphology of _kann_ then constructs S, while _verstehen_ does not literally construct anything, even though it is a category that can potentially construct VP. One could imagine an alternative syntactic structure for (6.22a) with an embedded VP within the higher VP, the lower one constructed by _verstehen_ and the higher one by _kann_, and in this
structure the accusative would be in the Construction Domain for *verstehen. But there are various arguments against this (cf. also Steedman 1984 for a literature summary of the arguments in favor of a verb-cluster structure in Dutch, corresponding to that of German, even though the order of main and subordinate verbs is inverted in Dutch), and in any case Jacobs argues that this will not solve the problem of (6.22b), in which the accusative *das has been rearranged to the leftmost position in S, preceding the nominative subject *ein kluges Kind. Therefore, (6.22b) does not satisfy the Construction Domain requirement for case assignment formulated in (6.4).

I consider it significant that these clause-final verbs are immediately adjacent to one another and form a tightly integrated verbal group. In fact, Jacobs (1992) uses the term "integration," and he points out that these clusters function like a single word that has undergone morphological compounding (cf. the similar analysis of Bierwisch 1990). The phrase *ohne weiteres in (6.22), for example, cannot intervene between *verstehen and kann: *weil das ein kluges Kind *verstehen ohne weiteres kann. I suggest that what is going on in (6.22) is an extension of the Construction Domain for *verstehen by virtue of its immediate attachment to kann. This latter, as a finite verb, can potentially construct both a VP and S. If *verstehen is attached to make a compound higher verb, V, then this V will construct the matrix VP in the usual way (by MNC), and the finiteness and agreement features on kann will construct S. The subcategorization and case requirements of kann can therefore be satisfied throughout the whole S. But so too can those of *verstehen, by virtue of AtC. Because of its attachment and quasi-morphological-compound status, *verstehen alone does not construct any Construction Domain within this structure, even though it is a constructing category. However, by being immediately attached to the finite V, kann, it is free to satisfy its subcategorization and case-assignment requirements anywhere within the S constructed by kann, and hence structures such as (6.22b) are completely productive. Moreover, this process is iterative. The constituent [verstehen kann] is a compound finite verb by virtue of the finiteness of kann, and can be combined with another non-finite verb whose co-occurrence requirements can also be satisfied within the highest node constructed by finiteness, and so on. The Construction Domain for *verstehen is therefore extended to that of kann by AtC.

Jacobs (1992) also discusses an example of subcategorization and case assignment by a preposition in German that points to a similar application of AtC. There are dialects of German in which a limited form of Preposition Stranding is grammatical, converting sentences such as (6.23a) into (6.23b):
Grammaticalized node construction

(6.23) a. weil s[die Leute vP[pp[damit] rechnen]]
   "because the people on-it count," i.e.
   because people are counting on it

b. weil s[da die Leute vP[y[mit rechnen]]]

In (6.23b) the components of the PP, da and mit, are separated, and when this happens the preposition mit must be immediately attached to the verb rechnen. Jacobs points out that there cannot be any intervening adverbials, such as sicher "certainly": *weil da die Leute mit sicher rechnen, i.e. because people are certainly counting on it. Once again we can say that the finite and agreement-marked verb rechnen constructs both VP and S, and the subcategorization and case-assignment requirements of the preposition mit, a constructing category, can then be satisfied anywhere within the highest S constructed by the verb to which it is immediately attached.8

Obviously a lot more needs to be said about both Attachment Promotion and Construction. My point is simply that the explanatory basis for these kinds of syntactic and semantic phenomena appears to be the capacity of certain types of categories to render quite abstract constituent structure accessible and predictable on-line, whereupon an immediately attached category can be interpreted relative to this higher structure. Adjacency appears to provide the signal for such promotional processes, but it remains to be established whether such adjacency is always required, and what kinds of categories can exemplify X and Y in our definitions, and with what consequences.

For example, infinitival embeddings (or Equi-NP Deletion structures) such as John wants to go to the movies provide a further example of Attachment Construction, whereby the Construction Domain projected from to go is extended to that projected from wants, i.e. the whole S. As a result, the subcategorization requirements of the infinitival verb can be satisfied by John, the matrix subject, as well as by material within the embedded VP. But these infinitival embeddings can also occur with three-place matrix verbs such as I persuaded John to go to the doctor and even I promised John to go to the doctor, in which the infinitival verb is not directly adjacent to the matrix verb, the category that constructs the higher VP and S. In order to capture the obvious descriptive generalization that underlies all these Equi-NP structures, we must increase the possibilities for extending a subordinate Construction Domain beyond the set defined in (6.21). For example, the Construction Domain of the subordinate Y, Y a PNCC, might be extended to that of X, X a PNCC, if there is no intervening Z that constructs a higher
phrase distinct from the higher phrases constructed by both X and Y. According to this revision, it would be the adjacency of to go with I persuaded John and I promised John that would extend the Construction Domain of the infinitival to include all the nodes of these matrix clauses.

6.2.5 The axiom of constructability

After this discussion of parsing principles that can construct nodes other than an immediate mother, we need to return to the axioms of MNCC Existence and Uniqueness which motivated the principle of Mother Node Construction:

(3.5) Axiom of MNCC Existence

For each phrasal mother node M there will be at least one daughter category C that can construct M on each occasion of use.

The Axiom of MNCC Uniqueness stipulated that each such category C will construct a unique M (cf. (3.6)).

It should now be clear that we need to modify (3.5) in order to take account of the possibility that M may be constructed on some occasions not by a daughter, but by some granddaughter node, either on the basis of its morphological properties (e.g. the finiteness of V) or on the basis of its categorial identity (e.g. Aux in English). The crucial insight to be captured in our parsing axioms, therefore, is that each phrasal node P must be constructable by some dominated word of category C that will render P recognizable on each occasion of use. According to our theory, abstract syntactic structure must always be recognizable on-line, on the basis of terminal elements in the parse string. The axiom of MNCC Existence was proposed in the context of a parser whose only construction principle was MNC. Now that additional possibilities have been defined, we must replace (3.5) with the more fundamental axiom of Constructability:

(3.5') Axiom of Constructability

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.

The axiom of MNCC Uniqueness can remain, since any MNCC must still consistently construct just one unique mother. It can be generalized to refer to the additional possibility of unique grandmother node construction, however, as follows:
Grammaticalized node construction

(3.6') Axiom of PNCC Uniqueness

Each PNCC will consistently construct a unique M, and possibly a unique G, on each occasion of use.

The claim of (3.6') is effectively that a category will construct a grandmother only if it also constructs a mother. All the GNCCs considered in this chapter seem to have this property, and it makes sense, since a mother must always be present if a grandmother is to be superimposed on it. It remains to be seen, however, whether a given category could uniquely determine a grandmother without also constructing the mother at the same time. In this case the mother would need to be constructed independently.

Notice that I have continued to insist in (3.5') that P must be constructed by a category that is dominated by P, i.e. it cannot be constructed by a sister alone. The motivation for this is the same as that which was discussed in connection with projection from a daughter for MNCCs (and heads) in section 6.1.1.

Notice also that the axiom of Constructability leaves open the possibility of yet other parsing principles, such as Great-Grandmother Node Construction, as long as the phrasal node P is recognizable on the basis of the relevant lower category. It would be interesting to see how far constructability can proceed up the line of dominating nodes within a tree in different languages. In general, the greater the distance between a dominating and a dominated node, the less predictable the former will be on the basis of the latter, on account of the greater number of selectional options that can intervene. This is why construction of a mother by a daughter is highly productive, cf. MNC; construction of a grandmother is less productive and often requires additional morphological material to supplement the syntactic categories themselves; while construction of even higher phrasal nodes appears to be either non-existent or very rare (though cf. sections 6.3.2 and 6.3.4). This hierarchy follows from the declining constructability and uniqueness that accompany increasing distance between dominating and dominated nodes. The limits of constructability are not predicted, however, and remain to be investigated. Depending on the outcome, (3.6') may need to be generalized still further.

6.3 Some predicted category and ordering alternations

In section 6.2.1 I predicted a possible ordering alternation within the NP for languages that are otherwise left-mother-node constructing: either Det or N
could be the leftmost daughter of NP, since both are MNCCs for NP. Now that we have increased the set of construction principles to those discussed in section 6.2, the potential for this kind of ordering alternation between functionally equivalent categories increases also. So does the potential for actually grammaticalizing (i.e. possessing) alternative constructing categories. I shall assume here, in accordance with the axiom of Constructability (3.51), that every phrase contains at least one dominated word and category that can construct it. What can vary is whether this phrase is constructed by MNC, GNC, AgP, etc. We have seen, for example, that both a non-nominative NP and V can potentially construct a VP, and that the root S can be constructed by a topic marker within a TopP or by a second-position finite verb. In languages that contain such alternative constructing categories, the principle of EIC is just as compatible with construction by a non-immediate daughter as it is with an MNCC, as long as these categories occur in the position predicted. Either way the phrase will be constructed, and optimal IC-to-non-IC ratios can be guaranteed. Hence we predict alternations between identically or similarly positioned categories constructing one and the same higher node in parsing, both within a language and across languages. In what follows I shall present some further examples of these alternations.

### 6.3.1 Alternative constructors of sentential nodes

The clause, S, is a vital constituent in parsing, and there are a number of ways of constructing it. In section 6.2.2 we saw that it could be constructed by a [+ Nom] NP or pronoun, or by a finite verb or auxiliary. In English, both a finite lexical verb within VP and an auxiliary within AuxP will construct S by the same principle (GNC), and typically at the same relative point in the parse string, immediately to the right of the subject (which may also construct S if it is [+ Nom]). Finite lexical verbs and auxiliaries are in complementary distribution: if there is no auxiliary, the lexical verb is finite; if there is an auxiliary, the lexical verb is non-finite. This complementarity, coupled with their identical construction site, follows from their functional equivalence in our terms. They are both S-constructors.

In this section I shall consider S-construction by a verb in more detail, and in relation to S-construction by other categories. In the simplest case there may be a unique position within S for finite verbal material, e.g. the second position for the finite auxiliary in Warlpiri (cf. Kashket, 1988), activating S-construction by MNC. The finite verb may also move into this mother-node-construction position from a different position, e.g. from within the VP. This
latter possibility is widespread throughout the Celtic and Germanic languages. In Welsh the highest V within VP is moved into the leftmost position of S (cf. e.g. Sadler 1988). The moved word may be a lexical verb, as in (6.24a), or an auxiliary, as in (6.24b) (from Sadler 1988):

“see-PAST-3Sg Mary the accident,” i.e.
Mary saw the accident.

b. Gwnaeth Siôn ennill.
“do-PAST-3Sg John win,” i.e.
John won.

The first exemplifies the order VSO(X), the second AuxSV(OX). The fronted word is always finite in surface structure and constructs S on the left periphery of that constituent, making S consistent with all the other phrasal categories of Welsh, which are also left-mother-node constructing. Thus, Welsh has prepositions within PP, adjectives precede their complements within AdjP, the NP is initiated by Det or N, a non-finite V is leftmost within VP, and so on. The left-peripheral positioning of finite verbs can therefore be motivated by the fact that all constructing categories are left-peripheral, and that Vf is the MNCC for S (without the need for category constancy, as usual, cf. section 6.1.2). It is further motivated by the fact that a finite verb is a single-word item. By positioning it to the left of the subject, which is a possibly multi-word phrase (cf. ch. 5.3), the CRD for S achieves an optimal (left-to-right) IC-to-word ratio (as long as the fronted verb is extracted from the VP and is not discontinuously attached to it).

Some support for this analysis comes from embedded clauses in which the verb can be either finite or uninflected. If the verb is uninflected it can construct a VP (by MNC), but it cannot construct S since it is no longer unique to S. We accordingly predict that verb fronting will not apply in these cases, and that these embeddings will maintain the basic SVO order that is assumed for Welsh, with the verb(s) under the domination of VP. This prediction is correct. The finite S-constructing verb ennillai in (6.25a) undergoes fronting, its uninflected counterpart ennill in (6.25b) does not (cf. Sadler 1988: 31–32):

(6.25) a. Disgwyliais i [yr ennillai Siôn]
“expect-PAST-1Sg I that win-COND-3Sg John,” i.e.
I expected that John would win.
Some predicted category and ordering alternations

b. Disgwyliaisiai [i Sîôn ennill]
   "expect-PAST-1Sg I for John win," i.e.
   I expected John to win.

The two embeddings are distinguished both by the form of the subordinate verb and by the choice of complementizer. Yr constructs a finite clause as right sister (by SNC), i constructs a non-finite clause, and they occur in complementary distribution. There is much debate about the precise constituent structure of (6.25b), but from the perspective adopted here, the uninflected verb is not a constructor of S, and for this reason it does not get preposed.

The continental Germanic languages have a very similar verb-fronting rule moving the highest V from the VP to left-peripheral position within S. This verb can be preceded by an initial topicalized constituent, which I assume to be immediately dominated by a separate constituent, namely the root S_r; or else this initial constituent may be empty. The result is the so-called "verb-second" and "verb-first" constructions of main clauses in Germanic. Examples from Danish are given in (6.26) (examples from Bredsdorff 1962):

(6.26) a. I går så jeg ham på gaden.
   "yesterday see-PAST I him on street-the," i.e.
   Yesterday I saw him on the street.

b. Har De været i Danmark?
   "Have-PRES you been to Denmark," i.e.
   Have you been to Denmark?

As in Celtic, the fronted verb must be finite, and it originates in VP. Also as in Celtic, the left-peripheral construction of S makes this node fundamentally similar to the other phrasal categories of Danish and Scandinavian generally. Thus, the verb is otherwise initial within VP, there are prepositions in PP, the adjective precedes its complements within AdjP, and so on. Left-peripheral node construction is the norm, therefore. The fronting of V_f is also motivated by EIC within the S domain, and it shortens the CRD for S_r in the event that there is a topicalized initial constituent, as in (6.26a), which has the structure $s_r[IC-1 s[V_f . . .]]$. V_f indicates the onset of S in this structure, which is the second IC of the S_r domain. The earlier S is constructed, the smaller the CRD for S_r will be, and the better its EIC ratio.

There is one major difference between verb fronting in Welsh and Danish. The Danish rule does not apply in finite subordinate clauses corresponding to
Grammaticalized node construction

(6.25a) in Welsh. Thus, if (6.26a) is embedded under a higher verb, as in (6.27), finite-verb fronting does not take place:

(6.27) a. Jeg ved [at jeg så ham på gaden i går]
   "I know that I see-PAST him on street-the yesterday," i.e.
   I know that I saw him on the street yesterday.

b. *Jeg ved [at i går så jeg ham på gaden]

c. *Jeg ved [at så jeg ham på gaden i går]

We can account for this by stating that a second- or first-position finite verb within a Danish sentence constructs a grandmother node in addition to the mother node S, namely Sr, the same main or root clause that is constructed by the topic marker wa in Japanese, again by GNC (cf. (6.13) above). If an initial constituent is present, as in (6.26a), it is attached to Sr by ICA. Because the structure of (6.27a) involves a subordinate clause constructed by the complementizer at ("that"), verb fronting cannot apply: the grandmother constructed by the first or second position Vf conflicts with the mother constructed by the complementizer. Let us abbreviate this mother, a finite subordinate clause, as Ss within the present context.

We have, therefore, the following construction principles for Welsh and Danish respectively:

(6.28) Welsh
   Vf: constructs S (MNC)
   Vnf: constructs VP (MNC)
   yr: constructs Ss (MNC); right S (SNC)

(6.29) Danish
   # (IC-1) Vf: constructs S (MNC); Sr (GNC); attach IC-1 to Sr (ICA)
   Vf elsewhere: constructs VP (MNC); S (GNC)
   Vnf: constructs VP (MNC)
   at: constructs Ss (MNC); right S (SNC)

(6.28) and (6.29) illustrate S-construction through MNC and also through SNC. For Danish we have the additional possibility of S-construction through GNC in the event that the finite verb remains within the VP, as in the subordinate clause of (6.27a). We do not want the fronted Vf to construct a VP in first or second position, i.e. in main clauses, since Vf serves as a direct MNCC for S in these cases, constructing Sr by GNC. Elsewhere, i.e. in subordinate clauses, Vf is an MNCC for VP and constructs S by GNC.
Summarizing, there are a number of distinct categories in Welsh and Danish that can construct an S: lexical verbs with the appropriate morphological inflections; non-lexical or auxiliary verbs with the appropriate morphological forms; and complementizers such as *yr* and *at*. What is significant from the point of view of ordering is that each S is consistently constructed at its onset by at least one of these categories, and that they conspire to create the same effect. Both main and subordinate clauses are consistently recognized on their left peripheries, by a V<sub>f</sub> and/or a complementizer. Moreover, the two languages contrast at exactly the point at which there is a functional redundancy in their parsing routines, making alternative grammaticalizations both possible and predictable. Subordinate clause complementizers like *yr* and *at* will immediately construct a right sister S by SNC. Hence, it is no longer necessary to front the V<sub>f</sub> in order to construct this S on its left periphery. According to our axioms, there has to be some construction of S by a dominated constituent, but the positioning of this constituent does not have to change when left-peripheral node construction is already being accomplished by SNC. A grammar could therefore respond to this functional redundancy by limiting the fronting of V<sub>f</sub> to main clauses only, in which there is no overt complementizer. This is what Danish has done. Welsh, on the other hand, has simplified and generalized its verb-fronting rule so as to apply in all S-environments, whether dominated by S<sub>r</sub> or S<sub>s</sub>, but at the expense of a certain redundancy. And whereas verb fronting then becomes a crucial component of the surface signaling of the S<sub>r</sub>/S<sub>s</sub> distinction in Danish, it does not have this function in Welsh. Instead, S<sub>r</sub> appears to be constructed by a topicalization particle similar to that of Japanese (cf. Sadler 1988).

As in Danish, the S node in German is also consistently constructed on its left periphery by finite lexical or auxiliary verbs and by complementizers such as *dafi*. In fact, the parsing routines of German are identical to those of Danish (cf. (6.29)), despite the difference between the two languages in the grammar of their basic verb positions. The fronted V<sub>f</sub> of German originates in final position within the VP, making the contrast between main clause verb-first and verb-second orders and subordinate clause verb-final orders superficially more striking than the corresponding contrast in Danish, with and without verb fronting. But verb fronting itself applies to identical lexical and auxiliary categories, and under identical conditions. And because the fronted V<sub>f</sub> in German constructs both S (by MNC) and S<sub>r</sub> (by GNC), it is just as incompatible with the S<sub>s</sub> constructed by *dafi* as its Danish counterpart is. Compare (6.27) with (6.30):
Grammaticalized node construction

(6.30) a. Ich weiß [daß ich ihn gestern auf der Straße sah]
   "I know that I him yesterday on the street see-PAST-1Sg," i.e.
   I know that I saw him on the street yesterday.

b. *Ich weiß [daß gestern sah ich ihn auf der Straße]

c. *Ich weiß [daß sah ich ihn gestern auf der Straße]

Despite the fact that German is basically verb-final, therefore, its S-nodes are systematically recognized by a left-peripheral daughter (for main clauses) or sister (for subordinate clauses). Moreover, I shall suggest in section 6.3.3 that the VP in German is also systematically constructed well in advance of the verb, by non-nominative case-marked NPs. S and VP seem to be at variance with the other phrasal nodes of German which are predominantly left-mother-node constructing, and this creates the appearance of considerable inconsistency in this language. Thus, NPs are uniformly recognized by a left-peripheral Det, inflected Adj or N; PPs are predominantly prepositional; most complements of an Adj within AdjP follow the constructing Adj (apart from NPs). Hence, the rest of German grammar has predominantly left-peripheral MNCCs. When we realize, therefore, that S and VP, the categories that seem most inconsistent when viewed from a narrowly grammatical point of view, are actually constructed at or near their left peripheries by additional construction principles such as SNC, GNC, and possibly AgP as well, the typological inconsistency disappears. German is fundamentally a left-peripheral constructor of higher phrasal categories, just like Danish and English. The difference consists in the set of categories that actually do the constructing. The V in Danish and English is the first on-line daughter category to construct the VP; in German it may not be and cases may do the job instead.

From a typological perspective we therefore see that German is not, as is often thought, a Japanese-type language, despite the position of the verb. Japanese is a consistent right-peripheral constructor of higher nodes, with the exception of the matrix S_r constructed by the topic marker (cf. (6.13)). German is fundamentally a left-peripheral constructor, and it exemplifies the variation space that is available to grammars that have additional devices, particularly case marking, for the grammaticalization of left-peripheral node construction. The fact that case is assigned to the left in this language is a consequence, as I see it, of the fact that case is a higher node constructor, just like V, or like P within PP. And because German is a left-peripheral node constructor, one or the other higher node constructor must occur at or near
the left periphery within each phrase. German is therefore fundamentally English-like in the positioning of its node-constructing categories (cf. further sections 6.3.3 and 6.3.4).

Finally, notice how the reduced productivity of the verb-fronting rule of Danish and German within English has resulted in more complex parsing routines in this language. The major grammatical details are summarized in Hawkins (1986). Notice in this context only that the regular verb-fronting rule of Danish and German has become more limited in the range of environments in which it applies, and that the relative positioning of subject NP, lexical verb, and direct-object NP has become fixed. As a result, the initial IC-1 that triggers fronting of the finite verb and construction of both S (MNC) and S_r (GNC) will not contain a subject or object NP. It may be a PP (down the hill ran John) an Adv (never have I done this before), etc. (cf. Hawkins 1986: 180). The grammar of English has also separated the categories of Aux and V and introduced a new phrasal category of AuxP (cf. Lightfoot 1979). Despite these grammatical changes, the parsing routines for S in English are actually similar to those of the other Germanic languages, and we can now supplement the principles of (6.10) above with some extra details, following the principles proposed for Danish (and German) in (6.29). I shall simply present the construction principles of English without further discussion:

(6.31) English
# (IC-1) V_f/Aux_f: constructs S (MNC); S_r (GNC); attach IC-1 to S_r
[-NP] (ICA)
V_f elsewhere: constructs VP (MNC); S (GNC)
V_nf: constructs VP (MNC)
Aux_f elsewhere: constructs AuxP (MNC); S (GNC)
that (unstressed): constructs S_s (MNC); right S (SNC)

6.3.2 Alternative subordination indicators

In the last section we saw a number of alternative ways of constructing the sentential nodes, S, S_r, and S_s, via our various parsing principles, in languages that are otherwise left-mother-node-constructing. These alternatives all have the effect of constructing clauses on their left peripheries, and so contribute to a very general parsing regularity: all phrasal nodes in these languages are recognized at or near their left peripheries, by daughters, granddaughters, or adjacent sisters. Right-mother-node-constructing languages have at their
disposal similar alternatives for constructing higher phrases, in the form of identically positioned and functionally equivalent constructing categories. I shall illustrate the right-peripheral regularity on the basis of alternative types of (prenominal) relative clauses. Recall that a number of these languages have gone over to postnominal relative clauses, for reasons of Immediate Matrix Disambiguation as discussed in ch. 5.6.1. In this section I shall consider the alternative ways of constructing relative clauses in languages that maintain the right-peripheral-node-construction regularity in their NPs, i.e. with head-final nouns and prenominal clauses.

There are a number of different types of prenominal relative clauses, according to C. Lehmann (1984: 49–72). From a parsing perspective we need to distinguish at least the following four partially distinct structures and parsing routines. What unites all of them is optimally efficient IC-to-word ratios within the CRDs for NP.

The first type is exemplified by Lahu and Chinese. These languages have a nominalizing particle, ve in Lahu, de in Chinese, positioned on the right periphery of the relative clause and to the immediate left of the head noun. These particles are not uniquely indicative of relative clauses as such. They occur in subordinate clauses of other types, and with non-clausal modifiers of a head noun, such as adjectives or genitives. Examples from Lahu are given in (6.32) (taken from Lehmann 1984: 61–62):

(6.32) a. NP[våʔ-ö + qō thåʔ cō tā ve yā + mi + ma]
   "pig's-head ACC cook PERF NOML woman," i.e.
   the woman who has cooked the pig's head

b. NP[daʔ jā ve ò + li]
   "nice very NOML custom," i.e.
   a very nice custom

c. NP[ò + šī tā ni ve thåʔ nō mà ḡa mō laʔ]
   "blood out come NOML ACC you not PART see QU," i.e.
   Have you not seen blood coming out (of it)?

Following Lehmann's discussion, a plausible analysis for these structures of Lahu is (6.33):

(6.33) NP[ S Noml (N] 
       AdjP
       PossP
       . . .
]
The nominalizer does not appear to form a constituent with the material to its left, and when a head noun is absent (cf. (6.32c)) it is the unique constructor of NP. One parsing function of ve is, therefore, to construct NP, by MNC. Hence Lehmann's term "Nominalisator" (i.e. nominalizer). A second function is to indicate co-constituency with the IC to its left, i.e. to attach this IC under the NP constructed by ve and thereby indicate the sisterhood of the appropriate IC with ve (cf. the "attribution" function in Lehmann 1984).

For the relative clause construction of (6.32a) we can posit the structure of (6.34), and the construction principles of (6.35):

(6.34) \[ NP[S[...V_f] Noml N] \]

(6.35) **Lahu**
- N: constructs NP (MNC)
- Noml: constructs NP (MNC); attach left IC to NP (ICA)
- V_f: constructs S (MNC) [or S (GNC) and VP (MNC) if V_f is under VP]

There is no evidence in this language for assuming in (6.32a) a subordinate S, \( S_s \), distinct from S or \( S_r \), as there is in other languages, because there is no dominated constituent that will uniquely construct it. \( V_f \) constructs S alone, and ve constructs NP. Hence \( S_s \) is not constructable, and the subordination of S is indicated only by its attachment under NP (through ICA), and not on the basis of any uniquely constructing dominated constituent such as a complementizer in English.

What the structure of (6.34) makes clear is that the ordering of the three categories, \( V_f \), Noml, and N, makes for an optimally efficient IC-to-word ratio. The CRD for NP contains three ICs: the first is initiated by \( V_f \) on the right periphery of S, and the other two are single-word categories. That is, three ICs are recognized by three adjacent words.

The second type is exemplified by Basque. In this language the prenominal relative has the form of a clause containing a right-peripheral finite verbal form, to which is attached an -n suffix. This suffix occurs not only in relative clauses (cf. (6.36a)), but also in other subordinate clauses, such as (6.36b), and it therefore serves to construct an \( S_s \) over the S constructed by \( V_f \).

(6.36) a. \[ NP[ama-k erra d-u-en liburu-a] \]
   "mother (DEF)-ERG burnt ABS-PRES-(ERG)-SUBDN book-DEF," i.e. the book that mother has burned
Grammaticalized node construction

b. Esa-ida-zu \( \_s_2[z_e r\ ari\ z-e_r-a-n] \)
   "tell(IMP)-DAT-ERG what do ABS-AUX-SUBDN," i.e.
   tell me what you are (in the process of) doing

Relative clauses appear to have the gross structure of (6.37), therefore:

\[(6.37)\quad \text{NP}[S_2[S\ldots V_f + n]] N]\]

The \( V_f \) constructs \( S \), \(-n\) constructs \( S_s \), and \( N \) constructs NP, by the following parsing principles:

\[(6.38)\quad \text{Basque}\]

\begin{align*}
N: &\quad \text{constructs NP (MNC)} \\
-n: &\quad \text{constructs } S_s \text{ (GNC)} \\
V_f: &\quad \text{constructs } S \text{ (MNC)}
\end{align*}

I assume that if there is a VP in this language, it will be constructed by non-finite verbal material to the left of the finiteness marking. Finiteness in this analysis would then construct \( S \) by GNC, attaching it over VP, and \(-n\) would construct \( S_s \) over \( S \) by a principle of Great-grandmother Node Construction.

According to (6.37), the NP is a binary branching structure consisting of \( S_s \) and \( N \). \( S_s \) is a non-branching structure dominating \( S \), and both \( S_s \) and \( S \) are uniquely constructable on the basis of morphological affixes on the right peripheral verb. \( V_f + n \) therefore initiates the first IC of NP, \( S_s \), and \( N \) provides the second. As a result the CRD for NP is just as optimal as it is in Lahu: its two ICs are recognized on the basis of two adjacent words, making a 100% ratio.

The \(-n\) suffix of Basque finds a close parallel in the \(-a\) suffix of the Dravidian languages (cf. Lehmann 1984: 50–51), which is added to tense marked verb stems and which appears to construct \( S_s \). A structural analysis along the lines of (6.37) seems appropriate for these languages as well, in conjunction with parsing principles corresponding to (6.38).

The third prenominal relative clause type is exemplified by Turkish. Instead of a rich verbal form comprising finite morphology plus an affix constructing \( S_s \), this language disposes of non-finite verbal forms that appear to be uniquely indicative of a subordinate clause. An \(-en\) suffix is the unmarked form, according to Lehmann (1984: 52). An example is (6.39):

\[(6.39)\quad \text{NP}[mekteb-e gid-en]\quad \text{adam}
   \quad "school-DAT go-SUBDN man," \text{ i.e.}
   \quad a \text{ man who is going to school} \]
Since these non-finite verbal forms are unique to $S_s$ and are not found in non-subordinate clauses dominated by $S$ (and possibly $S_r$ as well), I shall assume that the gross structure for (6.39) is (6.40), without the embedded $S$ of Basque (6.37):

(6.40) $NP[S_s \ldots V_{nf}] N$

The hypothesis here is that $S$ is not constructable. In earlier chapters, however, we have seen evidence for a VP in Turkish (cf. ch. 4.1.4 and 4.1.7). This suggests that we have the following construction principles for Turkish:

(6.41) *Turkish*

$N$: constructs NP (MNC)

$V_{nf}$: constructs VP (MNC), $S_s$ (GNC)

Despite the absence of $S$, the Turkish NP still consists of two ICs, $S_s$ and $N$, the former constructed on its right periphery by $V_{nf}$. This verb initiates the CRD for NP, therefore, and the immediately following $N$ concludes it. Once again two ICs are recognized on the basis of two adjacent words.

Another language of the Turkish type is Korean, in which a special non-finite participial verb form containing an -n suffix constructs $S_s$ and stands in clause-final position immediately preceding the head noun. Contrast the main clause of (6.42a), which contains a finite verb, with the corresponding relative clause (6.42b) (examples from Jhang 1991: 269):

(6.42) a. $S_s[Totwuk-i pang-eyse nao-ass-ta]$

"thief-NOM room-from emerge-PAST-DECL," i.e.

A/The thief emerged from the room.

b. $NP[pang-eyse nao-n totwuk]$

"room-from emerge-SUBJN thief," i.e.

a/the thief who emerged from the room

The -n marked $V_{nf}$ constructs VP (by MNC), like its Turkish counterpart (cf. ch. 4.1.6 for supporting evidence for a VP in Korean), and $S_s$ by GNC. $S_s$ and $N$ are therefore recognized on the basis of two adjacent words, $nao-n$ and $totwuk$, which makes the CRD for NP optimally efficient.

Contrasting with this Korean structure is, interestingly enough, Japanese, which exemplifies our fourth major type. Japanese differs from all three prenominal relative clause types considered so far in not having an explicit subordination indicator within the relative clause. Instead, subordination is signaled by the adjacency of the clause to the immediately following head
noun, and by the consequent attachment of this clause under NP. A main clause containing a finite verb, and an NP containing a corresponding relative clause, are given in (6.43) (taken from Lehmann 1984: 70):

(6.43) a. _s[Ano hito-ga hon-o kai-ta]
   "that man-NOM book-ACC write-PAST," i.e.
   That man wrote a book.

b. _NP[ano hito-ga kai-ta hon]
   "that man-NOM write-PAST book," i.e.
   the book that that man wrote

The parallelism between main and subordinate clauses suggests that we assign to (6.43b) the gross structure of (6.44):

(6.44) _NP[sl[... V_J] N]

An _s will not be constructable, but _S will be, by _V_J as usual. _N will construct _NP, and _S will be attached under this _NP by _ICA:

(6.45) Japanese
   _N: constructs _NP (MNC); attach left _S to _NP (ICA)
   _V_J: constructs _VP (MNC), _S (GNC)

Once again, the two ICs of the _NP, _S and _N, are constructed and recognized by two adjacent words, _kai-ta and _hon in (6.43b), and the CRD for _NP is optimally efficient.

What we have seen in this section is that whether we are dealing with a structure like (6.34), (6.37), (6.40), or (6.44), the positioning of alternative node-constructing categories is always such as to provide the best possible IC-to-word ratio within the _NP CRD. There can be differences in the precise categorial status of the relative clause: it may be dominated by _s alone, by _S alone, or by both. There are corresponding differences in the constructing categories (_V_nf, _V_J, _V_f plus a subordinating affix), and in the higher nodes that they construct (_V_nf constructs _s, _Noml constructs _NP). There can be two or three ICs of the _NP. But the crucial similarity across all of these structures is that the phrasal node of the relative clause itself is constructed on its right periphery, and is followed by shorter (here exclusively single-word) ICs thereafter. The result is optimal EIC ratios. Some other possible prenominal relative clause structures are also predicted on this account, e.g. _NP[ss[sl[... V_J] Comp] N], with optimal ratios in its CRDs.
for both NP and S. But a large number of non-optimal structures are not predicted, and as far as I can tell such structures are non-occurring.  

6.3.3 Alternative constructors of VP

In ch. 6.2.2 I argued that the case particles of Japanese construct a feature-marked NP by MNC, and a VP or S by GNC (cf. (6.13)). The non-nominative case particles, -o and -ni, are therefore constructors of VP, in addition to V. Since they are encountered before V in the parse string, VP can be constructed in advance of the verb, and this shortens the CRD for S. We have also seen suggestive evidence in Turkish for early VP construction on the basis of non-nominative case marking in NPs (cf. chs. 4.1.7 and 5.2.2). Consider now the VP-constructing function of non-nominative cases in German. This language has a typologically anomalous verb-final VP for a language with predominantly left-mother-node-constructing categories, such as prepositions within PP. The anomaly disappears when we adopt a parsing perspective and consider the role of parsing principles other than MNC. This approach also makes a prediction for the sequence of VP-internal ordering changes in the history of Germanic.

Let us review the parsing principles for German discussed in section 6.3.1. I argued there that German shared with Danish the following principles of S and VP construction (cf. (6.29)):

\begin{equation}
\begin{align*}
\text{German} \\
\text{# (IC-1) } V_f: & \quad \text{constructs } S \text{ (MNC); } S_r \text{ (GNC); attach IC-1 to } S_r \\
& \quad \text{(ICA)} \text{ elsewhere: \quad \text{constructs } VP \text{ (MNC); } S \text{ (GNC)}} \\
V_{nf}: & \quad \text{constructs } VP \text{ (MNC)} \\
\text{dafi:} & \quad \text{constructs } S_\text{e} \text{ (MNC); right } S \text{ (SNC)}
\end{align*}
\end{equation}

We can now add an additional possibility for VP construction, triggered by accusative or dative case-marking. We do not want these cases constructing a VP when they occur in initial, topicalized position, i.e. IC-1 in (6.46), since I assume that this IC-1 is immediately dominated by S_r. Nor do we want a nominative-marked NP constructing S in this position, for the same reason. One way to capture this is to limit the construction of the relevant grand-mother nodes to NPs occurring in any position to the right of the sentence-initial finite verb, i.e. whether in a main or a subordinate clause:
In the case of VP-construction, this limitation will (correctly) construct VP only when VP is immediately dominated by S and when the CRD for this latter has already been initiated by some other constructing category, such as V_f or -NP.

Consider a sample tree structure (6.48) constructed in accordance with these principles, for the sentence *Gestern habe ich ihm das Buch zu seinem Geburtstag gegeben* “yesterday have I him the book for his birthday given,” i.e. yesterday I gave him the book for his birthday.
Our principles do not require a VP substructure in a VP, but they are compatible with this analysis as well as with a flat VP structure: each VP can be constructed, and the grammar of German must be consulted in the parse to see if it allows embedded VPs, or whether it is one and the same VP that is multiply constructed in this environment, according to the PS rules. In fact, there is evidence for embedded VPs in German from c-command asymmetries (cf. Primus 1989). There is additional motivation from EIC. If VP substructure were a flat structure, it would contain four ICs (NP, NP, PP, and V\textsubscript{nf}) and seven words, making an IC-to-word ratio of 4/7 = 57%. With the addition of a VP\textsubscript{2}, the aggregated IC-to-word ratio is much higher, namely 75%. VP\textsubscript{1} has 2/2 = 100%, and VP\textsubscript{2} 3/6 = 50%. I leave it open whether there is a VP\textsubscript{3} dominating PP and V\textsubscript{nf}. There is less obvious grammatical motivation for this node, even though it is constructable by V\textsubscript{nf}, in this particular tree at least.

The dative-marked NP within the PP, seinem Geburtstag, raises an interesting issue. According to (6.47), a dative-marked determiner should construct a dative-marked NP as mother, and a VP as grandmother. But it cannot construct a VP here because the preposition zu has already constructed a PP, and this phrase is the only admissible grandmother for Det in this environment, according to the PS rules. We therefore see that the application of GNC can be preempted in the event that a grandmother for the dative-marked category has already been constructed in the parse string, and (6.47) needs to be revised accordingly. More generally, it appears that dative (and accusative) case allow us to infer VP as a default option, in the absence of an alternative grandmother. If this is correct, it provides some rationale for the apparently schizophrenic case-assignment facts of German: prepositions assign case to the right within PPs; VP-dominated verbs (and also adjectives, cf. below) assign case to the left. In this way, VP can be reliably inferred when alternative grandmothers such as PP have not already been constructed, and the relevant case-marked NPs can then be used to construct VP on a left periphery of that node, in lieu of the verb. We can therefore revise (6.47) as follows:

\begin{equation}
(6.47') \text{German} = \begin{cases} 
\text{Pro} \\
\text{Det} \\
\text{N} \\
\text{[+Acc]/[+Dat]} \\
\text{[+Dat]/[+Acc]} \\
\end{cases}
\end{equation} 

\begin{align*}
\{ \text{construct NP (MNC); construct VP (GNC)} \}
\end{align*}

in the environment \# (IC-1) \text{VP}_f \ldots ,

when no alternative G has already been constructed in the left-to-right parse
This formulation appears to cover the leftward case-assignment rule within adjective phrases as well, e.g. AdjP[ihrem Vater ähnlich] “her-DAT father similar”, i.e. similar to her father. In sie wußte, daß sie ihrem Vater ähnlich war “she knew that she her-DAT father similar was,” i.e. she knew that she was similar to her father, the dative ihrem Vater will construct a VP under the S of the subordinate clause by GNC. When ähnlich is encountered, this adjective will construct an AdjP (by MNC) and will be immediately recognized as the case-assigner for ihrem Vater. This NP will be attached under AdjP, therefore, and AdjP itself will be attached under the VP that has already been constructed, by checking the PS rules and activating ICA. The finite verb war as a VP- and S-constructor (cf. (6.46)), will also be attached under this VP. At this point the dative-marked determiner will be dominated by a great-grandmother VP rather than a grandmother, i.e. vP[AdjP[NP[ihrem Vater] ähnlich] war]. It is still VP-dominated, however, and this VP was plausibly a grandmother at the time it was constructed. More generally, the AdjP is part of the VP in this example, and the ultimate grandmother node constructed by the case-assigning adjective has not yet been encountered when the dative is parsed, in contrast to prepositions. Hence, VP is constructed in the usual way, and case marking can still perform its VP-recognition function at the left periphery of VP. In the event that AdjP is not VP-dominated, as in die ihrem Vater ähnliche Tochter “the her-DAT father similar daughter”, i.e. the daughter similar to her father, I would hypothesize that the alternative G, here the NP constructed by the determiner die, will need to be constructed on-line prior to parsing the dative.12

Let us assume, then, that case marking can function to construct VP on-line in German, and that something like (6.47') is correct. With this assumption let us return to the VP-internal ordering regularities that we have established in previous chapters. I mentioned in ch. 5.2.2 that the leftmost position of the VP is filled by pronouns in German. Pronouns precede full NPs in the basic order, and in the absence of heavy stress on the pronoun, this is the only order that is grammatical; cf. (5.14)—(5.15). In ch. 4.3.3 we saw clear performance evidence for the positioning of NP before PP in the basic ordering conventions of German. We have, therefore, the following unmarked order in the VP: vP[Pro NP PP V].

The relative sequencing of Pro, (non-pronominal) NP, and PP is a consequence of their relative aggregate weights, and is exactly what we find to the right of V in English (cf. ch. 5.3): vP[V Pro NP PP]. But notice now that if case can construct VP, then both the German and the English systems can be equally optimal in providing early VP recognition within S. Pronouns are
Some predicted category and ordering alternations

Robust surface indicators of nominative versus accusative versus dative case; cf. Hawkins (1986: 16–18). If a pronoun is not present within the VP, then the leftmost daughter of the NP, typically a determiner or adjective, will also regularly be able to construct VP on the basis of accusative and dative case marking (cf. Hawkins 1986: 13–15 and section 6.3.4 below). Only the PP is not recognizably unique to VP on-line, and this phrase is often postposed to the right of V, the alternative category for constructing VP.

The early positioning of Pro and NP before PP in the German VP can therefore be motivated both by VP-internal weight considerations, comparable to those that hold for English, and also by S-internal considerations, specifically the capacity for (accusative and dative) case to construct VP. If we return to the tree in (6.48), we see that the CRD for S consists of three ICs, V, NP, and VP, which are recognized on the basis of three adjacent words, habe ich ihm, making a 100% ratio. This is just as optimal as the corresponding English sequence I have given, which constructs NP, AuxP, and VP within S. Meanwhile, within a VP the relative ordering of Pro, NP, and PP is optimal, while the potentially negative effects of having a single-word V in right-peripheral position are in large measure reduced by the following considerations: PPs can be (and regularly are) postposed to the right of V (cf. Hawkins 1986: 147–50); heavy S and VP phrases can be (and regularly are) extraposed (cf. ch. 4.3.5 and Hawkins 1986: 144–47); and the embedding of VPs within VPs (e.g. VP within VP in (6.48)) makes for more efficient CRDs and improves the aggregate IC-to-word ratio compared with an unstructured VP. To these can be added an additional, rather interesting fact of German grammar. The final verb is quite regularly not all that short. The frequent verb clusters, such as gesehen haben muß “seen have must,” i.e. must have seen, analyzed by Jacobs (1992) in terms of Incorporation into a single IC and parsed by Attachment Construction (cf. section 6.2.4), will add significantly to the aggregate weight of this IC. For all these reasons, CRDs for VP in the surface structures of German are probably as optimal in performance as their English counterparts, despite verb-final positioning; and so are those for S, assuming VP construction through case.

We can now give substance to our typological claim that German is fundamentally an English-type language, and is quite unlike Japanese, despite the superficial similarity of their verb positions. Both German and English are left-peripheral constructors of higher nodes. For English this has already been amply illustrated, cf. e.g. ch. 5.3. The IC construction patterns for German phrasal nodes are summarized in (6.49):
Prepositions construct PP on the left periphery of PP (cf. (6.49a)). NPs are immediately recognizable on the basis of Pro, or by a left-peripheral Det or N constructing NP by MNC, while Adj will construct NP by AgP (6.49b). Adjectives occur to the left of their complements within AdjP, in general, except for a preceding NP whose positioning has already been motivated for reasons of VP construction (6.49c). VPs are constructable by a left-peripheral dative or accusative pronoun; by a left-peripheral dative or accusative full NP whose case marking is in turn indicated on a left-peripheral determiner, adjective, or noun (cf. section 6.3.4); or VP may be constructed by a left-peripheral V in the event that there are no other arguments of V, or only postposed arguments such as PP (6.49d). Subordinate clauses have left-peripheral complementizers (6.49e). The S within a root clause is constructed by a left-peripheral Vf (which may or may not be preceded by a constituent in S, IC-1) (6.49f). And S may also be constructed (in subordinate clauses) by a left-peripheral nominative NP (6.49g).

Clearly, higher nodes in German are systematically constructed at or near their left peripheries, as they are in English, despite the fact that Vnf is final within VP, and Vf is clause-final in subordinate clauses. This anomaly disappears when it is realized that there are other devices for constructing VP and Ss on a left periphery, and that German makes systematic use of them. Accusative- or dative-marked arguments of the verb cannot follow V, unless they are extremely heavy (cf. Hawkins 1986: 149–150); and a complementizer must regularly be present in subordinate clauses, and cannot be removed with the ease of English That-deletion.

This is a very different pattern of higher node construction from Japanese, whose corresponding phrases and constructing categories are juxtaposed for
Some predicted category and ordering alternations

comparison in (6.50). Higher nodes in this language are constructed at or near their right peripheries, i.e. ICₘ, making verb finality fundamentally harmonic with the rest of the language. Even when the VP is constructed in advance of V by case particles, these particles are on the right peripheries of their NPs, and these NPs regularly stand adjacent to V following longer PPs (cf. ch. 4.1.5).

(6.50) a. PP[NP P]
   b. NP[ ... Particle]
      [oCase] [oCase]
   c. AdjP[ ... Adj]
   d. VP[ ... NP[ ... Particle] ... ], VP[ ... V]
      [+Dat/+Ace]
   e. Sₜ[S Comp]
   f. Sₜ[TopP[XP Top] S]
   g. S[ ... NP[ ... Particle] ... ]
      [+Nom]

Japanese is postpositional (6.50a); case particles construct feature-marked NPs on the right periphery of these NPs, cf. (4.9), (6.50b); adjectives stand to the right of their complements (6.50c); VP construction by NPs is still near a right periphery of the VP, or else VP is constructed by V itself in final position (6.50d); complementizers are final in Sₜ (6.50e); the topic marker is final within TopP (cf. (6.13)), and the leftward positioning of TopP relative to the much heaver S within Sₜ is motivated in ch. 5.6.3 (6.50f); and the nominative particle will construct S at the conclusion of this NP rather than at its onset and immediately adjacent to VP (6.50g).

The fundamental similarity between German and English now makes it easier to understand how Modern English evolved historically out of a system very like the German one, and it provides a possible explanation for variation in the Germanic VP generally. We have seen that there are two partially overlapping determinants of VP-internal orders: relative weight, and VP constructability. Since the ICs of the VP are left-mother-node constructing, i.e. ICₘ, grammaticalized orders will be sequenced short before long: Pro before full NP before PP. And there are two types of VP constructors: the verb; and accusative or dative case marking on pronouns and full NPs. German makes systematic use of case marking, cf. VP[Pro NP PP V]; English relies on the verb, cf. VP[V Pro NP PP].

Imagine now that a German-type system were to lose its surface case marking on some or all of its NPs and pronouns. This account would then
predict that the verb would have to take over as the constructor of VP, at the same time that the relative sequencing of arguments remained positioned according to their weights. The verb would accordingly move leftwards, and its arguments would move to the right. Since surface cases are never lost on all NP types immediately and simultaneously (for example, even in Modern English a distinctive nominative versus non-nominative marking is preserved on pronouns but not on full NPs), this account predicts further that the historical progression in postposing should reflect the capacity of the relevant argument to construct VP in advance of the verb. Thus, PPs can never construct VP, so they should be the first to postpose. NPs should then be progressively postposed as their uniqueness to VP is eroded. Since non-nominative case marking is lost on full NPs before pronouns, this predicts a historical stage in which \( \text{vp}[\text{NP } V] \) merges progressively into \( \text{vp}[V \text{ NP}] \), at the same time that pronouns remain preverbal, i.e. \( \text{vp}[\text{Pro } V] \).

Bean (1983) provides extensive word-order data from Old English confirming these predictions. PP was regularly postposed in Old English (as it is in Modern German). And in the ninth century there were no instances at all of \( \text{vp}[V \text{ Pro}] \), only \( \text{vp}[\text{Pro } V] \), whereas \( \text{vp}[V \text{ NP}] \) and \( \text{vp}[\text{NP } V] \) were both fully productive. SVO and SOV orders account for 49% and 31% respectively of all clauses containing a non-pronominal, full NP subject and object (figures from Saitz 1955, quoted in Bean 1983: 115). Bean also argues that it is the surface distinctiveness of subject and object marking that drives the preposing of the verb even with full NP objects, for when the VP is preceded by a pronominal subject (which is typically distinctively nominative), the distribution of \( \text{vp}[V \text{ NP}] \) to \( \text{vp}[\text{NP } V] \) is reversed compared with a full NP subject: only 33% have sVO and 60% have sOV. In our terms, the sequence \( s[\text{Pro } \text{vp}[\text{NP } V]] \) will render the NP recognizable as a direct object, even with weak or ambiguous case marking (or indeed with no case marking at all!), by virtue of the clear nominative case marking on the subject pronoun, and this will suffice to construct VP. Preposing of V is then less necessary than it is with a full NP subject, \( s[\text{NP } \text{vp}[\text{NP } V]] \), unless the language imposes a rigid subject-before-object ordering constraint, in which case the second NP in the sequence will be able to construct VP by virtue of its position.

Many facts of Old English begin to make sense, therefore, if we assume that this language was changing from a German system with VP construction through oblique case marking, in response to case syncretism. This explains the progression from (a) to (c) in (6.51):
Some predicted category and ordering alternations

(6.51)  

a. \( VP[Pro \ NP \ PP \ V] \)  
b. \( VP[Pro \ NP \ V \ PP] \)  
c. \( VP[Pro \ V \ NP \ PP] \)  
d. \( VP[V \ Pro \ NP \ PP] \)  

What about the last stage, (6.51d), the system of Modern English which was attained later, certainly by the twelfth century (according to Saitz’s 1955 figures)? Let us reflect on verb-initial VPs. What does our theory predict for these systems?

The main prediction it makes is that this option will always be available to left-peripheral node-constructing languages that have a VP, because V will always be able to construct it. Obviously if there is no case system, then VP construction through V is the only option, and if case is lost in a (6.51a) language, then verb fronting is required. Moreover, even if there is a rich case system, this may not be sufficient for VP recognition, and construction through V may still be the only option. Imagine, for example, that dative and accusative can be regularly S- as well as VP-dominated, as seems to be the case in Icelandic.\(^{15}\) Dative and accusative will then not serve as reliable constructors of VP. If these cases can construct VP, on the other hand, then both options are available.

A language with the sequence (6.51d) is therefore compatible with all logical possibilities for case marking, and this explains why case syncretism is not a necessary cause of verb fronting in general, but at best a sufficient one (as argued in Hawkins 1986: 49–51): verb-initial systems even in Germanic occur both with and without a case system, and can be acquired prior to any case syncretism. The present theory defines the precise causality of case syncretism: verb fronting will result if case-marked arguments precede V and function as VP constructors, and if case is then lost. To this we must then add a further condition. If subject-before-object ordering is rigidly fixed, then the second NP in an [NP NP] sequence may also serve to construct VP by a positionally defined application of GNC. This seems to be a regular strategy in Dutch and in Pennsylvania German (Mark Louden, personal communication). If [NP NP] is not unambiguously subject-object, however, then VP construction through verb fronting is again required.

Two questions remain. Pronouns in Modern English regularly distinguish nominative from objective case. Why, then, did the system of (6.51c) not remain, and why did the verb become initial? Secondly, if Old English (and German) had the capacity for constructing VP either by case or by V, what, if
any, reasons can be given for selecting, and in the case of German maintaining, the (6.51a) order?

As for the first question, (6.51c) is certainly a possible order for languages like Modern English, according to our theory, and indeed systems like (6.51c) are found in the clitic pronoun–verb sequences of Modern Romance and in Bantu (cf. ch. 5.2.1). The order of (6.51d) is equally possible, even when Pro can construct VP, for the reasons we have mentioned. The reason for the shift to (6.51d) may have something to do with the fact that the Old English pronouns did not become phonologically reduced and cliticized like their Romance counterparts, perhaps because of their physical separation from the verb in the verb-final patterns (6.51a) and (6.51b), which co-existed with (6.51c) (and later with (6.51d)) in Old English. These pronouns, being NPs, then became subject to NP-sensitive rules of ordering, and when full NPs became postverbal, there were grammatical advantages to positioning pronouns after the verb as well.

As for the second question, notice that a language of type (6.51a) marks both the onset and the offset of the VP, and this may have certain functional advantages for constituent recognition over marking just one periphery via a constructing category. Daughter categories of the relevant phrase are explicitly contained within a “brace” or “Klammer” (to use the German phrase for this phenomenon), and this may facilitate gathering up the ICs in processing, distinguishing them from the ICs of higher containing phrases, etc. Not all phrasal categories will contain multiple construction devices that make this option possible, and even if multiple constructing categories exist, there will generally be EIC reasons for not constructing the higher phrase at both ends, since this will often prevent dominated material from being excluded from the CRD for that phrase, which will lengthen CRDs and reduce their EIC ratios. Thus, the German VP brace is regularly broken by extrapositions, etc., in order to shorten the CRD for VP, and so was the Old English brace (cf. Stockwell 1977a). Nonetheless, the brace per se clearly has some advantages, and the reader is referred to Ronneberger-Sibold (1990) for a discussion of its functional motivation in Modern German.

At the very least, this theory of alternative VP constructors provides some new hypotheses for further testing on these data from Germanic, and it seems to make correct predictions for the data sample that we have considered briefly here.
6.3.4 Alternative constructors of NP

We have seen several daughters of NP that will uniquely construct this phrase by MNC: the head noun itself; pronouns; and names. I have also argued that determiners will construct NP by MNC. If determiners also construct a determiner phrase within this higher NP, then NP will be constructed by GNC instead. In ch. 6.2.3 I discussed yet another possibility: NP construction from an adjective via AgP. Latin was cited as an example, cf. (6.52):

(6.52)  bonarum feminarum
        "good-GEN-PL-FEM woman-GEN-PL-FEM," i.e.
        of good women

Adjectives agree with nouns in gender, number, and case in Latin, and this serves to activate the mother of the noun, NP, as a dominating node over the adjective, cf. (6.18). If bonarum also constructs an AdjP by MNC, as seems reasonable, then NP will be its grandmother. NP can therefore be constructed prior to encountering N in the parse string, and this was cited as one of several functional advantages for having a morphological agreement system.

Demonstrative and possessive determiners in Latin also exhibit noun agreement in gender, number, and case, as illustrated in (6.14) above. As a result, whether a noun phrase contains all three of these morphologically marked items, $\text{NP}[\text{Det Adj} \ N \ldots ]$, or just any one or two, $\text{NP}[\text{Adj} \ldots ]$, $\text{NP}[N \ldots ]$, $\text{NP}[\text{Adj} N \ldots ]$, etc., the first NP-dominated category encountered always provides all the syntactic information that can be projected from the morphology. Specifically, NP will be immediately constructable by either MNC, AgP, or GNC. And if case is a higher node constructor in Latin, constructing S and VP as it does in Germanic, Japanese, and Turkish, then these higher nodes will be immediately constructable as well.

The history of morphological syncretism in the Indo-European languages merits special investigation from the parsing perspective adopted here. The modern Romance and Germanic languages, for example, generally preserve a reduced system of gender and number agreement, sufficient for the construction of NP by AgP. But case inflections have been systematically lost, and this has consequences for higher node construction of the kind discussed in the last section. The precise sequencing of changes in inflectional leveling within the NP seems to be interesting.

Consider Germanic. Germanic sequences of $\text{NP}[\text{Det Adj} \ N \ldots ]$ are not as richly differentiated morphologically as their Latin counterparts. There are three genders and two numbers for AgP in the earliest dialects, as in Latin,
but fewer case distinctions overall and less differentiated forms for the cases. Without going into the details, notice only the following gross developmental traits. First, case syncretism applies earliest and most extensively to nouns. Second, one subclass of adjectives, the so-called weak class, undergoes early and extensive case leveling, before the strong class. And third, determiners are most resistant to leveling and preserve case distinctions long after they have been lost on nouns and on (weak) adjectives.

This relative sequence is clearly visible in Modern German and is summarized in Hawkins (1986: 13–16). Nouns are invariable for case, except for a dative plural -(e)n suffix, and an optional -e in the dative masculine and neuter singular. Weak adjective inflections are limited to (the same) two forms for all genders, numbers, and cases, -e and -(e)n, while strong adjectives encode a rich case distinctiveness and have at their disposal five forms, -er, -en, -es, -em, and -e. Determiners such as the definite article are also richly differentiated with six distinct forms; der, den, des, das, dem, die.

What could explain this relative progression in case syncretism? Notice first that it correlates with the left-to-right sequencing of these inflectionally marked categories, \( \text{NP}[\text{Det Adj N ... }] \): determiners are the most differentiated categories, adjectives next, and nouns least. This NP-internal order was the basic one in all the early West Germanic dialects (cf. Hawkins 1983: 224, which summarizes Smith 1971), and it is also the basic order of Modern German. Clear surface case marking is evidently skewed to the left. More precisely, notice that the strong adjective inflections must be used when there is no determiner, i.e. \( \text{NP}[\text{Adj N ... }] \). When a determiner precedes, adjective inflections are either exclusively weak (e.g. after the definite article) or primarily weak with occasional strong forms (e.g. after the indefinite article); cf. Hawkins (1986: 15). Case distinctiveness is therefore carried by the determiner in \( \text{NP}[\text{Det Adj N ... }] \) sequences, and of course in \( \text{NP}[\text{Det N ... }] \) as well, and by Adj in \( \text{NP}[\text{Adj N ... }] \). It is always the first item in the NP that reveals the case of the NP, if case is clearly marked at all. But this first item is also the category that constructs NP in our theory, by MNC, AgP, or GNC. So we have the following regularity: the category that first constructs NP in the left-to-right parse is the category that bears distinctive case inflections, if there are distinctive case inflections. Pronouns also conform to this regularity, being single-word ICs that are richly marked for case. They provide immediate NP construction and case recognition.

This regularity makes sense from several points of view. I have argued that NPs in German are feature-marked, \( \text{NP}_1 \) etc. Whatever category constructs NP is actually constructing a case-marked NP, therefore. All things being
equal it is more efficient if the case feature is assigned at the same time as the gross syntactic category, NP, is recognized. Otherwise there will be a delay in the assembly of information relevant to one and the same phrase. In addition, case marking is also a higher node constructor. A case-marked determiner or strong adjective will therefore construct a higher VP or S at the same time as it constructs the lower dominated categories, NP, AdjP, etc. This will have the advantage of providing early recognition of VP.\textsuperscript{16} It will also avoid leaving NPs unassigned to a higher phrase on-line, or misassigning them to S rather than VP.

For these reasons clear case marking is motivated on the left periphery of NP, and this explains the syntactic distribution of strong versus weak adjectives, and the morphological expressiveness of determiners, both in Modern German and in the history of Germanic. It may also explain another long-standing puzzle. The least expressive NPs are those without determiners or adjectives, i.e. $NP[N \ldots]$. What we notice in Modern German is an extension of the definite article to many NPs of this type that lack it in English, not for reasons of definiteness, but for case-marking purposes. E.g. \textit{Hast du den Johann gesehen?} \textit{"have you the-ACC John seen"}, i.e. Have you seen John? \textit{Er zieht den Rosen die Nelken vor} \textit{\textquotedblleft he prefers the-DAT roses the-ACC carnations,	extquotedblright} i.e. he prefers carnations to roses. A definite article is semantically redundant with a proper name, and is unnecessary (and generally optional) with generic plurals. It is traditionally believed in German grammar that these definite articles are present for reasons of case marking, and so they are; but one of the functions of case marking is immediate higher node construction, as we have seen. The hypothesis that these modern examples suggest for the history of Germanic is the following. Perhaps the very evolution of the definite article (out of a more strongly stressed deictic demonstrative) is to be explained in terms of its (originally) rich case-marking potential. In other words, an expansion of the pragmatically more restricted deictic meaning of the demonstrative determiner into the more general definiteness concept of the definite article (cf. Hawkins 1978, 1991) would have had the effect of converting many NPs of the form $NP[N \ldots]$ into a corresponding case-marked $NP[Det N \ldots]$. This could also explain the expansion of the numeral \textit{one} into the indefinite article. The more frequent use of determiners would then have compensated for the loss of case distinctions on nouns themselves.

It remains to be investigated whether this is a viable explanation for the expansion of determiners in the history of Germanic. At the very least our parsing approach has uncovered a descriptive regularity. Cases are clearly marked on the first category in the parse string that constructs NP (and other
Grammaticalized node construction

higher nodes), and historical changes in the class of determiners had the effect of removing a large number of the very structures, \( \text{NP}[N \ldots] \), that were increasingly not case-distinctive, replacing them with NPs that did conform to a parsing regularity that was otherwise threatened.

6.3.5 Alternative scope indicators

Finally in this chapter I want to tie together some observations that have been made in this book about the surface marking and processing of scope relations, and point out some more relevant data of a cross-linguistic nature.

In ch. 2.4 I introduced the Principle of Structural Determination, repeated here for convenience:

\[
\text{(2.36) Principle of Structural Determination}
\]

A node \( X \) can depend syntactically and/or semantically only on those nodes that structurally integrate it (as defined in (2.11)), i.e. on nodes within a Structural Domain of \( X \) in a higher constituent \( C \).

Scope is a form of semantic dependency. If a node \( Y \) has scope over a node \( X \), then the semantic interpretation of \( X \) is dependent on that of \( Y \) and is determined by it. Hence, according to (2.36), \( Y \) must be in the SD of \( X \) within a higher constituent \( C \). A WH-word or phrase can move into the Comp position of languages like English, where it asymmetrically c-commands all the nodes within the sister \( S \) and has scope over them. These nodes have the WH-node within their respective SDs, therefore. Alternatively I pointed out in section 6.2.4 that WH can move into the preverbal and so-called focus position of SOV languages like Turkish, and still have scope over the whole sentence, despite the deeply embedded nature of this focus position and despite the existence of a VP in this language (cf. ch. 4.1.7). The mechanism that was proposed as the explanation for this structural and semantic possibility was the parsing principle of Attachment Promotion. \( Y \), the WH-node, is immediately attached to \( X \), the verb, and is interpreted as a sister of the highest constituent constructed by \( V \), here \( S \). Other nodes dominated by \( S \) can accordingly have WH within their SDs so extended, even though they may asymmetrically c-command WH in terms of surface structure alone. In both English and Turkish, therefore, WH can have scope over all the constituents of \( S \), and these constituents will have WH within their respective SDs, preserving the Principle of Structural Determination.

In addition to surface syntactic configuration and Attachment Promotion, scope dependencies can also be signaled by surface case marking and linear
precedence. Surface case marking contributes to the definition of the SD of the relevant case-marked NP, as discussed throughout ch. 2. Hence, if two quantified NPs are case-marked, their relative scope interpretation will be determined by which one has the other in its SD, and this will generally be a function of both configuration and morphology, as Primus (1987, 1991a) has shown (cf. especially ch. 2.3 and her case hierarchy (2.25)). Linear precedence is also a highly productive indicator of relative scope, in languages such as Hungarian (cf. Kiss 1987). In our terms, if two quantifier phrases, or negation plus a phrase in the scope of negation, exhibit asymmetrical scope whereby Y, the constituent encountered first in the parse string, has scope over X, the constituent encountered second, then the SD for X will contain Y but not vice versa. Hence, the semantic interpretation of X can depend on Y, but not vice versa. There is variation across languages in the extent to which left-to-right priority has been grammaticalized as a requirement on the scope interpretations of sister nodes. In Hungarian this is a strong requirement; in English it is much less grammaticalized and there are frequent scope ambiguities among sisters. Each sister can have the other in its SD on these occasions, and the result is ambiguity. The ambiguity is removed if linear precedence is grammaticalized, and hence if the sister to the left cannot have the sister to the right in its SD.

It is interesting to note that this grammaticalization of left–right scope marking in Hungarian does not interfere with EIC’s predictions for the postverbal domain in this language, which were tested in ch. 4.1.2 and 4.1.8. Where ordering is grammaticalized for semantic reasons, there is of course no freedom, and hence there is no choice in language performance between equivalent structures differing only in processing efficiency: the ordering has to follow the relative scope relations. Nonetheless, EIC’s predictions for Hungarian were extremely well supported.

A possible explanation for this is that the number of sentences in which there are two phrases whose relative orderings would result in truth-conditionally distinct scope interpretations is too small to make a difference in any sample of language performance. If the great majority of orderings are semantically equivalent, then EIC will apply to this majority in the usual way, positioning them in accordance with their relative efficiency, and the visible effects of scope will be negligible and insufficient to offset EIC’s predictions. Moreover, some of the scope-conditioned orderings will also presumably be optimal for EIC as well, which further reduces the quantity of semantically determined orderings at variance with EIC. This interplay between scope and EIC needs to be further investigated by focusing on sentence types containing two or more scope-bearing elements. My prediction
is that weight will not be the determining factor in these cases, just as it does not
determine the positioning of phrases under Topic and Focus nodes in language
performance (cf. ch. 4.1.2 and n. 3 ch. 4). My further prediction, however, is that
EIC's predictions for performance will remain intact, despite scope grammati-
calization, and that optimal orderings will occur in the unmarked case through-
out a performance sample, with non-optimal orderings declining in accordance
with their ratios, just as they do in Hungarian.

The grammaticalization of scope-sensitive orderings also requires further
investigation from the parsing perspective adopted here, but again I believe
that this approach can make sense of some apparent mysteries. For example,
it is no accident that negative particles such as not in English almost invari-
ably precede the VP in VO languages (cf. Dahl 1979, Dryer 1992; this applies
outside of Africa, cf. Dryer 1992: 98). A negative particle or phrase is sig-
ificantly shorter than a VP and is predicted to precede a longer phrase
constructed on its left periphery, \( V_p[V \ldots ] \). OV languages exhibit both
postverbal and preverbal negative positioning, resulting in a left–right skew-
ing or asymmetry across all languages in favor of preverbal positioning for
this short particle. The details will vary in the different OV-language types. In
some, negation will be attached as a left sister to VP. In others, it may be a
right sister, in conjunction with a bottom-up parse. Probably the most fre-
quent option will be attachment of the particle to V, in conjunction with
Attachment Promotion. By positioning the negative element in the so-called
Focus position of SOV languages, the scope of negation will extend to all the
nodes dominated by the highest node constructed by V. The positioning of
tense and aspect particles across languages exhibits a very similar distribution
to that of negative particles (cf. Dryer 1992: 99), suggesting that the same
principles are at work here too.

These grammaticalized ordering variants for scope-bearing elements there-
fore appear to be driven by EIC and by additional parsing principles such as
Attachment Promotion, despite their direct coding of semantic relations. In a
similar way, it was argued in ch. 5.6.3 that the positioning of Topic and Focus
nodes, which are associated with pragmatic and/or semantic functions, was
also driven by processing efficiency, specifically EIC. The grammaticalization
of left-to-right topological priorities is also very naturally explainable in
terms of processing. The scope-bearing item that is received and processed
first will be available to influence the interpretation of items that are received
and processed subsequently, whereas the converse fails: the hearer cannot
guess what is to come later and interpret an item already received in terms
of it.
7 Conclusions

In the preceding chapters I have presented evidence and arguments in favor of the following generalization: many fundamental and abstract structural properties of grammars can be explained by simple considerations of processing ease. In other words, grammars are in large measure performance-driven, and we might use the term "Performance-Driven Grammar" to refer to this research program. This conclusion has been motivated by a close analysis of linear ordering in performance and in grammars. It is further motivated by the correspondences between on-line procedures for recognizing constituent structure and the grammatical devices that make such recognition possible. It is also supported by grammatical constraints on relativization, on movement, and on numerous other phenomena across languages.

This conclusion places performance at the core of an explanatory model of syntax, and departs from a long research tradition in which performance has been viewed as playing either no role, or only a peripheral role, in the core grammar. The role of an innate UG is correspondingly reduced, while that of (ultimately innate) processing mechanisms is increased. The processing mechanisms in question are those that make it possible for humans to recognize grammatical structure in a rapid and efficient manner. They can be seen as the syntactic counterparts of the highly efficient (and again innate) mechanisms for sound and word recognition that have been experimentally validated in numerous studies.

The methodology that I have employed in reaching this conclusion has consisted in integrating and extending ideas from three research traditions: generative syntax; typological studies of language universals; and psycholinguistics, specifically language processing. The relation between performance and competence has been discussed hitherto on the basis of an extremely limited range of languages, primarily English. By systematically juxtaposing psycholinguistic and generative-syntactic insights with comparative data from a wide range of languages, a lot of the issues appear in a new light.
Conclusions

Difficult structures that are unusable in English show up as difficult but nonetheless usable in other languages, suggesting that there are differences in grammatical conventions across languages with respect to difficult structures. Structures that are supposedly impossible to use according to current processing architectures do actually appear in some languages, suggesting that these architectures are incorrectly designed.

Above all, by taking a large-scale comparative perspective, the set of grammatical generalizations about human language can be extended from the absolute universals and parameters of the generative paradigm to the implicational and distributional universals of the typological approach. Since the notion of processing ease is not discrete but is largely gradient, with experimental studies and frequencies of occurrence indicating relative degrees of preference within sets of related sentence types, the search for any systematic correspondence between processing and grammar can now be extended to include grammatical data that are also non-absolute and gradient in nature. By matching these performance data with the implicational structuring of grammatical conventions, and with statistical preferences across grammars, a profound correspondence between performance and grammar becomes clearly visible. This correspondence is impossible to discern when the grammatical focus is on the conventions of just one, two, or even a handful of languages, and on the absolute universals and parameters that are derivable from them.

The explanation I have proposed for the implicational and distributional universals of this study is ultimately psycholinguistic. Implicationally related properties such as “if SOV, then postpositions in PP” co-occur in a grammar because they are optimal for processing and involve minimal complexity: fewer computations are required over smaller numbers of entities in order to reach a given parsing decision, involving in this case the recognition of constituent structure. Where there are implicational hierarchies, the higher positions on the hierarchy involve less complexity than the lower positions. Similarly, frequently occurring structures across languages involve less complexity than infrequently occurring structures, all things being equal. This processing basis for these universals explains why many of them have been systematically ignored within generative grammar: their explanation is not ultimately grammatical. However, the very same processing considerations that explain these universals can also explain many principles and parameters that have been proposed in a generative context. A source of explanation drawn from language typology and from psycholinguistics therefore leads to a rethinking of the explanatory axioms of generative syntax. But
equally, the explanation for typological universals proposed here would have been impossible without the structural insights of generative grammar, and the data of language performance provide strong support, as it turns out, for many abstract syntactic analyses of the kind generative grammarians have proposed. It is this integration of mutually reinforcing insights from different subdisciplines that has led me to the findings and conclusions of this book, which I shall now summarize. I first summarize the basic structure of the argument (section 7.1), and I then formulate the major hypotheses that have been proposed and consider some further issues that they raise (section 7.2).

7.1 Summary of the book

I began in ch. 1 with a summary and critical discussion of some major processing explanations for grammar that have been proposed hitherto. These explanations reject the "assumption of pure acceptability" of Chomsky (1965), whereby the competence grammar is viewed as being completely autonomous of the mechanisms of performance (except perhaps at the level of the evolution of the innate language faculty), and they explicitly endorse a causative role for performance with respect to certain specified rules or principles of the grammar. The trouble is that even though these proposals are generally well argued, they provide us with an anecdotal picture of the relationship between performance and competence. At least Chomsky's (1965) picture was clean and systematic: grammar is autonomous of performance. But now we are presented with evidence that performance does explain some rules and principles of the competence grammar, and this immediately raises two general questions: which grammatical areas and generalizations are so explainable? and by which performance principles?

The research literature summarized in ch. 1 provides little insight into these questions. There is no general discussion of the grammatical areas that can be predicted to be explainable by performance mechanisms. And the responsiveness of grammars to independently motivated sources of processing difficulty appears to be unsystematic: not all grammars block difficult structures; and not all sources of difficulty seem to result in a regular grammatical response. Attempts to circumvent these difficulties have resulted in processing theories that incorporate the "assumption of absolute unprocessability." But any such assumption is premature at this stage, and there are counterexamples to the individual proposals made.

The first chapter sets out a different approach to answering these questions. It is argued that the grammatical generalizations that are most revealing from
the point of view of the relationship between performance and competence are the implicational and distributional universals derived from extensive cross-language comparison. The present book identifies particular areas of grammar for which a performance explanation can be motivated: e.g. linear ordering, and hierarchies of relativization, movement, and related phenomena. And it proposes a new metric of complexity on the basis of which general predictions for the performance-grammar interface can be made. This approach still does not provide us with an overall theory of the relationship between performance and the other contributors to the competence grammar, including the innate UG. But it does, I claim, considerably generalize the discussion; and it identifies several new hypotheses for future investigation (cf. section 7.2). Above all, it makes some explicit proposals for, and tests some general predictions of, the processing module within an interacting, modular theory of language universals. Other modules will include innate syntactic universals, and innate semantic and cognitive universals. In addition, there will be functional principles derived from communicative needs and real-world constraints (cf. Hawkins 1988c and 1992a for further discussion). An ultimate theory of language universals can only come from specific proposals for the contents of these modules, together with explicit testing of the consequences of explaining a given body of facts in terms of one set of primitives rather than another. These issues are discussed in ch. 1.4 and 1.5.

Ch. 2 began by enumerating some of the major ways in which grammars can conventionalize processing mechanisms and preferences in their rules or principles. Processing is a major source of "grammaticalization" for linear ordering rules, for mother-daughter pairings in phrase-structure conventions, category selections in rules, exceptions and constraints on rules, etc. The remainder of the chapter was devoted to defining the notion of structural complexity, since it is the complexity of different structures that appears to be responsible for many of these grammaticalization responses. The crucial concept here is that of a "Structural Domain" (SD), defined as a grammatically or psycholinguistically significant subset of structurally related nodes in a tree dominated by a constituent C (cf. (2.8)). The SD of a given node X consists of those dominating and sister nodes that "structurally integrate" X in C (cf. (2.11)). In different language types the precise composition of a syntactically defined SD for X (cf. (2.9)) may be modified, depending on whether X is morphologically coded by case marking, or depending on whether the language grammaticalizes left-to-right topological priorities for certain grammatical operations applying to sisters (i.e. precedence relations).
It was then proposed that structural complexity is an additive function defined on SD sizes (cf. (2.12)). Some preliminary support for this was given from performance data on English relative clauses, which show that as SD sizes for "relativization domains" increase, relative clauses become less frequent in production and more difficult to comprehend (cf. ch. 2.2.1). This complexity metric was compared with that of Miller and Chomsky (1963) and that of Frazier (1985). Frazier had already criticized Miller and Chomsky's metric in terms of the global ratio of non-terminal to terminal nodes throughout a sentence. The insufficiencies of this metric are to be expected, as far as we are concerned, because the ratio between non-terminal and terminal elements is largely a constant function: the more terminal elements in a sentence, the more non-terminal structure (in general). Frazier's alternative limits the viewing window for assessing this ratio to three adjacent words in the parse string. But theoretical and empirical problems for this alternative were summarized in ch. 2.2.2. A three-word viewing window removes much relevant structure from consideration, and fails to account for the performance data involving processing ease in English relative clauses. The theory of SDs quantifies the amount of structure to be computed within domains that are less arbitrary than three successive words, and more grammatically and/or psycholinguistically significant. For example, it quantifies the number of nodes that need to be accessed when relativizing on a subject, or on an object, or on a genitive within an NP, etc. It quantifies the number of nodes (and terminal elements) that need to be accessed in order to recognize all daughter ICs of the VP, and so on.

This theory was then applied to the Accessibility Hierarchy (AH) of Keenan and Comrie (1972, 1977). Using the concept of a Minimal SD (cf. 2.23)), it was shown that the ranked grammatical positions down the AH are associated with increasing structural complexity, even when allowance is made for different surface codings of these positions across languages. The syntactic environments in which relativization "cuts off" in different languages can therefore be explained in terms of complexity: if relativization is possible in a more complex environment, it will be possible in a less complex environment, but the converse fails. The retention of a pronoun in the position relativized on exhibits the opposite implicational pattern, because pronoun retention makes the processing of relative clause structures easier: if a pronoun is retained in a less complex environment, it will be retained in all more complex positions that a language can relativize on. By recasting Keenan and Comrie's cross-linguistic generalizations in these terms, we achieve not only an explanation for the AH, but also an improved description.
of the generalizations themselves (cf. ch. 2.3). It was also suggested that there is a hierarchy for movement constraints and an implicational structuring to the selection of bounding nodes across grammars.

Finally, in ch. 2.4, I argued that as SDs increase in size and become more complex, so the corresponding processes of semantic interpretation become more complex and "context-dependent". By the Principle of Structural Determination (cf. (2.36)), syntactic and semantic dependencies can hold only between X and a node within its SD. This explains subject–object asymmetries, since the more complex node, the object, has the subject in its SD, but not vice versa. It explains the increasingly restricted referential interpretations assigned to generic plurals in increasingly complex syntactic environments, cf. *beavers build dams with branches*. Such asymmetries in dependency and in the compositional process of assembling the meaning of a sentence from its parts were claimed to be necessary in order to avoid infinite regresses into mutual delimitation in interpretation (the dams that are built by the beavers that build the dams that are built by the beavers that ...). This results in hierarchies of argument positions, in which the lower positions have SDs that contain the higher positions and properly include their SDs, and in asymmetrical priorities for syntactic dependencies and semantic interpretation. This in turn requires clear surface coding for these positions by syntactic, morphological, or topological devices (cf. the theory of Primus 1987, 1991a). Increasing structural complexity therefore involves not just accessing more nodes in processing, but also computing the possible impact of each node within the SD for X on every X within the sentence. It is this increasing complexity that makes sense of hierarchies such as those of Keenan and Comrie (1972, 1977) and Primus (1987, 1991a) and that explains why grammatical operations should cease to apply on lower positions, why more semantically overt and disambiguating material (such as pronouns in the positions relativized on) should favor lower positions, and so on.

Ch. 3 argued that different orderings of immediate constituents (ICs) and words are also associated with different levels of structural complexity in language processing. Different orderings result in constituent recognition domains (CRDs) of varying sizes (cf. (3.2)), i.e. in different numbers of words and nodes that need to be processed in order to recognize the ICs of a given phrase. A CRD is an SD comprising, on this occasion, the terminal and non-terminal nodes of a phrase C that are relevant for the recognition of C's structure. CRDs typically consist of a proper subset of the nodes dominated by C, and some orderings reduce the size of this proper subset
compared with others and result in faster and more efficient IC recognition. This is because all daughter ICs that are phrasal constituents are typically recognized on a left or right periphery by "mother-node-constructing categories" (MNCCs), just as the containing category C is typically recognized by an MNCC in leftmost or rightmost position. By placing ICs in different positions, CRD sizes may be increased or decreased. Some preliminary evidence taken from the published psycholinguistic literature indicates a clear performance preference for orderings that reduce CRD sizes (cf. ch. 3.3). Alternative methods of quantifying this preference were considered in ch. 3.4, and arguments were given for defining a principle of Early Immediate Constituents (EIC) as follows: the human parser prefers linear orders that maximize the IC-to-non-IC ratios of its constituent recognition domains (cf. (3.14) and (3.14')).

Some predictions made by EIC were then formulated for linear ordering selections in performance and in grammar. For performance it is predicted that IC orderings will be preferred in the unmarked case whose ratios are most optimal, for all different sets of words and daughter nodes assigned to the ICs of a given phrase. In the marked case, orderings with non-optimal ratios would be tolerated just in case these ratios are still reasonably good, and more precisely they will be tolerated in proportion to the magnitude of their ratios (cf. (3.17)). These preferences will be reflected in psycholinguistic experiments, in acceptability judgments, and in text frequencies. For these latter EIC predicts that the great majority of orderings in texts will be those that have the best possible EIC ratios, with the minority of non-optimal orders being frequent in proportion to the magnitude of their ratios (cf. (3.18)). A further performance prediction was formulated for temporary ambiguity resolutions (cf. (3.20)): when the parser has a choice between two alternative structural attachments for an IC, with different consequences for the overall EIC score for a sentence, it will prefer the attachment site that gives the best score.

For grammars EIC makes a corresponding unmarked and marked case prediction for basic orders (cf. (3.21)). It is assumed that if a grammar conventionalizes a rule or principle of order, it will do so in response to EIC and on the basis of aggregate weights for the ICs of a phrase. Basic orders will be those with optimal EIC scores, whose grammaticalization emerges very naturally from a preceding historical stage in which the most frequent orders were exactly those that are subsequently grammaticalized, given the weight aggregates. It cannot be the case, however, that all assignments of non-ICs to ICs will always be optimal in a basic order, and this
explains the existence of rearrangements such as Extraposition and Heavy NP Shift. Some predictions for grammatical rearrangement rules were formulated in (3.23). Finally, a prediction was formulated for the existence of word order (and center embedding) hierarchies among the non-optimal CRDs for a given phrase (cf. (3.24)).

After a summary of EIC’s predictions for typologically different structures in ch. 3.7 (cf. table 3.2), I discussed the relationship between EIC and various principles of generative grammar that have been proposed for linear ordering (cf. ch. 3.8.1). I argued that EIC provides an explanation for these principles — where they are correct. They lack descriptive adequacy, however: numerous ordering constraints in different language types are not predicted; significant differences in productivity and frequency are unaccounted for among the orders that are permitted; and these generative principles are arbitrary and do not follow from anything else in the grammar, precisely because their ultimate explanation is not grammatical. In ch. 3.8.2 I criticized proposed explanations for word order in terms of pragmatics. The pragmatic principles that have been appealed to are contradictory (given before new and new before given!), and I argued that there is no dichotomy between pragmatic-word-order languages and grammatical-word-order languages. There are different degrees of free versus fixed ordering across languages and across structures (cf. section 7.2), but all order is regulated by a single principle, EIC, whose ultimate motivation is to make the recognition of grammatical structure more efficient. This was argued to be a more basic function for order than signaling the discourse status of entities. On the other hand, there are certain grammaticalized nodes, such as Topic nodes, that seem to be motivated ultimately by pragmatic information status and that determine the movement of constituents into these nodes in performance. The actual ordering of these “discourse-configurational” nodes, however, is regulated by EIC. Finally, in ch. 3.8.3, I summarized and criticized some earlier “weight-based” principles of syntax.

Ch. 4 tested the performance predictions of EIC. Multiple branching structures were first selected from a variety of languages that exemplify the different structural types of table 3.2 and for which there is evidence of grammatically free ordering. The text-frequency prediction of (3.18) was tested on written samples from these languages, and both the unmarked and marked case predictions were found to be consistently well supported (cf. ch. 4.1). Data from binary branching structures were considered in the next section (4.2). EIC’s performance predictions apply to all performance data, whether order is grammatically free or fixed, so if a basic order is conventionalized, it was predicted that it must be maintained or rearranged.
in such a way that optimal EIC orders are produced in performance in the unmarked case, etc. Some data involving basic orders and their rearrangements were examined in ch. 4.3. Again, the performance predictions are well supported, even in those cases where the basic order has to be rearranged in many or most sentences in order to have optimal ratios (cf. e.g. English Particle Movement, ch. 4.3.1, and German Extraposition from NP, ch. 4.3.5). Overall, data from ten typologically and genetically diverse languages support EIC’s predictions for performance.

Ch. 4.4 compared the predictions of EIC with those of pragmatic theories, using the quantified text methodology of Givon (1983, 1988) applied to some of the same textual material on which EIC was tested. The pragmatic theories tested were: Givon’s Task Urgency; New before Given ordering; and the Prague School’s Given before New. None of these theories is well supported: IC orderings exemplifying all these pragmatic principles are productive in all the languages investigated. But since they contradict each other, there is no significant pragmatic generalization for linear ordering. There are, however, interesting correlations with EIC. In languages and structures for which EIC predicts a short-before-long preference, there is a correlation with given-before-new orders; long-before-short preferences correlate with new before given. This correlation makes sense when it is considered that constituents that are pragmatically given generally require less linguistic material for referent identification than those that are new, a generalization that was independently tested on NPs and their discourse status, regardless of their ordering, and found to be confirmed. The overall success of the pragmatic predictions is less than that of EIC, however, and the fact that EIC can actually explain the existence of two distinct patterns and preferences across languages (short before long versus long before short) indicates that syntactic processing is the primary generalization in this area. Pragmatic information status then correlates with whichever pattern is predicted by EIC. This explains why two (mutually contradictory) pragmatic principles appear to be motivated, neither of whose predictions is sufficiently strong to preclude significant instances of the other, with the result that both principles are productively supported in all languages. I concluded that pragmatic ordering principles are epiphenomenal and derivative of EIC. The only permitted exception to this is if a language has grammaticalized a discourse-configurational node, such as a Topic node, in which case performance data are driven by the relevant pragmatic generalization, while the Topic node itself is positioned by the grammar in accordance with EIC.
In ch. 4.5 EIC’s predictions for temporary ambiguity preferences were tested on some sentence types that had been investigated by Frazier (1979a, 1985) and that support her Late Closure principle. They also support the role of EIC in on-line ambiguity resolution. Some modifications were proposed to her explanation for certain garden paths involving Late Closure and Minimal Attachment, and it was argued that all her parsing preferences reduce to a single generalization: when it has a choice, the parser prefers syntactic analyses with less complex SDs (cf. also ch. 2.2.2 and 3.5.3).

Ch. 5 tested EIC’s predictions for grammars. Section 5.1 examined the most up-to-date cross-linguistic data on the cross-categorial word-order combinations first formulated by Greenberg (1966): “if prepositions in PP, then genitives follow N in NP”; “if postpositions in PP, then genitives precede”; etc. EIC’s unmarked and marked case predictions for grammars are well supported in relation to the binary branching structures that are implicit in these implicational universals. The great majority of languages exemplify one or other of the two optimal orders, [CmIC] and [ICmC], with [C ICm] being either slightly preferred or equal to [mICC] in the marked case. The statistical correspondences between optimal EIC orders in the performance data of ch. 4, and the frequencies for optimal grammaticalized orders are quite striking (typically in the 85%-95% range). An alternative explanation for these universals in terms of the limited viewing window of the first stage in Frazier and Fodor’s (1978) parser was criticized (cf. Frazier 1979b, 1985). Such a proposal is too absolute, and is disconfirmed by the existence of languages that can regularly exceed the limit, and by the existence of structures for which EIC predicts that it is even preferable to have a longer rather than a shorter viewing window for maximally efficient structure recognition (cf. ch. 5.1.3).

The logically possible orderings for the multiple branching structure \( \text{NP}\{\text{N Adj}_{S}\{C S\}\} \) were considered in ch. 5.2.1 (cf. table 5.8). Four of the 12 possibilities are optimal for EIC, and these are exactly the four that are productively attested across languages. Two further orders have good but not optimal scores, and these two are attested, but no orders with lower scores appear to have been grammaticalized as basic orders. Grammatical data involving the relative positioning of NP and PP within VP provided a further test of EIC’s predictions for grammars in ch. 5.2.2. These grammaticalized orders within multiple branching structures provide further evidence for the strong correlation between optimal EIC orders in performance (cf. ch. 4.1) and in grammars.

Ch. 5.3 took a look at English, a language with strongly grammaticalized word orders, and proposed some simple performance-based regularities for a
number of hitherto unexplained ordering conventions. For the VP the following holds: if an IC is necessarily single-word within the CRD for VP, then it precedes all possibly multi-word ICs (cf. (5.25c)). For the NP: if an IC is possibly single-word within the CRD for NP, then it precedes all necessarily multi-word ICs (cf. (5.28c)). The ordering of complement categories, NP, PP, and $\tilde{S}$, has been grammaticalized in all phrases in accordance with their respective weight aggregates, from short to long, e.g. $[V \text{ NP PP } \tilde{S}]$, $[\text{Adj PP } \tilde{S}]$, $[P \text{ NP PP }]$, etc.; cf. ch. 5.3.3. These regularities all follow from EIC.

EIC's predictions for grammatical rearrangement rules were considered in ch. 5.4. For heavy categories with left-peripheral recognition, such as $\hat{S}[\text{Comp } S]$, EIC predicts postposing; for the corresponding right-peripheral recognition categories ($\hat{S}[S \text{ Comp}]$), preposing is preferred, though postposing can also yield smaller EIC increases. Data from several languages were given that appear to support these predictions, and were followed by a more detailed discussion of Extrapolation in German. Light categories, such as pronouns or single-word particles, were predicted to be rearranged to the left of their CRDs. Illustrative rules were considered from English and German that support this prediction. Many subtle conditions of application on all of these rules are also explained by EIC.

Ch. 5.5 tested EIC's predictions for word-order hierarchies (cf. (3.24)). As the CRD for a given phrase becomes less optimal on account of the center embedding of increasingly complex constituents, there is a diminishing tolerance for the relevant orderings across languages and a greater preference for their more optimal counterparts. This explains the implicational structuring of noun-phrase-internal word orders in Hawkins's (1983) Prepositional Noun-Modifier Hierarchy, which was shown to correlate precisely with increasingly complex CRDs for the PP. The classic center-embedding environments that are commonly discussed in the literature involving a clausal embedding within a higher VP or $S$ were also shown to be hierarchically arranged ($NP > \overline{VP} > \tilde{S}$, cf. (5.66)): a more complex constituent can be center-embedded only if all less complex constituents can be center-embedded as well. Languages exhibit cut-off points down these hierarchies, just as they do for relativization and movement down the hierarchies of ch. 2.3. In all these cases, lower positions involve more complex SDs than higher positions.

Ch. 5.6 discussed left–right asymmetries across and within languages, i.e. asymmetrical skewings in the distribution of categories throughout the parse string that are motivated ultimately by the temporal or left-to-right sequencing of speech, in conjunction with EIC’s efficiency metric. A number of different examples from earlier sections were first summarized, for example
Conclusions

the leftward skewing for single-word pronouns, the rightward skewing for heavy 3[Comp S] structures, the asymmetrical distribution of single-word versus multi-word ICs in the English VP and NP, etc. Ch. 5.6.1 then examined some more general typological asymmetries between head-initial and head-final languages. These two language types exhibit productive mirror-image orderings across common phrasal categories, e.g. PP[NP[N PossP]] versus PP[NP[PossP N] P], etc. For certain sentential nodes, however, there are asymmetries. The sequence 3[Comp S] is productive in both head-initial and head-final languages; the reverse 3[S Comp] occurs only in the latter. Derivatively, [V 3] orders are found in many head-final languages that would otherwise be predicted to have [3 V], namely (primarily) when the 3 is Comp-initial. Similarly [N 3] is a highly productive relative clause ordering in head-final languages, and the exclusive order in head-initial ones.

An explanation was proposed for these typological asymmetries in terms of an independent parsing preference for immediate matrix disambiguation. In a head-initial language, the order 3[Comp S] will be predicted by EIC and this will simultaneously guarantee immediate matrix versus subordinate clause disambiguation based on the presence versus absence of a complementizer. The reverse order predicted for head-final languages, 3[S Comp], results in on-line indeterminacy over matrix versus subordinate status. This indeterminacy can be resolved by inverting the position of the complementizer (among other solutions), and as a result 3[Comp S] is found in many head-final languages; this 3 is then regularly postposed within its clause, etc. This account generalizes to explain the [N 3] asymmetry as well. The absence of an asymmetry within the relative clause, i.e. the absence of 3[C S] N structures corresponding to 3[Comp S] V in head-final languages, is explained on the grounds that the independent parsing preference for immediate matrix disambiguation which motivates these asymmetries is more adequately satisfied by preposing the head noun within NP, i.e. NP[N 3], than by preposing the lower category C that constructs the relative clause itself.

Ch. 5.6.2 examined the relative positioning of subject, object, and verb from the perspective of EIC. Tomlin (1986) and others have established clear and statistically significant frequency differences in the distribution of the six logically possible orderings across languages, cf. (5.72): SVO and SOV account for the great majority (87%) of languages; among the remainder, VSO is the most productive (9%); VOS and OVS are both attested (accounting collectively for 4%); and OSV is either unattested or exceedingly rare. EIC's predictions were calculated both on the assumption that there is a
VP in these transitive clauses, and without the VP assumption. Calculations were made for different assumptions about word-length aggregates within subject and object NPs, and performance data involving these aggregates were considered from both VO and OV languages. It was argued that EIC predicts the distributional facts of Tomlin's ranking exactly, specifically: the unmarked preference for SVO and SOV; and the declining frequencies within the marked orders. Just one explanatory principle can replace the three that Tomlin proposed. The asymmetrical preference for subject before object falls out automatically from EIC's predictions and does not need to be stated or explained separately: the two unmarked orders, SVO and SOV, and the most preferred marked order, VSO, all have subject before object.

Finally in this chapter it was argued that topic and focus nodes also exhibit asymmetrical orderings in response to EIC (ch. 5.6.3). The existence of these discourse-configurational nodes is ultimately driven by pragmatic and/or semantic notions, but their ordering is best explained by EIC, just as the relative ordering of their syntactic sister nodes is explained by EIC.

Ch. 6 considered higher node construction from a parsing perspective. I argued that principles such as Mother Node Construction have been grammaticalized in the phrase-structure conventions of languages, so that structural groupings of words can be recognized in language use. Speakers and hearers have lexical entries for individual words in their mental dictionaries, and these can be accessed by matching the phonetic input of the parse string with the appropriate phonological representations in the dictionary. But for whole phrases generated by the grammar there is no such listing in the dictionary, and the challenge for the parser is to recognize the appropriate abstract structure on the basis of material that is present in the parse string. This imposes strong constraints. In ch. 6.1 I argued that the head of phrase generalization, so prevalent in most current syntactic theories, is actually motivated by the requirements of parsability, and that the defining properties of heads can be given a simple functional explanation in these terms. Thus, properties such as obligatoriness, distributional equivalence, the uniqueness of a head to its mother, projection from a daughter, as well as the head as morphosyntactic locus, as subcategorizand and as governor, were all argued to follow from this theory. At the same time, certain proposed properties of heads that do not follow from the need to recognize higher nodes in parsing, namely bi-uniqueness, category constancy, and the semantic functor status of heads, were argued to be independently questionable and unsupported. The grammar of heads is therefore shaped by MNC, and is rendered less stipulative in the process: all heads are MNCCs.
The converse fails, however: not all MNCCs need be heads. Our theory allows for a possible plurality of MNCCs for any M, e.g. Det and N can both construct NP (cf. the axiom of MNCC Existence in (3.5)), and this explains why bi-uniqueness and category constancy are not defining properties of constructing daughters. Our phrase-structure predictions are still highly constrained, however, by the requirement that a dominating category must contain at least one dominated category, most typically an immediate daughter, that can construct the dominating one. Our theory is compatible with some of these MNCCs being heads, and even with all of them being heads (within appropriately reformulated rules, rather than the fairly standard rule set we have assumed throughout this book.) It is also compatible with none of them being heads, and hence with a theory of phrase structure that dispenses with the head of phrase generalization altogether.

Ch. 6.2 defined and illustrated some additional principles for constructing and activating abstract nodes in parsing, other than the phrasal mother. These are Sister Node Construction (ch. 6.2.1), Grandmother Node Construction (6.2.2), Agreement Projection (6.2.3), and Attachment Promotion and Construction (6.2.4). In ch. 6.2.5 the axiom of MNCC Existence was replaced by the more general axiom of Constructability (cf. (3.5'), which asserts that: 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.

The axiom of Constructability makes available to grammars a number of new ways for constructing higher nodes, and predicts some alternations across and within grammars in the selection of categories that do the constructing at or near the appropriate left or right periphery for that structure. Thus, in addition to V constructing a mother VP, we now allow for appropriate case-marked NPs to construct VP as well, by Grandmother Node Construction; verbs that agree with S-dominated subjects can construct S as well as VP, by Agreement Projection; and so on. Some sample predictions for similarly positioned alternative constructors of higher nodes were considered in subsequent sections of ch. 6.3, with particular reference to sentential nodes (ch. 6.3.1), prenominal relative clauses (6.3.2.), the VP (6.3.3), and the NP (6.3.4). Ch. 6.3.5 summarized alternative ways of indicating scope relations in surface structure that are made possible by the parsing principles of this book.
7.2 Major findings and further issues

In this final section I shall briefly discuss some further issues raised by the major findings of this book, which have not been discussed so far. In the process I shall define what I think the major findings are.

Our first major finding involves the basic ordering conventions of grammars:

(7.1) The basic ordering conventions of grammars maximize ease and efficiency of processing, in the unmarked case, specifically the ease and efficiency with which phrases and their ICs are recognized.

Not all phrases have a basic ordering defined on their daughter ICs. Those that do have grammaticalized EIC.

This explanation amounts to the claim that linear ordering is subservient to constituent-structure recognition. Different orderings of ICs result in more or less efficient structure recognition, as we have seen repeatedly. In my earlier monograph (Hawkins 1983) I offered a more narrowly grammatical account of word-order rules and universals, using certain principles and formal conventions of X-bar theory. This approach could make sense of the head-ordering parameter and of the resulting cross-categorial universals for binary branching structures (cf. ch. 5.1), but it did not generalize to predict other regularities, such as the left-right asymmetries of ch. 5.6, and it had to be supplemented with some additional principles of an ultimately performance nature (e.g. the Heaviness Serialization Principle). The same will be true, I believe, for any purely grammatical approach, precisely because grammatical principles on their own give us no good reason to expect distinctions between long and short ICs, or the absence of certain relative orderings of \(N, \text{Adj}, \bar{S}\), or different directionalities for rearrangement rules, or typological asymmetries between head-initial and head-final languages. By contrast, a performance approach subsumes all of these facts – in addition to the cross-categorial universals for which X-bar theory, or directionality of case and \(\theta\)-role assignment, are well suited – and is therefore more general and explanatory. I would accordingly argue that the direction of explanation in this area goes clearly from performance to competence, and that the generative principles that have been proposed are simply the descriptive rules or principles of particular grammars.

What I did not discuss in ch. 3.8.1 was a further possibility for descriptive grammars which may explain some language-particular differences in ordering conventions that have been reported in the published literature and that
Conclusions

appear to be conditioned by non-syntactic factors. It is conceivable that languages could conventionalize their responses to EIC in partially different ways, using different entities of the grammar as the primitives in terms of which the rules of basic order are defined. This could be expected, according to our theory, if certain subtypes of a given phrase like NP, e.g. semantically or morphologically-defined subtypes, were to possess aggregate weight differences that yielded the same kinds of EIC advantages in a basic order that we find with syntactic subtypes such as pronouns versus non-pronouns. If there are different positioning rules for pronouns, why not for other grammatically defined subtypes of NP as well?

Jacobs (1988) has proposed some precedence principles for German that include Agent before Non-agent and Dative before Patient, in the sense of Fillmore’s (1968) case or semantic roles, i.e. ordering principles in terms of θ-roles. Hawkins (1992c) examines the textual distribution of these roles in German and argues that they exhibit a clear correlation with weight: agents are significantly shorter than non-agents; datives are significantly shorter than patients. If Jacobs's precedence rules are correct descriptive principles of German grammar, therefore, they will follow from the same EIC-motivated short-before-long principle as explains Pronoun before Full NP, another of his precedence principles. By contrast, in the Bantu language Sesotho (cf. Morolong and Hyman 1977) precedence rules are not sensitive to the dative or patient status of an entity, but to animacy. If dative and patient are both animate, or both inanimate, they can occur in either order. But if they differ in animacy, the animate one must always come first. I believe that animate entities will be shorter on aggregate than inanimate ones and will require less linguistic material for referent identification, much as given entities are shorter than new ones. If this is the case, Sesotho will have grammaticalized a short-before-long principle for these NPs (within an otherwise short-before-long language), using animacy as a primitive rather than θ-roles.

These generalizations require further examination and testing. Nonetheless, if we think about cross-linguistic differences in these terms, we may be able to explain the existence of certain non-syntactic ordering conventions within particular phrases and grammars, against the broader background of purely syntactic principles that hold for all other phrases. The ordering of all phrasal subtypes would be explainable by EIC on this view, on the basis of the relevant aggregate weights, and this could provide the missing link between syntactic and non-syntactic generalizations within particular descriptive grammars. Alternatively, this kind of reduction may
Major findings and further issues

not always be possible, and a grammaticalized order such as Agent before Non-agent may need to be explained in some other way. Clearly, however, the more apparently disparate generalizations that can be subsumed under a single unifying principle, the better.

This also puts a new perspective on our discussion of pragmatic ordering principles. I showed in ch. 4.4 that when order is grammatically free, EIC makes good predictions, while various pragmatic principles do not. I explicitly allowed, however, for the existence of pragmatically (and/or semantically) motivated grammatical nodes, such as Topic and Focus nodes of the kind found in Hungarian, and I presented some performance evidence for their reality (ch. 4.1.2): NPs are not positioned into these preverbal discourse-configurational nodes of Hungarian on the basis of their weights, as they are in the postverbal, grammatically free domain. Now, it is possible that some languages may have selected pragmatically-defined NP subtypes for their ordering conventions, involving identifiability or other subtypes of definiteness. If so, these pragmatic nodes can also be assigned a basic order by EIC, in the same way that Topic and Focus nodes are (ch. 5.6.3). The result will look like pragmatically determined order in the grammar, but again it will not be. It will be EIC-determined order of pragmatically motivated phrases, within grammars that have such phrases. Languages will differ over their grammaticalization of these pragmatic distinctions, just as they differ over the relevance of animacy. If NP subtypes are to be distinguished pragmatically and referred to by ordering conventions, however, it is essential that the relevant pragmatic distinctions should be grammaticalized in the syntax, and it is predicted that the particular ordering rules in question will be based on weight aggregates and optimal efficiency within the containing CRD for the language in question, as usual. The kind of language on which this hypothesis might be tested is Papago, to judge by the discussion of pragmatic factors alone in Payne (1987). The evidence for grammaticalized orders in terms of pragmatic entities will need to be compared with Hale's (1992) ordering arguments in terms of syntactic entities. Performance data will then need to be examined for evidence of grammaticalization, in the manner of ch. 4.3, and specifically for evidence of the best grammatical theory. Which grammaticalized categories or subcategories have been assigned a fixed order should be visible in the data of language performance; cf. further (7.3) below.

A further issue raised by (7.1) is of a psycholinguistic, rather than a grammatical, nature. It is being claimed that words and ICs are positioned in ways that benefit recognition, i.e. comprehension by the hearer. I have been non-committal throughout this book about the precise benefits of EIC for the
speaker, i.e. for production. There is, of course, a general benefit for the producer if his or her speech is optimally packaged for the hearer, since communication will then be effective. It is also probably the case that there is a significant overlap in processing loads associated with low EIC ratios for both the speaker and the hearer. For example, increasingly complex center-embedding environments of the kind discussed in ch. 5.5 require more simultaneous processing of nodes within a given higher constituent from both perspectives. However, there are some apparent discrepancies between production and comprehension, and in these cases it appears to be the needs of the hearer that are reflected in cross-linguistic ordering conventions.

Consider head-final languages like Japanese. The preposing of heavy S complements with right-peripheral complementizers is advantageous for the hearer, in conjunction with a bottom-up parse. It is not obviously of benefit to the speaker, who does not share the hearer's vagueness about the precise status of each clause being processed, and who knows from the outset that a given \( \tilde{S} \) is a subordinate clause within a given matrix S. If we were to calculate "construction production domains" from the speaker's perspective, rather than CRDs from the hearer's, we should begin registering the presence of the matrix S from the moment its first dominated constituent appeared in the left-to-right production string, and it should make no difference whether a dominated \( \tilde{S} \) was preposed or center-embedded. It is only from the hearer's perspective that preposing is preferred, because it shortens the CRD for the matrix by including within this CRD only the right-peripheral complementizer of \( \tilde{S} \), i.e. the first constituent dominated by \( \tilde{S} \) that makes it clear to the hearer that this \( \tilde{S} \) is a subordinate IC within a matrix S, by constructing it.

For English-type languages in which heavy phrases are postposed to the right, it has been argued by De Smedt (to appear) that there is an explanation for ordering in terms of production. According to his Incremental Parallel Formulator, syntactic segments are assembled incrementally into a whole sentence structure, and the relative ordering of segments can reflect both the original order of conceptualization and the processing time that is required by the Formulator for more complex constituents. The late occurrence of heavy NPs is explicitly accounted for in this model in terms of the greater speed with which short constituents can be formulated in the race between parallel processes in sentence generation. The existence and productivity of languages like Japanese, however, appears to be in direct conflict with this explanation, because heavy constituents are preferred before shorter ones in these languages.
Nothing hinges on this in the present context. Indeed, if EIC can be systematically generalized from a model of comprehension to a model of production, by showing that the Formulator also benefits from a rapid and efficient positioning of categories that construct phrasal nodes, then so much the better. In the meantime it appears that EIC is a comprehension-oriented principle of production, with numerous but not completely systematic correspondences between ease of comprehension and ease of production. I leave a resolution of this issue to the psycholinguists.

The second major finding of this book is given in (7.2):

(7.2) Grammatically free and grammatically fixed linear orderings of constituents are determined by the same principle (EIC), in performance and grammar respectively.

This claim rejects the view that constituents whose order is not fixed by grammatical conventions will be positioned by pragmatic principles in performance. Instead freely ordered constituents are arranged by EIC, on the basis of the weights that each constituent happens to have in each performance instance, in conjunction with the structural characteristics of the language in question. This theory accounts naturally for historical change. When linear orderings of ICs become grammaticalized, the orders that are fixed are those that were most frequent at the immediately preceding free order stage. What has not yet been discussed is the question: when do grammars conventionalize fixed orders, as opposed to leaving them subject to performance principles of arrangement? What other changes in the grammar will result in the acquisition (or loss) of fixed-order conventions?

This question is complex because it gets at the issue of the interrelatedness of different “modules” or components of the grammar. For example, Keenan (1978: 120–121) points out that “case-marking and word order restrictions on major constituents of basic sentences have somewhat the same function of coding major grammatical relations, e.g. ‘subject-of’, ‘direct object-of’, etc.” And he proposes the “Principle of Covariation of Functional Equivalents” according to which “the more we assign a language overt case marking the freer can be its basic word order and conversely.” This insight also underlies historical explanations for the change from the more freely positioned subject and object in an SOV language to SVO word order in response to morphological syncretism (cf. e.g. Vennemann 1975). In SVO languages subject and object occupy fixed positions in relation to the verb. Keenan’s principle appears to be correct, in general, though there are some languages without case morphology that appear to have very little fixed ordering, e.g. Bulgarian
Conclusions

(Donca Minkova, personal communication); and conversely languages with case marking that have quite extensive grammaticalized orders, e.g. German.

There are a number of issues here that may be worth pursuing from the current perspective. First, the distinction between fixed- and free-word-order languages is not an all-or-nothing thing. It is a matter of degree. Some phrases are associated with more freedom of order among their daughter ICs than others. For example, the relative ordering of P and NP within PP is typically fixed and diachronically stable, whereas the ordering of V in relation to many of its arguments and modifiers, and of these latter in relation to one another within the VP or S, is much freer and more diachronically variable (cf. Hawkins 1983). The reason for the difference between PP and VP is probably none other than the reason for the variation that has been observed within the VP in English and other languages (cf. ch 5.3 and 5.4). For example, the positioning of pronouns in VP can be highly constrained and fixed, whereas the relative ordering of (non-pronominal) NPs, PPs, and S is more variable, making it often difficult to determine whether there is a basic order at all (cf. ch. 5.3.3). In German, a non-subject S complement is grammatically fixed (or obligatorily rearranged) in postverbal position; the (less complex) V \overline{P} complement occurs both before and after the verb (cf. ch. 5.4.1).

The hypothesis suggested by these facts is that, in the transition from free to fixed word order, languages first grammaticalize IC orderings that yield fully optimal scores, or scores with the largest differential between the most optimal order and alternative orders. Thereafter IC orders will be grammaticalized with smaller differentials between the most optimal and less preferred alternatives (all these scores being calculated on aggregate weights, as usual). Thus, the fixing of pronominal orders such as [V Pro Part], [V Pro PP], etc., is a grammaticalization of optimality. Similarly, the grammaticalization of [P m NP] and of [NP m P] within binary branching structures will generally be optimal within the relevant language type, and so is [V m NP] and [NP m V], which explains the absence of VO/OV alternations in many languages. The grammaticalization of [... V \bar{S}] in German is motivated by the fact that the reverse [... S V] has a significantly worse EIC score. The differential between [... V \bar{VP}] and [... \bar{VP} V] is not so great, so both orders are possible.

These and other predictions made by EIC for what appears to be a grammaticalization continuum from free to fixed ordering need to be explicitly tested. If EIC can, in fact, explain such a continuum, this will require a rethinking of the conditioning factors elsewhere in the grammar that give rise to grammaticalized orders. The fixing of [... V \bar{S}] in German will be
explainable irrespective of whether there is a case system or not. On the other hand, the verb-final rule of that language is plausibly sanctioned by the fact that the case system does provide early recognition of the VP, and this, coupled with the brevity of pronouns, can explain why these items have a basic order positioning to the left of all other daughters of VP (cf. ch. 6.3.3). In English, length alone explains the grammaticalization of [V Pro PP]. Case marking must therefore be viewed in the context of the role that it plays in parsing, in order to predict what consequences (if any) its loss will have for the fixing of word order. At the same time many fixed orders can be explained whether a language has a case system or not. Case marking is one way of indicating constituent structure and co-constituency. As this form of constituent structure identification is lost, in those languages that depend on it, certain changes in linear ordering conventions can be predicted and explained. There may be a reordering of constructing categories, such as the positioning of the verb. Strict adjacency can be imposed as a signal of co-constituency, and scrambling and discontinuities will be outlawed. The complementarity of case morphology and fixed word order to which Keenan (1978) refers would then follow, on this view, from their functional equivalence as surface signals of phrase structure. Cases do not just code subject-versus-object status. They construct the respective phrasal categories (e.g. S versus VP) in which these NPs are positioned, and they indicate co-constituency with whatever category assigns the case. Reorderings and strict adjacency can compensate for these functions when surface cases are lost. This also needs to be further investigated.

Our third finding is related to (7.2):

(7.3) The distribution of variant orderings in performance can provide evidence for the existence and nature of any grammatical ordering conventions.

In effect, performance data involving frequency of occurrence, and correlations between the distribution of differently ordered ICs and their respective weights, can supplement purely grammatical data in providing arguments about grammatical conventions. Grammatical data involve the opposition between grammaticality and ungrammaticality, or in this context between the occurrence and complete non-occurrence of certain orderings: John \( v_P [gave \ N_P [it] \ P_P [to \ the \ student]] \) is grammatical, \( *John \ v_P [gave \ P_P [to \ the \ student] \ N_P [it]] \) is not; Mary \( v_P [called \ N_P [him] \ up] \) is grammatical, \( *Mary \ v_P [called \ up \ N_P [him]] \) is not. The claim is, therefore, that patterns of occurrence in
Conclusions

performance can also reveal the existence and nature of grammaticalized orderings such as [V NP PP] and [V NP Part] in English.

One very common assumption in the literature is that basic orders will be retained in performance in the unmarked case in preference to any rearranged orders, and hence that basic orders will be most frequent in texts. This is certainly the case in general, because according to our theory basic orders are those that are most optimal for EIC and that are grammaticalized on the basis of weight aggregates. [V NP PP] accounts for 95% of the data in table 4.22, and [V PP NP] for only 5%. But not all basic orders are more frequent than their rearranged counterparts: a limited number of non-optimal orders are permitted in grammars, just as they are in performance. The basic order [V NP Part] accounts for only 42% of the data in table 4.21, and the rearranged [V Part NP] for 58%. Collectively these two orders satisfy EIC's predictions for performance, but this means that greater textual frequency is not a necessary correlate of grammatical basicness.

Nor is it sufficient to establish basicness. The greater frequency of [V NP PP] over [V PP NP] would be predicted by EIC whether there was a grammaticalized order or not, since 73% of these structures have PP > NP, compared to only 13% NP > PP (and 14% NP = PP). The true test for grammaticalization must therefore come from a more sensitive examination of performance. Specifically it can come from examining those (minority) subsets of the data whose weights are not in accordance with the aggregates, and seeing whether they still maintain any grammatical ordering convention, or whether they respond immediately to EIC in the absence of a convention to the contrary.

There are two types of revealing situations: cases in which EIC scores for two orders would be equal and both would be productive in a free-order language, yet only one occurs or is productive on account of grammaticalization; and cases in which EIC would actually define a weak preference for the non-grammaticalized order, yet the grammaticalized basic order is again the only productive one. The number of these retained basic orders in the face of weak contrary EIC pressures is limited by EIC's predictions for performance, which require the most optimal orderings in the unmarked case. Nonetheless, they do exist, and these in conjunction with the cases where EIC's preferences are equal are highly revealing for the grammar.

We saw numerous examples in ch. 4. [V NP PP] is clearly the basic order in the English VP, since it is the only productive order when NP = PP, and it is also productively retained when NP > PP by one to three words (cf. table 4.22). [NP PP V] is the basic order in German, because it is the productive one
when NP = PP and when NP > PP by one or two words (cf. table 4.23). \([V NP \text{ Part}]\) is basic because it is the only productive order when NP = Part, and because there is still a significant number of these orders when NP > Part by one word (32%). By contrast, in the postverbal domain of Hungarian, in which there are no grammatical arguments for basic ordering, there is an immediate sensitivity to EIC and no lingering grammaticalized ordering: cf. table 4.4. The productivity of certain orders in the face of equal or weak contrary EIC preferences therefore provides evidence for grammaticalization.

A further argument for basicness can come from the very existence of alternative orderings. For example, if \([V NP \text{ Part}]\) is the basic order in Modern English, we expect productive rearrangements to \([V \text{ Part NP}]\) whenever NP > Part, which is what we get. If, in Future English, \([V \text{ Part NP}]\) becomes the basic order, we predict no corresponding rearrangements to \([V NP \text{ Part}]\) because the basic order will be optimal. Hence, the productivity of both orders in Modern English provides an argument for the grammaticalization of \([V NP \text{ Part}]\). If \{NP, Part\} were freely ordered, then \([V \text{ Pro Part}]\) and \([V \text{ Part Pro}]\) should both be grammatical, which they are not. All sources of evidence (grammaticality distinctions, performance distributions, and rearrangement considerations) therefore conspire to reveal \([V NP \text{ Part}]\) as the basic order of the English VP, despite its relative infrequency!

Performance data involving the distribution of variant orderings can also provide evidence about constituent structure:

(7.4) The constituent structure of a set of grammatical categories will be reflected in the attested orderings of these categories in language performance: different constituent structure groupings will define possibly different optimal orderings in conjunction with EIC.

Hence, if there are different grammatical analyses of constituent structure in a given case, then by comparing the predicted optimal orderings of each analysis with the attested data, performance will often provide evidence that can choose between them.

For example, the ordering of an accusative-marked NPo in Japanese relative to a nominative NPga exhibits a distributional pattern that was argued in chs. 4.1.5 and 5.6.2 to provide evidence for a VP. These data from Japanese (and similar data from Korean) corroborate arguments of a purely grammatical nature for the existence of VP. Performance data from Hungarian support the basic sentence structure proposed for that language by Kiss (1987, 1991; cf. (4.3) and ch. 4.1.2).
As for competing grammatical analyses, there are different proposals in the literature for the derived constituent structure of Extraposition from NP in English and German (cf. ch. 4.3.5). A complex NP, $N_P[N \tilde{S}]$, appears to exhibit identical grammatical behavior involving c-command effects and WH-extraction, whether the $\tilde{S}$ is adjacent to $N$ or extraposed into clause-final position, inviting the hypothesis that this $\tilde{S}$ is still immediately (but discontinuously) attached to NP after extraposition (cf. McCawley 1982, 1987). Other grammatical analyses are opposed to crossed branching, in principle it seems. Performance data from German provide strong support for discontinuity, for it can be shown that the pattern of application versus non-application in texts is exactly what EIC would predict, on the assumption that the $\tilde{S}$ remains attached to NP; cf. tables 4.26–4.29. The larger the distance between $N$ and $\tilde{S}$, i.e. the more intervening material and the greater the discontinuity, the fewer instances of Extraposition from NP there are. In particular, there is a threshold at which the EIC advantages for whatever dominating constituent contains the complex NP (e.g. VP) are offset by the disadvantages for the NP itself when $\tilde{S}$ is extraposed. Extraposition from NP is highly productive in structures that fall below this threshold and in which discontinuity is motivated, and practically non-existent in those structures that exceed the threshold. This prediction cannot be made if $\tilde{S}$ is extracted from NP and attached to, say, a higher S (cf. Reinhart 1983), because there will then always be an EIC advantage to extraposing a heavy $\tilde{S}$ within this higher constituent, and removing an $\tilde{S}$ from NP will leave this latter constituent with an optimal ratio as well. Hence, Extraposition from NP in this analysis is wrongly predicted to be productive out of complex NPs in all environments, and to pattern just like Extraposition proper, in which the rearrangements are always highly preferred to their corresponding basic orders (cf. table 4.25). The correct prediction for Extraposition from NP relies on the conflicting EIC advantages for the containing domain and for the NP, and on the discontinuity assumption, and so provides direct evidence for the best derived constituent structure.

It is my belief that many more such arguments for discontinuity will emerge from performance data of this type. For example, the data from Japanese and Korean support not only a VP, but also a discontinuous VP analysis for OSV structures, as proposed by O'Grady (1987), and do not support alternative analyses involving, for example, the extraction of $N_P$ out of VP and attachment to $S$, with or without Chomsky-adjunction (cf. ch. 4.1.5). OSV structures are extremely infrequent in performance in both languages, and this massive preference for SOV, even when O is longer than S, is
predicted by the discontinuous analysis, but not by the other analyses (which
all prefer a long O before a short S). Other structures and languages need to
be examined from this point of view, especially languages with rich morphol-
ogies that appear to exhibit regular discontinuities in surface structure
between e.g. NP constituents, i.e. Latin, Australian languages, etc. If these
constituents are really discontinuous and share a common dominating NP
node, then their distribution and frequency should pattern like Extrapolation
from NP: the discontinuity is predicted only for those surface patterns in
which the EIC benefits for the phrase(s) containing the NP outweigh the
disadvantages for the NP itself resulting from the discontinuity.

More generally, the adjacency of daughter ICs in the unmarked case, which
we observe across languages and which has given rise to the axiom of no
crossed branching in many grammatical models, is a simple consequence of
EIC: if non-NP-dominated material intervenes between the first IC of NP
and the last, the ratio for NP will be non-optimal. However, the advantages
for containing CRDs can also motivate some NP discontinuities, as we have
seen. Hence, this explanation for adjacency in the unmarked case also pre-
dicts certain instances of non-adjacency, and the axiom of no crossed branch-
ing is therefore unmotivated. The performance explanation for the effects of
this axiom rules out the validity of the axiom itself! Extrapolation-type rules
are very productive in many languages. Other forms of discontinuity may
require additional morphological features, such as agreement or case mark-
ing, that guarantee recoverability of constituent structure. If these features
are present, this will make available a greater range of structures for rearran-
gement, free ordering, and for discontinuity in response to EIC.

Our fifth finding also involves constituent structure:

(7.5) Heads of phrases are constructing categories in parsing whose gram-
matical properties are motivated by the need to recognize the nature of
the phrasal category that dominates a set of ICs in the parse string.

The grammatical properties of heads, and the parsing motivation for them,
were discussed extensively in ch. 6. I argued that there are numerous ways in
which a grammar could render dominating phrase structure recognizable on
the basis of uniquely determining daughter categories, and that this expan-
atation is even compatible with grammars that make no use of the head of
phrase generalization at all and in which mother–daughter relations are
constrained by our parsing axioms. This approach can explain why a num-
ber of grammatical properties cluster round these so-called heads, instead of
just stipulating that they do, and it can resolve some disagreements in the literature.

There is a related issue in this context. Theories of grammar incorporating the head-of-phrase generalization are typically associated with a rich phrase-internal branching structure consisting of dominating nodes that are projections of the head. Recent theories also postulate a relation of bi-uniqueness between constructing categories and their mothers or maximal projections; cf. ch. 6.1.2. The data and theoretical arguments of this book make me suspicious about the reality of this amount of phrase-internal structure and suggest instead that syntax is much flatter than is often assumed.

Notice first that by the metric of complexity developed in ch. 2.2, the postulation of such phrase-internal constituency adds considerably to the complexity of Structural Domains, for there are now many more dominating nodes than there would be in a flatter structure. Imagine the tree structures of (3.1'a) and (3.1'b) with all these projections added under the NP and PP ICs of the matrix VP. Such structures will be very complex indeed, and the CRDs for VP will contain much lower IC-to-non-IC ratios.

Our theory is based on the proposition that more nodes means greater complexity. To the extent that grammar is driven by performance in the areas we have examined, this theory predicts that grammars will be parsimonious in their phrase-internal structure, all things being equal, and will prefer flatter to richly structured phrases. All things may not be equal, of course. Nonetheless, this consideration provides an antidote to what seems to be the unmarked assumption in syntactic analysis at the present time, which is to postulate as much structure as possible!

Secondly, many of the arguments for rich constituency, involving scope and binding asymmetries, do not necessarily require an asymmetric c-command analysis in terms of constituent structure, because these kinds of asymmetries are often formally coded in languages in terms of left–right precedence, or in terms of differential case marking, etc. (cf. Keenan 1988; Primus 1989, 1991a; and ch. 2.2.1 and 2.4).

Thirdly, whereas the uniqueness of a constructing category to its mother is motivated by parsability, bi-uniqueness is not (cf. ch. 6.1.2). Indeed, bi-uniqueness is at variance with the fact that there is often a plurality of daughter categories capable of constructing the same mother phrase, and leads to the loss of the descriptive and cross-linguistic generalizations presented in ch. 6.3. Variation in the ordering and selection of these different daughter categories is correctly predicted on the assumption that they can each construct the mother.
For all of these reasons flat structures should be assumed in the unmarked case, unless we absolutely have to postulate a phrase-internal constituency grouping and an additional dominating node. The existence of the phrasal categories themselves, on the other hand, is plausibly motivated ultimately by the semantic and syntactic relations that unify groups of words. Where there appears to be cross-linguistic variation, e.g. over the VP node, it was argued in ch. 5.6.2 that this phrase can improve EIC ratios for the sentence in many language types (e.g. SVO), though it results in no increase for VSO, and in a decrease for OSV. Phrasal categories can be motivated by a variety of considerations, therefore; phrase-internal structure is more problematic, and should be assumed sparingly. This is what we have done here (cf. the working assumptions about constituent structure in (3.12), ch. 3.4), and the ease with which phrase-internal groupings could be dispensed with in making correct EIC predictions throughout this book intensifies my suspicions about their non-existence.

Our sixth general finding involves hierarchies of grammaticalized word orders across languages. There is an emerging generalization uniting these hierarchies with other syntactic hierarchies, and with hierarchically arranged phenomena in other areas of language as well. We have argued for the following:

(7.6) Implicational hierarchies of center embeddings within CRDs (cf. ch. 5.5), of relative clause formations within Relativization Domains (cf. ch. 2.3), and of WH-extractions within Movement Domains (cf. ch. 2.3) are all syntactic complexity rankings explainable in terms of relative degrees of processing ease and efficiency.

A common metric of structural complexity unifies these distinct areas of the syntax and gives substance to the intuition that processing ease and efficiency decline down these hierarchies (see (7.7) below). The rather fascinating correlate of this for a theory of grammaticalization is that certain grammatical options, such as the center-embedding of a heavy \( \tilde{S} \), or relativization on a low position of Keenan and Comrie's Accessibility Hierarchy, or WH-extraction out of a finite clause, are only permitted in a language if all less complex options within the same grammatical domain are also permitted. These hierarchies define the sequence in which grammatical variants are selected within each grammatical domain, and the claim is being made that this sequence involves increasing complexity, and that cut-off points represent a conventionalized response by speakers of each language not to tolerate processing difficulty or inefficiency below that point. How far down a given
Conclusions

hierarchy a grammar goes will presumably reflect the availability or otherwise of alternative grammatical expressions within other domains that can achieve the same or similar expressive power and meet speakers' communicative and conceptual needs. To the extent that these needs are similar across cultures, we may expect to see a similar mix of more and less complex grammatical options selected across all grammatical domains. Another relevant factor that appears to determine selections made across different grammatical domains is again processing. The selection of verb-final word order is argued in Hawkins (1994) to have consequences for the surface coding of 0-roles, and for restrictions on WH-extractions and raisings, because of the way in which argument–predicate structure is recognized in such languages.

The other fascinating point about these syntactic hierarchies is that we now have an emerging synthesis with the empirical and theoretical results that have been reported for hierarchies elsewhere in the grammar. In ch. 1.2 I referred to the existence of phonological hierarchies of expanding vowel and consonant inventories, and I mentioned that an explanation has been offered for them in terms of added processing effort (cf. Lindblom et al. 1984; Lindblom and Maddieson 1988). Richer inventories are associated with more elaborated and complex phonetic variables. A processing explanation has also been offered by Kay and McDaniel (1978) for the implicational hierarchy of basic color terms established by Berlin and Kay (1969) (cf. n.1 ch. 1). Kay and McDaniel argue that the ranking of this hierarchy is determined by the neural anatomy of our color vision system, and that languages lexicalize basic color terms for the most contrastive stimuli first, i.e. those that produce the most distinctive neural response and that are most readily perceptible. There is also evidence for increasing conceptual complexity in many of the semantic markedness hierarchies discussed by Croft (1990).

Clearly, increasing complexity and processing effort appears to be a major, if not the major, source of explanation for implicational hierarchies across languages. The precise nature of the complexity and processing effort will vary depending on the grammatical or lexical domain in question, but it does seem that implicational hierarchies owe their existence to processing, just as individual implicational universals such as “if SOV, then postpositions within PP” define co-occurrences of phenomena that are driven by processing. These two word orders are optimal for EIC when they co-occur. The hierarchies define increasing degrees of difficulty among possibly co-occurring options within a common grammatical or lexical domain.

The definition of syntactic complexity that we have proposed has already been amply discussed:
(7.7) Syntactic complexity is an additive function defined on Structural Domains; cf. ch. 2.2.1 (and 7.1 for a summary).

The Structural Domains that we have been assuming are defined on surface structures, enriched with empty nodes for relativization and WH-movement constructions (cf. the definition of a Relativization Domain in (2.15) and of a Movement Domain in (2.29)). This theory continues a tradition of research and an insight initiated by Miller and Chomsky (1963), whereby complexity is equated with the amount of structure in a sentence. We have argued that the quantification of this amount needs to be relativized to certain grammatically and/or psycholinguistically significant sets of nodes in a tree, in order to assess how much structure is relevant to particular grammatical operations and parsing decisions. The more such structure, the more complexity.

One question that arises is whether our metric is sufficient on its own to account for all cases of syntactic complexity for which there is performance evidence. For example, the Derivational Theory of Complexity quantified the degree of distance between deep and surface structures, and so defined an additional relevant variable beyond surface structure alone. This theory was abandoned on account of its failure to make correct predictions for performance (cf. Prideaux 1984 for a summary of the issues), but it has recently been revived by Pritchett and Whitman (1994) in relation to more current grammatical assumptions which result in very different predictions for performance. The inclusion of empty categories within the surface structures for relativization and movement considered here goes some way towards accommodating earlier derivational levels. But it remains to be seen whether other aspects of derivational complexity are reflected in performance, and whether the definition of surface structure complexity proposed here needs to be supplemented in this way, and if so how.

I have argued against Frazier's (1985) alternative domain for assessing complexity, in terms of three successive words in the parse string. I have also criticized the five-to-six-word viewing window of Frazier and Fodor's (1978) first-stage parser. Some languages have linear orderings that will regularly exceed these limits, while some of EIC's preferences assert themselves well within these putative bounds on assembling phrasal packages (e.g. \textit{vP}[gave it to the man] versus \textit{*vP}[gave to the man it]). These kinds of proposals for absolute limits on the parser's viewing window, and the related claims about working memory limitations beyond these processing spans, are conjectures that do not strictly follow from any general psycholinguistic or
grammatical principles that I am aware of. In fact, one could argue that it is implausible that the human parser would have any built-in limitation defined on something as uninteresting as the number of words in the parse string, given that the object of parsing is to build a whole structural representation, and given that languages vary radically in the internal complexity of their words. Building structure on-line requires accessing whatever number of words are necessary for the fast recognition of that structure. Sometimes the efficiency of recognizing a whole sentence structure is increased if one phrasal node has a very long recognition domain, e.g. VP, in order to keep the recognition domain of another short, e.g. NP, as we saw in the case of Extraposition from NP, cf. ch. 4.3.5. So instead of thinking in terms of absolute word limits in parsing, I believe it makes more sense to think of the structure that needs to be identified, and to have an efficiency metric that sets a premium on extracting as much structural information as possible from as little input as possible. This gives us, in effect, a theory of relative complexity and of ranked preferences. It defines no upper bounds and absolute limits, because we have no way of knowing at present what these are, and indeed whether they exist at all.

The kind of ultimate explanation for these preferences that I find plausible is a neurological one. In more complex CRDs, Relativization Domains and Movement Domains, there are more structural nodes to process, and more neurons are being activated in the mapping from sound to structural representation. If the same amount of structural information can be derived from the processing of fewer nodes and terminal elements, as a result of selecting one linear ordering over another, then there will be fewer simultaneous computations to be made, and fewer neurons will be firing at the time the relevant phrase and its ICs are recognized. If relative clause formation is made on a subject rather than an oblique NP, fewer computations will be required to recognize the Relativization Domain, and again fewer neurons will be firing. But whether the brain has some built-in limit on the number of neurons that can fire simultaneously in its relevant neural networks is simply unknowable at the moment. Certainly it doesn't have the limits that have been proposed.

Further evidence for this approach to complexity comes from temporary ambiguity preferences in performance:

(7.8) When the parser has a choice, it prefers to resolve temporary ambiguities by postulating smaller rather than larger SDs.
I have argued that Frazier's (1979a) three parsing preferences, Minimal Attachment, Late Closure, and the Most Recent Filler all reduce to this generalization (cf. chs. 3.5.3 and 4.5).

Finally, I can summarize the thrust of this whole book in the following generalization:

(7.9) The impact of processing on grammatical universals of both the absolute and variation-defining kind has been fundamental.

It may ultimately be possible to derive many more syntactic universals from processing considerations. Syntactic categories may owe their existence to the need to reduce a multiplicity of real-world semantic distinctions to a manageably finite set of entities that can be readily manipulated in real time, whose structural relations with other entities in a sentence can be rapidly computed, etc. This would result in a number of abstract semantic correlations with syntactic categories that are often difficult to define, and in what are plausibly hierarchically arranged differences over category membership across languages. Constituent groupings of categories may also be driven ultimately by various semantic relations between entities, again reduced to a manageably finite set of grammatical co-occurrence restrictions with only rather approximate correlations with meaning and subsuming many different semantic relations and subtypes. More generally, the very autonomy of syntax may owe its existence to the need to make semantics processable for both the speaker and the hearer, and it remains to be seen whether any precision can be given to the formula: semantics + processing = syntax. The role of processing as an explanation for many syntactic universals has been extensively illustrated in this work. The precise integration of efficiency and performability considerations with the other major determinant of language structure, meaning, is an important and exciting research goal for the future.
One of the great scenes in world literature contains an amusing discussion of word order preferences. The philosopher in Molière's *Le Bourgeois Gentilhomme* is giving a lesson in French composition to the unwitting Monsieur Jourdain, who is anxious to discover the best possible arrangement for his amorous message to *la belle marquise*. The EIC principle will now be put to the ultimate test. Can it explain the philosopher's stylistic preferences, and the spontaneous speech of Monsieur Jourdain? It appears that it can!

MONSIEUR JOURDAIN: Je suis amoureux d’une personne de grande qualité ... Je voudrais donc lui mettre dans un billet: *Belle marquise, vos beaux yeux me font mourir d’amour; ... Je ne veux que ces seules paroles-là, ... mais tournées à la mode, bien arrangées comme il faut. Je vous prie de me dire ... les diverses manières dont on les peut mettre.*

LE MAÎTRE DE PHILOSOPHIE: On les peut mettre ... comme vous avez dit:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[belle marquise]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s [2/4 = 50%]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ou bien:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s [VP[PP[d’amour]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mourir] me font]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[belle marquise]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP [3/3 = 100%]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s [2/5 = 40%]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agg. = 87.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ou bien:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s [NP[vos beaux yeux] VP[VP[PP[d’amour]] me font [belle marquise] VP[mourir]]</td>
<td>3/3 = 100%</td>
<td>2/3 = 67%</td>
<td>3/3 = 100%</td>
<td>2/3 = 60%</td>
</tr>
<tr>
<td>s [2/6 = 33%]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agg. = 77</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ou bien:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s [VP[VP[mourir]] NP[vos beaux yeux] [belle marquise] VP[VP[PP[d’amour]] me font]]</td>
<td>3/3 = 100%</td>
<td>2/7 = 29%</td>
<td>3/10 = 30%</td>
<td>2/8 = 25%</td>
</tr>
<tr>
<td>s [2/8 = 25%]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agg. = 47</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Ou bien:

\[ s[vp[me font] \text{ NP[vos yeux beaux]} \text{ vp[vp[mourir] \text{ belle marquise]} \text{ vp[pp[d'amour]]}]] \]

\[ np[3/3 = 100\%] \quad vp[2/4 = 50\%] \]

\[ s[2/3 = 67\%] \quad \text{Agg.} = 67 \]

Monsieur Jourdain: Mais de toutes ces façons-là, laquelle est le meilleure?

Le maître de philosophie: Celle que vous avez dite: *Belle marquise, vos beaux yeux me font mourir d'amour*. [i.e. BEST EIC RATIO!]
Notes

1 Introduction

1 Here and throughout this book I shall use the term "self-embedding" for structures of the type \( A[X A Y] \), i.e. a constituent of type \( A \) is embedded within a higher constituent of the same type, in such a way that there is non-null material to both the left and right of the lower \( A \). In example (1.2) in the main text the relative clause that gave bad milk is self-embedded within the higher relative clause that the cow . . . kicked. The term "center-embedding" will be used for any structure of the type \( a[X A Y] \), i.e. a constituent of type \( A \) is embedded within a higher constituent of any type, \( a \), and there is again non-null material to both the left and the right of \( A \). Self-embedded structures are therefore also center-embedded, but the converse fails; a constituent \( A \) will be center-embedded and not self-embedded in the event that \( a \) is a constituent of a different type from \( A \). I have referred to example (1.1a) in the main text as center embedding rather than self-embedding. This is because it is not obvious that we should regard a subordinate clause as a constituent of the same type as a main clause (cf. ch. 6.2.2 and 6.3.1 below for further discussion). A number of current grammatical models would disagree, of course, and would assign the same dominating category to both, and for these models (1.1a) would be another example of a structure that is both self-embedded and center-embedded. Either way (1.1a) will be center-embedded, and embedding a relative clause within a relative clause in (1.2) is equally clearly a self-embedding.


3 Consider also the famous implicational universal from the lexicon involving basic color terms and established by Berlin and Kay (1969):

\[
\begin{array}{cccc}
\text{white} & \Rightarrow & \text{red} & \Rightarrow \\
\text{black} & \Rightarrow & \text{green} & \Rightarrow \\
& & \text{blue} & \Rightarrow \\
& & \text{brown} & \\
\end{array}
\]

If a language has just three color terms, they will be white, black, and red. If there is a fourth, it will be green or yellow, etc. Kay and McDaniel (1978) have proposed a processing explanation for this hierarchy, in terms of ease of perception. They argue that the focal points for basic color terms are determined by the neural anatomy of our color vision system. Focal colors are those that produce the most distinctive neural
response, and the above ranking encodes the most contrastive stimuli first. Again we have a processing explanation for the structuring of cross-linguistic co-occurrences, and for correlated frequencies.

2 The grammaticalization of processing principles

1 It appears that case morphology can also sometimes expand an SD defined syntactically, as well as reducing it, though this option appears to be rather rare across languages. Imagine that an accusative were to occur as a sister of a VP that contained a nominative. This structure has been proposed, in fact, for VP-nominative languages like Toba Batak: cf. Schachter (1984), Keenan (1988). This language is verb-initial, so the proposed structure is S[V NP-Nom] NP-Acc]. Toba Batak also possesses VP-accusative structures, i.e. S[VP [V NP-Acc] NP-Nom], and the two are distinguished on the basis of different verb prefixes, a D-prefix for the VP-nominative structure, and an M-prefix for VP-accusatives. Imagine also that a dative were to occur as a sister of some projection of V that included an accusative which was lexically required to co-occur with the dative, as has been proposed for German in Primus (1987, 1989, 1991a). The structure here would therefore be VP[VP [NP-Dat VP[NP-Acc V]], with the different cases being distinguished via nominal morphology. In the Toba Batak VP-nominative structure, the accusative will not have the nominative in its SD, by the definition in (2.9), whereas the nominative does have the accusative in its SD. We could therefore say that the SD for the accusative is expanded to include the grammatically required nominative which it c-commands, whereas the SD for the nominative has the accusative removed, as before. The accusative will then contract the usual structural relations with the nominative, reflected in anaphoric binding asymmetries, etc. (cf. Schachter 1984, Keenan 1988), since it has the nominative in its SD, even though it c-commands it. The accusative in the VP-accusative structure will also have the nominative in its SD, though not vice versa, as usual. Similarly, the SD for the dative in German will be expanded to include the grammatically required accusative, which it asymmetrically c-commands.

The anaphoric binding possibilities of these Toba Batak and German structures exhibit an intriguing difference that is worthy of note. In Toba Batak the binding possibilities seem to be completely determined by the (verbal) morphology, and it is as if the syntax didn't exist at all, whereas both appear to be relevant in German. The Toba Batak [[V NP-Nom] NP-Acc] structure permits only the accusative to be a reflexive anaphor to a nominative antecedent, not the other way round, even though the accusative asymmetrically c-commands the nominative (cf. Schachter 1984b; Keenan 1988):

(i) [Di-ida si Torus] dirina
   "see Art Torus self," i.e.
   Torus saw himself.

(ii) *[Di-ida dirina] si Torus
   "see self Art Torus"
The alternative VP-accusative structure, formed with the M-prefix, again permits only the accusative to be the anaphor. The structure is now \[[V NP-Acc] NP-Nom\], and the nominative asymmetrically c-commands the accusative, so this is expected:

(iii) \[\text{Mang-ida dirina\ si Torus}\]
     "see self Art Torus," i.e.
     Torus sees himself.

(iv) *\[\text{Mang-ida si Torus\ dirina}\]
     "see Art Torus self"

Comparing both sets of structures, we see that there is no correlation between grammaticality status and positioning the anaphor \textit{dirina} inside or outside the VP. Syntactic c-command relations do not determine these anaphoric relations. Rather, the NPs designated as accusative by the relevant verbal morphology must be the anaphors, and the NPs designated as nominative must be the antecedents. In our terms, all accusatives will have nominatives in their SDs, regardless of their syntactic positioning, whereas nominatives will not have accusatives in their SDs, and if there is an accusative in the syntactically defined SD of the nominative (in the VP-nominative structure) it will be systematically removed.

In German, on the other hand, the rules of anaphoric binding seem to be sensitive to both morphology and syntax, as argued in Primus (1989). If the dative is a reflexive anaphor and the accusative is its antecedent in \[\text{NP-Dat [NP-Acc V]}\], the sentence is ungrammatical:

(v) *Ich habe \[\text{sich selbst [den kranken Mann im Spiegel gezeigt]}\]
     "I have himself the sick man in-the mirror shown," i.e.
     \[\text{DAT \quad ACC}\]
     I showed himself the sick man in the mirror. (i.e. the sick man to himself)

This is predicted by syntactic c-command alone. What is not predicted by c-command is that (vi), in which the dative is the antecedent and the accusative the anaphor, is not fully grammatical either, even though the anaphor does not c-command the antecedent:

(vi) ?Ich habe \[\text{dem kranken Mann [sich selbst im Spiegel gezeigt]}\]
     "I have the sick man himself in-the-mirror shown"
     \[\text{DAT \quad ACC}\]

In our terms, this could be accounted for by saying that the dative extends its SD to include the grammatically required and co-occurring accusative. Since the accusative is an anaphor in (vi), the dative antecedent thereby acquires an anaphor in its SD, which is disallowed (by the Principle of Structural Determination in (2.36) below). The accusative anaphor, on the other hand, contains the dative antecedent within its SD defined syntactically, which is permitted. In (v), however, the accusative antecedent contains an anaphor in its SD defined syntactically, which explains the ill-formedness, even though the dative anaphor contains a (permitted) accusative antecedent within its extended SD. These examples indicate that both syntactic configuration and
morphology contribute to well-formedness in German, as Primus (1989) shows, and they suggest that an antecedent–anaphor pair will be fully grammatical in this language only if (a) the anaphor contains the antecedent in its SD (defined syntactically and/or morphologically) and (b) the antecedent does not contain the anaphor in its SD (defined syntactically or morphologically). Both conditions are met in (vii), for example, which is fully grammatical:

(vii)  Er hat [den kranken Mann [sich selbst anvertraut]]
     "he has the sick man himself entrusted," i.e.
     \[\text{ACC DAT}\]
     He entrusted the sick man to himself.

The fact that (v) is judged more ungrammatical than (vi) appears to reflect the fact that SDs defined syntactically have priority in German binding rules. Even though they may be fine-tuned morphologically, the violation of a binding condition involving a syntactically defined SD results in a much worse ungrammaticality than one involving a morphologically constituted SD.

The clear difference between Toba Batak and German can be seen in the contrast between (i) and (v). (i) is grammatical in Toba Batak because the c-commanding anaphor acquires a morphologically signalled antecedent in its SD, Torus, while the anaphor is removed from the syntactically-defined SD of Torus. In short, the structural domains for the grammar of anaphora in this language are entirely morphologically constituted. In German they are both morphologically and syntactically constituted, with the syntax having priority. The last remaining major possibility is for SDs to be entirely syntactically constituted, and for the morphology to be ignored. This happens in most languages with ergative–absolutive morphology; cf. n. 7 ch. 2.

2 Since the precise role of morphology and topology in constraining SDs in different languages will not be a major focus of this book, the definition of (2.11) is kept deliberately general. The reader is referred to Primus (1987) and (1991a) for a detailed exposition of how morphology and topology can impact structurally sensitive grammatical operations, such as antecedent-anaphor relations across languages.

3 The percentage total for the relative clauses in (2.16) comes to 90%. The remaining 10% cover locatives with a where relative pronoun, e.g. the place where, temporals with a when, e.g. the time when, and a miscellaneous group.

4 The percentages of (2.17) are derived from the data of table 3.1 in Keenan and S. Hawkins (1987: 70). I have omitted relatives formed on objects of comparison both in section 2.2.1 and in 2.3, for the reasons discussed in n. 6 ch. 2. Keenan and S. Hawkins also tested relative clauses formed on derived subjects, i.e. the subjects of passive predicates, and on genitives in derived subjects. These relative clauses have also been omitted here. All the subjects referred to in (2.17) are therefore undervived subjects, i.e. NPs that have maintained their subject role throughout the derivation.

5 If the difference between peaks of complexity is small across two sentences, then this obviously reduces the significance that any such difference can have in explaining the acceptability differences between them. For example, the initial three-word sequence in the sentence That John washed the dishes surprised Mary has a non-terminal-to-terminal ratio of 6\(\frac{2}{3}\) according to Frazier's metric, whereas the same three-word sequence has a 5/3 ratio in the extraposed version It surprised Mary that
John washed the dishes. (For a summary of Frazier’s major assumptions in calculating ratios, cf. the discussion of (2.18)–(2.20) in the main text.)

(i)

\[
\begin{array}{cccccc}
S \\
S \\
NP & VP & NP \\
That & John & washed \\
3 & 2 & 1 & 1 & 0 & 1 & 1 \\
\end{array}
\]

\[6\frac{1}{2}\]

(ii)

\[
\begin{array}{cccccc}
S \\
NP & VP & NP & S & S \\
It & surprised & Mary & that & John & washed \\
2\frac{1}{2} & 1 & 1 & 1\frac{1}{2} & 2\frac{1}{2} & 1 & 1 & 0 \\
\end{array}
\]

\[5\]

The difference between \(6\frac{1}{2}\) and 5 is highly significant for Frazier, since it purportedly explains the preference for (ii). Notice, however, that if S and \(\bar{S}\) were counted simply as 1 rather than \(1\frac{1}{2}\) (which is, of course, a rather questionable numerical assignment), the difference in ratios would be just 5/3 versus 4/3, i.e. a difference of 1 only, which is less significant.

Moreover, it cannot be regarded as insignificant that in the other portions of (i) and (ii), i.e. the words other than that John washed, it is the extraposed version that exhibits more complexity than the sentence with a sentential subject, according to Frazier’s criteria, and this added complexity should surely offset the slight advantage for the three-word sequence that John washed in the extraposed sentence. For example, the three words in the matrix clause of (ii), It surprised Mary, have a ratio of 4\(\frac{1}{2}\)/3, i.e. 1.5, whereas the two words of the matrix in (i), surprised Mary, have a ratio of 2/2, i.e. 1.0. More precisely, if we aggregate over all five adjacent three-word viewing windows for the seven words of (i), and over all six for the eight words of (ii), the aggregate ratio for
(ii) turns out to be higher, 1.36 versus 1.13 for (i), which suggests, counterfactually, that (ii) should be more complex.

All of these considerations suggest that Frazier's metric is not the correct account of the preference for (ii) over (i) in performance. An alternative approach to Extraposition will be presented in ch. 3 and will be tested empirically in ch. 4.3.4.

6 The lowest position on the Keenan–Comrie AH, objects of comparison (OCOMP), is ignored in this context, for two reasons. The actual syntactic expression of OCOMP is highly variable across languages, as Stassen (1985) has shown, and the predictions made by my theory will depend on the precise coding device employed. I have not examined the coding device of each language in relation to its other relativization options, and moreover Keenan and Comrie's own data have many gaps for this position, which means that they lack relevant information. In the present context it is therefore safer to concentrate on the clearer syntactic positions of AH, i.e. all but OCOMP, and to draw our theoretical conclusions from these.

7 There is an interesting partial discrepancy between Primus's hierarchy for ergative–absolutive languages and the basic grammatical relations of Keenan and Comrie's AH. As is well known, the ergative case is reserved for subjects of transitive verbs only, whereas the subjects of intransitives are assigned the absolutive case, as are the direct objects of transitive verbs. The absolutive case therefore subsumes both subjects and direct objects, the ergative case subjects only. For these languages Primus (1987, 1991a) has proposed a hierarchy on which the absolutive is placed higher than the ergative; cf. also Croft (1990: 104–105). Hence, the absolutive corresponds to the nominative on the Nominative–Accusative hierarchy, and the ergative to the accusative. One piece of evidence for this involves zero coding, i.e. the absence of overt case marking. If any argument is zero-coded in a nominative-accusative language, it is always the nominative, the highest ranked case on the hierarchy. Similarly, in an ergative-absolutive language, it is the absolutive argument that favors zero coding. For example, ergative arguments in Yup'ik are marked with an -m suffix, absolutes with -∅ (cf. Reed et al. 1977). In both case-marking systems there is also a correlation between zero marking and frequency of occurrence. Intransitive subjects are much more frequently occurring than transitives; hence intransitive subjects are more frequent than either transitive subjects or transitive objects. Nominative arguments are accordingly much more frequently occurring than accusatives, and absolutes much more than ergatives. The positioning of absolutes above ergatives on a morphological hierarchy also makes sense in terms of our theory of structural complexity and Structural Domains. The presence of ergative case marking signals that the accompanying predicate is transitive and that there will be an absolutive DO, i.e. the minimal SD for the ergative requires a co-occurring DO. An absolutive, on the other hand, does not necessarily require any additional arguments within its minimal SD. Similarly, accusative marking signals transitivity and an accompanying nominative, but a nominative does not actually require accompanying arguments in its minimal SD. In both cases, the highest position on these hierarchies makes no co-occurrence requirements, and the next position down requires the co-occurrence of the highest position. What differs is the selection of the transitive argument that is to be morphologically united with the intransitive subject. The very existence of both systems has been given a semantic
motivation in Keenan (1984). But it is well known, of course, that there are numerous syntactic processes in almost all ergative-absolutive languages that ignore surface morphology and that treat ergative-marked NPs and intransitive subjects as a unit: cf. Comrie (1978), Dixon (1979). These languages therefore assign priority to syntactic over morphological criteria in their rule formulations, and this results in the discrepancy between morphology and syntax, and derivatively in the discrepancy between grammatical relations defined syntactically and the morphological hierarchy for ergative-absolutive languages that places a mixed case for subjects and objects above a case for subjects only (the ergative). The only point I wish to stress here is that the relative ranking of positions on the ergative-absolutive hierarchy makes sense in terms of our complexity theory, just as the relative ranking of positions on the nominative-accusative hierarchy can be explained in this way. All of Primus's hierarchies appear to be explainable, though what remains to be explained is the manner in which languages resolve conflicts between them, e.g. by assigning priority to syntactic over morphological properties in ergative-absolutive languages.

8 The full predictions that I would propose on the basis of structural complexity and processing ease for the grammar of relativization are the following, therefore:

(i) if a language permits relativization (using any strategy) on a low position on AH, it permits it on all higher positions; if any language has a gap in its relativization positions, then it must have an obligatory promotion rule converting the position that is not relativizable into a higher one that is (cf. the data of (2.22) in the main text);

(ii) if a language permits relativization using a pronoun-retaining strategy on a high position on AH, it permits it on all lower positions that the language permits relativization on (cf. the data of (2.27)).

A pronoun-retaining strategy is a particular instance of a [+ Case] relative clause, one which contains a nominal element that unequivocally expresses which NP position is being relativized on. If pronouns facilitate the task of semantically interpreting the role of the head within the relative, then other [+ Case] strategies should do so as well, hence:

(iii) if a language permits relativization using a [+ Case] strategy on a high position on AH, it permits it on all lower positions that the language permits relativization on.

Conversely, we expect that a [- Case] strategy, which does not facilitate semantic interpretation in this way, should be more difficult to process as we descend the hierarchy:

(iv) if a language permits relativization using a [- Case] strategy on a low position on AH, it permits it on all higher positions.

Prediction (iii) is well supported in Keenan and Comrie's (1977) data; there is just one exception, as shown:
[+ Case] strategies

I exception (Tongan: SU and IO and OBL and GEN, not DO)

SU and DO and IO and OBL and GEN: Hindi, Slovenian, French
DO and IO and OBL and GEN: Arabic (Classical), Hebrew, Chinese
IO and OBL and GEN: Batak, English, Hausa
OBL and GEN: Fulani, Minang-Kabau, Roviana
GEN: Japanese, Turkish

Prediction (iv) is exceptionlessly supported:

[– Case] strategies

SU only: Malagasy, Maori, Arabic (Classical)
SU and DO only: Fulani, Hausa, Finnish
SU and DO and IO: Basque, North Frisian
SU and DO and IO and OBL: Korean, Turkish
SU and DO and IO and OBL and GEN: Japanese

These predictions are well supported and are stronger and more restrictive than the predictions made by Keenan and Comrie themselves. Moreover, these predictions are now theoretically motivated, by the complexity theory that underlies AH. Prenominal versus postnominal order is not necessarily significant from the point of view of Relativization Domains and their structural complexity; nor is the fact that some strategy happens to be used for subjects (the primary one). But [+ Case] is significant since [+ Case] eases performance difficulty, whereas [– Case] does not. Hence, we formulate our predictions on strategies that are theoretically significant from the perspective of complexity (from the rich set of imaginable possibilities, cf. the discussion in Maxwell 1979, Comrie and Keenan 1979). As a result, we limit the application of Keenan and Comrie's Hierarchy Constraint 3 (“strategies that apply at one point may in principle cease to apply at any lower point”). [+ Case] strategies work in the opposite direction: they are productive in lower positions and may cease to apply in higher ones. We keep this prediction for [– Case], however, and for relativization in general. Finally, we remove altogether their Hierarchy Constraint 2, the continuous segment prediction (“any RC-forming strategy must apply to a continuous segment of the AH”). In the present context this would allow a [– Case] strategy to apply to e.g. IO and OBL and not to SU and DO, or a [+ Case] strategy to apply to e.g. DO and IO and not to OBL and GEN. Empirically, the continuous segment prediction does hold for separate strategies as Keenan and Comrie define them, but it is questionable whether this prediction actually follows from an implicational hierarchy, since it allows for strategies to apply to some implicational antecedent positions on AH, and not to a consequent position, thereby going against what is usually understood by an implicational dependency.

9 Cf. Reinhart (1983: 188–198) for a discussion of some apparent counterexamples to the claim that a quantified expression QP_i cannot be under the scope of another quantifier phrase QP_j if QP_i asymmetrically c-commands QP_j. In our terms, QP_j would have QP_i in its SD, and QP_i would not have QP_j in its SD, hence QP_j would be interpretable as being under the scope of QP_i, but QP_i could not be under the scope
Both our analysis and Reinhart's define the same prohibition on quantifier scope interpretations under these circumstances, and hence the kinds of considerations she discusses in order to preserve her analysis are equally relevant in this context. Reinhart argues that apparent counterexamples of this sort can generally be explained on the grounds that there is an entailment relationship between the two quantifier scope readings, or because the apparent ambiguity is really a case of logical vagueness. On other occasions her proposed tree structures do make the right predictions. These arguments support the proposed correlation between syntactic and semantic dependencies for which I am arguing here, using the notion of a Structural Domain rather than c-command. The two notions are similar except that the SD of X in C consists of nodes that dominate X as well as nodes that c-command it.

3 Early Immediate Constituents

1 The numbering of nodes, \( S_1, S_2, NP_1, NP_2, \) etc., requires comment. Here and throughout this book I will generally number nodes only if there is more than one relevant occurrence in a given tree. The numbering will then reflect the order in which these nodes are constructed in the left-to-right parse. Thus, in (3.1'a) the matrix \( S_1 \) is recognized before the subordinate \( S_2 \), etc.

2 The trees (3.1'a) and (3.1'b) incorporate fairly traditional assumptions about phrase structure. The performance theory of this book does not, in general, depend upon the syntactic assumptions of any one current model, and I would like my results to be as compatible as possible with different models. All things being equal, I prefer flatter structures to rich phrase-internal branching, and I do not postulate numerous (non-branching) projections of head categories. Such devices add greatly to the complexity of phrase structures, as defined here (cf. (2.12)), and my theory defines a clear preference for the most minimal possible trees. Some empty categories are assumed here (cf. e.g. (2.15) and (2.29) above), but in the context of word order the main emphasis will be on the ratio of immediate constituents to (phonetically realized) words, making intermediate structure irrelevant, including any empty categories. Empty categories will therefore be ignored in this context.

3 Kimball's (1973) New Nodes principle was defined as follows: the construction of a new node is signaled by the occurrence of a function word. The class of constructing categories is rather broader than this, however, as discussed in the main text.

4 Here and in (3.4) I wish to remain as neutral as possible between competing syntactic theories of grammatical competence. For this reason I refer in these parsing principles to the traditional notion of a Phrase-Structure Grammar. I do not wish to imply by this that there are not significant improvements in more recent conventions for defining phrase structure, licensing conditions, etc., or that there are no significant differences between the proposals of Government-Binding Theory or Generalized Phrase Structure Grammar. My purpose in this context is to argue that fundamental properties of ordering and constituency have their explanatory origin in language processing, regardless of differences in the precise format in which they are described, and I do not wish to become embroiled in issues of grammatical theory whose significance lies elsewhere. All models have to define linear ordering in some
form, and they all have to generate constituent structures, and I claim that fundamental aspects of these two grammatical phenomena are determined by processing.

5 In ch. 6.2.5 this axiom will be revised and generalized, after a detailed discussion of additional parsing principles such as Grandmother Node Construction. The more general axiom will be referred to as the axiom of Constructability.

6 In this constituent structure for Japanese I assume the existence of a VP node, following the grammatical arguments of Saito (1985) and Saito and Hoji (1983). There is also performance evidence for the reality of a VP in our Japanese data (cf. ch. 4.1.5). For discussion of a possible analysis of (3.9b) in which S is discontinuously attached to the VP, see section 3.4 below.

7 The numbering of the S nodes differs in (3.9a) and (3.9b) because the relative order in which these nodes are constructed differs (cf. n. 2 ch. 3). In (3.9b) the subordinate S is recognized first and is labeled S₁, the matrix S being S₂. But in (3.9a) the nominative-marked Mary-ga is a constituent of the matrix S, and hence a daughter of the matrix precedes daughters of the subordinate S. I believe that this nominative marking is sufficient to construct the matrix S, by the principle of Grandmother Node Construction (cf. ch. 6.2.2), and so the matrix is constructed first and is labeled S₁, the subordinate S being S₂.

8 This use of words alone as an indicator of the amount of non-IC structure in a CRD will be more problematic in languages that have very rich word-internal structure, i.e. incorporating and polysynthetic languages (cf. Comrie 1989: 45 for a definition and illustration of these terms). If a single word can house morphological material that is indicative of several nodes within a syntactic tree, then words alone become a less reliable quantificational unit for non-IC structure, since some words in a text will presumably be very rich in structure, while others will not. Predictions for alternative orderings will then depend crucially on the internal composition of these words, rather than simply on their number. The languages in my performance sample (cf. ch. 4) are not, in general, incorporating and polysynthetic. Where there are rich compounding processes, some attempt has been made to reflect this in the IC-to-word ratios by assigning to certain morphemes the status of separate words, e.g. various morphemes within verb compounds in Japanese and Korean, and case particles in these languages. In this way the amount of non-IC structure is more aptly reflected in the “word” count. Ultimately, the IC-to-non-IC ratios for the words and phrases of a text will need to be calculated on the basis of precise assumptions about what the tree structure for each word and phrase is. When this is done, differences between languages with regard to the amount of permitted word-internal structure will be less significant, and the same quantificational measure and predictions will be appropriate for languages of all morphological types.

9 In these ordered sets representing a CRD, e.g. VP(Mary-ga, it-ta), the subscripted categories S and V are attached to the words that actually construct these categories within the CRD. Since Mary-ga does not construct an IC within the CRD – indeed it is not even dominated by VP at all – it has no subscript attached to it.

10 This left-to-right calculation procedure is a revision of the one that was used in Hawkins (1990). As I pointed out in that paper (p. 233), most alternative CRDs for the orders I am comparing can be distinguished without doing the calculations on a
left-to-right basis, since the actual number of non-ICs or words varies within the respective CRDs. As a result IC-to-non-IC (or to-word) ratios vary as well, and there is no need to do the more elaborate calculation all the time. The left-to-right procedure is therefore generally reserved in the present work for those structures in which finer discriminations are needed among CRDs that would otherwise be identical according to the simpler procedure, as illustrated in the main text. The precise procedure of (3.15) differs from that in Hawkins (1990) in that it no longer involves calculating a (left-to-right IC-to-word) percentage for each word in the CRD, but only for each IC, as illustrated in (3.16'). In this way the new procedure can readily compute ratios for ICs to non-ICs as well as for ICs to words. It also removes a somewhat dubious aspect of the previous calculation procedure that was pointed out to me by Philip Lieberman (personal communication), namely: if an IC is recognized on its right periphery, e.g. [S Comp] in (3.16b) in the main text, then the previous method began calculating the IC-to-word ratios for all daughters of this S, 2/2, 2/3, etc., even in advance of the parser's recognizing that these words were contained within the relevant IC, here S. This was defended on various grounds in Hawkins (1990: 239-40, n. 4); for example, it correctly quantifies the processing load at the time that the relevant constructing category is encountered in the on-line parse. But it was an unattractive corollary nonetheless, and the arguments that motivated it are now, I believe, better captured in the definition of (3.15).

11 I am grateful to Erica Garcia (personal communication) for reminding me over the years that there is typically an explanatory step missing in linguistic discussions of structural options and their attested frequencies. Why exactly do the alternants have the frequencies they do? The answer I am going to propose and test for word-order alternations goes as follows. The vast majority of attested word orders are those that are easiest to use, i.e. more efficient according to EIC than all other competing orders with their respective categories and weights. For the small minority of orders that are non-optimal, EIC ratios are still predicted to be good, and my claim is that it doesn't make a whole lot of difference in these cases whether the optimal or the nearly optimal order is chosen. As the differential increases between optimal scores and their non-optimal counterparts, however, it does make an increasing difference. The processing load of the non-optimal structures is increasingly dispreferred, and the calculation that the speaker must make in real time (unconsciously) when choosing between alternatives becomes more clear-cut: differences between alternatives are more readily perceptible and calculable, and more is at stake. Hence, text frequencies follow, on this account, from a definition of optimality and structural simplicity (in processing), coupled with the assumption that non-optimal alternatives will occur only if not much is at stake, and in proportion to the closeness of fit to the corresponding optimum.

12 The texts on which I am going to test the EIC Text Frequency Prediction in ch. 4 are all written texts. This prediction applies equally, however, to transcripts of spoken data. Spoken data will be more complicated to test, because of false starts and interruptions in CRDs, and also because longer and more complex phrases will be less numerous than in written texts. Nonetheless, I make the same predictions for them. Spoken data will be shaped by EIC, just like written data. If anything, I expect that spoken data will be more (or equally) in conformity with EIC, because the
processing load associated with less efficient CRDs will not be helped by the additional source of visual information that comes with written material. Hence, if we can establish the correctness of EIC on the basis of written materials, we should be able to establish it at least as well, and possibly even more convincingly, on the basis of spoken data.

13 In fact, there are three examples in my data (out of a total of 417) in which PP $\geq$ NP that have the rearranged order [V PP NP]. These rearrangements are not accounted for by EIC. Notice, however, that they account for only $3/417 = 0.72\%$ of the data. Cf. table 4.22 for a tabulation of all these Heavy NP Shift data.

14 This separation of linear precedence rules from the rules defining immediate domination is an important feature of a number of grammatical models, including Generalized Phrase Structure Grammar (cf. Gazdar et al. 1985, and Hawkins 1985: 578 for a brief summary of the literature).

15 Hale (1992) argues that there is a grammaticalized basic word order in Papago, despite its considerable word-order freedom. Tonal phrasing, determiner allomorphy, and extraction rules all motivate a head-final order within all phrases at d-structure, and hence a basic OV order within the clause. He can find no comparable arguments in Warlpiri, however, to justify a basic order within that free-word-order language.

16 Mithun (1992) makes the important point that many languages may not have a grammaticalized basic order at the sentence level, and that this has significant consequences for typological classification and for predictions made in terms of basic orders. She discusses three languages in detail, Cayuga, Ngandi, and Coos, all of which have considerable word-order freedom, and argues that order is driven by pragmatic principles in these languages, not by syntax. What I am suggesting instead is that order is driven by syntactic recognition, in all languages including these. There is certainly variation, within and across languages, in the extent to which there are grammaticalized basic orders and in the degree of word-order freedom (cf. n. 17 ch. 3). But I believe that we should re-examine the performance data from these languages, in order to test the predictions of EIC and to compare them with the predictions of different pragmatic theories, as is done for some of the languages in my sample in ch. 4.4. This will require the formulation of explicit, and independently motivated, proposals about phrase structure. It does not necessarily entail buying into any one current grammatical model in toto. But alternative orders affect the sequence in which structural decisions can be made in on-line processing, and this could be, at the very least, a major determinant of ordering alongside pragmatic principles, even if one does not accept that these latter are purely epiphenomenal, as I shall argue here.

There are some hints in the literature on free-word-order languages that syntactic processing is a much more pervasive influence on language performance than is currently recognized. For example, Mithun (1992: 50) points out that heavy constituents can be postponed in Coos to prevent them from "blocking the flow of discourse." She gives the following example (taken from Frachtenberg 1913: 16):

(i) \textit{MiL halt! e$^E$ne xle$^I$tc e$^E$L$^a$s teq xL$^e$yis.}

"please now thou with-it-with speak-2nd Pers this-my with-language," i.e.

Please now speak my language.
The constituent corresponding to "my language" is postposed behind the verb, and is presumably co-indexed with the (shorter) cataphoric pronoun before the verb. (This structure is reminiscent of German patterns such as Ich bin davon überzeugt, daß . . ., "I am of it convinced that," i.e. I am convinced that.) One could view this postposing as a discourse device. But it does not just facilitate the flow of discourse. If there is a VP in Coos, the recognition domain for this phrase is being shortened by bringing forward the verb in the left-to-right parse string (just as it is in the German example). If there is no VP, the CRD for S will be shortened. There is an alternative performance motivation for these data, therefore, and we need to know to what extent efficient syntactic recognition could account for orderings that are currently being attributed to pragmatics alone. Mithun (1992: 57) points out another suggestive detail, from both Coos and Cayuga. There are nominalizing particles in these languages (lE/hE in Coos, ne' in Cayuga) that can precede simple NPs containing nouns, as well as more complex nominalizations of other phrases. The following example is from Cayuga:

(ii) akaqnihnató:k ne' nóne': ne:kye ne' kowiyáqtatre
    "they-noticed the you-know this the she-is-getting-a-baby," i.e. They noticed that she was expecting.

These particles are fixed in their positioning, and Mithun points out that "if these particles floated throughout sentences, their function as cues to the roles of following constituents would be compromised." In our terms, ne' in Cayuga is a mother-node-constructing category, constructing NP. There are plausibly two ICs in the matrix clause of (ii), the verb and NP. The positioning of the italicized ne' guarantees that these two ICs are recognized in an optimally efficient two-word sequence. The second ne' constructs an NP within the embedded clause and appears to be equally efficiently positioned. These kinds of details suggest that EIC may be just as operative in these languages as it is in the languages of the present study.

The hypothesis in Dryer (1989b) is also very suggestive in this context. Dryer argues, albeit tentatively, that the most frequent clause-level order (OV or VO) does predict other implicationally related word-order characteristics (within the NP and the PP, etc.), even though there may not be a basic grammaticalized order in the clause, and even though orders at this level may be discourse-conditioned. This correlation makes no sense, of course, if there are two quite distinct determinants of order, pragmatic and grammatical. It makes perfect sense, however, if there is just one, EIC. For whether a particular order is grammaticalized or not within the clause, we expect that the selections made in performance will be harmonic with the orderings that have been fixed in other categories, so that a predominance of OV over VO will indeed co-occur with GenN, postpositions in PP, etc., as he suggests (cf. Dryer 1989b: 85).

17 Notice that all languages have at least some fixed word orders (even in Australia, cf. Blake 1987 for an overview and Bowe 1990 for a case study), and all languages have at least some free word orders (even in English, cf. ch. 5.3.3). Ordering within the sentential nodes, S and VP, is generally freer than in the more referential nodes, NP and PP, so if there is considerable freedom in the latter, there will be considerable freedom in the former. One might explain this by saying that it is the sentential nodes that dominate referential constituents to which pragmatic theories
apply, whereas they do not apply within referential nodes, and hence more freedom is expected in the former. The trouble with this is that one then has no way to explain the ordering alternations that do occur among the daughter ICs of NP, e.g. in Extraposition from NP structures (cf. ch. 4.3.5). More generally, I believe that the explanation for the greater word-order freedom at the clause level has to do with the greater number of ICs within S and VP, on average, coupled with the intrinsic variability in the weights of these ICs, for expressive reasons. This then results in considerable EIC-induced ordering variation. If there are only regularly two ICs in a PP, for example, one of which is a single-word P, then there is always going to be an optimal ordering, \([P_m NP]\) or \([NP_m P]\), as long as NP is consistently recognized on one or the other periphery. The potential for free ordering is much less, therefore, within the less complex categories, NP and PP, than it is within the larger and more complex VP and S, dominating more material.

18 I do not wish to imply by this that discourse considerations will always be less fundamental determinants of, and explanations for, grammatical structure. They may explain the origin of topic and focus nodes, for example. In Hawkins (1978) I argued that there was a chain of explanation from the pragmatic properties of definite referring expressions in English to their corresponding semantic and syntactic properties (cf. Hawkins 1991 for a more recent statement of the same explanatory intuition). And in Hawkins (1988c) I summarized other examples of plausible pragmatic explanations for grammatical phenomena. Ordering is different, however, and I am sceptical about the explanatory role of pragmatics here, for the reasons given in the main text.

19 Behaghel’s (1932: 6) Gesetz der wachsenden Glieder claims, in the German original, that “von zwei Gliedern, soweit möglich, das kürzere vorausgeht, das längere nachsteht,” i.e. of two elements, as far as possible, the shorter precedes, the longer follows.

20 The Relator principles claim that Relators have their preferred position: (a) at the periphery of the relatum with which they form one constituent (if they do so); (b) in between their two relata. For a particularly lucid summary and discussion of these and other linearization principles of Functional Grammar, see Siewierska (1991a).

4 Testing EIC’s performance predictions

1 There are some sequences of two PPs for which it is undecidable, even in context, whether PP₂ is embedded under PP₁ or is a sister of PP₁. These PPs will be grammatically permutable and truth-conditionally equivalent and PP₂ will be interpretable either as a sister of PP₁ or as a modifier of the N within PP₁. An example from my data is (i), which is readily permutable with, and equivalent to, (ii):

(i) The raven slept \(\text{PP₁}[\text{on a perch}] \ \text{PP₂}[\text{behind the back door}]\)
(ii) The raven slept \(\text{PP₂}[\text{behind the back door}] \ \text{PP₁}[\text{on a perch}]\)

but which is also analyzable as (iii), with PP₂ functioning as a complement of perch:

(iii) The raven slept \(\text{PP₁}[\text{on a perch}] \ \text{PP₂}[\text{behind the back door}]\)
These cases were not distinguishable, according to my operational test, from those in which PP\(_2\) can only be analyzed as a sister of PP\(_1\), either at all or in context, i.e. cases such as *The raven flew [through the barn] PP\(_2\)[into the open air]*, discussed in the main text, which is not analyzable as *The raven flew PP\(_1\)[through the barn PP\(_2\)[into the open air]]*. Since the permutable analysis is always available for *The raven slept on a perch behind the back door*, I counted these cases as [PP\(_1\) PP\(_2\)] rather than PP\(_1\)[NP[N PP\(_2\)]]

The justification for this is as follows. If the lower attachment should actually be preferred on-line in some cases, according to EIC's Performance Attachment Prediction (3.20) for temporary ambiguities (cf. chs. 3.5.3 and 4.5), then it would be quite possible for the embedded PP\(_2\) to be either longer than or shorter than the remaining material within the higher PP\(_1\), and to have optimal ratios in either case. Hence, by analyzing these structures as sisters only, we conduct a more critical test of EIC. If there is still a pronounced preference for a shorter PP\(_1\) before a longer PP\(_2\), despite the fact that there will be a subset of the structures that admit of two analyses, one of which prefers short before long ([PP\(_1\) PP\(_2\)]), while the other can be optimal with both short before long and long before short ([PP\(_1\)[NP[... N PP\(_2\)]]]), then EIC's predictions will have been fulfilled despite certain motivated exceptions. We will see in table 4.2 that EIC's predictions are well supported, and that they are not adversely affected by this procedure. If anything, the success rate should actually be higher than that shown in the table.

In future textual and experimental testing of this structure, it would be desirable to eliminate PP sequences from the data that are ambiguous between (i) and (iii) above, in addition to removing all the non-permutable, non-truth-conditionally-equivalent PPs illustrated in the main text. All the structures remaining would then be of the type *The raven flew [through the barn] PP\(_2\)[into the open air]*, i.e. those with permutable, truth-conditionally-equivalent alternations, that are unambiguous with respect to whether PP\(_2\) is embedded under PP\(_1\). Only the different possible attachments under VP or S would then remain at issue, which is largely irrelevant for EIC since the same predictions are made in the two cases, as discussed in the main text. For discussion of the relationship between EIC's linear ordering predictions and temporary ambiguity/garden-path avoidance in these data, see further ch. 5.3.3.

2 The subscripts i and j attached to PP are used to distinguish the PP that occurs first in the linear string (i) from the one that occurs second (j). These different orderings are also associated with different structural attachments in (4.1c) and (4.1d), though not in (4.1a) and (4.1b). The numbered subscripts, PP\(_1\) and PP\(_2\), are used in this chapter and throughout the remainder of this book to refer to the length difference between the two constituents, regardless of their positioning. Specifically, PP\(_2\) is the PP that is greater than, or equal to, PP\(_1\) in syntactic weight.

3 The precise figures for [(AdvP)\(_m\) NP V\(_m\) NP (XP)] and [(AdvP)\(_m\) NP \(_m\) NP V (XP)] collected by Stephen Matthews can be summarized as follows. In the former case (one NP preposed), only 172/232 = 74\% of all orders would satisfy the (short-before-long) unmarked case prediction, if this were to be applied to these cases, which is significantly lower than the 91.4\% in table 4.4. The marked case predictions would also receive only partial support: 39\% of orders are non-optimal for one-word differentials; 23\% are non-optimal for two-word differentials; but 43\% for three words! For the structure in which both \(_m\)NPs are preposed, only 42/63 = 67\% would satisfy the
unmarked case prediction. For the marked case prediction, 49% of orders are non-optimal for one-word differentials; 45% for two words; and 0% for three words. These figures clearly indicate that the preverbal domain is very different in structure from the postverbal domain, and plausibly in content as well.


5 I am grateful to Charles Kim and Kaoru Horie for suggesting to me the appropriateness of this consistent structural analysis for Japanese (and Korean), and for investigating its consequences in their respective data.

6 I assume throughout these calculations that the different constituent structures of (a)–(d) will all be necessarily recognizable in parsing, and that the CRDs for S and VP will proceed from the word that constructs the first IC to the word that constructs the last IC of each, as usual. What can potentially vary is the precise time course in which the VP node is constructed in the on-line parse. I assume for present purposes that VP is constructed by V, even though it will be argued in the main text that some daughters of VP are unique to VP and will construct it in advance of V. Nonetheless, other daughters are not so unique. In table 4.7’ some additional calculations will be run for VP structures that are plausibly recognizable in advance of V. In this context notice only that VP is always constructable by V, whether or not it is also recognizable by an additional daughter of VP, and that the preference for long before short ICₘₛ remains in either case. In section I, for example, the scores for the (a) and (b) structures are completely unaffected by whether VP is recognized in advance of V or not, and the long-before-short orderings preserve their 17% preferences in the CRDs for S and VP respectively. For Ic and Id the calculations would be as follows, assuming that the VP-dominated ICₘ constructs VP on the basis of a right-peripheral case particle such as -o and in advance of the verb:

1 ₁ICₘ = 3 ₂ICₘ = 5

<table>
<thead>
<tr>
<th></th>
<th>S ICₘ V</th>
<th>VP ICₘ V</th>
</tr>
</thead>
<tbody>
<tr>
<td>IC</td>
<td>S CRD: 2/4 = 50%</td>
<td>S CRD: 2/6 = 33%</td>
</tr>
<tr>
<td></td>
<td>VP CRD: 2/2 = 100%</td>
<td>VP CRD: 2/2 = 100%</td>
</tr>
<tr>
<td></td>
<td>Aggregate = 75</td>
<td>Aggregate = 67</td>
</tr>
<tr>
<td>Id</td>
<td>S[VP[ICₘ] ₁ICₘ V]</td>
<td>S[VP[ICₘ] ₂ICₘ V]</td>
</tr>
<tr>
<td></td>
<td>S CRD: 2/4 = 50%</td>
<td>S CRD: 2/6 = 33%</td>
</tr>
<tr>
<td></td>
<td>VP CRD: 2/5 = 40%</td>
<td>VP CRD: 2/7 = 29%</td>
</tr>
<tr>
<td></td>
<td>Aggregate = 45</td>
<td>Aggregate = 31</td>
</tr>
</tbody>
</table>

The preferences are still for the longer ₂ICₘ to precede the shorter ₁ICₘ in each case, and with a degree of difference that is now almost identical to the differentials in the flat structures Ia and Ib. Hence, the long-before-short preference holds constant across each of these parallel structures, whether VP construction precedes V or takes place at V.
7 The calculations of table 4.7' are based on the assumption that the case particle -o can construct VP in the structures in which NPo is VP-dominated, i.e. $s[NPga \backslash V]$ and $s[VP[NPo] NPga \backslash V]$. In analyses that permit discontinuity, NPo will always be able to construct VP. In analyses that do not, an NPo that follows NPga will be able to construct VP (since it signals the onset of this higher node), but an NPo that precedes will not, since it is no longer assumed to be VP-dominated in analyses (3) and (4). It is important to point out that the predictions we will derive from table 4.7' in table 4.8 section II do not depend crucially on the assumption that -o can construct VP in advance of V. In other words, if V alone constructs VP, as illustrated in table 4.7, very similar predictions are made, as discussed in the main text. In fact, analyses (3) and (4) come out worse on this assumption, while the flat structure and the discontinuous analyses do equally well. Since individual IC-to-word ratios are affected if V alone constructs VP, I illustrate the differences in scores in table 4.7'' (next page).

8 The precise predictions made by the four analyses of table 4.8 section II on the assumption that V alone constructs VP (cf. n. 7 ch. 4 and table 4.7'') are as follows:

**EIC predictions (1):** as in table 4.8 section II, 82% success rate in unmarked case (70/85 orders = most optimal);

**EIC predictions (2):**
Optimal orders: X where NPga $\geq$ NPo  
Y where NPo $\geq$ NPga  
82% success rate in unmarked case (70/85 orders = most optimal);

**EIC predictions (3):**
Optimal orders: X where NPga $>$ NPo $:+3$  
X or Y where NPga $>$ NPo $:+2$  
Y where NPga $>$ NPo $+1$ and NPo $\geq$ NPga  
36% success rate in unmarked case (31/85 orders = most optimal);

**EIC predictions (4):**
Optimal orders: X where NPga $>$ NPo $:+6$  
X or Y where NPga $>$ NPo $:+5$  
Y where NPga $>$ NPo $+1-4$ and NPo $\geq$ NPga  
19% success rate in unmarked case (16/85 orders = most optimal)

The success rates of table 4.8 section II, by comparison, are: 82% (prediction (1)); 95% (prediction (2)); 40% (prediction (3)); and 40% (prediction (4)).

9 It is important to stress, here and elsewhere, that these are illustrative IC-to-word calculations only. Because the number of relevant structures in the data is not huge, I calculate all weights in terms of the difference between two constituents, as discussed in section 4.1.1 of the main text. If there is a four-word differential between $\_2NPga$ and $\_1NPo$, therefore, this could be because NPga = 6 NPo = 2, or NPga = 7 NPo = 3, or NPga = 8 NPo = 4, etc., and each of these word-total assignments will result in slightly different EIC differentials between competing orderings. In any subsequent and more detailed investigations of Japanese, it would be desirable to collect a large enough number of each of these different subsets of word-total
Table 4.7" IC-to-word ratios for Japanese [{NPga NPo}V]

Assume: V constructs VP;  
V = 1 word

<table>
<thead>
<tr>
<th></th>
<th>NPga</th>
<th>NPo</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>NPga = 10 NPo = 2</td>
<td></td>
</tr>
</tbody>
</table>
\[
\text{s[NPga} \text{ VP[NPo V]]} \\
\text{S CRD: 2/4 = 50\%} \\
\text{VP CRD: 2/2 = 100\%} \\
\text{Aggregate = 75}
\]
|   | NPga = 6 NPo = 2 |  
\[
\text{s[NPga} \text{ VP[NPo V]]} \\
\text{S CRD: 2/2 = 100\%} \\
\text{VP CRD: 2/12 = 17\%} \\
\text{Aggregate = 58.5} \\
\text{Agg. differential = (-)16.5}
\]
|   | NPga = 4 NPo = 2 |  
\[
\text{s[NPga} \text{ VP[NPo V]]} \\
\text{S CRD: 2/2 = 100\%} \\
\text{VP CRD: 2/12 = 17\%} \\
\text{Aggregate = 62.5} \\
\text{Agg. differential = (-)12.5}
\]
|   | NPga = 2 NPo = 2 |  
\[
\text{s[NPga} \text{ VP[NPo V]]} \\
\text{S CRD: 2/2 = 100\%} \\
\text{VP CRD: 2/12 = 17\%} \\
\text{Aggregate = 73} \\
\text{Agg. differential = (-)2}
\]

II NPga = 7 NPo = 2

<table>
<thead>
<tr>
<th></th>
<th>NPo</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Aggregate = 75</td>
</tr>
</tbody>
</table>
\[
\text{(2) Aggregate = 61} \\
\text{Agg. differential = (-)14} \\
\text{(3) Aggregate = 66.5} \\
\text{Agg. differential = (-)9.5} \\
\text{(4) Aggregate = 75} \\
\text{Agg. differential = 0}
\]

III NPga = 6 NPo = 2

<table>
<thead>
<tr>
<th></th>
<th>NPo</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Aggregate = 75</td>
</tr>
</tbody>
</table>
\[
\text{(2) Aggregate = 62.5} \\
\text{Agg. differential = (-)12.5} \\
\text{(3) Aggregate = 69} \\
\text{Agg. differential = (-)6} \\
\text{(4) Aggregate = 76} \\
\text{Agg. differential = (+)1}
\]

IV NPga = 4 NPo = 2

<table>
<thead>
<tr>
<th></th>
<th>NPo</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Aggregate = 75</td>
</tr>
</tbody>
</table>
\[
\text{(2) Aggregate = 66.5} \\
\text{Agg. differential = (-)8.5} \\
\text{(3) Aggregate = 75} \\
\text{Agg. differential = 0} \\
\text{(4) Aggregate = 80} \\
\text{Agg. differential = (+)5}
\]

V NPga = 2 NPo = 4

<table>
<thead>
<tr>
<th></th>
<th>NPo</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Aggregate = 66.5</td>
</tr>
</tbody>
</table>
\[
\text{(2) Aggregate = 75} \\
\text{Agg. differential = (+)8.5} \\
\text{(3) Aggregate = 87.5} \\
\text{Agg. differential = (+)21} \\
\text{(4) Aggregate = 89} \\
\text{Agg. differential = (+)22.5}
\]

etc.
assignments to allow the calculations and predictions to be carried out for each. The precise number of words by which O can exceed S in the SOV structure and still remain optimal for EIC compared with the (discontinuous) OSV structure will then vary for the different subsets. What will not vary is the general principle that is involved here, namely that the discontinuous OSV structure will only be optimal when O is significantly longer than S. As a result, EIC will prefer some SOV structures that have a short S before a longer O.

10 This simplifying assumption is supported by the very suggestive distribution of orders in the \(PP_m = NP_o\) column in table 4.10 section II. There are two possible structures in which order X can be optimal, compared to just one for order Y. Matching this, the distribution of X to Y orders is also two to one in favor of X (60 to 31). If \(PP_m\) is assigned with equal frequency to VP or S, then EIC predicts exactly this predominance of X orders. Even if the frequencies are not equal, there are still two possible structures in which the X orders can be optimal, so we expect at least a predominance of X over Y, in general. This is what we find in the corresponding Korean data, cf. table 4.12 section II.

11 The relative frequencies of SVO, VOS, SOV orders, etc., are taken from Siewierska's table 1 (1991b: 110). This table is based on an analysis of 1450 main declarative clauses in her data base, of which 308 contain both an overt subject and an overt object (lexical, pronominal, or sentential). The remaining clauses were either intransitive, impersonal, or transitive clauses with the subject expressed solely by agreement on the verb. The frequencies of table 4.19 are therefore percentages of this total of 308 for each transitive structure type. The mean lengths of each argument in each order are taken from Siewierska's table 10 (1991b: 116), and are based on an expanded corpus of transitive structures, which increases the reliability of each of these means. Her expanded corpus could not be used for the relative-frequency figures, however, since she deliberately collected a larger number of the minority types. Thus the 308 transitive clauses within the 1450-clause corpus are an "undoctored" text sample that can provide an indication of relative frequencies within spontaneous texts. Siewierska's textual sources in Polish are listed in Siewierska (1991b: 138).

12 Here and throughout this book I assume that the rearrangement referred to as "Heavy NP Shift" (whose very existence as a rule of grammar was questioned in ch. 3.5.2) does not change constituent structure. That is, the NP and PP are still VP-dominated when rearranged out of the basic order. This is supported by a standard VP-constituency test, illustrated in (i)-(iv):

(i) They said that he would \( VP [\text{give the very valuable book to Mary}] \) and he did \( VP [\text{give the very valuable book to Mary}] \)

(ii) They said that he would \( VP [\text{give the very valuable book to Mary}] \) and \( VP [\text{give the very valuable book to Mary}] \) he did

(iii) They said that he would \( VP [\text{give to Mary the very valuable book}] \) and he did \( VP [\text{give to Mary the very valuable book}] \)

(iv) They said that he would \( VP [\text{give to Mary the very valuable book}] \) and \( VP [\text{give to Mary the very valuable book}] \) he did
The VP still behaves like a single constituent, and undergoes preposing, whether Heavy NP Shift has taken place or not. Hence the postposed PP is still a VP constituent.

What the precise internal constituency of this VP is, before and after rearrangement, is a matter of debate. Here I assume a simple flat-structure alternation \( \text{vp}[V \text{ NP PP}] \Rightarrow \text{VP}[V \text{ PP NP}] \). Another possibility is \( \text{vp}[V[V \text{ NP PP}]] \), with a binary branching VP-internal structure, alternating with either a discontinuous \( V' \), i.e. \( \text{vp}[V[V \text{ PP V}[\text{ NP}]]] \), or with \( \text{vp}[V[V \text{ PP NP}]] \). Our performance data are, in fact, equally compatible with the flat structure and with the binary branching structure for this grammaticalized order. This is illustrated with the word-total assignments NP = 4 and PP = 2.

\[
\begin{align*}
\text{NP} &= 4 \quad \text{PP} = 2 \\
\text{vp}[V[V \text{ NP PP}]] &\Rightarrow \text{VP}[V[V \text{ PP NP}]] \\
V' \text{ CRD:} &\quad 2/2 = 100\% \\
V' \text{ CRD:} &\quad 2/4 = 50\% \\
\text{VP CRD:} &\quad 2/6 = 33\% \\
\text{Aggregate} &= 66.5 \\
\text{vp}[V[V \text{ NP PP}]] &\Rightarrow \text{VP}[V[V \text{ PP NP}]] \\
\text{VP CRD:} &\quad 3/6 = 50\% \\
\text{VP CRD:} &\quad 3/4 = 75\% \\
\text{Aggregate} &= 87.5 \\
\end{align*}
\]

For this two-word differential, the postposing of the heavier NP is motivated by a higher EIC score both in the binary branching and in the flat structure. Hence, our data do not discriminate between these structural alternatives. All things being equal, our general theory defines a preference for simpler structural groupings with fewer nodes; see ch. 7.2 for further discussion.


14 There seems to be just one minor exception to this. If the VP consists of no more than *to* plus a verb, while the VP comprises *is* \( A_{\text{adj}} \), e.g. *to sing is easy*; and if we assume that the VP is constructed by *sing* rather than *to* (because of the ambiguity of the infinitival *to* with the prepositional *to*); then the ratio for both the S and VP domains in the basic order would be 100%. The extraposed *it is easy to sing* would have 100% for S (2/2), but only 75% for VP (3/4), meaning that there was a preference for the basic order here. However, as soon as the VP acquires even one more daughter, e.g. *to sing songs is easy*, the ratio for S drops to 67% (2/3), while that of the VP in the extraposed order remains at 75%, so that the direction of preference is now the same as for all other cases of the alternation \( S[XP \text{ vp}[is_{A_{\text{adj}}[Adj]]}] \) and \( S[S_{vt} \text{ vp}[is_{A_{\text{adj}}[Adj]} XP]] \).

15 The corresponding German sentences to (4.29b) and (4.30b) in English, in which Maria asymmetrically c-commands sie, are of course grammatical, e.g. *Ich habe Maria, das Buch geliehen, das sie bestellt hat.*

16 Notice that none of the words dominated by an extraposed \( S \) is counted in the ratios for VP in the Extraposition from NP structures. This is because of the discontinuity of the complex NP. There are two ICs of VP, NP, and V, and by the time the V is reached, this domain has been fully recognized. The \( S \) is the last IC of NP, so the CRD for NP proceeds from its first IC, Det, to its last, as usual, and the \( S \) is constructed, as usual, by a left-peripheral complementizer. The extraposed \( S \) is still dominated by VP.
(and by NP), therefore, and this is reflected in the bracketings in our structural representations. But it stands outside the CRD for VP.

17 The E domain in these structures consists of two ICs, NP and S, just as it does in table 4.27. The CRD for E therefore proceeds from the first word to construct NP (Det) to the first word to construct S (Vf). Because Extrapolation from NP is a structure-changing operation, according to the S-attachment analysis, the S domain contains two ICs in the input structure and three in the output. Hence this CRD must now be added to our EIC calculations, because it no longer remains constant. The NP loses a constituent, of course, so it must also be included. Three CDs are now impacted by Extrapolation from NP, in contrast to just two in the discontinuous analysis, as illustrated in table 4.27.

18 Givon (1983, 1988) mentions another factor that affects the predictability of an entity in discourse, in addition to referential distance: the number of semantically compatible referents in the directly preceding discourse (the preceding three clauses in his quantification of this “potential interference” factor). The referential distance measure can be quantified independently of the potential interference count, however, and Givon himself regularly uses referential distance alone as an indicator of relative predictability. This is what we shall do here as well.

19 Cf. the axiom of Constructability, defined in ch. 6.2.5.

5 Testing EIC's grammatical predictions

1 Cf. further the discussion of Dryer’s (1989b) hypothesis in n. 16 ch. 3.

2 In general, such topicalizations front a single constituent only; cf. Hawkins (1986: 178–179) for discussion.

3 Heavy stress on him makes this sentence better; cf. ch. 4.3.1 and 4.3.2 for discussion.

4 Even if we take into account the (largely semantically based) proposals for internal constituent structure in the NP, the short-before-long preference defined by EIC for English-type languages still appears to make the right predictions for IC structures that include non-maximal projections of the head, such as the following: [Det [Num [Adj [N PP]]], or [Det [Num [Adj [N S]]]]. In general, I am sceptical about the reality of such rich phrase-internal syntactic groupings. They add enormously to the complexity of a tree structure, and so should not be postulated unless they absolutely have to be. Yet linguists do postulate them, at the slightest hint of grammatical support, without considering sufficiently whether the facts could not be equally well handled within the context of a flatter constituent structure. Thus, the “structural domains” of relevance to scope and semantic dependencies (cf. ch. 2.4) could ultimately reflect linear precedence or morphological case marking rather than syntactic structure as such. Many of the trees linguists propose have atrocious IC-to-non-IC ratios, and a principle like EIC will prefer any grammatical model that can gain adequate descriptive coverage with the maximum “tree pruning” (to use Ross’s 1967 picturesque phrase). Cf. further ch. 7.2.

5 The CRD for S here proceeds from Mary-ga, which constructs the subject NP, to it-ta, which constructs the VP. The S domain is not initiated in this discontinuous analysis by the complementizer of S, because these elements do not construct VP.
on-line, as the case particle -o was argued to do in ch. 4.1.5. When the verb is encountered, the daughter ICs of VP are recognized and the cost of the discontinuity can be meaningfully assessed within this CRD. But for the processing of the S domain, the calculation proceeds from the first recognizably S-dominated constituent, Mary-ga, to the last, it-ta.

6 The IC-to-word ratio for the center-embedded (5.39), measured left to right and without the VP assumption, is 63.3% for the matrix \(S_1\) CRD. For (5.41a) the aggregated left-to-right ratio for \(S_1\) is 83.3%. The left-to-right aggregate for (5.40a) remains 100%.

7 The reader is referred to Hawkins (1992c: 203–204) for a tabulation of these infinitival VP data in German, whose aggregate length is 3.4 words.

8 Notice that the implicational prediction for PrNMH derived from (3.24) is logically independent of the frequency prediction made by the Basic Order Prediction of (3.21). The frequency prediction is defined on sets of grammars; the implicational prediction of (3.24) is defined on each individual grammar. The former could still be true at the same time that the latter was false, e.g. if there were languages without the predicted co-occurring word orders, even though the total number of languages remained greater or equal down each adjacent position of the hierarchy. If the latter is true, however, and every individual language conforms, then the former must be true also.

9 There are four prepositional languages out of 148 in the Expanded Sample of Hawkins (1983) that combine \{PossP N\} with \{N Adj\}, rather than with the predicted \{Adj N\} (cf. Hawkins 1983: 67). 4/148 (= 2.7%) is significantly less than the 25% that would have been assigned to this combination of \{PossP N\} and \{Adj N\} from among the four logically possible combinations (GenN and AdjN, NGen and AdjN, etc.), if the assignment were random. Hence, “if GenN then AdjN” is a statistically significant implicational generalization in prepositional languages. Cf. Dryer (1993) for a more recent quantification, which confirms the statistical significance of this implication.

10 An exception to this is, for example, the prepositional language Karen (Sino-Tibetan), in which genitives are the only IC daughters of NP to precede N, thereby providing a constituent that intervenes between P and the first MNCC of the NP, N (cf. Hawkins 1983). In this language, therefore, GenN co-occurs with NDem, in violation of the PrNMH in (5.60). Dryer (1993) also cites a small number of prepositional languages in which AdjN co-occurs with NDem (e.g. Sango), which necessitates a conversion of this implication from exceptionless (as it was in Hawkins 1983) to statistical.

11 It would be interesting to examine the handful of prepositional languages that combine \{Adj N\} with \{N Dem\} (or \{PossP N\} with \{N Dem\}, cf. n. 10 ch. 5) to see if these leftmost daughter categories of NP can function as unique constructors of NP by one of the additional construction principles proposed in ch. 6.2. For example, if the adjective is marked for agreement with a noun, then that adjective will be able to construct an NP by Agreement Projection (cf. ch. 6.2.3). Since adjectives can be just as short as determiners and nouns, positioning them on the leftmost periphery of NP can be optimal for EIC ratios within the containing PP CRD, and within the NP itself, if they can uniquely construct NP.
Recall some of the other implicational generalizations and hierarchies discussed in ch. 1.2 and n. 3 ch. 2, e.g. the expanding vowel inventories or basic color terms across languages. These data have been explained in terms of increasing perceptual complexity. There is also evidence for increasing conceptual complexity in many of the semantic markedness hierarchies discussed in Croft (1990). Clearly, increasing complexity is a major source of explanation for implicational hierarchies; see further ch. 7.2.

Another way to resolve the vagueness of the main clause/subordinate clause distinction is to mark main clauses explicitly in some way. I shall argue in ch. 6.2.2 that the Japanese topic marker wa does this, since the occurrence of this marker is limited to main clauses. Those clauses that contain a wa-phrase are unambiguously main clauses, therefore.

Many of these [[Comp S] V] (5.5D) and [[N Š] V] (5.6D) structures will be rearranged, obligatorily or optionally, into their preferred counterparts, such as [V [Comp S]], as we pointed out in sections 5.1.2 and 5.4.1. The dispreferred structures will often be motivated as basic orders in the relevant grammars, and they will remain in performance samples, EIC permitting.

There is yet another logical possibility for VSO languages: they might have the structure s[Vp[V [m S m O]]], in which the V alone is VP-dominated, and subject and object are immediately S-dominated. In the alternating SVO patterns in these languages, both V and m O would be VP-dominated. This analysis differs from the one assumed in table 5.12 in that it avoids a VP-discontinuity, while still assuming a VP. The EIC score for this structure is shown in (i):

(i) \[s[Vp[V [m S m O]]]\]
   S CRD: 3/4 = 75%
   VP CRD: 1/1 = 100%
   Aggregate = 87.5

The aggregate is now 87.5, which is higher than that for SVO (84). Hence, this analysis makes the counterfactual prediction that VSO languages should be more frequent than SVO. Only the discontinuous VP analysis, or the flat-structure analysis without a VP, result in EIC ratios for VSO languages that correctly predict their lesser frequency. Alternatively, at least some of the VSO languages that exist may have structure (i) alternating with SVO (Welsh and Irish?), and there may be independent reasons for why a basic VSO of this type should be less frequent, e.g. the lack of consistency in the domination of m O across different sentence patterns.

If V alone were to construct VP, the scores for SOV, OVS, and OSV would be as shown in (i)–(iii) (with word-total assignments as in table 5.13):

(i) \[s[S_m Vp[O_m V]]\]
   S CRD: 2/4 = 50%
   VP CRD: 2/2 = 100%
   Aggregate = 75
The aggregate for SOV would be lower than it is in table 5.13 (84), and both SOV and OVS would be equally preferred over OSV. Clearly, the assumption of VP construction through O_m or V makes better predictions, and accounts for the considerable preference for SOV.

17 Recall also that in both Japanese and Korean, which were argued to have a VP in ch. 4.1.5 and 4.1.6, the number of OSV structures in performance is extremely small: just 3% of transitive clauses in Japanese, and 4% in Korean. This massive preference for SOV over OSV is most consistent with an analysis in which O and V are still dominated by VP in the OSV structure, as argued by O'Grady (1987) and as illustrated in ch. 4.1.5. The advantage for SOV is not correctly captured if the object is extracted from the VP and attached to S (= clause), since the prediction is then made that SOV will regularly be converted into OSV whenever the object is longer than the subject, yet these rearrangements happen only rarely. For example, compare the EIC scores for s[S_m Vp[O_m V]], s[lvp[O_m] S_m Vp[V]], and s[lvp[O_m] S_m V] when S_m = 2 words and O_m = 4:

(i) \[s[S_m Vp[O_m V]]\]
S CRD: 2/5 = 40%
VP CRD: 2/2 = 100%
70% EIC aggregate

(ii) \[s[lvp[O_m] S_m Vp[V]]\]
S CRD: 2/3 = 67%
VP CRD: 2/4 = 50%
59% EIC aggregate

(iii) \[s[lvp[O_m] S_m V]\]
S CRD: 3/4 = 75%
VP CRD: 1/1 = 100%
88% EIC aggregate

The discontinuous structure predicts no rearrangement to OSV ((ii)), the structure with O_m extracted from the VP predicts rearrangement ((iii)). It is the discontinuous structure that is supported by the data, both in performance and in grammars.

6 Grammaticalized node construction

1 For example, one of Ernst's (1991: 195) arguments involves the distribution of adverbs such as even and mostly in English. Adverbs of this class can be adjoined to the left of any XP, as shown in (i)–(iii):

(i) Vern likes \[NP[even the flowers that his mother planted]\].
(ii) Vern may like the flowers his mother planted.

(iii) Vern seems ecstatic about his mother's flowers.

However, they cannot be adjoined to what would be an NP within a DetP in the DetP analysis:

(iv) *Vern likes the flowers his mother planted.

which suggests that the only maximal projection here is NP, as in (i).

2 A necessary and sufficient definition of Government must refer also to the barriers on this relation, so that a verb does not govern an NP across another maximal projection such as PP in structures like V_PP[P NP]. The full definition given in Riemsdijk and Williams (1986: 231) is: X governs Y if Y is contained in the maximal X-projection of X, X\(\text{max}\), and X\(\text{max}\) is the smallest maximal projection containing Y, and X c-commands Y.

3 These are again necessary conditions, not necessary and sufficient conditions; cf. n. 2 ch. 6. Appropriate barriers will need to be incorporated on these relations. What is at issue in the present context is how far up, not how far down, the tree these relations can apply.

4 I assume that the parser recognizes the identity of these two S nodes by consulting the grammar in the usual way. An adjacent subject NP and VP can only be dominated by a single S, not by two distinct S nodes. The second S constructed is therefore assumed to be the same as the first, and a tree structure is built that is well-formed according to phrase-structure principles. GNC and ICA are, in effect, collapsed on these occasions.

5 Notice that if this very simple generalization for agreement is correct, it supports some proposals for constituent structure and provides evidence against others. In particular, if both Det and N are heads (cf. the discussion in section 6.1.3), then if Det agrees with N, it will not be bijacent to it, since it will asymmetrically c-command an N within an NP, i.e. DetP[Det NP[... N ...]]. On the other hand, if both Det and N are sisters within a flat NP structure, then bijacency will be satisfied. I consider this to be a further argument in favor of a flatter NP structure. If N is a head and Det a specifier within the X-bar rule format of (6.1) in the main text, then there will need to be an extension of the sisterhood concept to include ICs at different bar levels within the same maximal projection, if the bijacency generalization is to be captured.

6 Notice that agreement, as a form of syntactic dependency, is subject to the predictions of the Principle of Structural Determination formulated in (2.36) (cf. ch. 2.4). The agreeing category B has to have A within its SD. The bijacency of B to A guarantees that this is the case, and limits the containing node C that defines the SD for B, within which A must be contained, to either a mother or at most a grandmother of B.

7 Cf. Moravcsik (1978) for a discussion of the cross-linguistic facts and impicational regularities in the area of agreement.

8 Preposition-stranding examples corresponding to (6.23b) are also found productively in Dutch, in which preposition and verb must again be adjacent; cf. Donaldson (1981: 64).
9 Single-word adjectives are also predicted to follow $S_s$ or $S$ in prenominal relative clause constructions, and (as shown in ch. 5.2.1 and table 5.8) they do. Genitive NPs, which are also shorter than relative clauses on aggregate, should also follow these prenominal relative clauses and precede the head in the unmarked case, and this also appears to be the case, cf. C. Lehmann (1984).

10 Recall also the explanation proposed in ch. 5.6.1 for why subordination indicators within $S_s$ or $S$ are not preposed within relative clauses for reasons of immediate matrix disambiguation, as they often are within other subordinate clauses. The proposal is that if a grammar is going to respond to immediate matrix disambiguation in these cases, it will prepose the higher c-commanding constituent, i.e. the head noun, and achieve consistent top-down structural recognition, rather than preposing the lower constructor of $S_s$.

11 I assume that the structure of a sentence such as *Ich habe ihm das Buch gegeben* "I have him the book given," i.e. I have given him the book, is approximately the following, with $S$ a sister of $NP$ :

12 In other words, there will need to be an initial determiner or other $NP$-constructing category preceding these prenominal relative clauses, or *erweiterte Attribute* (cf. Weber 1971), as in the example given in the main text.

13 German does have a small number of postpositions as well, e.g. *dem Haus gegenüber* "the house opposite," i.e. opposite the house.

14 If the complementizer is removed, the verb position reverts to "main clause" word order again, and so the "subordinate clause" is again constructed on its left periphery, e.g. *Ich weiß, er ist ein braver Junge* "I know, he is a good lad," in lieu of *Ich weiß, daß er ein braver Junge ist* "I know that he a good lad is."
15 See Askedal (1992) for a discussion of the difference between Icelandic and German with regard to the isomorphism between morphological case and configurationality.

16 Notice that early recognition of VP within S via case will not lengthen the CRD for VP at the same time that it shortens that for S. The CRD for VP will be initiated by whatever category constructs its first IC, e.g. Det constructing NP. If that Det carries a case capable of constructing VP, and if VP is the last IC of S, as it generally is, then early construction of VP will be advantageous for S and will have no impact on the CRD for VP. If case were carried only on the right periphery of NP, this would lengthen the CRD for S, but would again have no necessary impact on VP.

7 Conclusions

1 The communicative needs of different cultures are not always similar, of course, and the overall complexity level of one whole language may not be similar to that of another. It is useful in this context to heed the very sensible words of Comrie (1992: 194): "I... conclude that there is probably no difference among individual humans or among human groups in the capacity for handling languages of different degrees of complexity, but that there may well be differences in the degree of complexity of individual languages, a view that is, incidentally, quite compatible with respect for the dignity of all humans and the right to cultural diversity."

Envoi

1 The syntactic structures and EIC calculations in this Envoi have obviously been added into Molière's text as written. They incorporate the following assumptions:

(i) All EIC calculations are made in terms of IC-to-word rather than IC-to-non-IC ratios (cf. ch. 3.4), though cf. point (viii) below.

(ii) Only those Constituent Recognition Domains (CRDs) are considered that undergo an ordering change in at least one of these sentences – the others remain constant and do not affect the aggregate for the sentence as a whole.

(iii) The four phrases whose daughter ICs undergo an ordering change in at least one version are: NP (consisting of three ICs, vos, beaux, and yeux); VP (i.e. the matrix VP consisting of three ICs, me, font, and the embedded VP); VP (the infinitival VP comprising two ICs, the infinitive mourir and PP); and S (which consists of the subject NP and the matrix VP).

(iv) I assume that all these phrases maintain their structure after movement; discontinuous ICs are therefore allowed and are associated with an EIC-defined cost compared with the corresponding adjacent ICs, cf. the extensive discussion of discontinuity in chs. 3 and 4.

(v) [Belle marquise] is assumed to be a parenthetical phrase that does not constitute an IC of any phrase within this S (cf. Espinal 1991 for one theoretical approach to such "disjunct constituents"); these two words may nonetheless contribute to the length of a CRD and to its IC-to-word count, depending on their position-
(vi) The determiner *vos* can construct NP in our theory, being unique to it.
(vii) The accusative-marked clitic pronoun *me* can construct VP in our theory, being unique to it.
(viii) The preposition *d' in pp*[d'amour]* is counted as a separate word in these EIC calculations, in order to reflect the existence of an additional (Prep) constituent within this (PP) daughter of VP.
References


References

(1990) "Verb cluster formation as a morphological process." MS, Max-Planck-Institut für Psycholinguistik, Nijmegen.


Working Paper 2 Theme Group 2, European Science Foundation Programme in Language Typology, ESF Office Strasbourg, 1–33.


References


(n.d.) "On the function of agreement."
References


References 479


References

Order, Working Paper 2 Theme Group 2, European Science Foundation Programme in Language Typology, ESF Office, Strasbourg.


References


Index of names

Abe, Y., 364
Abney, S., xv, 62, 72, 291, 344, 349, 352, 353
Andersen, E., xv
Anderson, S.R., 55, 352
Antinucci, F., 323
Aoun, J., xv, 1
Akedal, J.O., 468
Baker, C.L., 213
Baltin, M.R., 1
Bean, M.C., 400
Behaghel, O., Ill, 118-120, 210, 455
Berlin, B., 436, 442
Berwick, R.C., 6, 7, 9, 21
Bever, T.G., 6, 324
Bierwisch, M., 48, 377
Blake, B., 1, 39, 42, 368, 454
Bloomfield, L., 1
Bowe, H.J., 454
Bowerman, M., 213
Bredsdorff, E., 383
Carlson, G., 243
Chesterman, A., 54
Chomsky, N., 1, 7, 10, 11, 12-13, 16-17, 25, 32-34, 62, 73, 75-76, 88, 107, 321, 343, 344, 346, 348, 411, 413, 437, 442
Chung, S., 332
Clancy, P.M., 323
Clark, H.H., 113, 209
Cohen, M.S., xvi
Corbett, G.G., 343, 357
Croft, W., 2, 15, 17, 436, 447, 464
Crothers, J., 5
Culicover, P.W., 184-188, 208, 367
Cutler, A., xv, 4, 5, 116, 321, 323
Dahl, Ö., 408
DeLancey, S., 2, 17
Derbyshire, D.C., 332
De Smedt, K.J.M.J., 426
Dik, S.C., 119-120
Dixon, R.M.W., 55, 273, 448
Donaldson, B.C., 466
Dowty, D., 55
DuBois, J.A., 2
Dunlea, A., xv
Ebert, R.P., 304-306
Emonds, J.E., 88, 92, 180, 201, 246, 309, 310-311, 324, 362
Enc, M., 374
Erdmann, P., xiv, xv, 66, 118, 190-195, 341
Erguvanli, E.E., 374
Ernst, T., 352, 353, 465-466
Espinal, M.T., 468
Fillmore, C.J., 424
Firbas, J., 111-117, 215
Fodor, Janet D., xv, 2, 4, 6, 9, 11, 52, 327, 418, 437-438
Fodor, Jerry, A., 3, 4, 5, 62, 64, 68, 358
Frachtenberg, L.J., 453
Frazier, L., xii, xiv, 2, 4, 5, 6, 8, 9, 11, 16, 22, 25, 33-37, 66, 67, 76, 95, 242-249, 265-269, 294, 323, 324, 413, 418, 437-438, 439, 445-447
Gadler, H., 189
Garcia, E., xv, 452
Gazdar, G., 1, 107, 343, 354, 453
Gee, J., xv
Gell-Mann, M., xv
Gil, D., 118
Gilligan, G., 5, 323, 369

483
Index of names

Gorrell, P.G., 244
Greenberg, J.H., 2, 5, 6, 16, 253, 255, 258, 328, 418
Grice, H.P., 106
Grosu, A., 117
Guernsey, S., 243
Gundel, J.K., 115, 339, 340, 364

Haiman, J., 2
Hakulinen, A., 169, 170
Hale, K., 71, 112, 114, 425, 453
Hall, C., 5
Harris, Z.S., 1
Hawkins, S., 31-32, 445
Hewitt, B.G., 351
Hinds, J., 139
Hoekstra, T., 1
Hoji, H., 139, 451
Hopper, P., 2
Horie, K., xiv, xv, 137-154, 229-237, 457
Hornstein, N., 16
Hudson, R.A., 343, 345, 346, 351, 353, 356
Hunter, P.J., 64-65
Hyman, L.M., 337, 424

Inoue, A., 4, 11, 67, 249, 325
Jackendoff, R., 22, 73, 107-109, 111, 253, 269, 284, 287, 289, 293-296, 343, 354
Jacobs, J., xv, 376-378, 397, 424
Jespersen, O., 1
Jhang, S.-e., 1

Kashket, M.B., 381
Kay, P., 436, 442-443
Kim, A.H.-O., 374, 375
Kim, C., xiv, xv, 155-157, 211-212, 457
Kim, N.-K., 155
Kimball, J., 60, 450
Kiss, K., xiv, 114-115, 130, 133-134, 210, 407, 431
König, E., xv
Kooij, J.G., 1
Koopman, H., 107, 109, 253, 273
Kornai, A., 73, 344
Krifka, M., xv
Kroch, A.S., 1
Kuno, S., 5, 11, 138, 139, 363-364

Lascaratou, C., xiv, 178-180
Lasnik, H., 7, 12, 442
Lehmann, W.P., 253
Lenerz, J., 188, 189
Levelt, W.J.M., 3
Levin, H., 66
Levinson, S.C., xv, 106, 114
Lieb, H.-H., xv
Lieberman, P., 4, 452
Lightfoot, D.W., 16, 107-109, 387
Lindblom, B., 5, 436
Lötscher, A., 189
Louden, M., 401
Lust, B., 67
MacDonald, M.C., 243
Maddieson, I., 5, 436
Mallinson, G., 133
Marcus, M.P., 6, 61, 62
Marshall, C.R., 113, 209
Marslen-Wilson, W.D., xv, 4, 5, 116
Mater, E., 39
Matthews, P., xv
Maxwell, D., 449
Mazuka, R., 67
McCawley, J.D., 81, 197, 432
McCloskey, J., 332
McDaniel, C., 436, 442-443
Miller, G.A., 25, 32-34, 75-76, 321, 413, 437
Minkova, D., 428
Mithun, M., 111, 112, 453-454
Mooshammer, C., xiv, 198-204
Moravcsik, E.A., 466
Morolong, M., 337, 424
Newmeyer, F.J., 2, 7, 442
Nichols, J., 332, 367, 371
O'Grady, W., 81, 432, 465
Payne, D.L., 111, 425
Pearlmutter, N.J., 243
Pinker, S., 213
Polinskaja, M.S., 332
Pollard, C., 343
Prideaux, G.D., 64-65, 92, 437
Prince, E.F., 341
Pritchett, B.L., 62, 437

Index of names 485

Pullum, G.K., 73, 332, 344
Radford, A., 291
Rayner, K., 11, 66, 67, 249
Reed, I., 447
Reinhart, T., 1, 26, 50, 52–54, 69, 191,
197–198, 207, 298, 353, 364, 432, 449–450
Reinholtz, C., xiv, 134–137, 160–164,
278–279
Riemsdijk, H.C. van, 7, 466
Rochemont, M.S., 184–188, 208
Ronneberger-Sibold, E., 402
Ross, J.R., 20, 43, 49, 197, 462
Ruhlen, M., 121
Sadler, L., 332, 382–383, 385
Sag, I., 52, 309, 343
Saito, M., 80, 139, 145, 451
Saitz, R.L., 400, 401
Sasse, H.-J., 115, 117, 339
Saussure, F. de, 1, 26
Schachter, P., 443–445
Seuren, P.A.M., xv
Shannon, T., xiv, 198–210
Siewierska, A., xiv, xv, 174–178, 238, 330,
455, 460
Smith, C.S., 288
Smith, J.R., 404
Smith, N.V., xv, 9
Sperber, D., 106, 113, 209, 216
Stallings, L., xv
Stassen, L., 447
Steedman, M., 6, 376, 377
Steele, S., 366
Stockwell, R.P., 69, 117–118, 196, 402
Tannenhaus, M.K., 243
Testelec, Y., xiv, xv
Thompson, S.A., 2, 111, 112, 117, 345, 363
Tomlin, R.S., 279, 329–339, 420–421
Travis, L., 107, 109, 253, 273
Trueswell, J.C., 243
Tyler, L.K., xvi, 4, 5, 116
Ueda, M., 67, 249
Vennemann, T., 253, 427
Vergnaud, J.-R., xvi
Vilkuna, M., xiv, 169–171
Vincent, N.B., 368
Wanat, S.F., 66
Weber, H., 290, 467
Weinberg, A.S., 6, 7, 9, 21
Whitman, J.B., 437
Wilkins, W., 367
Williams, E., 7, 288, 466
Wilson, D., 106, 113, 209, 216
Zipf, G.K., 216
Zwicky, A., 273, 343, 351, 356, 357
Index of languages

Abkhaz, 351, 367, 369-370, 371
Amharic, 316
Arabic, 44, 272, 316, 449; see also
Egyptian Arabic
Australian languages, 114, 433, 454; see
also Dyirbal; Maung; Ngandi;
Warlpiri
Bantu languages, 275, 280, 402; see also
Kinyarwanda; Luganda; Sesotho
Basque, 37, 265, 272, 371, 389-390, 449
Brahui, 320
Bulgarian, 427-428
Burmese, 272
Catalan, 37
Cayuga, 453-454
Celtic languages, 332, 382; see also Irish;
Welsh
Chinese (Mandarin), 44, 109, 110, 268,
270, 271, 272, 274, 281, 388, 449
Coos, 453-454
Danish, 383-385
Dravidian languages, 271, 390; see also
Tamil; Telugu
Dutch, 6, 210, 376, 377, 401, 466
Dyirbal, 265, 270, 272, 273-274
Egyptian Arabic, 299
English (Modern), 5, 9, 10, 11-12, 19-20,
21, 22, 24, 26-28, 29-32, 34-36, 37,
42, 44, 48-50, 51, 54-55, 57, 64-66,
68, 79, 89-90, 92, 94, 100-101, 102,
108, 109, 110, 121, 123-129,
159-160,180-182,182-188,190-196,
197-198, 208, 212-214, 219-221,
226-227, 230, 238-241, 244-248, 251,
263, 265, 270, 275, 276, 282-296
passim, 299, 309-312, 319-320, 324,
328-329, 330, 341, 347, 348, 360,
361-363, 367, 370, 372, 381, 387, 399-
402, 405, 406, 408, 426, 428, 430-431,
432, 449, 455-456, 460-461, 465-466
Old English, 196, 258, 353, 400-402
Eskimo, see Yup'ik
Ewe, 272
Finnish, 40, 121, 160, 169-171, 272, 339,
449
French, 275, 440-441, 449, 468-469
Frisian (North), 37, 449
Fulani, 44, 449
Galla, 267
German, 11, 12, 39-40, 42, 48-49, 109,
110, 118, 121, 188-190, 196, 197-210,
223-224, 226, 229, 232, 238-240, 256,
258, 263, 264, 266-267, 275, 279-280,
290, 299, 302-308, 313-315, 320, 323,
325, 351, 352, 353, 366, 367, 370,
376-378, 385-387, 393-399, 401, 402,
404-405, 424, 428-429, 430-431, 432,
443-445, 454, 461, 462, 463, 467
Germanic languages, 121, 272, 382, 383,
393, 399-402, 403-405; see also
Danish; Dutch; English; Frisian;
German; Icelandic
Greek, 121, 178-180, 272, 316, 330, 333
Hausa, 37, 44, 449
Hebrew, 43, 44, 449
Hindi, 449
Hixkaryana, 332
Hungarian, 42, 105, 114, 121, 129-133,
160, 164-168, 172-173, 210, 222-224,
226, 228, 231, 238-240, 241, 339, 407,
425, 431, 456-457
Hurrian, 272
Icelandic, 42, 468
Indonesian, 37
Irish, 464
Jalalte, 271

Javanese, 44

Karen, 463
Kera, 37
Kinyarwanda, 299

Kpelle, 109, 110

Lahu, 265, 270, 272, 274, 388–389
Lakhota, 266–267, 302
Latin, 299, 367, 368–369, 403, 433
Luganda, 37
Luiseno, 366
Lushei, 272

Malagasy, 37, 196, 299, 449
Maori, 37, 449
Masai, 316
Maung, 316
Maya, 316
Minang-Kabau, 44, 449

Ngandi, 453

Papago, 112, 114, 425, 453
Persian, 11, 12, 68, 263, 265, 299, 302, 323, 325–326

Polish, 121, 174–178, 330, 333, 460
Quechua, 320

Romance languages, 121, 272, 273–274, 275, 280, 402, 403; see also Catalan; French; Latin; Rumanian; Spanish
Roviana, 449
Rumanian, 121, 133–134, 457
Russian, 48–49

Sango, 463
Sesotho, 337, 424
Slovenian, 449
Spanish, 316

Tamil, 270
Telugu, 271, 272
Thai, 299, 316
Toba Batak, 38, 44, 443–445, 449
Tongan, 449
Tuscarora, 299

Urhobo, 44

Warlpiri, 366, 368, 369, 381
Welsh, 382–385, 464

Yaqui, 11, 263, 272, 299, 320, 325–326
Yavapai, 327
Yoruba, 348
Yup'ik, 477
absolute unprocessability, assumption of, 8–9, 266, 411
absolute unprocessability, assumption of, see absolute unprocessability
of absolute unprocessability, assumption of
acceptability, versus grammaticality
intuitions, 7, 10–11; see also pure
acceptability
acceptability, versus grammatically
intuitions, 7, 10–11; see also pure
acceptability
Accessibility Hierarchy (of Keenan and
Comrie), xii, 14, 32, 36, 37–45, 413,
414, 435, 447, 448–449; see also
relative clauses
Absolutive Case Marking, 447–448
accusative case marking, 27–29, 39, 42–43,
161, 365, 393, 443–445; see also case
adjacency of ICs, see Attachment
Construction; Attachment Promotion;
 discontinuous constituent structure;
heads, adjacency of
discussing the construction of higher nodes
Agreement Projection, see Agreement
Projection (AgP)
Agreement Projection (AgP),
definition, 368–369
discussion, 366–373, 403, 404
anaphoric binding, 28, 51, 53–54, 145, 434,
443–445
Animated First Principle (of Tomlin),
337–338
argue (semantic argument categories),
see heads, defining properties of;
functors
aspect,
auxiliary positioning, 259–260
particle positioning, 408
asymmetries, see left–right asymmetries
Attachment Construction (AtC),
definition, 376
discussion, 376–378
Attachment Construction (AtC),
definition, see also relative clauses
Attachment Construction; Attachment Promotion;
 discontinuous constituent structure;
heads, adjacency of
attachment construction,
definition, 376
attachment promotion, see Attachment
Construction; Attachment Promotion;
 discontinuous constituent structure;
heads, adjacency of
auxiliary verbs,
as constructing categories, 361–362, 381,
382, 387
ordering of, 259–260, 366
Axiom of Constructability,
definition, 379
discussion, 379–380, 422
see also construction of higher nodes
Axiom of MNCC Existence,
definition, 63
discussion, 63–64, 344–345, 346–348,
358, 379–380
see also heads; Mother Node
Construction
Axiom of MNCC Uniqueness,
definition, 63
discussion, 63–64, 344–345, 358, 379–380
see also heads; Mother Node
Construction
Basic Order Prediction (of EIC),
definition, 96
discussion, 95–99, 251, 328
testing
on basic orders in English, 282–296
on binary branching structures,
255–265, 265–269, 328–339
on multiple branching structures,
269–274, 274–281
see also EIC
Bijacency, 367, 466

Subject index 489

binary branching structures,
EIC predictions for, 103–104
testing EIC's predictions for grammars,
see Basic Order Prediction
testing EIC's predictions for performance, see Text Frequency Prediction
bi-uniqueness (of heads and maximal projections), 349–350, 421, 434; see also construction of higher nodes; heads
bottom-up (versus top-down) processing, 66–67, 157, 249, 252–253, 289
case
assignment (syntactic), 355, 376, 377–378, 386; see also Construction Domain; heads
hierarchies (of Primus), 42–43, 414, 447–448
marking (morphological), 23, 26, 56, 468 and Grandmother Node Construction, 147–148, 161, 278, 362–363, 365, 393, 398, 404–405; see also construction of higher nodes; Grandmother Node Construction and word order freedom, 427–429, 433; see also freedom of word order
loss of, and verb-position change, see case syncrétism in the VP; see also absolutive; accusative; dative; ergative; nominative
syncrétism,
in the NP, 403–406
in the VP, 258, 399–402, 429
category constancy (of heads and maximal projections), 349–350, 421; see also construction of higher nodes; heads
center embeddings, 5, 8, 10–12, 14, 19–20, 23–24, 66, 68–69, 117, 138, 185, 303–308, 316–321, 323–325, 419, 435, 442; see also Hierarchy Prediction; self-embeddings
clitic pronouns, 275, 402; see also pronouns
cognitive explanations for grammar, 2, 17, 117
cohort model (of Marslen-Wilson and Tyler), see word recognition
communicative dynamism, see Prague School
Comparative Substitution (in English), 311 complement phrases (of X-bar theory), 344 relative ordering in English, 293–296, 322
see also heads; left-right asymmetries; X-bar theory
complementizer,
as a constructing category, 350, 360–361, 383, 385
positioning of, 263, 325–326
see also construction of higher nodes; functional categories; heads; Sister Node Construction
Complex NP Constraint (of Ross), 43, 197 and pronoun retention, 43–44
complexity, see Derivational Theory of Complexity; hierarchies; metrics of complexity; processing complexity, degrees of; structural complexity
comprehension versus production (for EIC), 425–427
configurationality parameter (of Hale), 71–72
hierarchies of c-command asymmetries, 42
see also Structural Determination;
Structural Domain
Constituent Recognition Domain (CRD) definition, 58–59
see also EIC; Performance Attachment Predictions; Structural Domain
Construction Domain
definition, 355
construction of higher nodes, 60–62, 246–247, 359
principles for
Basque, 390
Danish, 384
English, 363, 387
German, 393, 394, 395
Japanese, 365, 392
Lahu, 389
Turkish, 391
Welsh, 384
see also Agreement Projection;
Attachment Construction;
Attachment Promotion; Axiom of Constructability; Axiom of MNCC Existence; Axiom of MNCC Uniqueness; Construction Domain;
Grandmother Node Construction; Immediate Constituent Attachment, Mother Node Construction; Sister Node Construction
conversational implicature, 106
Covariation of Functional Equivalents (of Keenan), 15, 427

dative case marking, 42–43, 365, 393, 444–445; see also case

Dative–Accusative Shift, in English, see Dative Movement in Korean, 211–212

Dative Movement (in English), 100, 212–214, 311–313

dependent marking (of Nichols), see head marking

Derivational Theory of Complexity, 92, 437; see also complexity
determiner (phrase), 62, 288, 318, 348–349, 352–353, 403–406, 465–466; see also functional categories; heads; NP node, construction of
direct objects, and case marking, 28, 39–40, 138
ordering of, 328–329, 420–421
relativization on, see relative clauses and semantic interpretation, 51–56
sentential, see center embedding; Extraposition; sentential nodes Structural Domain of, 27, 39–40
verb agreement with, see agreement see also case marking
directionality,
of case assignment, 107, 109–111, 253, 273, 423
of theta-role assignment, 107, 109–111, 253, 273, 423
see also case; theta roles
discontinuous constituent structure, 81, 118–119
in English, 80–81, 126
in Extraposition from NP structures, 196–208, 432
in German, 196–208, 432
in Japanese, 80–81, 147–149, 151, 154, 457–460
in Korean, 155
in Polish, 175
in the VP, 126, 147, 155, 175, 300–301, 330, 335, 432
discourse, 4, 217, 455; see also conversational implicature; discourse-configurational nodes; New before Given; pragmatic principles of linear ordering; Prague School; Relevance Theory; Task Urgency
discourse-configurational nodes, 114–115, 130, 241–242, 339, 417, 425; see also focus, topic position

Early Immediate Constituents (EIC) comprehension versus production, 425–427
definition of, 77, 78–79
left-to-right measurement of, 81–83, 451–452
measurement of, 69–83
predictions for grammatical rearrangements, 100; see also Rearrangement Prediction
predictions for grammaticalized basic orders, 96, 254–255; see also Basic Order Prediction
predictions for linear orderings in performance, 84, 86–87; see also Performance Ordering Prediction; Text Frequency Prediction
predictions for temporary ambiguity preferences; 94; see also Performance Attachment Prediction
predictions for typologically different structures, 104
predictions for word order hierarchies, 102; see also Hierarchy Prediction
preliminary evidence for, 64–69
testing the grammatical rearrangement predictions, 296–315
testing the grammaticalized basic order predictions, 251–265, 269–274, 274–281
on basic orders in English, 282–296
testing the performance ordering predictions, 122–214; see also Text Frequency Prediction
testing the word order hierarchy predictions, 315–321

ergative case marking, 368, 447–448
Extended Mother Node Construction (EMNC), definition of, 359
extraction, see Movement Domain; subjacency; WH-movement
Extraposition, 20, 100, 262, 263–264, 298, 433
in German, 302–308, 432
Extraposition from NP, 262, 264
in English, 197–198
in German, 198–210, 266, 267, 304; see also discontinuous constituent structure

finiteness, and constituent structure
recognition, 354, 361–362, 366, 372, 381–385, 389, 390; see also agreement;
auxiliary verbs; Grandmother Node Construction; inflectional morphology
First Analysis Constraint (of Frazier), 67, 243–244, 246, 248; see also garden-path sentences; Late Closure; Minimal Attachment; Most Recent Filler; temporary ambiguities
fixed word order, 87–93; see also freedom of word order; grammaticalization
focus
constructions, 184–188, 189–190, 208
positions, 339–341, 406, 408, 425
in English, 184–188, 341
in Hungarian, 130, 186
in Turkish, 374, 406
see also discourse-configurational nodes
freedom of word order,
degrees of, 428
EIC's performance predictions for, see EIC
historical transition from free to fixed, 95–96, 426
testing EIC's performance predictions for, see EIC
versus fixed word order, 114, 416,
427–429, 454–455
within a performance module, 87–93
see also case marking;
grammaticalization
freezing principle (of Culicover), 185
functional categories (in X-bar theory), 73,
346; see also heads; X-bar theory
functional explanations for syntax, 1–2; see also cognitive explanations for grammar; pragmatic explanations for grammar; processing explanations for grammars
Functional Grammar (of Dik), 119-120,
455
functors (semantic functor categories),
356–357, 370, 421; see also heads, defining properties of
government, 353–356, 466; see also
Construction Domain; Government-Binding Theory; heads
Government-Binding Theory (of Chomsky), 343, 353, 372, 450
grammatical rearrangements, see
Rearrangement Prediction;
rearrangement rules
grammatical relations, see case marking;
direct objects; indirect objects;
Relational Grammar; subjects
grammaticalization,
of processing principles and preferences, 19–24; see also EIC Basic Order
Prediction; EIC Hierarchy
Prediction; EIC Rearrangement Prediction; freedom of word order versus fixed word order;
performance-grammar correlation;
processing explanations for grammars
Grandmother Node Construction (GNC)
definition, 361
discussion, 62, 148, 161, 314, 354,
361–366, 380, 384, 387, 390, 392,
393, 403, 404
head marking (of Nichols), 367, 369–370
Head Proximity (of Dik), 120
heads
adjacency of, 265–269
defining properties of, 343–348, 351–357,
421, 433
ordering of, 107, 108, 265–269, 273
see also head marking; Head Proximity;
X-bar theory
heavier element principle (of Stockwell),
117–118
Heavy NP Shift (in English), 20, 57–60,
75, 89, 117, 182–188, 214, 262, 298,
460–461
hierarchies,
of basic color terms, 436, 442
of cases, see case hierarchies
of c-commanding nodes, 42
of center embeddings, 14, 435
of expanding consonant inventories, 5,
436
of extraction environments, 47–50, 435
of grammatical relations, see Accessibility Hierarchy
of relative clauses, 37–41, 435
of semantic markedness, 436
of theta-roles, 42
of vowel systems, 5, 436
of word orders, 316–321
Hierarchy Prediction (of EIC),
  definition, 102
discussion, 101–102
testing the prediction, 316–321
see also EIC

Immediate Constituent Attachment (ICA)
definition, 62
discussion, 60–62, 116, 384
Immediate Matrix Disambiguation (IMD),
  324–326, 364, 388, 467
Impermissible Ambiguity Constraint (of Frazier), 6, 294, 324
implicational universals, xii, 14, 253, 410, 412–413, 418, 436, 442; see also hierarchies
importance, see pragmatic predictions for linear ordering
indirect objects,
  and case marking, 39–40, 138
  relativization on, see relative clauses
Structural Domain of, 29–30, 39–40
verb agreement with, see agreement see also case marking
inflection (phrase), see finiteness;
  functional categories; heads;
  inflectional morphology; X-bar theory
inflectional morphology, 351–352, 361–362; see also agreement;
  Agreement Projection; finiteness;
  Grandmother Node Construction;
  morphology
information structure, see discourse;
  discourse-configurational nodes;
  focus; pragmatic principles of linear ordering; topic
innateness of the language capacity, 1–2,
  3–5, 7, 12–13, 16–17, 110–111, 442
evolution of the innate grammar, 7,
  12–13, 442
see also modularity thesis
Inversion after Locative Phrase Preposing
  (in English), 311
in focus positions, 131, 339, 341
in the positioning of sentential nodes,
  264, 322, 323–328, 420
in the positioning of relative clauses,
  264–265, 322, 323–328
in subject–object ordering, 328–339, 420–421
in topic position, 115, 131, 339–341
involving short ICs, 321–322, 420
within the relative clause, 264–265, 326–328
left-to-right IC-to-non-IC ratios, 81–83, see also EIC
logical problem
  of language acquisition, 16
  of language performance, 16
Meaning–Form Dependency Principle (of Keenan), 370
metrics of complexity,
  Frazier’s, 33–37, 76, 413, 437, 445–457
Hawkins’s, see structural complexity;
  Structural Domain
Miller and Chomsky’s, 32–33, 75–76, 413, 437
see also Derivational Theory of Complexity
Minimal Attachment (of Frazier), 36, 67, 95, 247–249, 323, 326, 418, 439; see also temporary ambiguities
Minimal Structural Domain,
  definition, 39
in relative clauses, 39–41, 413
see also Structural Domain
modularity thesis (of Jerry Fodor), 4, 5,
  12–13, 62, 64, 68
morphology,
  processing of, 5
  and Structural Domains, 27–29
see also inflectional morphology
Most Recent Filler (of Frazier), 36–37, 95, 248, 439; see also temporary ambiguities
Mother Node Construction (MNC),
  definition, 62
discussion, 60–62, 116, 246, 344–358 passim, 381–382, 384, 387, 389, 392, 393, 403, 404, 421
see also Axiom of Constructability;
  Axiom of MNCC Existence; Axiom of MNCC Uniqueness; Extended Mother Node Construction; heads
Mother Node Constructing Category (MNCC), see Mother Node Construction
Movement Domain, 37, 47–50, 248, 437
definition, 47
see also Structural Domain, WH-movement
multiple branching structures, EIC predictions for, 103-104
testing EIC's predictions for grammars, 269-281
testing EIC's predictions for performance, 122-171
negative
auxiliary positioning, 259-260
particle positioning, 408
Nested Dependency Constraint, 6
neurons, 438
New before Given,
defining the predictions for ordering, 225-226
discussion, 224-237, 237-242, 417
testing the predictions,
on English: [NP V (PP₁ PP₂)], 227
on German: NP-PP rearrangements, 229
on Hungarian: [V {mNP₁ mNP₂}], 228
on Japanese: [{NP O PP^} V], 236-237
see also pragmatic principles of linear ordering; Prague School; Task Urgency
New Nodes Principle (of Kimball), 60, 450
newness, see pragmatic predictions for linear ordering
nominalizing particles, 388-389, 454
nominative case marking, 27-29, 39, 42-43, 161, 362, 365, 443-445; see also case
non-terminal to terminal node ratio, see metrics of complexity
NP node,
universals of ordering in the, 259, 261, 264-265, 269-274
construction of, 403-406
ordering in the English NP, 287-293
objects, see direct objects; indirect objects
old information, see pragmatic principles of linear ordering
parameters of linear ordering, 107, 109-111, 253; see also Government-Binding theory
Parsing principles, see Agreement Projection; Attachment Construction; Attachment Promotion; construction of higher nodes; First Analysis Constraint; Grandmother Node
Performance Attachment Prediction (of EIC),
definition, 94
discussion, 93-95, 242-249
see also garden-path sentences; EIC; temporary ambiguities
Performance-Driven Grammar, xii, 409, 439
Performance Ordering Prediction (of EIC),
definition, 84
discussion, 84-87, 95-96
testing, see Text Frequency Prediction
see also EIC
phonology,
processing of, 5, 436
universals of, 5, 436
see also stress; weight
possessive phrases, 29-30
see also genitives
PP node,
grammaticalization of adposition ordering, 256
PP ordering within VP, 96-97, 107-108, 255-257
universals of ordering within PP, 259
see also Prepositional Noun-Modifier Hierarchy
pragmatic explanations for grammar, 455
(n. 18)
pragmatic principles of linear ordering, defining their predictions, see New before Given; Prague School; Task Urgency
testing the predictions on English, 220, 227, 230
on German, 223, 229
on Hungarian, 222, 228, 231
role of pragmatics in linear ordering, 237–242
Prague School, defining Prague School predictions for ordering, 226–227
testing the predictions on English: [NP V {PP1 PP2}], 230
on German: NP–PP rearrangements, 229
on Hungarian: [V {mNP1 mNP2}], 231
on Japanese: [(iPPm 2PPm) V], 232;
[[NPm PPm] V], 234–235
see also New before Given; pragmatic principles of linear ordering; Task Urgency
predictability, see pragmatic principles of linear ordering
preposing of heavy constituents, 8, 66–67, 80–81; see also rearrangement rules
Prepositional Noun-Modifier Hierarchy, 316–319, 419, 463
principle of least effort (of Zipf), 216
Priority of the Most Immediately Containing CRD, 277–278
processing, complexity, see complexity
explanations for grammars, 1–9, 13–14, 15–16, 18, 19–24, 46, 49, 110–111, 411–412, 423, 438, 439, 455
see also absolute unprocessability; construction principles; grammaticalization; parsing principles; performance
Pro-drop languages, 369
production versus comprehension (for EIC), 425–427
Pronoun Fronting (in German), 313–315
pronouns, cliticization of, see clitic pronouns
in rearrangement rules, 100–101, 308–315
in relative clauses, 43–45
in WH-movement structures, 49–50
pure acceptability, assumption of, 7, 10–12, 15, 411
quantifier scope, 51–52, 145, 407, 449–450; see also scope relations
Rearrangement Prediction (of EIC),
definition, 100
discussion, 99–101
testing the prediction on heavy categories, 297–308
on light categories, 308–15
see also EIC
rearrangement rules, 90–91; see also Comparative Substitution; Dative–Accusative Shift; Dative Movement; Extraposition; Extraposition from NP; Heavy NP Shift; Inversion after Locative Phrase Preposing; Participle Preposing; Particle Movement; preposing of heavy constituents; Pronoun Fronting
referential independence, and Structural Domains, 52, 54
Relational Grammar, 39, 42
relative clauses
in English, 30–32, 66
in a performance experiment on English, 31–32, 445
positioning of constructing category, 265, 270–271, 326–328
pronoun-retaining, 43–45, 413, 448–449
strategies of, 43–45, 448–449
types of pronominal relative clauses, 270–271, 388–393
universals of, 37–45, 448–449
Relativization Domain, 31, 38, 45, 437
definition, 31
see also relative clauses; Structural Domain
Relator principles (of Dik), 120, 455
Relevance Theory (of Sperber and Wilson), 216
right-mother-node-constructing categories, definition of, 103–105
scope relations, 51–52, 145, 374, 406–408, 434
self-embedding, 5, 8, 11–12, 442
semantic explanations for grammar, 17;
see also cognitive explanations for grammar; pragmatic explanations for grammar
semantic functors, see functors
sentential nodes,
construction of, 381–393 passim
see also Extrapolation; Extrapolation from NP; preposing of heavy constituents; left-right asymmetries; relative clauses
Sister Node Construction (SNC),
definition, 359–360
discussion, 62, 347–348, 359–361, 383
specificity, degrees of, 52
specifiers (of X-bar theory), 344; see also X-bar theory
spoken versus written texts (for testing EIC), 452–453
Standard Theory of generative grammar (Chomsky 1965), 88, 107, 348
stress, 182, 187, 209
definition, 29
predictions for grammatical variation, 37–50
and its semantic correlates, 50–56
see also complexity; metrics of complexity; Structural Domain
Structural Determination, principle of
definition, 51
see also Structural Domain
Structural Domain (SD),
definition, 25
SD of a node X in C, 28
syntactically defined SD of a node X in C, 26
and semantic interpretation, 50–56
and temporary ambiguity preferences, 95, 248–249
see also Constituent Recognition Domain; Construction Domain; Minimal Structural Domain; Movement Domain; Relativization Domain; structural complexity; structural integration; Structural Determination
structural integration, of a node X in C, 26
subcategorization, 353–356, 376, 377–378;
see also Construction Domain; heads
subjacency, 7, 9, 21; see also Movement Domain; WH-movement
Subject–Simple Verb Inversion (in English), 310–311
subjects,
and case marking, 39–40, 138
ordering of, 196, 328–329, 420–21
relativization on, see relative clauses and semantic interpretation, 51–56
sentential, 196; see also sentential nodes
Structural Domain of, 27, 39
verb agreement with, see agreement see also case marking
syntactic weight, see weight
Task Urgency (of Givon),
defining the predictions for linear ordering, 217, 218
testing the predictions
on English: [NP V {PP, PP}], 220
on German: NP–PP rearrangements, 223
on Hungarian: [V {mNP, mNP}], 222
see also New before Given; pragmatic principles of linear ordering; Prague School
temporary ambiguities,
contextual resolution of, 243
semantic resolution of, 243
and structural complexity, 94–95, 242–249; see also structural complexity; Structural Domain
see also garden-parth sentences;
Impermissible Ambiguity Constraint; Late Closure; Minimal Attachment; Most Recent Filler; Performance Attachment Prediction
tense,
auxiliary positioning, 259–260
particle positioning, 408
Text Frequency Prediction (of EIC),
definition, 86–87
discussion, 84–87, 294, 452–453
testing the predictions,
on English: [NP V {PP, PP}], 123–129; Dative Movement, 212–214; Extrapolation, 190–196;
Subject index

Heavy NP Shift, 182–188;
Particle Movement, 180–182
on Finnish: [NP [NP IC]], 169–170,
[V {NP, NP}], 171
on German: Extraposition from NP,
196–210; NP–PP rearrangements,
188–190
on Greek: [NP v{V NP}], 178–180
on Hungarian: [V {m NP}], 129–133; [V {PP, PP}],
164–168; [V {m NP}], 172–173
on Japanese: [{iNP m 2 NP m} V],
140–149; [{iPP m 2PP m} V],
149–150; [{NP m PP m} V], 150–154
on Korean: [(NP m 2 NP m) V],
155–157; [{iPP m 2PP m} V], 157;
[{NP m PP m} V], 157–159; Dative–
Accusative Shift, 211–212
on Polish: [s(NP v{V NP})], 174–178
on Rumanian: [V {m IC, m IC}],
133–134
on Turkish: [(NP m 2 NP m) V],
134–137; [{NP m NP} V],
157–164
see also EIC

Theme First Principle (of Tomlin), 337–338
theta-roles, 22, 42, 424, 436; see also
directionality of theta-role assignment
topic position, 115, 339–341, 408, 425
in Finnish, 169
in German, 201
in Hungarian, 130
in Japanese, 139, 361, 363–364
in Rumanian, 133
see also discourse-configurational nodes;
focus
unprocessability, see absolute
unprocessability

Verb Object Bonding (of Tomlin), 279,
337–338
verb-second, 366, 383
vowel systems, 5, 436
VP node,
and basic word-order predictions for
subject, verb and object, 328–339
construction of, 393–402
in Japanese, 139–140, 145, 431, 451
in Korean, 155
ordering in the English VP, 282–287,
293–296, 311–313
universals of ordering in, 274–281
see also case marking; configurationality
parameter; discontinuity of
constituent structure; Grandmother
Node Construction

weight
aggregate weights for categories, 97–99,
275, 330, 334
correlations with pragmatic information
status, 238–240
weight-based principles of linear
ordering, 117–120
see also EIC; Gesetz der wachsenden
Glieder; heavier element principle
WH-movement, 9, 22, 25, 46–50, 373, 436;
see also Movement Domain
word recognition, 4, 116
written versus spoken texts (for testing
EIC), 452–453

X-bar theory, 73, 107–109, 344, 423; see
also complement phrases; heads;
specifiers