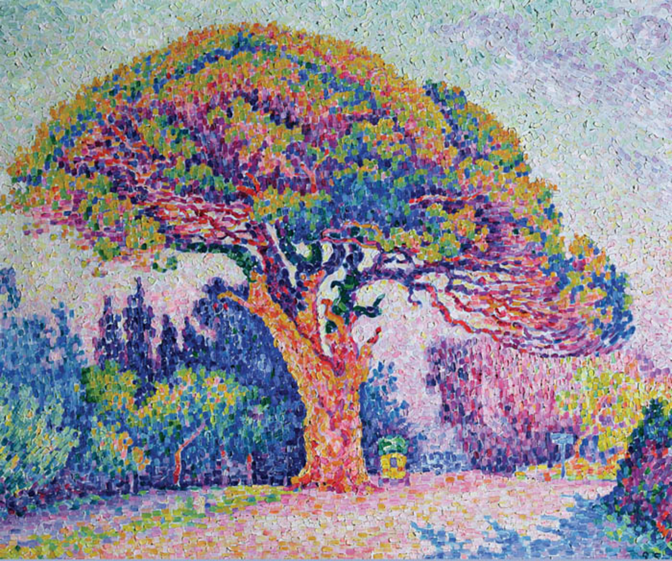


GRAMMAR & COMPLEXITY

Language at the Intersection of Competence and Performance

PETER W. CULICOVER

OXFORD LINGUISTICS

Grammar and Complexity

This page intentionally left blank

Grammar and Complexity

Language at the Intersection of Competence and Performance

PETER W. CULICOVER

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

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

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

First Edition published 2013

Impression: 1

British Library Cataloguing in Publication Data

Data available

ISBN 978–0–19–965459–8 (Hbk)
978–0–19–965460–4 (Pbk)

Printed by the MPG Printgroup, UK

Contents

Acknowledgments ix
Preface xi

Part I: Theoretical background

1 Varieties of grammatical complexity 3
1.1 Tolerating complexity, or Squinting at Dali's Lincoln 4
1.1.1 Resolution 4
1.1.2 Core and periphery 7
1.1.3 On explanation in linguistics: beyond the core and the periphery 8
1.2 The logic of complexity 10
1.3 Formal complexity—competence 14
1.3.1 Markedness 14
1.3.2 Derivational economy 16
1.3.3 Syntactic complexity 19
1.4 Processing complexity 20
1.5 Other notions of complexity 24
1.5.1 Surprisal 24
1.5.2 Optimality theory 25

2 The architecture of constructions 26
2.1 Constructions 26
2.2 Correspondences 29
2.2.1 Representing constructions 29
2.2.2 VP 33
2.2.3 Meaning 36
2.3 Constructional complexity 37
2.3.1 Regularities and idiosyncrasies in the constructional lexicon 37
2.3.2 Coverage 47
2.3.3 Strain 53

Part II: English constructions

3 Isolating constructional complexity: two case studies 61
3.1 English relatives 62
3.1.1 Inheritance and complexity 62

3.1.2 Relatives as constructions—Sag (1997) 63
3.1.3 Simplifying the description 69
3.1.4 Relative correspondences 72
3.1.5 Appositive and free relatives 84
3.1.6 Summary 85
3.2 Focus inversion 86
3.2.1 Inversion 87
3.2.2 Comparative inversion 91
3.2.3 The subject is 'low' 94
3.2.4 Focus inversion as a construction 100
3.2.5 More focus inversion constructions 102
3.2.6 Conclusions 106

4 Constructions and the notion 'possible human language' 108
4.1 Sluice-stranding 109
4.1.1 Correspondences for questions 110
4.1.2 Interpreting sluicing 112
4.2 Comparative correlatives 114
4.3 Concessives 121
4.4 Imperatives 126
4.5 Not-topics (*not in my car (you won't)*) 131
4.6 Summary 135

Part III: Processing complexity and grammar

5 Reflexes of processing complexity 139
5.1 Universals 140
5.2 Where do universals live? 142
5.3 Islands 146
5.3.1 Complexity factors 146
5.3.2 Grammatical constraint violations 147
5.3.3 Computing correspondences 150
5.4 Parasitic gaps 153
5.4.1 Parasitic gap constructions 153
5.4.2 Complexity factors in parasitic gaps 157
5.4.3 Some impossible parasitic gaps 162
5.4.4 Summary: processing gaps 164
5.5 Complex ellipsis 166
5.6 Summary 173

Part IV: Acquisition, change, and variation

6 Explaining complexity: the learner in the network 177
6.1 The paradox of complexity 177
6.2 Growing complexity 179
6.2.1 Infinitival relatives 179
6.2.2 Preposition-stranding 181
6.3 Maintaining complexity 186
6.3.1 Learning constructions 187
6.3.2 Constructions in contact: viruses in the body linguistic 194
6.4 Conclusions and perspectives 207

7 Constructional complexity and change 209
7.1 Change as explanation 209
7.2 Derivations 213
7.2.1 Affix hopping 213
7.2.2 The shift to structure: I^0 215
7.3 English constructions 221
7.3.1 Some core constructions 221
7.3.2 Ordering the correspondences 227
7.3.3 Constructional domains 230
7.3.4 *do*-support 231
7.4 Tracing the changes 233
7.4.1 Periphrastic *do* 234
7.4.2 Growth of English *do*-support 237
7.4.3 Transition to SAI 243
7.5 Summary 251

8 Integrating constructions, complexity, and change 252
8.1 CWG verb clusters 253
8.2 Derivational accounts 259
8.3 Optimality accounts 264
8.4 A constructional account 267
8.5 Two biases 270
8.6 Clusters in the network 273
8.7 Summary 281

References 285
Index 307

This page intentionally left blank

Acknowledgments

Much of the research reported on here was initiated or carried out during several visits to the University of Tübingen. I am most grateful to the Alexander von Humboldt Stiftung for their generous support, and to my sponsors, Erhard Hinrichs, Marga Reis, and Susanne Winkler, all colleagues and dear friends. I owe a special debt of gratitude to Susanne Winkler for her unflagging energy and innumerable contributions to our ongoing collaboration, and to Katharina Schmeh for her continuing friendship and support. I am also grateful to the Ohio State University for a sabbatical leave that has made it possible to complete the work in a timely fashion. Thanks to my colleagues for making the Department of Linguistics at Ohio State a very special place in which to teach and pursue research. And thanks to Beth Hume for everything that she's done to help me be as productive as I can be.

I first conceived of this project in a series of conversations with Vladimir Sloutsky, whose good-natured and broad-ranging skepticism about domain specificity I took as a special challenge. This challenge has occupied me for much of the past half-dozen years, and will likely continue well into the future.

I benefited enormously from discussing many of the ideas touched on in this book in a seminar at Ohio State during the spring of 2008 with Pilar Chamorro, Dina Kapetangianni, Lia Mansfield, Vedrana Mihalicek, John Pate, Cory Shain, Oxana Skorniakova, Laura Wagner, and Chris Worth. A timely visit to Freiburg gave me the opportunity to sharpen up some of these ideas with Guido Seiler, Daniel Jacob, and their students.

As always, thanks to my friend and collaborator Ray Jackendoff, who has helped me so much and in so many ways. A special thanks to my editor, John Davey, for his friendship, support, encouragement, and keen enthusiasm for pushing the boundaries of linguistic theory for these many years. I am grateful to two reviewers for their helpful comments and suggestions. Thanks to Wolfgang Sternefeld, who read and commented on a substantial part of the manuscript.

For helping me to find balance, I am deeply indebted to Anne Comarda, Abby Dorn, Alex Ferm, Chris Kirk, Julia McSheffery, Fenella Roche, and Donna Winters. Om Shanti.

Parts of this text have been adapted from other pieces that I have published, and appear here through the kind permission of the original publishers. These are:

Culicover, Peter W. (2008), "Squinting at Dali's Lincoln—How to think about language," *Proceedings of the Forty-First Annual Meeting of the Chicago Linguistics Society* (Chicago: Chicago Linguistics Society), 109–28: portions of Chapter 1.

Culicover, Peter W. (2008), "The rise and fall of constructions and the history of English *do*-support," *Journal of Germanic Linguistics*, 20: 1–52: Chapter 7.

Culicover, Peter W. (2011), "A reconsideration of English relative constructions," *Constructions*, 2, available online at <http://elanguage.net/journals/index.php/constructions/index>: first half of Chapter 3.

Culicover, Peter W. and Susanne Winkler (2008), "English focus inversion constructions," *Journal of Linguistics*, 44: 625–58: second half of Chapter 3.

Salvador Dali's Gala Contemplating the Mediterranean Sea which at Twenty Metres becomes a Portrait of Abraham Lincoln (1976) is reproduced by permission of the Fundació Gala-Salvador Dali.

And finally, all my thanks and love to Diane and Dan, for making this and everything else possible and worth doing.

Preface

This book is an investigation of the idea that the grammar of a natural language is determined in part by complexity; hence the title, *Grammar and Complexity*. This is an idea that has been prominent in linguistic theorizing right from the beginning of generative grammar. It is found in the early work on markedness, in more recent proposals regarding derivational economy and constructional inheritance, and it runs through argumentation that appeals to some notion of "naturalness" that draws upon simplicity of description.

The subtitle, *Language at the Intersection of Competence and Performance*, is intended to invoke the idea that the grammar is a product of two forces: (i) pressure to reduce the formal complexity of the grammar, and (ii) pressure to avoid complexity in applying the grammar to actual linguistic expressions, that is, processing complexity.

In order to illuminate the relationship between grammar and complexity, we need to have a reasonably precise idea of what a grammar is, and how the complexity of a grammar is measured. Both topics are rightfully contentious, since in neither case are there any widely accepted *a priori* principles to guide us, beyond good scientific method. Consequently there are a lot of proposals on the market about what constitutes grammar, and about what constitutes complexity, both formal complexity and processing complexity.

In order to keep the investigation within more or less manageable bounds, I adopt the following framework.

- Form–meaning correspondences: I focus on the relationship between syntactic structure and interpretation, leaving aside for the most part questions having to do with phonological form. Hence when I refer to "correspondences between form and meaning," I mean syntactic form, including constituent order, morphological form, and hierarchical structure, on the one hand, and literal meaning, information structure, and discourse representation, on the other.
- Constructions: Within the domain of syntax, I take 'constructions' to be a useful level of description to account for these correspondences.
- Grammatical complexity: I assume, following most contemporary work, that the grammars of natural languages are less complex, on some well-defined measure of complexity, than arbitrary collections of logically possible constructions.

- Grammaticality: I take “grammatical” to mean “licensed by the grammar” in a narrow formal sense. Assuming that a grammar consists of a set of constructions that express form–meaning correspondences, a grammatical sentence is one whose structure satisfies the syntactic description given by some subset of these constructions, and whose meaning conforms to the corresponding representation given by that set of constructions.
- Acceptability: Crucially, grammaticality is different from “acceptability,” which has to do with the subjective judgment of how “good” or “odd” a sentence sounds. A sentence may be grammatical in formal terms, but less than fully acceptable, for reasons having to do with processing complexity, semantic or pragmatic ill-formedness, prosodic constraints, and so on. An ungrammatical sentence is (usually) unacceptable, but not all unacceptable sentences are ungrammatical.

Part 1 sets out the theoretical background in which these points are elaborated. Chapter 1, “Varieties of grammatical complexity,” reviews a number of recent prominent views about complexity. It highlights the two types of complexity that I focus on in this book, formal complexity and processing complexity. I take formal complexity to be the measure of the amount of idiosyncrasy in a grammar, a measure that corresponds with the extent to which there are phenomena in a language that cannot be accounted for as special cases of other, more general phenomena. Processing complexity is a measure of the resources required to compute the correspondences between particular forms and their meanings, and, I argue, corresponds to subjective judgments of unacceptability when it exceeds some threshold.

Chapter 2, “The architecture of constructions,” lays out the view of a grammar as a set of constructions, and contrasts it with familiar derivational alternatives. The perspective of Chapter 2 draws heavily on ideas presented in my earlier work, especially Culicover 1999 (*Syntactic Nuts*) and Culicover and Jackendoff 2005 (*Simpler Syntax*).

Part II deals with particular English constructions. Chapter 3, “Isolating constructional complexity: two case studies,” analyzes in some detail two cases of formal complexity. The first case concerns relative constructions. I show that relative clauses and other relatives share an interpretation that each particular syntactic configuration corresponds to in a systematic way. This correspondence simplifies and unifies the description of the set of constructions. As a consequence of establishing this constructional uniformity, it is possible to isolate precisely the respects in which particular special cases are idiosyncratic. This exercise lends support to the view that grammars are less complex than arbitrary collections of logically possible constructions.

The second case concerns focus inversion in English. There is a class of constructions in English in which the subject follows the main verb. These constructions constitute counterexamples to the general pattern of the language, which is that the subject precedes the main verb. I argue that these are distinct, special-purpose constructions whose function is to mark the subject as contrastive. Furthermore, I argue that it is no accident that these constructions have this property—VP-final position is a typical position for focus in English, and the focus inversion constructions take advantage of this fact.

Chapter 4, "Constructions and the notion 'possible human language'," continues in the same vein. I show that in a diverse set of constructions identified in *Syntactic Nuts* and elsewhere, there is significant systematicity accompanied by particular aspects of idiosyncrasy. But the idiosyncrasy is not entirely arbitrary. While idiosyncrasy, and thus formal complexity, exists in natural languages and must be accounted for in grammars, a constructional approach allows us to see how the idiosyncrasy is constrained by its relationship to the more general.

Processing complexity is the focus of Part III. One of the most significant consequences of the Acceptability thesis stated above is that not every case of unacceptability is necessarily a grammatical phenomenon. Some unacceptability is the consequence of the complexity of computing the form–meaning correspondence. What this means in practical terms is that certain principles that have been formulated over the past fifty years or so to account for the "ungrammaticality" of classes of sentences of English and other languages may in fact have nothing to do with grammar or grammaticality per se. Rather, they are arguably descriptions of syntactic configurations that lead to processing complexity. Removing these phenomena from the domain of grammar opens up the possibility of defining more precisely and narrowly the scope of grammatical theory, and thus the notion 'possible human language.'

Chapter 5, "Reflexes of processing complexity," extends the explanatory role of processing complexity to a range of phenomena that have been accounted for by putative grammatical constraints. Drawing heavily from the recent literature, I argue that such constraints are not part of the theory of grammar. Rather, they are descriptions of syntactic configurations, and perhaps also form–meaning correspondences, that lead to processing complexity, and thus unacceptability.

Part IV turns to the question of why complexity exists at all in language, given that there are strong pressures to reduce and even eliminate idiosyncrasy in grammars. The answer given in Chapter 6, "Explaining complexity: the learner in the network," is that complexity can arise as a consequence of language acquisition and can be sustained against the pressure to eliminate it by the structure of the social network. As before, the constructional framework proves to be central to the explanation. Learners are predisposed to acquire constructions and generalize them beyond the available evidence in very restricted ways. Competing constructions reside in the social network, and even in individuals. The less complex construction will drive out the more complex construction, unless there is sufficient support for the more complex one. This support can take the form of high frequency of occurrence, limited contact between individuals with different constructions, or geographical isolation of the more complex variety. Similarly, cases of processing complexity can be acceptable in a given language where they are strongly supported by high frequency of occurrence and lack of competition from less complex alternatives.

In Chapter 7, "Constructional complexity and change," I track the history of a number of related constructions in English. I argue that the changes can be understood as the consequence of the pressure to reduce complexity, expressed in terms of the formal complexity of constructions. I show how this approach can be applied to the shift from fully general V2 to more specialized subject aux inversion and to the rise of fully general *do*-support.

Chapter 8, "Integrating constructions, complexity, and change," attempts to bring together the ideas about constructions, social networks, and complexity in the service of explaining some of the fine details of language variation and change—the phenomena concern the variability and distribution of verb clusters in Continental West Germanic. I argue that the variation can be seen as the effect of competing complexity biases that favor different constructions; the competition takes place in a social network that favors one or the other alternatives for non-linguistic reasons.

PART I

Theoretical background

This page intentionally left blank

1

Varieties of grammatical complexity

This chapter highlights the two types of complexity that I focus on in this book, formal complexity and processing complexity. Formal complexity is the measure of the amount of idiosyncrasy in a grammar, that is, phenomena that cannot be accounted for without special stipulations. Processing complexity is a measure of the resources required to compute the correspondences between particular grammatical forms and their meanings.[1] I argue that when it exceeds a certain level, processing complexity corresponds to subjective judgments of unacceptability. These two types of complexity are distinguished below from a number of other ideas about what constitutes complexity in language, some of which may turn out to be related to formal and processing complexity.

These two conceptions of complexity are central to our understanding of what it means to know a language, what knowledge of language consists of, and how we describe the grammars of natural languages. The status of idiosyncrasy in grammar is by no means a settled question in grammatical theory, and how we handle it has fundamental consequences for the form of the theory. Processing complexity, on the other hand, bears on the data that we take into consideration in formulating grammatical descriptions, and so, by extension, grammatical theory. As shown by Miller and Chomsky (1963), it is possible for a particular sentence to be unacceptable due to processing demands, but grammatical, that is, it meets the conditions of well-formedness imposed by the grammar of the language.

Such phenomena are well-known, but their fundamental importance is not always recognized. The basic fact is that we do not in general have distinguishable subjective judgments about *why* a sentence is unacceptable in a language.

[1] I use the term processing complexity here to refer to the complexity of processing by the human mind. Thus I leave aside more abstract questions of computational complexity, i.e., those that bear on the complexity of processing by an ideal computational device, such as a Turing machine. Quite possibly the first could be formalized in terms of the second, but I do not pursue such a formalization here.

A sentence may be unacceptable because of some processing complexity, because it violates some rule of grammar, because it expresses a meaning that is strange in some way, because the way it packages information does not conform to the structure of the information already present in the discourse, or because of some other infelicity or mismatch.

Speakers are sometimes able to recognize the source of the unacceptability, even without formal training. For example, we can usually recognize when the words are in the wrong order in a string. But in many cases, all we know is that the sentence is odd. *Why* it is odd requires a theory.

As a consequence, the description of what speakers know, that is, the grammar, cannot be arrived at simply by observing and cataloguing speakers' judgments about which sentences are acceptable and which are unacceptable. Sentences may be grammatical, that is, well-formed, but unacceptable for reasons that have nothing to do with grammar per se. If we are trying to understand what grammar is, and how to properly construct a grammatical theory, it is circular to suppose that all acceptability judgments bear directly on the grammar. To do so may be a useful idealization, and such an idealization has in fact proven very fruitful in much of grammatical theorizing over more than half a century. But serious consideration has to be given to the possibility that there are cases of unacceptability for which the grammar is not responsible.[2]

In the next few sections I look at some of the ways in which complexity has been addressed in linguistics and the role that it has played in formulating grammatical theories. This review is not intended to be comprehensive or exhaustive, but seeks to situate the cases of formal and processing complexity that I address in subsequent chapters within the broader discussion of 'linguistic complexity.'

1.1 Tolerating complexity, or Squinting at Dali's Lincoln[3]

1.1.1 Resolution

One important aspect of grammatical complexity has to do with the relationship between idiosyncratic detail and the backdrop of systematic regularity. For an analogy in a non-linguistic domain, consider the painting by Salvador Dali in Figure 1.1.

For most viewers familiar with American history, the image is easily identified—it is a painting of Abraham Lincoln. It is possible to see the

[2] For additional discussion of this point, see Culicover and Winkler (2010), as well as the papers in Culicover and McNally (1998).

[3] Portions of this section appeared in Culicover (2008a).

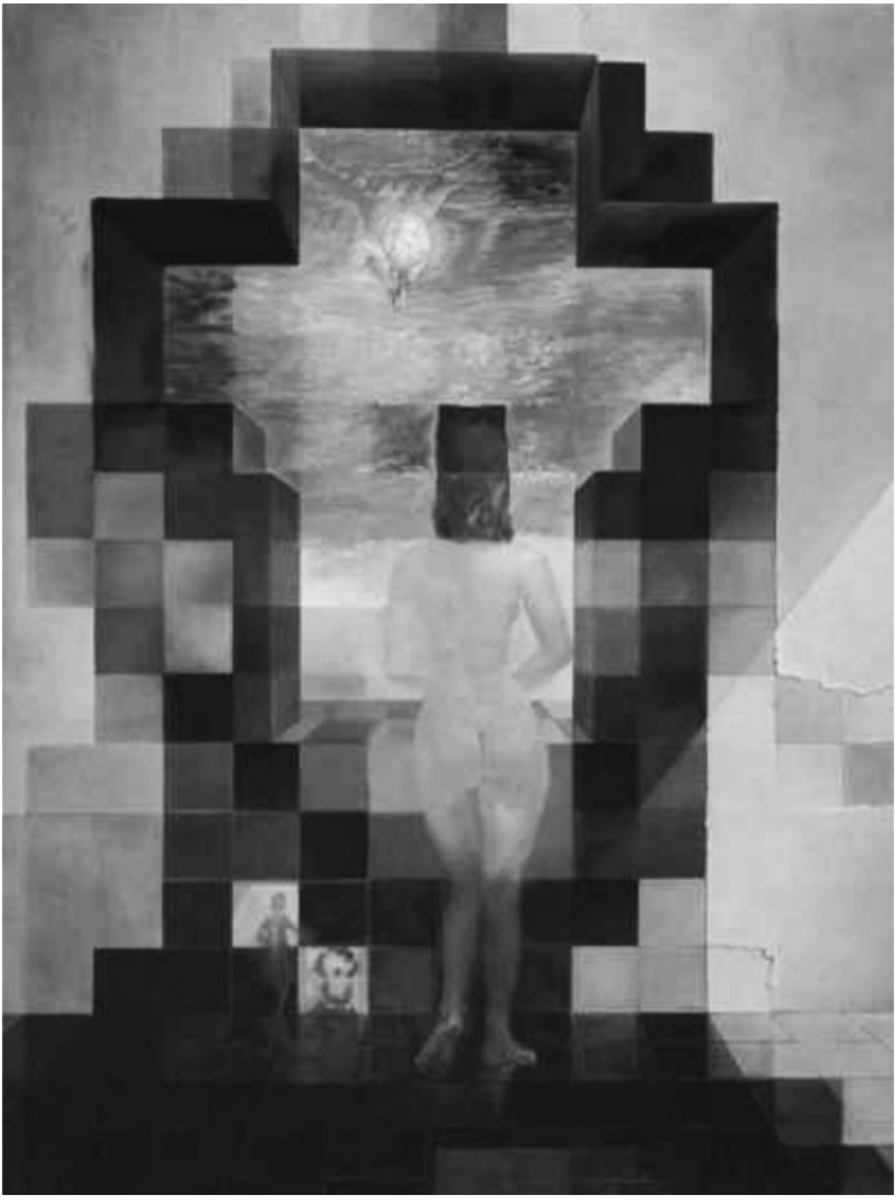


FIGURE 1.1. *Gala Contemplating the Mediterranean Sea Which at Eighteen Metres Becomes the Portrait of Abraham Lincoln*, (Homage to Rothko), 1976, by Salvador Dali, (1904–89)/ Museo Dali, Figueres, Spain/Index/The Bridgeman Art Library. © Salvador Dali, Fundació Gala-Salvador Dali, DACS, 2012

FIGURE 1.2. Small Dali's Lincoln. © Salvador Dali, Fundació Gala-Salvador Dali, DACS, 2012

image of Abraham Lincoln more clearly by squinting, and the more one squints, the more it looks like Lincoln. Alternatively, if we make the image smaller, the image of Lincoln is also easier to see (Figure 1.2).

In either case, that of squinting or resizing the image, we are able to see the image by eliminating information, in fact, a considerable amount of information that is not essential to the image of Lincoln.

But, note that there is *so much information* in the Dali painting that does not conform to the image of Lincoln that even with a lot of squinting, or making the image very small, the image is at best suggestive of Lincoln.

The basic point here is that when there is a mass of detail, squinting may bring out the structure, while losing the detail. In large measure, whether this actually happens depends on the organization of the information. In Dali's Lincoln, the image of Lincoln emerges from dark and contiguous regions, while the information that is irrelevant to the image of Lincoln is light in color and soft in outline, or very small (e.g. the small Lincoln in the lower left region of the image). This information can be ignored or simply is lost when we squint or when the image is made very small.

There are several conclusions to draw. One is that the structure is really "there" in some sense, but it is not "there" in another sense. It is emergent from the ways in which the parts of the whole interact with one another, and cannot be isolated from it. A second is that the "irrelevant information" is as real as the emergent structure, but less well organized. Nevertheless, it must be taken into account as part of a description of the whole. More importantly, its presence is intimately tied to the apparent structure, and contributes to recognition of that structure. This perspective is, in effect, Dali's philosophy of science.

It is also a perspective that is directly applicable to syntactic theory. Syntactic theories in the mainstream tradition, that is, *Syntactic Structures* (Chomsky 1957) through the Minimalist Program (Chomsky 1995), have progressed largely through the equivalent of large-scale squinting.

The argument made in chapters 2 and 3 of *Simpler Syntax* (Culicover and Jackendoff 2005) is, essentially, that if there is a significant amount of important detail, as there is in the case of Dali's Lincoln, massive squinting gives a distorted view of what the object that we are looking at really is.

1.1.2 Core and periphery

The analog in language to Dali's Lincoln is, in effect, the well-known 'core–periphery' distinction. On the mainstream view, which is an idealization, the core is comprised by Universal Grammar (UG) (subject to parametric variation), while the periphery is a less well organized and less well understood body of phenomena such as marked exceptions and irregularities that arise from processes such as "analogy" and historical change.

The core–periphery distinction is related to 'markedness.' For Chomsky (1965) (*Aspects*), markedness is a graded phenomenon that reflects relative centrality, naturalness, simplicity, ease of learning, and related notions. The introduction of the core–periphery distinction can be seen as a distillation of the notion 'markedness hierarchy' into a binary distinction. It dispenses with more/less marked, by locating the unmarked in the core and the marked (or more marked) in the periphery.

The consequence is a dramatic conceptual simplification (corresponding to our "squinting"), which connects naturally to the characterization of UG in terms of parameters. On the mainstream view, the core and the parameters are part of the human biological endowment for language. The value of a parameter is set by the learner on the basis of minimal linguistic input.[4]

One important consequence of the core–periphery distinction, particularly in syntactic theory, is that it has focused considerable attention on understanding the formal devices that languages do and do not use to realize universal phenomena such as argument structure and *wh*-interrogatives. Another is that it has led to the uncovering of a wide range of empirical phenomena in the attempt to integrate apparent exceptionality, idiosyncrasy, and counterexamples into a general framework of universals and parametric restrictions.

Despite the methodological value of the core–periphery distinction, it is clear that relegating all exceptionality and idiosyncrasy to the periphery is an overly strong simplification. Chomsky himself notes that "we do not expect to find chaos in the theory of markedness, but rather an organized and structured system, building on the theory of core grammar" (Chomsky 1981b: 216), and "marked structures have to be learned on the basis of slender evidence

[4] For discussion of just how much input is sufficient, see Fodor (1998).

too, so there should be further structure to the system outside of core grammar" (Chomsky 1981b: 8). Chomsky does not say, however, how this aspect of the system works.

It is in fact not clear that the core–periphery distinction can be sustained as a principled distinction (even if it turns out to have been a useful methodological device) nor that it reflects anything beyond generality of function and frequency of use. As argued in Culicover (1999) (*Syntactic Nuts*), for example, apparent syntactic idiosyncrasies beyond the level of individual words are learned, they display various degrees of specificity, and native speakers have sharp and reliable intuitions about them, just as they do for putative core phenomena. Furthermore, Occam's Razor demands that it be shown that a learning mechanism that can acquire the peripheral cases cannot also acquire the core. Hence, core–periphery may indeed be nothing more than a "rough and tentative distinction," one drawn "for working purposes (and nothing more than that)" (Chomsky 1993: 17–18).

The possibility that core–periphery is merely a distinction that has been drawn for practical reasons is certainly consistent with the idea that the structure that we see in language is not "there" in the sense that holds in the idealization. But if this is the case, then these questions arise (again), as they do in Dali's Lincoln: Where is this structure? Why can we see it? What is the nature of the cognitive faculty that gives rise to the appearance of structure? In order to begin to answer these questions, we have to consider a bit more closely the relationship between the simple structure of the core and the more complex detail in the periphery.

1.1.3 On explanation in linguistics: beyond the core and the periphery

In its earlier incarnations, mainstream syntactic theory had the goal of accommodating the higher degree of detail actually found in the data. This accommodation is often called "the theory of markedness." Here's a quote from Chomsky (1977: 76), referring back to the 'Conditions' framework (Chomsky 1973) that captures the essential idea in discussing the role of constraints in a theory of grammar:

> [O]ne might construct a rule to 'violate' the A-over-A condition, but only at a cost: the rule would have to make explicit the relevant structures so that it can apply without failing under the condition. 'The logic of this approach . . . is essentially that of the theory of markedness.' That is, the conditions become an integral part of an evaluation measure, rather than imposing absolute prohibitions.

On this view, which appears as early as *Aspects*, there are many possibilities in natural language that are complex in some sense; some are too complex to actually occur, while others are relatively rare. It is possible, in principle, to rank various phenomena in terms of their complexity, using the "cost" associated with the descriptions.

Chomsky subsequently set aside markedness theory in favor of a more sharply etched approach, relegating the complexity to the periphery and focusing on the core part of the core–periphery pair. The complexity was dealt with through a radical idealization, that of an ideal speaker-hearer in a homogeneous environment with instantaneous learning, who chooses among a small set of parameter values on the basis of the linguistic experience. And the conditions on structures and derivations in fact came to be interpreted as "absolute prohibitions."

One measure of Chomsky's insight is that he saw at the outset how dealing directly with the complexity in terms of a rich markedness theory would constitute a substantial if not impenetrable barrier to research progress, and he found a straightforward and effective way around the barrier, through the radical idealization of core–periphery. Ironically, perhaps, the consequence of this radical idealization has been not only a series of important insights into the structure of what we know when we know a language, but also a more refined appreciation of the true complexity of the phenomena that we are dealing with, but are perhaps unable to account for satisfactorily within the idealization.

It is the contention in *Simpler Syntax* (Culicover and Jackendoff 2005) that the focus on a narrow (and increasingly narrower) set of phenomena to get at the latent structure excludes from consideration a vast amount of the detail that exists in natural language. In effect, we choose not to see the smaller images in Dali's Lincoln. These images are real and can be ignored only at the risk of possibly seeing something that is not really "there."

In the case of language, the fine detail is not ignored, it is real. It is effectively acquired by children and represented in the minds of adult speakers, and therefore must be accounted for by linguistic theory. Our intuitions as native speakers are as sharp and reliable about the periphery as they are about the core. The picture that *Syntactic Nuts* and *Simpler Syntax* present is one in which there is no principled core–periphery distinction. The core is the extreme end of a continuum of form–meaning relations that is characterized by maximal generality and semantic transparency, e.g. what can be described in terms of phrase structure rules, robust filler-gap dependencies of the *wh*-movement variety, and so on. But there is a vast array of less than general detail along this continuum to account for as well, from the individual word, to idioms, to semi-regular derivational morphology, to fixed expressions with open slots (like *to put one's money where one's mouth is*), to constructions that reflect the basic structures of a language but have special idiosyncratic properties (like *to* V(erb) *one's way* PP_{path}), to syntactic nuts, which are semi-productive but idiosyncratic syntactic constructions (like sluice-stranding: *She said she was talking with someone, but I can't remember who with*).

The evidence suggests, in fact, that learners are capable of learning very specific details, small-scale generalizations, idiosyncratic exceptions, and lexically restricted but semi-productive syntactic and morphological constructions. Generalization, when it occurs, may go beyond the actual experience, but does not appear to be so radical as to induce maximal uniformity on the basis of limited experience. We need an architecture for language acquisition and processing and the representation of knowledge that captures this state of affairs.

What follows in this book is an exploration of the possibility that genuine explanation can be found for the properties of grammar and of grammars in terms of complexity. It consists of detailed explorations of particular linguistic phenomena in English and other languages, of arguments for particular architectural features of grammar, of arguments for a particular conception of the term 'complexity' as it applies to grammar, and of scenarios that suggest the role that complexity might play in the course of language acquisition, language interactions in social networks, and the language change that arises as a consequence of these interactions.

1.2 The logic of complexity

A persistent theme in linguistic theorizing has been that 'complexity' (or the related notions of 'simplicity,' 'economy,' and 'markedness') form an important part of the explanation for why the grammars of particular languages are they way they are, and why grammar in general is the way it is. This idea appears to be relatively straightforward when it is applied in a commonsense way to particular phenomena, for example the following.

- A general "rule" is intuitively less complex than a list of special cases, perhaps because the rule makes fewer demands on memory or demands less of the learner in terms of hypothesis testing, confirmation, and correction.
- Children acquire and produce words that refer to everyday objects such as trucks sooner than words that refer to time and modality, perhaps because the latter concepts are more complex in some sense than the former, perhaps because they are more "abstract."
- Languages arguably lack certain syntactic configurations because they require too many resources (e.g. memory) to parse and to relate to meanings, and in this sense are "complex."
- A language change takes place because the resulting grammar is "simpler" than the one that it replaced. (This is a very old idea.)

While such ideas are intuitively plausible, they demand a deeper understanding of what counts as more or less complex, and why. 'Complexity' must be defined independently of what is rare or non-existent; otherwise, our account is circular.[5] And, ideally, our notion of complexity should account for language processing, language acquisition, language change, and variation.

But we do not yet have an independent measure of complexity. Hence we have to make inferences about what is more or less complex based on our understanding of language processing, language acquisition, language change, and variation. Moreover, we need to understand the architecture of the system that underlies these dynamical aspects of language, in order to explain why a linguistic structure that requires significant resources to process (that is, to parse and interpret), is also relatively difficult to learn, why it would also be relatively rare in languages and in corpora, and why languages would change so as to minimize its occurrence.

Here is an example that illustrates the difficulty of constructing non-circular explanations in this arena. Given that the word order VSO is somewhat common and OSV is very rare in the world's languages,[6] we could speculate that the order OSV is more complex for the purpose of expressing thematic roles than the order VSO. One plausible scenario is that if the arguments S and O are encountered before the verb, it is necessary to hold them in memory until the verb comes along to indicate what roles to assign to them. But if the verb is encountered first, it comes with its roles, and each argument can be assigned a role immediately.

This scenario immediately runs into problems with the observation that SOV is the most common order.[7] Perhaps we would then say that there is less memory required when the first argument in a sentence is always a subject. That is, SV & SOV is less costly in processing terms than SV & OSV. Or perhaps case-marking in V-final languages allows a hearer to assign default thematic roles before the verb is encountered, so having the S always first somehow facilitates this default assignment. But without independent evidence about how memory is allocated in processing, or whether there are default roles, or whether case-marking triggers default role assignment, there is no genuine explanation. And beyond this, it is by no means obvious why

[5] For discussion of the claim that rare and complex correlate, see Newmeyer (2007).

[6] In Map 81A of the *World Atlas of Linguistic Structures* (http://wals.info/chapter/81), there are 95 languages with the dominant order VSO and 4 with the order OSV (Dryer 2011).

[7] In Map 81A of the *World Atlas of Linguistic Structures* (http://wals.info/chapter/81), there are 565 languages with the order SOV (Dryer 2011). Cf. footnote 6.

complexity for the hearer should explain typological generalizations (see Wasow 1997).[8]

For functionalists, the relationship between complexity and typological generalizations goes in both directions, and draws on the idea that language users seek to minimize the "effort" of using language (Newmeyer 2000). On one hand, what is universal should be simple, because it is something that every language has to do and therefore should require minimum effort. On the other hand, what is complex will not be common, if it occurs at all, because its complexity will lead users to avoid using it.

For non-functionalists, e.g. Chomsky (1995), the relationship is not as explicit, but interestingly, still draws on the assumption that languages seek to achieve simplicity—in Chomsky's case even ultimately "perfection" in some absolute sense. "Perfection" in this sense is achieved by maximizing economy of derivations, where the cost of a derivation is measured in terms of the number of formal operations required to describe it, as well as properties of these operations.

These two examples illustrate two perspectives on linguistic complexity. Miestamo (2006, 2008) characterizes them as follows:

(1) **absolute**—complexity as an objective property of the system
relative—complexity as cost/difficulty to language users

Absolute complexity has to do with the amount of information required to describe a system, that is, formal complexity in the sense introduced earlier. The more information that it takes to describe the system, the more complex it is. This is essentially the approach taken by Chomsky at least until *Aspects*, taking markedness as the measure of complexity. Relative complexity, that is, processing complexity, has to do with the amount of resources required to compute linguistic representations.

Dahl (2009) calls relative complexity "agent-related" complexity. For Dahl this notion of complexity is associated with cost or difficulty, while absolute complexity is complexity in a formal sense. (See also Dahl 2004.)

To take another apparently quite different case, McWhorter (2001) argues that creoles are less complex than other languages in virtue of having simpler descriptions. On this view, a language that makes more morphological

8 It is certainly possible that what is more complex for the hearer is also more complex for the speaker. In order to show that this is indeed the case, we would have to know a lot more than we do about the actual mechanisms used in production and comprehension.

distinctions in a given category, for example, grammatical number, is more complex than one that makes only two such distinctions.[9]

But in both cases, it is not obvious what the formal complexity of description has to do with what we want complexity to explain, as noted above. This is not to say that there is no intuitive sense to the notion that, for example, a language with simpler descriptions should be easier to learn or easier to process—after all, there is a lot less for people to remember if the language doesn't have any inflectional morphology. But without a reasonably precise characterization of the learner, memory requirements, the dynamics of accessing and retrieving idiosyncratic items, the mechanisms for computing general rules, and the processor that constructs interpretations given items of various degrees of morphological specificity, this notion is no more than an intuition.

McWhorter's complexity metric is given in (2).

(2) a. An area of grammar is more complex than the same area in another grammar to the extent that it encompasses more overt distinctions and/or rules than another grammar.
b. A syntax is more complex than another to the extent that it requires the processing of more rules, such as asymmetries between matrix and subordinate clauses (e.g., Germanic verb-second rules), or contains two kinds of alignment rather than one.
c. A grammar is more complex than another to the extent that it gives overt and grammaticalized expression to more fine-grained semantic and/or pragmatic distinctions than another.
d. Inflectional morphology renders a grammar more complex than another one [that lacks inflectional morphology - PWC] in most cases.

Except for (2b), these are all connected to the complexity of the grammar as a whole, that is, global complexity. We may contrast global complexity with the local complexity of the form–meaning correspondences themselves, measured in terms of the formal statement of individual correspondences and the alternative correspondences that express the same meaning. The main differences between the two types of complexity are summarized in (3).

(3) a. **global (grammatical) complexity**
i. devices for conveying explicit meaning
ii. marking semantic distinctions

[9] I discuss this idea further in Chapter 6.

b. **local (constructional) complexity**
   i. many constructions for same correspondence
   ii. idiosyncrasies and exceptions
   iii. specific conditions on correspondences
   iv. formal conditions on well-formedness that do not correspond to meaning

As the comments on McWhorter (2001) attest (see *Linguistic Typology*, 5: 167–87), as well as the papers in Sampson et al. (2009), the question of relative global complexity is very difficult if not impossible to resolve. There are issues about whether and how the proposed measures actually correspond to complexity, whether it is possible to actually apply these measures effectively, whether the assessment of the complexity of particular aspects of particular grammars is accurate, and so on.

Much of the disagreement revolves around the difficulty of defining, measuring, and comparing global complexity. For some discussion of whether languages are overall more or less complex than other languages, see Sampson et al. (2009), Gil (2008), and the papers in Miestamo et al. (2006) and Everett (2005). Gil (2008) argues that languages that are morphologically simpler do not necessarily compensate by having more a complex mechanism of semantic interpretation. The additional question of whether there are social and cultural correlates of complexity is taken up in Trudgill (2011) and Everett (2005).

With the foregoing as background, I review in the next two sections the main proposals in the literature for characterizing local complexity.

1.3 Formal complexity—competence

1.3.1 Markedness

An important thread in the development of theories of grammatical complexity is the perspective of markedness that Chomsky sketched out in the sixties but did not pursue extensively in later work (e.g., Chomsky 1964, 1965; Chomsky and Halle 1968). The general idea is that something that is more complex in formal terms is dispreferred by learners, other things being equal, and consequently less frequently found in languages.[10] Markedness is essentially a measure of the naturalness of a linguistic construct *in relation to the grammar*. In phonology it is realized as an ordering of feature values, possibly in a context—e.g., voicing is more natural intervocalically than non-voicing. In syntax it has been tied to relative generality—e.g. extraction of a *wh*-phrase

[10] For extensive discussion of the logic and history of markedness, see Battistella (1996).

from all syntactic contexts is less marked than extraction of a *wh*-phrase from a restricted set of contexts.

In Chomsky's early work, markedness is assumed to correspond to complexity of the description. On this view, the more information that is required to define some relation, the more marked it is. This perspective is in fact quite explicit in Chomsky's (1964) discussion of markedness, in the context of the A-over-A condition. In Chomsky's formulation, departure from a general principle (in this case extraction of a *wh*-phrase from an A-over-A configuration in violation of the A-over-A condition) requires a more specific statement, compared to the general rule. The approach assumes that there are many conceivable configurations of linguistic elements that can constitute grammars of natural languages, but not all have the same cost. Those that are too costly are those that are not hypothesized by a learner on the basis of the available evidence. Chomsky (1965: 53–4) discusses the "feasibility" of a theory of grammar in terms of the extent to which the available hypotheses are "scattered" with respect to evidence.

In fact, Chomsky allows for the possibility that something can be described within the formal framework of grammar that does not, and cannot, occur in a natural language. He writes (Chomsky 1965: 35), "In fact, the real problem is to restrict the range of possible hypotheses [about a grammar—PWC] by adding additional structure to the notion 'generative grammar'. For the construction of a reasonable acquisition model, it is necessary to reduce the class of attainable grammars compatible with given primary linguistic data to the point where selection among them can be made by a formal evaluation measure." That is, there are grammars that are definable within the formal system, but not "attainable," because they do not satisfy certain conditions imposed by the evaluation metric (that is, universal grammar).

On this view, a construction may be too marked to be acquired by a learner, or it is correspondingly too complex to survive in the competition among alternatives in the course of acquisition. Crucially, the markedness computation is part of the learning mechanism for language, and it guides the learner in formulating hypotheses about what the grammar is for a given body of linguistic data. The learner is itself oriented to the maximally unmarked formulation consistent with the data. This is the logic of markedness in *Aspects*.

This notion of complexity clearly shares conceptual underpinnings with those found outside of linguistics such as Kolmogorov complexity (e.g. Cover et al. 1983) and minimum description length (e.g. Grünwald 1996). It is central to work on the computational modeling of language evolution (e.g., Brighton et al. 2005). Minimum description length is a measure of the redundancy in a grammar (Brighton 2005; Juola 2008). The more redundancy there is, the simpler the description of the grammar. The more idiosyncrasy there is, the more complex the description of the grammar. A typical source of complexity in this sense has to do with paradigmatic uniformity, where the form of every member of a category can be described simply in terms of the paradigm, but exceptions (gaps and suppletions) must be specified separately. It is even possible that exceptions themselves follow a simple pattern, a situation that is in turn less complex than random exceptionality (Sims 2009).

However, markedness in linguistics is not just about formal complexity, but about the privileged status of some forms or properties over others, as in the work of Trubetskoy (1939), that is, 'bias' (cf. Chapter 7). It is possible that in some cases, the privileged status of the unmarked can be reduced to its formal properties. But this is not always the case, as can be seen in the case of the core–periphery distinction.

The introduction of the core–periphery distinction (Chomsky 1981a) reformulates markedness by putting those aspects of grammar that are putatively universal in the core, while relegating to the periphery those that are exceptional, idiosyncratic, and language-specific. By definition, core phenomena do not have to be learned, and are therefore cost-free from the perspective of markedness or complexity. Those in the periphery have to be acquired, and may require special-purpose learning mechanisms and possibly special machinery in order to be accounted for in a grammatical description.

As with many of the distinctions made in contemporary mainstream generative grammar, the core–periphery distinction is essentially on the right track. However, because it is categorical, it is not sufficiently refined to be able to play an explanatory role in accounting for the form of actual grammars. I argue in Chapters 3 and 4 that some constructions in natural languages express fundamental functions or aspects of meaning, while others are less fundamental. The relative centrality of a construction may not have anything to do with its formal complexity in the sense of markedness, however.

1.3.2 Derivational economy

The notion of 'economy' in the Minimalist Program (Chomsky 1995) is stated as an abstract version of processing complexity. The complexity of derivations is minimized, other things being equal, and the less complex alternative is implemented in the grammar. Going beyond the core–periphery distinction, a notion of economy is introduced in Principle and Parameters Theory (PPT) and the Minimalist Program (MP), to account for what does and does not occur, within the general framework of core grammar. On this approach, derivations that require fewer steps or are shorter (as measured in terms of syntactic structure) are more economical than those that require more steps or are longer. Economy is thus a way of imposing Chomsky's

(1965) feasibility criterion on core grammar, to rule out grammars that are otherwise capable of being formulated in the theory.

Since economy of derivation is intended to be understood in terms of the computational requirements of an abstract device, it appears to be a shift from absolute to relative complexity in the sense discussed above. But the relationship between this abstract device and actual human language processing is obscure at best (Johnson and Lappin 1997, 1999).

Linking complexity to the length of a derivation is reminiscent of the derivational theory of complexity of classical transformational grammar (Brown and Hanlon 1970; Fodor et al. 1974). On this theory, complexity of acquisition and processing of a grammatical construction corresponds directly to the number of transformational operations required to derive it. Thus, for example, subject aux inversion in English (*Are you my friend?*) is more complex than the simple declarative (*You are my friend.*). But a derivational theory of complexity does not connect to anything that might reasonably contribute to processing complexity, e.g. memory.

In MP a shorter move is simpler than a longer move, other things being equal. A move is assumed to be involved in the derivation of *wh*-questions, as in (4). The canonical position of the *wh*-phrase is marked by the trace *t*, and the two form the chain <what,*t*>.

(4) What are you eating *t*?

Mainstream approaches to *wh*-questions assume that movements in more complex structures are accomplished by stringing together short moves, as in (5).

(5) What did you say [*t* you were eating *t*]?

The two movements together constitute a long-distance displacement. Alternatively, we could assume that there is a single "long" movement, that is, one without an intermediate landing site in the initial position of the embedded clause. This is shown in (6).

(6) What did you say [you were eating *t*]?

Such a movement out of the embedded sentence would be more complex than the short movement in (4), measured in terms of the amount of structure intervening between the two parts of the chain.

The relevance of formulating complexity in terms of the length of moves on this approach is suggested by the fact that a sentence such as (7) is ungrammatical. It is a so-called 'superiority' violation, as well as a '*wh*-island' violation (Chomsky 1973).

(7) *What did you say [**who** was eating *t*]?

The type of explanation advanced in the MP for the ungrammaticality of (7) is based on the observation that there is another *wh*-phrase, namely boldfaced *who*, that is higher in the structure than *what* and therefore closer to the landing site in sentence-initial position than the original position of *what*. In fact, movement of *who* to sentence-initial position produces a well-formed sentence.

(8) Who did you say [*t* was eating what]?

So on this account we rule out (7) by requiring that the shorter move excludes the longer move. Just as the movement of *what* in (4) is shorter than the move in (5), so is the move in (8) shorter than the move in (7). (4) does not exclude (5) because they are different structures, and hence incomparable. But (8) and (7) are competing derivations on the same structure. By the definitions given, (8) is the more economical alternative and blocks the less economical (7).

Putting the technical details aside, the important point about accounts such as this one is that the computation of economy is not grounded in actual human processing or acquisition of language. It does not try to justify "shortest move" or "fewest steps" in terms of the computational requirements of a plausible model of human processing. Nor does it make specific predictions about the difficulty of processing or the course of acquisition, or about change and variation, beyond allowing certain derivations and excluding others in principle. This dissociation is in part due to the fact that MP is a program that seeks to work through the implications of the assumption that the grammatical system approximates as closely as possible a "perfect" system, one that employs only those devices that are absolutely necessary to map between sound and meaning (Chomsky 2005).

However, it is reasonable to ask whether there might not be a connection between the idealized description of a perfect system and the actual system that exists in the minds of human beings. I argue in Chapters 5 that the various constraints on natural language that have been proposed over the years in mainstream generative grammar (MGG), although not, for the most part, in the MP, are in fact radically idealized versions of what actually does exist in nature, that is, human language processing. The primary difference is that the MGG constraints are absolute prohibitions against certain formal properties of grammatical representations, couched in derivational terms. The actual constraints reflect the relative complexity of processing certain grammatical configurations, compared with others that express the same meanings. So while very complex representations are not ruled out, on this latter view, their complexity constitutes a bias against them which leads to their relative rarity or non-existence, as discussed in Chapter 6.

This is an outcome that I believe conforms better to the original conception of markedness, as well as to the facts. In the examples discussed here in connection with the MP, for example, we might understand short movement as the simplest variety of chain formation in computational terms that relate to the dynamics of processing syntactic configurations, while leaving open the possibility of more complex, longer chains that may occur only in certain languages and under certain conditions (see Cinque 1990; Postal 1998).

1.3.3 Syntactic complexity

The term 'syntactic complexity' is often used to refer to the hierarchical structure of sentences. Other things being equal, it is assumed that flatter structure is less complex, while greater embedding constitutes greater syntactic complexity. Thus, on this view coordination is less complex than subordination, and nominal expressions are less complex than clausal ones (Givón 2009; Givón and Shibatani 2009). Similarly, a construction in which a relative clause is realized as the sister of N or NP (9a) is syntactically more complex than a construction in which the relative clause is a coordinate or paratactic S (9b).

(9) a. I met [$_{NP}$ the man [$_{S}$ who you saw]]
b. [$_{S}$ I met the man] (and) [$_{S}$ you saw him]

A similar view is developed by Progovac (2009), who suggests further that embedding complexity has increased in the course of language evolution. This notion of complexity also plays a role in debates about whether or not creole languages, many of which lack highly embedded structures, are less complex than non-creoles (McWhorter 2001, 2005).

The question of relative syntactic complexity is part of a broader debate about whether all languages are equally globally complex, in the sense that relative simplicity in one component is compensated for by greater complexity in another component. E.g., if the syntax is simpler, so that certain relationships between phrases are not expressed in terms of syntactic configurations, then some or all of the pragmatics, the inflectional morphology, or the phonology will be more complex in order to account for these relationships.

There is a widespread intuition in linguistics that all languages must be equally globally complex, because they all have to do the same work.[11] This intuition touches on the fact that there is more to the description of a language than the syntax per se. The work that the language does is that of expressing a correspondence between form and meaning, where meaning comprises at least truth-conditional and information structure interpretation and discourse representations. While languages may differ in terms of the complexity of the system that constitutes form (that is, the grammar), it

[11] See the papers in Sampson et al. (2009).

might be argued that languages are essentially the same in terms of the overall complexity of the form–meaning correspondence.

This view of "equi-complexity" is accepted wisdom in linguistics, and is often repeated in introductory texts (see Kusters 2003 for a review). Kusters argues convincingly that there can be no empirical test of the equi-complexity hypothesis, because there is no bound on the possible components and subcomponents that one might define in order to measure complexity. Along related lines, Shosted (2006) provides evidence that there is no statistically significant correlation between phonological and morphological complexity. Significantly, he leaves out syntactic and pragmatic complexity on the grounds that there is no body of reliable data on the basis of which a useful comparison can be made, and no reasonable consensus about what would count.

I therefore leave the question of equi-complexity aside in this study. I am concerned with aspects of local complexity—the formal complexity of individual constructions and the relative processing complexity of configurations that express the same or closely related form–meaning correspondences. What matters in this case are the factors listed in (3b), as well as processing requirements.

However, the related question remains of whether some syntactic structures are more complex than others in some useful sense. For practical purposes I assume that the relative complexity of a given syntactic structure is not an inherent property of the structure, but derives from the role that the structure plays in the correspondence with meaning. That is, the particular structure may contribute to complexity as a consequence of its relationship to the broader set of form–meaning correspondences in the grammar, or it may contribute to complexity as a consequence of the processing mechanisms that compute the form–meaning correspondence, or both.

1.4 Processing complexity

Another approach takes the amount of memory and the properties of what is in working memory as factors that contribute to the complexity of processing dependencies (Gibson 2000; Hawkins 2004; Lewis et al. 2006). For example, the longer it is necessary to hold a constituent in memory before linking it to a subsequent lexical item that is dependent on it, the more complex the configuration, other things being equal. Properties of intervening constituents that require additional processing may also contribute to complexity. At issue are questions such as what the relationship is between the target constituent and the intervening constituents, and what it is about working memory that produces processing complexity when two or more constituents are being processed at the same time.

This notion of complexity plays a central role in the typological work of Hawkins (Hawkins 1994, 2004). Hawkins argues that languages evolve so as to have grammars that minimize processing costs. Consider, for example, two forms, A and B, which express the same meaning. If B requires more processing than A, then speakers (or hearers) will prefer A to B, other things being equal. This preference may be manifested in corpora by a greater frequency of A compared to B. If A is used by speakers more than B, then learners will have greater evidence for A than B, and will assign a preference to A compared to B. If there are no frequent contexts in which for other reasons, B is preferred to A, we might expect that over time B would become very rare in the language, or even that A would drive out B completely (an idea that I explore in Chapter 6). Or perhaps B would take on a different function and no longer compete with A (as suggested in Chapter 7).

To take a concrete example, if V precedes its complements in VP, then on Hawkins's view, the cost of identifying the complement, say NP, is less if the head N of the complement precedes everything else in NP, as shown in (10a).

(10) a. V [$_{NP}$ N XP] (e.g. *eat (the) bagel [you stole from me]*)
b. V [$_{NP}$ XP N] (e.g. *eat (the) [you stole from me] bagel*)

This difference in cost is due to the fact that as we go from left to right, the processor must keep more temporary structure in memory between *eat* and *bagel* in (10b). The reason for the asymmetry in processing cost between these two alternatives is that the sooner the head of the complement is identified (in this case, *bagel*), the sooner the processor can construct a semantic representation of its projection and begin to integrate its arguments and adjuncts. For example, if the verb assigns a thematic role to its complement, as in *eat the bagel (that you toasted)*, it takes a bit longer in the processing of (10b) to establish a representation of *bagel* and assign the Patient role to it, as measured by the number of words or complexity of structure that has to be processed between *eat* and *bagel*. And in some cases, XP might be either a constituent of VP or a constituent of NP, but which it is can't be determined until N is or is not encountered.

In a language with free ordering between heads and modifiers, we would expect a preference for the ordering in (10a) to be reflected in corpus frequencies. In the limit, we would expect the ordering in (10b) to disappear entirely. Consequently, we would expect languages that show the ordering (10b) to be very rare, if they exist at all.[12]

[12] I am assuming here that the questions raised earlier about why small differences in memory load would produce differences in production can be satisfactorily resolved.

Reduction of processing cost along these lines thus leads to branching harmony, which is the tendency for heads to appear in the same position in phrases across categories. Such an account falls under "relative complexity." In fact, there is nothing in the formal description of the structure (10a) per se that would lead us to prefer it to that in (10b) as being more complex in absolute terms.

But it is interesting to note that there is an interpretation of the greater simplicity of (10a) in terms of the formal complexity of the description if we measure this complexity not in terms of the structures in (10), but against the general pattern $[_{XP}\ X^0\ YP]$, along the lines discussed above in connection with the core–periphery distinction. While VP and NP both show this pattern in (10a), only VP does in (10b). So it is at least logically possible that the preference for branching harmony arises out of the pressure to generalize everything as much as possible, including branching direction.[13] We get this result if we reinterpret the absolute requirement of unidirectional binary branching of Kayne (1994) as markedness measured against a maximally efficient correspondence with meaning.[14]

Another case studied by Hawkins where the distance between related parts appears to play a role in determining the acceptability of a sentence involves lexical items such as *look up*. There is a preference for the two parts to appear close to one another. Compare the following—(11a) is clearly preferred over (11b).

(11) a. Leslie looked up the information that we recommended.
b. ??Leslie looked the information that we recommended up.

In the preferred example (11a), *look* and *up* are adjacent. Until we encounter *up*, we don't know that the lexical item is *look up*.[15]

The intuition behind such accounts is that the amount of memory that is required to process a string of words in time plays a role in how the words are ordered by speakers. The more memory that is required, the more complex is the processing. More precisely, what contributes to complexity are words that are held in memory but not fully translated into semantic representations.

[13] This observation raises the interesting possibility that there is a relationship in general between relative complexity, that is, complexity of processing, and absolute complexity, that is, formal complexity of grammars. The issue is not a simple one, however, in part because there are many factors that may contribute to complexity of each type; due to space limitations, I do not explore it further in this book, except for a few remarks in Chapter 8.

[14] This point was made originally in Culicover and Nowak (2002); I elaborate further in Chapter 6.

[15] Although I find no discernible subjective difference between *look up the information* and *look the information up*, the theory predicts that the latter should be slightly dispreferred and there is evidence that this in fact the case: cf. Hunter and Prideaux (1983) and Hawkins (1994: 64–5).

Hawkins does not actually measure this cost experimentally; rather, he measures the complexity of the syntactic structure that is built as the sentence is processed, and assumes that it correlates with the memory requirements.

To see how this works, consider the examples in (11). In (11a) *look* and *up* are adjacent, so that it is possible to construct a syntactic unit consisting just of *look up* immediately. When this unit is identified, it can be used in constructing the interpretation.

(12) $[_{\text{VP}}\ [_{\text{V}}\ [_{\text{V}}$ look $]\ [_{\text{PRT}}$ up $]\ ]\ ]\ldots$

On the other hand, when *up* is far from *look*, more structure has to be built before *look* and *up* can be interpreted as a unit.

(13) $[_{\text{VP}}\ [_{\text{V}}$ look $]\ [_{\text{NP}}\ [_{\text{NP}}$ the information $]\ [_{\text{S}}$ that we recommended $]]$ $[_{\text{PRT}}$ up $]\ ]\ldots$

Here, the partially interpreted *look* has to be held in memory, and the NP *the information that we recommended* cannot be assigned a fixed thematic role until *up* is encountered.

While this story is plausible, and appears to account for an impressive range of typological generalizations, it raises questions. When we talk about the complexity of the mapping between form and meaning, we typically (but not necessarily) think of the complexity of the process of computing this mapping. However, as Miestamo (2009) points out, it is no simple matter to define precisely what is more or less complex for the processor. Among the reasons for this are: (a) the architecture of the processor is in many respects undefined; (b) it can be argued that what makes production easier makes comprehension more complex and vice versa; (c) few if any processing costs for various processes have been independently established; (d) the capacity of memory for linguistic representations and the cost of using this memory is largely unknown.

It is for these reasons that Miestamo (2009: 82) says that "cross-linguistic studies of grammatical complexity should adopt an absolute definition of complexity." From a practical perspective this might be a reasonable position to take, but there does not appear to be any principled basis for ruling out the possibility that processing complexity is relevant to the form of grammars.

Explanations in terms of processing complexity ultimately rely on the specifics of the processing mechanisms. For example, using memory load imposed by the processing of discontinuous dependencies requires that the meaning of the first expression is available during the course of processing. There is no reason to believe that this is not the case when what is being processed is a lexical item such as *look up*; there may be other aspects of processing that are insulated from such knowledge (Bever and Townsend

2001). In fact, Lewis et al. (2006) and Vasishth and Lewis (2006) argue that in some cases, a greater distance between the verb and the head of the argument may facilitate processing. They suggest that this may be because in verb-final constructions, at least, "the interposed material helps to strengthen the representation of the noun-phrase argument by reactivating it through modification" (Lewis et al. 2006: 451–2) and "the interposed material somehow allows an easier anticipation of the upcoming verb" (Vasishth and Lewis 2006: 768; see also Vasishth 2011). Levy and Keller (2011) argue that expectation competes with syntactic processing complexity even in the same structure, interacts with it, and typically dominates it; see also Demberg and Keller (2008) and §1.5.1.

In another study, Wasow 2002 examined a number of properties of lexical items and phrases that potentially could have an effect on their temporal ordering in the English VP. He showed that there are a number of arguably independent properties, each of which imposes an ordering preference on constituents. For example, there is a preference for ordering given information before new information, and a preference for ordering "heavy" constituents after "light" constituents. Since these preferences do not always align with one another, it is at best possible to establish tendencies in the data—one preference does not always outrank the others.

In another empirical domain, a number of experimental studies of sentence processing have argued that the processing complexity of filler-gap dependencies increases as a consequence of intervening definite NPs (e.g. Gibson 1998, 2000; Gibson and Warren 2004; Grodner and Gibson 2005; Kluender 1992, 1998, 2004). While the evidence for the contribution of intervening structure to processing complexity appears to be secure, the explanation for *why* definite NPs contribute to complexity is unclear. One possibility is that the definite NP that intervenes between filler and gap introduces some type of incongruence between syntactic structure and discourse structure (cf. Erteschik-Shir 1997). Such an account requires an independently motivated and well-articulated picture of the dynamics of how the discourse structure and the syntactic structure interact in time, which we do not yet have, although Erteschik-Shir's proposals are suggestive.

1.5 Other notions of complexity

1.5.1 Surprisal

It has been shown that the relative unpredictability (the 'surprisal') of the continuation of a sentence at a given point in the string contributes to processing complexity, as measured for example by reading times, eye-tracking, vowel reduction, and tempo (Levy 2005; Jaeger 2010; Hale 2003, 2011; Demberg and Keller 2008). Jaeger (2010) has shown that omission of the complementizer *that* in English is correlated with the predictability of a sentential complement.

Results such as these suggest a type of processing complexity that is different from that attributed to memory load. Levy (2005: 40) points out that the surprisal model correctly predicts the effect noted by Lewis et al. (2006) and Vasishth and Lewis (2006) that in head-final languages processing complexity is less when the number of dependents preceding the head is greater, as compared with a resource-based theory in which the greater the length of a dependency, the more difficult the processing. Levy's proposal is that the more preverbal dependents there are, the more predictable is the position and identity of the verb.

1.5.2 Optimality theory

Optimality theory (OT) accounts for well-formed expressions by evaluating all possible expressions against a ranked set of constraints. Roughly speaking, given a set of expressions that form a natural class, i.e. they are the possible orderings of some set of elements, the expressions in this set that satisfy the highest ranked constraint are those that are "grammatical," while the others are not.

The role of complexity in determining what the constraints are is not always defined explicitly in OT.[16] A constraint exists presumably because of the complexity of the form, operation, or relationship associated with it. However, in standard OT the complexity associated with a set of constraints is not additive, in contrast with the proposal of Kluender cited above. Rather, constraints are ranked with respect to one another, in most cases contingently, so that some sources of complexity are relevant to the distribution of forms in a language, and others, those associated with lower-ranking constraints, are not. However, there are many instantiations of OT, and it would require a full-scale study in its own right to explore how complexity is, or should be, treated in OT. I leave open the question of to what extent any of the proposals in this book can be successfully implemented in some version of OT.[17]

[16] The complexity of OT grammars, especially with respect to acquisition, is a major focus of research. See, among many others, Frank and Satta (1998); Jäger (2002); Tesar and Smolensky (1998).

[17] I do note in Chapter 8 that some processing complexities can be naturally formulated as OT constraints. However, for the specific phenomena discussed in Chapter 8 it appears that the OT accounts that been proposed are more complex than the phenomena that they propose to explain, precisely because all of the data has to be characterized in terms of OT constraints even for properties whose complexity does not appear to be relevant.

2

The architecture of constructions

This chapter outlines the basic assumptions in this book about what constitutes a construction, how constructions are related to one another, and what constitutes complexity in a grammar formulated in terms of constructions. It relates these views about constructions to other constructional approaches, although I make no attempt to provide a comprehensive review. It also lays out several key assumptions about syntactic configurations in English, in particular that phrase structure is as flat as possible, and that English has the category VP. The descriptive apparatus developed here forms the basis for the analyses and theoretical proposals in subsequent chapters.

2.1 Constructions

In *Syntactic Nuts* (Culicover 1999), I argued that the core–periphery distinction (Chapter 1, §1.1.2) does not correspond to any architectural property of the human language faculty. That is, a native speaker's knowledge does not consist of a universal core grammar composed of abstract minimal operations (e.g. *Merge* and *Move*), principles and parameters, and a large body of phenomena that reside in the periphery.

In contrast, I assume that knowledge of language consists of constructions of varying degrees of generality and idiosyncrasy. The most general constructions are typically those that are recruited to perform the everyday expressive functions of a language—making statements, asking questions, issuing requests and orders, referring to objects in the world, and attributing properties to them, and so on. The less general constructions typically have more specialized functions.

Fillmore et al. (1988) is probably the earliest clear specification of the scope of a grammar that takes constructions explicitly into account. On this view, a construction such as passive, or pseudo-cleft, or *let alone*, or *What's NP doing XP*, is a distinguishable component of the grammar and constitutes an explicit part of the grammatical description. While the constructions of a language are individuated, they may and typically do share properties. Some constructions are basic, while others are complex, being composed of more basic constructions themselves. The most basic constructions are individual lexical items whose properties cannot be derived from anything more basic.

This view of constructions contrasts with the position taken in mainstream generative grammar, which is that constructions as such are not part of grammatical descriptions. They are held to be *informal* descriptions of the effects of interacting rules and principles, and nothing more. The classical mainstream analysis of passive, for example, derives the passive construction in English from the interaction of Move α, Case Theory, and requirements on feature discharge (Jaeggli 1986; Baker et al. 1989). For a general critique of this approach to grammar, see chapters 1 and 2 of *Simpler Syntax* (Culicover and Jackendoff 2005).

I understand a construction here to be the specification of those formal properties of syntactic structure and/or the corresponding interpretation that are not predictable from universal principles or general properties of the language, and thus must be independently stated in the grammar. Thus, a construction may be:

- a general structural configuration, e.g. VP, with its associated general interpretation.
- an idiom with slots for variables, such as *sell NP down the river* or *V one's way PP*, with its associated interpretation.
- an idiom with a fixed form and meaning, such as *kick the bucket.*
- a correspondence between a particular syntactic form, such as the English double object configuration [$_{VP}$ V NP_1 NP_2], and a meaning, in this case 'convey NP_2 to NP_1 by V-ing.'
- a restriction on the form of a sentence, such as the requirement that the tensed verb of a declarative and certain embedded clauses in German must appear in second position, so-called 'V2.' Such a restriction does not correspond to any particular meaning.[1]
- the licensing of a particular form that may appear in a range of correspondences, e.g. subject aux inversion in English.[2]

[1] Hence a construction in the sense I am using it here is not a 'sign' in the sense of Sag (forthcoming). For arguments that V2 in German in fact corresponds to an identifiable semantic function, see Truckenbrodt (2006).

[2] I assume here a rather straightforward notion of correspondences as involving literal and perhaps discourse meaning. There is good reason to believe, however, that otherwise functionally equivalent constructions may also be individuated in terms of their social meaning. See for example Elsig and Poplack (2009).

Even an individual lexical item is a construction, in that it expresses an irreducible, unpredictable correspondence between form and meaning.

While I do not assume UG in the sense of core grammar discussed in Chapter 1, this is not to deny that there may be universal properties of form or of the form–meaning correspondence. So, for instance, the generalization that V is the head of VP, and that more generally X is the head of XP, is probably not something that has to be stated for a particular language—the default is that phrases are headed.[3] The interpretation of NP as the direct object of V in a hierarchical structure in which they are daughters of VP is possibly a universal, although the order of V and the direct object in a given language is constructional.[4] On the other hand, the specification that the subject of a tensed S in English is the left sister of the VP headed by the inflected verb is a constructional property of English, albeit very general for the language. (And in Chapter 3 I discuss some English constructions in which this property of subject does not hold.)

Following *Syntactic Nuts*, I assume that the theory of grammar makes available an infinite space of logically possible constructions. But most of the logically possible constructions do not occur, and this is something that needs to be explained. The same or very similar constructions occur in language after language, a fact that also has to be explained. Within a given language, the specialized constructions typically replicate the properties of the more general constructions. Again, this is not an accident, and it needs to be explained. The fact that grammars composed of constructions are not completely random and chaotic is due to the fact that formal complexity is dispreferred, other things being equal. §2.3 makes more precise this notion of complexity as applied to constructions.

Our grammatical theory must also take into account the ability of learners to acquire not just the most general constructions, but also the most specialized ones. A central argument of *Syntactic Nuts* is that a learner that is able to acquire specialized constructions of varying degrees of generality has the computational resources to acquire fully general constructions when the evidence warrants. The existence of fully general constructions gives rise to the appearance of "core" rules, those that are found in many if not most languages and argued to be part of UG. But, *Syntactic Nuts* argues, UG in this sense is an illusion. What *is* universal, among other things, are some of the expressive functions and the overall architecture of meaning, that is, conceptual structure (Jackendoff 1983, 1990, 1997, 2002b). *Syntactic Nuts*

[3] It is possible to argue that headedness is necessary for learnability—see Culicover et al. (2008).

[4] Although the inverse, that direct object is always the sister of V in a VP may not hold.

argues that the universality of conceptual structure, and the requirement that a language must be able to express it, produces the illusion of universals of syntactic structure. *Simpler Syntax* suggests that once a formal device (e.g., case-marking) is available, it becomes part of the universal language "toolkit" and can be used for more or less idiosyncratic correspondences.

2.2 Correspondences

2.2.1 Representing constructions

The constructional approach to describing the form–meaning relation has been formulated in various ways in the literature: Construction Grammar (Fillmore 1988; Kay 2002; Kay and Fillmore 1999; Michaelis 2006), Sign-Based Construction Grammar (SBCG) (Sag 1997; forthcoming), the Parallel Architecture (Jackendoff 1975, 1983, 1990, 2002b), and *Simpler Syntax* (Culicover and Jackendoff 2005), and others. I adopt here the general perspective common to all of these approaches, setting aside a number of formal issues that are orthogonal to the main points that I want to make in this book.[5]

Hence I adopt the consensus view in constructional approaches to grammar that there is no principled distinction between syntactic constructions and the lexicon—see Michaelis (2006) for a review. The particular implementation of this view that I adopt here is based on the Parallel Architecture of Jackendoff (2002b: chapters 5 and 6); see also *Simpler Syntax* and Culicover (2009).

In order to state constructions with some precision, we have to make assumptions about the representations that participate in correspondences. Focusing first on the syntax, I take the perspective of *Simpler Syntax*: the minimal syntactic representations are the flattest ones possible. Phrases are typically (but not necessarily) projections of heads; there are NPs, PPs, APs, VPs, and Ss. Beyond this, we assume the minimal syntactic structure sufficient to account for the correspondence with meaning. This means that branching structure within the phrase is possible, but it must be motivated by the actual facts of the language, and not by a general principle.

[5] To give just one example, Müller (2006) argues for a lexical as opposed to a phrasal characterization of constructions. He argues that the meaning of the phrasal construction can be effectively represented in the corresponding lexical entry, and that the grammar is thereby simpler. The key is that the meaning can be associated with the head, which in turn determines the grammatical properties of the phrase. The phrase built around the head may appear in whatever variant forms such a phrase may adopt, while the meaning is preserved. He argues that the less attractive alternative is to associate the meaning with the phrase, which requires us to associate the same meaning individually with all syntactic variants of the phrase.

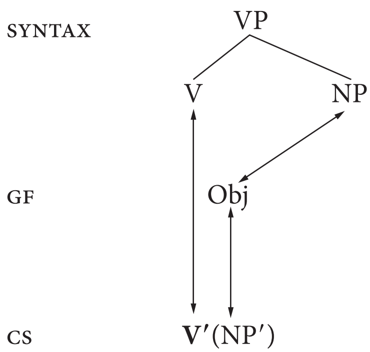
The most problematic phrasal category is VP. VP is included in the constructions that I introduced here, and in §2.2.2 below I give some arguments to support the assumption that English has a VP.

Every construction involves a syntactic representation and in most but not all cases, correspondences between the syntactic representation and conceptual structure (CS), and information structure (IS). In sentences these correspondences are mediated by links to the grammatical functions, minimally Subj(ect) and Obj(ect). These representations that constitute constructions are understood as follows.

SYNTAX: the representation of the structure of the expression, consisting of the constituents, their linear arrangement in time, their categories, and their hierarchical organization.
GF: an independent tier of syntactic representation that corresponds on the one hand to constituents in SYNTAX according to their configurational and/or morphological properties, and on the other hand to components of the CS representation.
CONCEPTUAL STRUCTURE (CS): the representation of the meaning of the expression, including its thematic and logical properties.
INFORMATION STRUCTURE (IS): the representation of the information structure properties of the expression, i.e. new vs old, contrast, focus, etc.

As noted, constructions range in generality from the very idiosyncratic and idiomatic to the very regular. Consider VPs. A typical general VP construction is shown in (1). This construction expresses the correspondence in English between a syntactic structure in which an NP is the complement of V, and the CS representation.

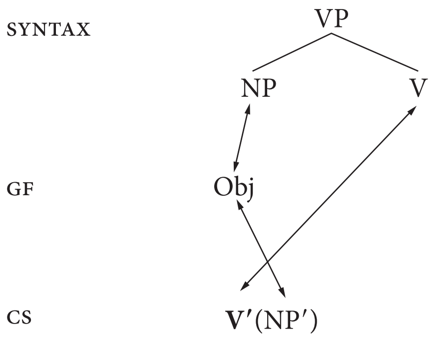
(1) *OBJECT (ENGLISH)*



V′ in (1) is the CS representation of V, and NP′ the CS representation of NP. Boldface of **V′** indicates that it is a function, of which NP′ is the argument. Precisely what this function is may be further specified by the lexical entry of V. The arrows in the schema function in lieu of subscripts to show the correspondences between the elements of the various representations.

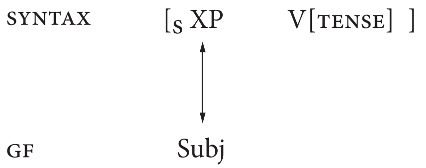
The counterpart to (1) in a V-final language like German has [$_{\mathrm{VP}}$ NP V] in SYNTAX and the same CS representation; cf (2).

(2) *Object (German)*



The syntactic realization of the subject GF in English is also specified constructionally, as shown in (3). The formal condition is that XP immediately precedes V[TENSE]. By default only what is relevant is mentioned, so there is no need for a variable following V[TENSE] or preceding XP.

(3) *Subject*



V may be an auxiliary verb or a main verb. Note that this construction simply licenses a particular syntactic form in the language; it does not specify a correspondence with a particular meaning as well. Moreover, it does not exclude the possibility that there are other constructions in the language that also express a correspondence with the Subj function (see §3.2).

A classic example of the constructional view is the English dative, discussed in detail by Goldberg (1995). The basic idea is that the core meaning of transfer of possession associated with a sentence such as (4) —

(4) Sandy gave Terry the money.

—is not solely a property of the verb, but is a property of the configuration (5).

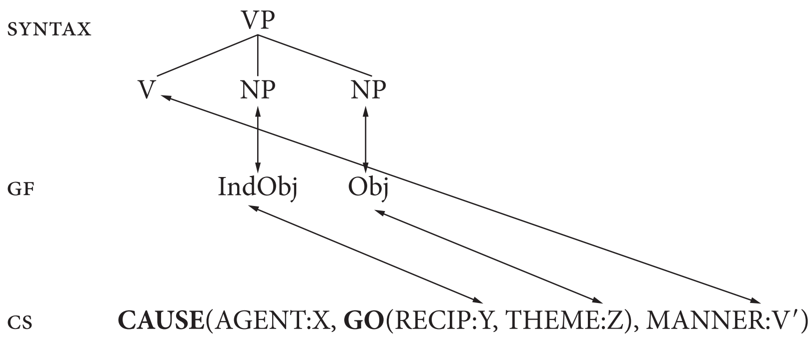
(5) $[_{VP}\ \mathrm{V}\ \mathrm{NP}_1\ \mathrm{NP}_2]$

If the verb expresses the meaning of transfer of possession, then of course its meaning and the meaning of the construction are compatible. But if the verb does not, the meaning is still present and can be coerced, in virtue of the construction, as in (6).

(6) Sandy $\left\{\begin{array}{l}\text{mailed}\\ \text{shipped}\\ \text{handed}\\ \text{tossed}\\ \text{bounced}\\ \text{kicked}\\ \text{rolled}\end{array}\right\}$ Terry the money.

In the lexical entries, the verbs denote the manner of moving the object, while the construction denotes change of possession.

The dative construction can be represented in this framework roughly as in (7), leaving out some details.

(7)


The construction thus imposes the interpretation of transfer of possession by means of the action denoted by the verb.

There is, in fact, a spectrum of generality for VP constructions. On the other end of the spectrum from (1) we have VP idioms like *kick the bucket* and *bite the dust*, meaning 'die.' In the middle we have constructions like V-*one's-way*-$\mathrm{PP}_{\mathrm{path}}$, as in *fight your way to the top* (Jackendoff 1992). These (and many others) are ordinary VPs as far as structure is concerned, but have conceptual structures associated with them that cannot be constructed compositionally. *Fight your way to the top*, for example, cannot be interpreted taking *your way* to be the Patient of *fight*, as in *fight the enemy*. Rather, the correspondence is that in (8).

(8) SYNTAX [$_{VP}$ fight [$_{NP}$ your way] [$_{PP}$ to the top]]

CS **GO**(AGENT:YOU, GOAL:TOP,MEANS:FIGHT)

The action FIGHT in this case corresponds to the means of achieving the goal.[6]

The lexicon is the repository of idiosyncratic information about the form–meaning correspondence. A word is a correspondence of a syntactically simple form with a lexical category, a sound and a meaning (Jackendoff 1990). A construction is a syntactically complex form with a lexical or phrasal category, depending on its structure, and possibly a corresponding meaning. Constructions may contain variable elements, to the extent that there are compositional parts of their interpretation. So the construction exemplified in (8) will have the lexical entry in (9).

(9) SYNTAX [$_{VP}$ V [$_{NP}$ pro's way] PP]

CS **GO**(AGENT:X, GOAL:PP′, MEANS:V′)

The meaning of the verb and the meaning of the PP will be composed with the constructional meaning as indicated. What is idiosyncratic, and requires lexical specification, is the correspondence of MEANS with the verb and the presence of the element **GO** in the CS representation.

There are many constructions in a language such as English, and to try to enumerate even a substantial subset here would take us far afield. I introduce and describe several individual constructions in subsequent chapters, as I deal with particular phenomena.

2.2.2 VP

Before moving on, I return briefly to the evidence for VP in English, and to the question of whether the structure of VP is in fact flat. Note first that it is not necessary to assume that there is a VP in English just in order to distinguish subject and object, although these have typically been distinguished in configurational terms in MGG, assuming the branching asymmetry in (10) (but see McCawley 1970).

[6] For completeness I show the correspondence between the Agent and the Possessor explicitly. In a more detailed representation, the Possessor of *way* might be notated as bound by the subject and therefore agreeing in person and number with it. Also, I omit the phonological component of constructions throughout.

(10) $[_{\mathrm{S}}\ \mathrm{NP}\ [_{\mathrm{VP}}\ \mathrm{V}\ \mathrm{NP}\ ]]$

But *Simpler Syntax* assumes that the grammatical functions Subj(ect), Obj(ect), and Ind(irect)Obj(ect) are primitives that correspond to syntactic configurations in language like English, but are not universally defined in configurational terms. So we might capture the same asymmetry directly in terms of linear order as in (11), without assuming that V and the direct object form a constituent.

(11)
$$\begin{array}{lcccc} \text{SYNTAX} & [_{\mathrm{S}}\ \mathrm{NP} & \mathrm{V} & \mathrm{NP} &] \\ & \updownarrow & & \updownarrow & \\ \text{GF} & \mathrm{Subj} & & \mathrm{Obj} & \end{array}$$

The assumption that the GFs are primitives is central to the *Simpler Syntax* account of so-called 'A-movement' constructions such as passive, raising to subject, and raising to object. These constructions are accounted for by assuming a single level of syntactic representation, and mapping one GF into another so that the first is realized syntactically as the second. For example, the Obj in a passive corresponds to the Subj of the clause. This approach borrows insights from Relational Grammar (Perlmutter and Postal 1983), including the idea that the syntactic structure is essentially flat. (For more discussion see Culicover and Jackendoff 2005: chapter 4.)

What then of the evidence for VP in English? There are two kinds of evidence. One consists of the fact that what appears to be a VP is moved or deleted. In example (12a) it appears that the VP *see Paris* is moved, in (12b) it is *see Paris in the spring* that is moved, while in (13), it appears that the VP *see(n) Paris in the spring* may be deleted under ellipsis.

(12) a. They said Terry would never see Paris in the spring, and see Paris, Terry never has at any time.
b. They said Terry would never see Paris in the spring, and see Paris in the spring, Terry never has.
[VP topicalization]

(13) Terry has never seen Paris in the spring, but Leslie has. [Ellipsis]

Additional evidence for VP comes from the fact that a phrase such as *see Paris in the spring* can be the focus of a pseudo-cleft, as in (14).

(14) What Terry did was [see Paris in the spring]. [Pseudo-cleft]

Only a single constituent may be in the focus of a pseudo-cleft. Hence *see Paris in the spring* cannot be analyzed as a sequence of constituents, e.g. [*see*]+[*Paris*]+[*in the spring*], that together fail to constitute a phrase.

Accepting the conclusion that English has a VP, the question then arises as to whether VP has any internal structure, so that arguments (such as the direct object) and adjuncts are not sisters. Ellipsis examples such as (15) suggest in fact that a VP may contain a VP. The missing material is indicated by strikeout.

(15) Terry has never seen Paris in the spring, but Leslie has [~~seen Paris~~] in winter.

On the basis of examples such as this, it appears that there is in fact internal structure in the VP, effectively [$_{\text{VP}}$ [$_{\text{VP}}$ *see Paris*] [$_{\text{PP}}$ *in the* $\left\{\begin{matrix} \textit{spring} \\ \textit{winter} \end{matrix}\right\}$]], undermining the assumption that the structure of the phrase is flat.

The evidence is equivocal, however. The examples in (16) suggest that it is possible to topicalize *put the beer* but that it cannot undergo ellipsis.

(16) a. They said that Terry would put the beer somewhere cold, and [put the beer] Terry did in the refrigerator.
b. *Terry put the beer in the refrigerator, and Leslie did [~~put the beer~~] in the oven.

(16b) can be used as evidence that *put the beer* is not a constituent, but (16a) seems to contradict this evidence.

The solution that I adopt here is that if there is a VP, its structure is flat—that is, [$_{\text{VP}}$ *put* [$_{\text{NP}}$ *the beer*] [$_{\text{PP}}$ *in the refrigerator*]]. Topicalization is a construction that allows the VP, starting with V, to be initial in the sentence under certain discourse conditions. The V combines with the following arguments to form a predicate interpretation. If an argument or adjunct is not adjacent to the topicalized constituent, the completion of the predicate interpretation is suspended until the argument or adjunct is encountered.[7] So in (16a), *put* requires two arguments, a Theme and a Location. The Theme requirement is satisfied by *the beer*, the Location requirement is satisfied by *in the refrigerator*. Since it is possible to only partially satisfy the requirements of the verb in the topicalized constituent, the string of phrases gives the illusion that *put the beer* is a constituent. In a sense, the second clause in (16a) contains a discontinuous constituent, *put the beer... in the refrigerator*. A similar analysis is applicable to adjuncts that are not selected by the topicalized constituent, but take it as a semantic argument.

[7] Here I am adapting ideas of Steedman (2000).

In the case of ellipsis, however, there are different licensing conditions at work. In order for an ellipsis to be interpreted, there must be something in the antecedent that corresponds to it syntactically that has an interpretation associated with it—the ellipsis is a pro-VP (see Culicover and Jackendoff 2012). Since *put the beer* does not correspond to a conceptual structure object (i.e. an action), this condition is not met and ellipsis is not possible.[8] So *Leslie did* $[_{\mathrm{VP}}$ $[_{\mathrm{VP}}$ Ø] *in the refrigerator*] lacks an interpretation. On the other hand, *see Paris* does correspond to an action. So it can be the antecedent of the ellipsis of *(has)*$[_{\mathrm{VP}}$ $[_{\mathrm{VP}}$ Ø] *in the winter*].

The crucial difference between the two cases is that in conceptual structure, *put* selects two VP arguments while *see* selects one. Only the fully specified antecedent can supply the interpretation of the ellipsis. The psycholinguistic evidence in fact points to this fundamental difference between arguments and adjuncts: the lexical representation of a verb includes information about its arguments, but not about its adjuncts (Boland 2005).

2.2.3 Meaning

I have already given some examples in which simple CS representations are shown in correspondence with syntactic structures. I assume, following Jackendoff (2002b), that CS is a distinct representation which contains the literal meaning of a sentence and the logical relations that it expresses. CS is thus more or less equivalent to the predicate calculus. The primitives that make up CS representations are concepts that correspond on the one hand to words and phrases, and on the other hand to entities in the world and their properties.

Nothing crucial in what is discussed in this book hangs on the specific details of CS. What is crucial is that there are correspondence rules that link specific syntactic characteristics to specific aspects of meaning. Most importantly, the primitive GFs Subj and Obj correspond to particular thematic roles associated with a CS relation, and also correspond to particular syntactic or morphological configurations. In a language such as English where Subj and Obj are expressed configurationally, the correspondence with CS has consequences for the linear order in which the phrases are expressed. For example, the Subj is the left sister of the inflected VP, in the typical active declarative sentence, and the Obj is the right sister of the main verb.

[8] The richer logical form associated with *put* is not reflected in a simple CS representation that simply notates the roles Agent, Theme, and Location. Thanks to Wolfgang Sternefeld for bringing this to my attention.

Another aspect of meaning, information structure, also interacts with linear order and syntactic structure. For example, B's utterance *In the spring* in (17) counts as a felicitous answer to A's question in virtue of the fact that the question is about TIME, as indicated by *when*, and *in the spring* denotes a time.[9]

(17) A: When will Terry see Paris?
B: In the spring.

B's answer could be more explicit, as in (18a). But the semantically identical (18b–d) are infelicitous (with default intonation) because they pick out the wrong part of the sentence as the answer to the question. That is, their information structure does not conform to that of the question.

(18) a. Terry will see Paris in the spring.
b. #In the spring Terry will see Paris.
c. #It is Paris that Terry will see in the spring.
d. #What Terry will do is see Paris in the spring.

I assume that each configuration of the sort illustrated here exemplifies a construction that expresses a correspondence between the particular configuration and some aspect of information structure.[10]

2.3 Constructional complexity

2.3.1 Regularities and idiosyncrasies in the constructional lexicon

A central idea in constructionalist approaches to grammar is that constructions are not arbitrary correspondences between form and meaning. Rather, specific constructions are likely to have properties of other, more general constructions—this relationship is called 'inheritance.' As we have seen, for example, English verbal idioms, such as *fight your way to the top*, *kick the bucket*, etc., typically have the same structure as literally interpreted VPs. What differs is the lexical specificity of the idiom and the corresponding idiosyncratic interpretation.

One of the central theses of this book is that grammars are structured so as to reduce the complexity of the set of constructions, where complexity is measured in terms of the degree of regularity and idiosyncrasy in this set. We can get some insight into how to characterize this type of complexity by

[9] For a more formal proposal about how this matching works, see Culicover and Jackendoff (2012).
[10] Rochemont and Culicover (1990).

looking at how regularity and idiosyncrasy in the lexicon have been dealt with in previous work.

Jackendoff (1975) showed how to capture regularity and idiosyncrasy in the lexicon by assigning a cost to those properties of lexical items that do not follow from general rules or paradigms. In fact, he extended the analysis to idioms with canonical syntax, arguing that the marginal complexity associated with them can be isolated in their non-compositional interpretation and lexical specificity. Jackendoff's basic argument is that the complexity of a lexicon is not determined by the number of terms needed to describe the entries. Rather, it is determined by the extent to which individual items have unpredictable idiosyncratic properties. If semantically related items have systematically related formal properties, for example, and differ only in a few specifics, the lexicon is less complex than if the properties of each of these items are completely *sui generis.*

On this view, all actual forms are listed in the lexicon with their full analysis. For example, the past tense of *have* is *had*, and the past tense of *bake* is *baked.* The first is irregular, and therefore incurs a cost, while the second is regular, because it follows the rule that realizes [PAST] as *-(e)d.* Jackendoff's measure of the information required to specify the lexical item is given here as (19).

(19) INFORMATION MEASURE

Given a fully specified lexical entry W to be introduced into the lexicon, the independent information it adds to the lexicon is

(a) the information that W exists in the lexicon, i.e. that W is a word of the language; plus
(b) all the information in W which cannot be predicted by the existence of some redundancy rule R which permits W to be partially described in terms of information already in the lexicon; plus
(c) the cost of referring to the redundancy rule R.

The cost of a lexical entry cannot be determined on this approach simply by examining it in isolation—it must be computed by referring to the redundancies. In the case of *had* and *baked*, for example, the lexical entries are simply as in Table 2.1, and the cost is computed through reference to the paradigm.

TABLE 2.1. Sample English verbal inflections

ROOT	PRES	PAST	PROGRESSIVE	...
have	have | has (*haves)	had (*haved)	having	...
bake	bake | bakes	baked	baking	...

Hence a form will be less costly in a language where there is a regularity that it corresponds to, and more costly in a language that lacks such a regularity.

Jackendoff's approach suggests that we follow a strategy of analyzing all constructions, from words to idioms to more general correspondences, with an eye towards representing the regular properties in a uniform way across all cases. That is, if we encounter what appears to be the same syntactic structure in two different constructions with different meanings, we do not want to say that the two constructions have a different (perhaps abstract) syntactic analysis. Rather, we want to say that they have the same syntactic analysis, to the extent that they appear to have the same structure. The difference in meaning is not attributed to a different syntactic structure, but to the fact that they are different constructions with different interpretations corresponding to them.[11]

In addition, we do not want to say that redundancies in the descriptions of these constructions are costly. To the contrary, redundancies are welcome, because they reflect the generalizations that hold between and across constructions. Placing a high value on redundancy across collections of lexical items, including constructions, allows us to capture the generalizations without insisting that the different constructions are all derived from a unique common underlying structure. (This type of complexity falls under the general category of 'minimum description length,' discussed in Chapter 1.)

To take just one example, consider subject aux inversion (SAI) in English, shown in (20). It is an alternative to the canonical *Subject* construction in (3) with respect to the position of the constituent that corresponds to the Subj GF.[12]

(20) *Subject aux inversion (SAI)*

SYNTAX $[_S V_{AUX}[\text{TENSE}]$ $\quad$ XP $]$

GF $\quad$ Subj

[11] For an early formulation of this idea, see Culicover (1971, 1973). This is also the general approach of *Simpler Syntax*.

[12] I assume here that anything that can be a subject in an uninverted sentence can appear in the inversion construction. Note, however, that non-NPs are much less acceptable than NPs in inversion. E.g.,

(i) a. [That it is raining] is obvious.
b. ??Is [that it is raining] obvious?

It is plausible to seek to attribute the lower acceptability of cases like (ib) to non-grammatical factors, perhaps related to those that are operative in cases of center-embedding (Chapter 6). For discussion of a range of cases, many of which do not involve subjects, see Kuno (1973).

As argued in Culicover (1971, 1973), SAI per se lacks a unique semantic function, just as SUBJECT does.[13] However, it plays a role in the description of yes-no questions, *wh*-questions (*what did you say*), and a number of relatively minor constructions, such as negative inversion (*Not even at that point did they realize that they had lost*), counterfactual conditional (*Had I known you were going to call, . . .*), and *so*-Inversion (*So angry had she become that she . . .*).

The statement of the yes-no question construction is given in (21) as an illustration of the redundancy. I set aside discussion of the others here in order to avoid introducing complications at this point in the exposition; the formulation of the *wh*-question construction is given in Chapter 7.

(21) *YES-NO QUESTION*

SYNTAX $[_{S}\ V_{AUX}[\text{TENSE}]\quad XP\ ]$

GF Subj

CS $\mathbf{Q}(S')$

The important point to recognize here is that not only does SAI not correspond to the same meaning across these constructions, but it applies under different syntactic conditions in each case. For instance, SAI in yes-no questions lacks any syntactic restrictions, as seen in (21). As shown in (22)–(25), SAI occurs in *wh*-questions with clause-initial *wh*-NP, but SAI with clause-initial *neg*-NP is decidedly marked—at best it has an archaic, literary flavor.

(22) a. Which of those books are you going to read?
b. ?None of these books am I going to read.

(23) a. What are you doing?
b. *Nothing am I doing.

(24) a. How many of those people did you talk to?
b. ?Not many of those people did I talk to.

[13] For an opposing view, see Goldberg and Del Giudice (2005), Goldberg (2006). I discuss this question in more detail in Chapter 3, §3.2.1.

(25) a. Whose books do you find most entertaining?
b. ?No one's books do I find particularly entertaining.

Another difference is that SAI in yes-no questions applies to the modals (including *do*), *have*, and *be*, while in the counterfactual conditional it applies only to *have* and (marginally) to *be* in contemporary English.

(26) a. Could you understand the text?
b. If I could understand the text, I wouldn't have asked you for help.
c. *Could I understand the text, I wouldn't have asked you for help.

(27) a. Did you finish your homework?
b. If I finished my homework, I would be asleep now.
c. *Did I finish my homework, I would be asleep now.

(28) a. Have you finished your homework?
b. If I had finished my homework, I wouldn't have called you.
c. Had I finished my homework, I wouldn't have called you.

(29) a. Are you finished yet?
b. If I were finished, I would be sleeping.
c. ?Were I finished, I would be sleeping.

On the basis of data such as this, it is plausible to take each of these constructions as independent, but incorporating the same formal device of SAI. The constructional lexicon is thus simpler than it would be if each construction specified a different position for the auxiliary verb (e.g. clause-initial, clause-final, reduplicated in situ, in situ and clause-initial, in situ and clause-final). The relationship between SAI and these constructions is essentially the same as the relationship between a particular inflection, for example, past tense *-ed* and progressive *-ing*, and the words that incorporate it, such as *baked*, *baking*, *constructed*, *constructing*, etc.

On the other side of the coin, different syntactic structures with what appears to be the same meaning should be analyzed as different constructions, in virtue of their different forms. This idea runs counter to the predominant strategy in mainstream syntactic theorizing, which takes sameness of meaning to be reflective of sameness of syntactic structure at an abstract level. (See chapter 2 of *Simpler Syntax* for extensive discussion.)

Consider now the complexity of English VP constructions, which were noted briefly in §2.2.1. Culicover and Jackendoff (2005) observe that in a language like English, it is possible to find VP structures with various degrees of generality. At the highest level, the English VP has roughly the form (30).

(30) $[_{\text{VP}}$ V (NP) PP* Adjunct*$]$

That is, VP is head-initial. The structure for VP in (30) is an instantiation of the UG principle (which is a central component of X theory) that in the unmarked case, phrases are headed, as in (31).[14]

(31) $[_{XP} \ldots X^0 \ldots]$

and therefore VP has the general structure in (32).

(32) $[_{VP} \ldots V \ldots]$

In a canonical English VP, such as *put the book on the table*, the verb is typically phrase-initial, and is followed by the direct object, oblique arguments, and adjuncts, as summarized in (30). But there are many specialized constructions that also draw upon this general pattern. One, sound +motion,[15] makes use of the structure (33)—

(33) $[_{VP}$ V PP ...]

—but imposes a particular meaning on the verb when the PP denotes a path, as in (34).

(34) The bus $\left\{\begin{array}{l}\text{screamed}\\ \text{roared}\\ \text{whistled}\\ \text{rumbled}\end{array}\right\}$ around the corner.

The meaning of the construction is given in (35).

(35) SYNTAX $[_{VP}$ V PP]

CS **GO**(AGENT:X,PATH:PP′,MANNER:V′)

Another construction, V-*the way*-PP, as in *fight your way to the top*, discussed above, expresses movement as well, with structure (36).[16]

(36) $[_{VP}$ V NP PP ...]

A related construction of the same form uses verbs derived from names of body parts or tools to refer to the means of accomplishing the goal, e.g. *shoulder, hammer*. The construction is given in (37).

[14] In stronger formulations, every phrase is headed. For discussion, see Culicover and Jackendoff (2005: 49–50, 110–11, 115).

[15] Jackendoff (2002b).

[16] Jackendoff (1992).

(37) SYNTAX [$_{VP}$ [$_{V}$ N] [$_{NP}$ pro's way] PP]

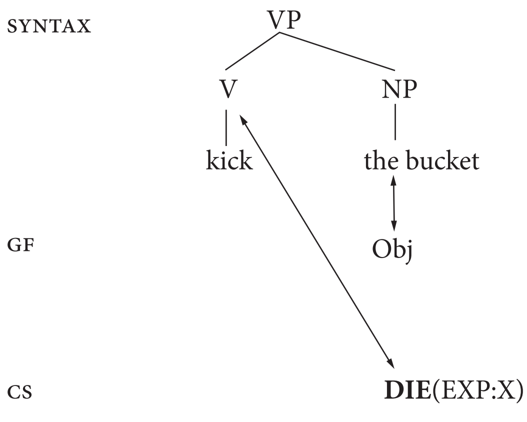
CS **GO**(AGENT:X, GOAL:PP′, MEANS:N′)

These are two distinct but related constructions. While they use the same syntactic form, the correspondence with meaning is somewhat different. And the V-NP-PP constructions and the V-PP constructions are all well-formed instances of VP, according to the general structure (36).

Culicover and Jackendoff (2005: chapter 1) suggest that (30) and the structures in (33) and (36) are specializations of the schema in (31), the UG principle of headedness. The patterns of (33) and (36) follow directly from the realization of arguments and adjuncts in the language.[17] They contribute only minimal complexity to the grammar beyond the baseline defined by the English instantiation of the UG schema.

A still more specific VP construction is the idiom *kick the bucket*. The reason why *kick the bucket* does not add significantly to the complexity of English is that it is a unique lexical item that expresses a correspondence with a canonical structure of the language. The essence of an idiom, in fact, is that it is an expression that has a perfectly good form and literal interpretation but, in the idiomatic case, it has an alternative unpredictable interpretation. (38) illustrates.[18]

(38)



[17] It is often claimed that the direct object must be adjacent to the verb, on the basis of examples like **I saw yesterday Terry*. But the V-Prt construction (*look up the information*) and heavy NP shift suggests an alternative view: the order in VP is relatively free but constrained on the basis of weight and stress, and their information structure correlates. See for example Hawkins (1994, 2001, 2004) and Wasow (1997, 2002).

[18] EXP is the Experiencer role. For an important proposal regarding the relationships between roles such as Agent and Experiencer and their internal composition, see Dowty (1991).

This idiom has precisely the structure specified by the more general construction, and the correspondence with GF is precisely what we would expect if it had a literal interpretation. The only difference is that the interpretation is not compositional, so that the meaning **DIE** corresponds to the entire idiom, headed by the V. Note that the representation in (38) captures the fact that *the bucket* is the direct object of the verb *kick* without assigning any thematic interpretation to it.

To the extent that the language tolerates unboundedly many lexical expressions, we would not want to count the existence of *kick the bucket* as adding significant complexity more than we would want to count the existence of *die* as adding significant complexity, beyond the fact that both are individual items in the lexicon. But we do want to say that a construction whose structure deviates in some way from the general structure of the language is more complex than one that doesn't.[19]

Another special case of a VP construction is the verb–particle construction of English, which takes two forms, as shown in (39).

(39) a. $[_{VP}$ V Prt NP ...] ~ b. $[_{VP}$ V NP Prt ...]
e.g., look up the information e.g., look the information up

A question that has been debated for a long time in syntactic theory is whether these two forms are variants of a single form or distinct but related constructions. Most derivational accounts take them to be variants of a single structure, because of synonymy.

However, there is some indication that the two-construction alternative is correct. Note first that the examples in (40) show that *right*+Prt can only appear after the NP (Emonds 1972: 551–2).

(40) a. She looked (*right up) the information (right up).
b. I wrote (*right down) their names (right down).
c. Leslie put (*right away) the groceries (right away).
d. Sandy ticked (*right off) the professor (right off).

Jackendoff (2002a) suggests that *right* is a specifier of the particle. A direct way to characterize the difference seen here is to say that the particle lacks a specifier when it precedes the NP, for some reason—it might be that the pre-NP particle is not phrasal, or is of a category that lacks a specifier, etc. Such a characterization of the difference treats (39a) and (39b) as distinct

[19] As discussed in more detail in Chapter 7, we would expect a structurally deviant construction to become increasingly rare in the language to the point that it disappears from usage, unless it occurs with high frequency or has some other salient property (such as a privileged interpretation) that effectively distinguishes it from more canonical alternatives.

constructions with slightly different structures, i.e. [V Prt NP] and [V NP [$_{\text{PrtP}}$ (Spec) Prt]].

On the traditional view, these two constructions are derived from the same source, by reordering the particle and the direct object.[20] For example, suppose that there are two possible input structures, one of which contains [*right* Prt] in VP-final position and one of which contains [Prt], as shown in (41).

(41) a. [V [v_1 [NP v_2 [$_{\text{PrtP}}$ [$_{\text{SPEC}}$ right] Prt]]]]
b. [V [v_1 [NP v_2 Prt]]]

In both derivations there is an empty head v_1 between V and NP that serves as a landing site for a moved Prt. In one structure the Prt is in a PrtP that is a complement of another v_2, shown in (41a), and in the other the Prt itself is the complement of v_2, as in (41b). On this analysis, Prt moves to v_2 and then to v_1 in (41b) because both moves are 'head-to-head,' but movement of PrtP is blocked in (41a) because it is phrasal, and hence cannot move to a head position.

The fact that a syntactic analysis can be formulated that captures the difference between *right*+Prt and Prt is a corollary of the fact that a syntactic analysis can always be formulated to capture a difference, if the right stipulations are made. In this case, it must be stipulated that head movement of Prt is blocked when there is an overt specifier of Prt, in order to avoid deriving **look up$_i$ the information right t$_i$*,[21] that Prt and PrtP originate in VP-final position, that there are two empty *v*'s through which Prt moves, that there are principles and features that license the movements, and that the branching structure is as shown in (41).

Given these assumptions, it is possible to express other properties of the construction in these terms as well. For example, if we assume that a pronominal NP cliticizes to V, it will be impossible for Prt to move to the left of the pronoun, which rules out cases such as (42).

[20] See Dehé (2005) for an OT account of the factors bearing on the two orders. Dehé assumes that they are categorically identical, and differ only on the order of the forms.

[21] There may not be a general prohibition on extracting a head over its specifier that explains this constraint, given the possibility of doing just that in the German split-NP construction exemplified in (i).

(i) Bucher hat er keine gelesen.
books has he no read
'As for books, he hasn't read any.'

See Ott (2009) and references cited there.

(42) *She looked up it.
*I wrote down them.
*Leslie put away some.
*Sandy really ticked off us.

However, there is evidence that this general direction is the wrong one, precisely because it does not treat the two constructions as distinct. There are V-NP-Prt idioms that do not have a V-Prt-NP counterpart, and vice versa. For instance, *work one's butt off* does not alternate with **work off one's butt.*

(43) a. Freddy worked his butt off. (Meaning: Freddy worked to an extreme degree.)
b. *Freddy worked off his butt.

The absence of an alternation here is plausible if we think of *work one's butt off* as having small clause interpretation, paraphraseable as *work (so hard that) one's butt (falls) off.*[22] The constructional meaning (what is in parentheses in the paraphrase) is nevertheless not explicit.

As expected, there are also idioms of the form V-Prt-NP that decidedly disfavor the V-NP-Prt counterpart. E.g.,

(44) a. Their actions set off a violent reaction.
b. *Their actions set a violent reaction off.

(45) a. He threw up what he had eaten for breakfast.
b. *He threw what he had eaten for breakfast up.

(46) a. We have given up smoking.
b. *We have given smoking up.

(47) a. Let's take out pizza tonight.
b. *Let's take pizza out tonight.

Whether the asterisked examples are simply ungrammatical, or unacceptable for reasons of processing or low frequency of occurrence is a question that I set aside here.

Given that both particle structures must be available for different idioms, it follows that both are available for the purpose of defining the two constructions where the particle can appear either before or after the NP. Permitting the independent existence of these two constructions in the grammar allows us to distinguish them with regards to the appearance of the specifier, since now the VP-final particle can be distinguished from the VP-internal particle

[22] Thanks to a member of the audience when I presented this material at the University of Cambridge for suggesting this interpretation.

simply in terms of its position in the phrase. While differences in the representation undermine the generality of the movement analysis, they contribute little in complexity to the constructional analysis.

The two constructions of (39) are specializations of the more general schema (30), if we take particles to be adjuncts—that is, intransitive prepositions heading prepositional phrases (as in Emonds 1976). These constructions are specific in at least two ways that do not follow from general principles: (i) they include particles of a sort that are not found in all languages, and (ii) they specify where the particles appear in the constituent ordering in VP.

The data involving verb–particle constructions therefore suggest that in general it is not possible to derive one construction from the other. Rather, they are two different (sets of) constructions. One is V-Prt-NP, and the other is V-NP-PrtP.

2.3.2 Coverage

Going back as far as Lakoff (1965) and Rosenbaum (1967), it has been recognized in syntactic theory that particular constructions are indexed to particular lexical items. For example, raising to subject, as exemplified in (48), applies to *seem*, *appear*, and a few other verbs and adjectives.

$$\text{(48)}\quad \text{There}\left\{\begin{array}{l}\text{seem}\\ \text{appear}\\ \text{*look like}\\ \text{are (un)likely}\\ \text{*are (im)probable}\\ \text{*are (im)possible}\end{array}\right\}\text{to be lots of surprises in this year's elections.}$$

The occurrence of this construction with the semantically related *seem* and *appear* and with *likely* and *unlikely* is surely not accidental. On the other hand, the fact that other semantically related items do not participate in the construction is an idiosyncrasy and therefore contributes to complexity of the lexicon. The idiosyncrasy seen here supports the view that cases such as raising to subject should be treated as constructions, and not as derivations from an abstract syntactic structure. While it is possible to index such derivations so that they are sensitive to the properties of the particular lexical items that they apply to, as Lakoff did, doing so is essentially a constructional analysis expressed in derivational terms.

Additional evidence that raising is in fact constructional is shown in Figure 2.1. The expression *there is unlikely* is non-existent before 1920, but begins to grow thereafter, as the trajectory of *there is likely* becomes steeper. It appears that the raising construction, formulated in terms of the lexical item *likely*, is generalizing over time to the lexical item *unlikely*. This lexical

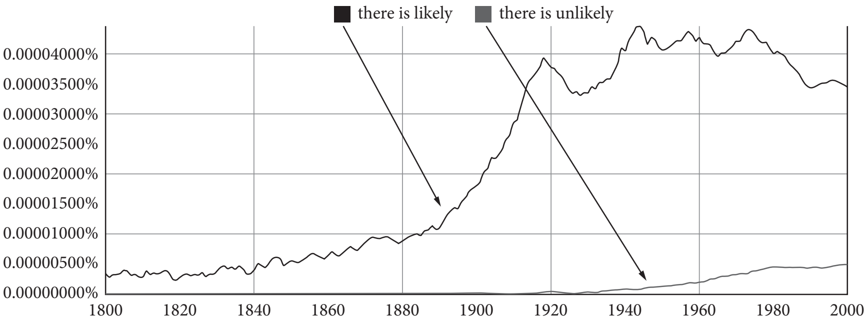


FIGURE 2.1. Occurrence of *there is likely* and *there is unlikely* in GoogleBooks N-gram viewer

diffusion may be the consequence of spread of the *unlikely* case to more speakers, or to a strengthening of the *unlikely* case in the lexicon of individual speakers, or both.

Lexical diffusion of this sort appears to be a natural phenomenon, which we may interpret as reducing complexity in the lexicon by simplifying the description of the set of lexical items in the construction. If a set of terms in a construction σ consists of two unrelated words, that is more costly than if σ consists of two related words, other things being equal. And if σ consists of many semantically related words, that is less costly than if it consists of only a few semantically related words and leaves out many others that are equally close in meaning.

I assume that the specification in a construction of an unprincipled exception is costly. There appear to be two properties of the set σ that are relevant, **similarity** of elements and **evenness** of coverage. Moreover, if the set is defined very narrowly, that is more costly than if it is defined very generally. Consider some examples.

- A construction that is restricted to the words *girl* and *television* is more costly than one that is restricted to the words *girl* and *boy*.
- A construction that is restricted to the words *girl* and *boy* is more costly than one that is restricted to all of the words that refer generically to male and female humans.
- A construction that is restricted to the words *girl*, *boy*, *woman*, *man* is less costly than one that is restricted to *girl*, *boy*, and *woman* but not to *man*.

In those cases where there is a strict taxonomic relationship among the words, it is possible to find simple ways of measuring the complexity of a set. For example, if the structure is a binary feature tree, the designation

[+HUMAN] is less complex than [+HUMAN,+YOUNG], and both are less complex than the set {girl, television}. Furthermore, [+HUMAN] is less complex than [+HUMAN, -ADULT, -MALE]. This complexity can be measured in rough terms by simply counting features and assigning a high cost to unanalyzed words.

But in general such a simple feature structure will not be available. Rather, we have an n-dimensional space in which words may be close to one another on one dimension but distant on another (Pustejovsky 1995; Pustejovsky and Boguraev 1993). For example, *apple*, *orange*, *lemon*, *tomato*, *cucumber* share the feature of being foods, the first three are fruits while the last two are vegetables, *tomato* and *apple* are prototypically red, *orange* and *lemon* are citrus, and so on. It is possible to order some of these features in a hierarchy, but prototypical color is a problem, since it cuts across all of the subcategories.

Suppose that the words in a language are distributed in an n-dimensional rectilinear space, corresponding to their semantic properties. The location of each word is represented by its coordinates in this space. Two words may be close to one another in meaning on one dimension, and far from one another on another. For simplicity, let us begin by assuming that there is no internal structure to this space; this is clearly an oversimplification.[23]

For any two words in the n-space, there is a smallest n-dimensional rectangle that contains these words. Using 2-space for the purpose of illustration, there is a smallest rectangle that contains the two points corresponding to these two words. Let us call this the 'local domain' of the set of words.

We may measure the complexity of any set of words by computing the number of words in the set as a fraction of the number of points in the local domain. For simplicity, let us assume that every point in the local domain is not only a possible word, but an actual word. In the case of Figure 2.2, the local domain is a 5x10 rectangle, so the number of points is 50. Since only two points are filled, the generality measure is 2/50=.04. If all of the points were filled, the measure would be 50/50=1.0. If all but one were filled, it would be 49/50=.98, and so on. This measure reflects the 'coverage' of the construction relative to the possible lexical items in the space.

For example, an English V-Prt construction that incorporates body parts is *to V one's Y off*. Y consists of *head* and *ass*. Not surprisingly, it is possible to find instances on the Web of *to laugh/work/sing/play/dance . . . one's Y off* for a

[23] Some dimensions simply are irrelevant to others. For example, AGE is relevant to humans, and we have different words for humans of different ages, but it is not a relevant dimension of variation for furniture. If it was, though, it would not be surprising to find words for old and young furniture in the languages of the world. This seems less plausible as we move into the realm of abstractions, although even there it is possible to predicate age—e.g. *new* (and metaphorically *young*), *old ideas*.

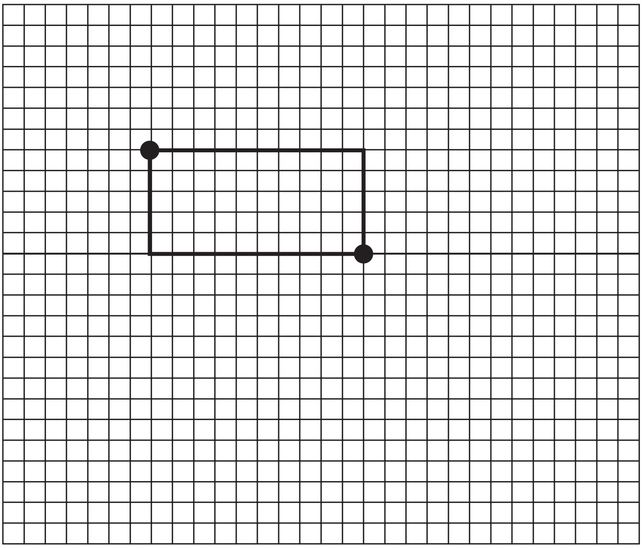


FIGURE 2.2. Smallest rectangle containing two words in 2-space

large number of synonyms of *ass*, including *buttocks*, *nates*, *arse*, *butt*, *backside*, *bum*, *buns*, *can*, *fundament*, *hindquarters*, *hind end*, *kiester*, *posterior*, *prat*, *rear*, *rear end*, *rump*, *stern*, *seat*, *tail*, *tooshie*, *tush*, *bottom*, *behind*, *derrière*, and *fanny*. These words form a compact region of the semantic space, and it would be surprising if the more common ones could not be used in this construction.

But what is more interesting is that other body parts can be use in this construction: *to V one's balls off*, *to V one's head off*, and *to V one's fingers off*, as well as the non-body part *to V one's socks off*. With some verbs the range of nouns is semantically more restricted, e.g. *sing one's head off* vs **sing one's feet off*, *run one's feet off* (not surprisingly) vs **run one's fingers off* (not surprisingly), and so on. The possibility appears to be constrained by the CS, roughly '*V* so hard that *one's N* fall(s) *off*,' where 'so hard' and 'fall(s)' are the constructional meaning.

In sum, increase in coverage is revealed in the extension of a particular construction to lexical items that share semantic features. In principle we can track changes in the use of a construction that has a lexical component to test this prediction.

To take another example in unpublished work, Culicover and Dellert (2008) looked at the distribution of English adjectives with verbs of change of state, meaning 'become.' The distributions are very diverse, for each

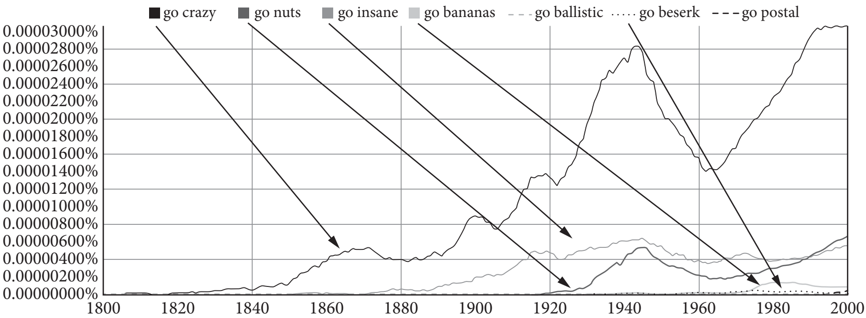


FIGURE 2.3. Frequency of occurrence of adjectives meaning 'crazy' with *go* from Google N-gram viewer

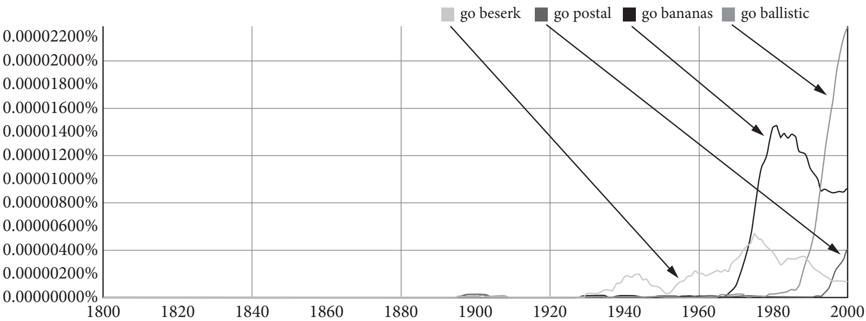


FIGURE 2.4. Frequency of *go bananas*, *postal*, *ballistic*, *beserk* from Google N-gram viewer

combination of verb and adjective. Figure 2.3 shows the occurrence of seven adjectives meaning 'crazy' with *go*: *go crazy*, *go nuts*, *go insane*, *go bananas*, *go postal*, *go ballistic*, and *go beserk* between 1800 and 2000, from Google N-gram viewer.

First *crazy*, then *insane*, then *nuts*, then *beserk*, then *bananas*, then *ballistic*, and finally *postal* are used with *go* with the meaning 'become crazy.' Each one gradually grows in usage after its initial appearance.

What we see here is the gradual increase in coverage of the construction of the form *go*+Adj, meaning 'become crazy.' The growth in the set of adjectives, particularly of the later adjectives, in this construction is seen more clearly in Figure 2.4, where the earlier ones are not shown. It is clear that the temporal order of emergence is first *beserk*, then *bananas*, *ballistic*, and finally *postal*.

The distribution of these adjectives in these constructions is of course nothing like their distribution in the language overall, as shown in Figure 2.5. *Bananas*, *nuts*, and *postal* have other meanings besides 'crazy.'

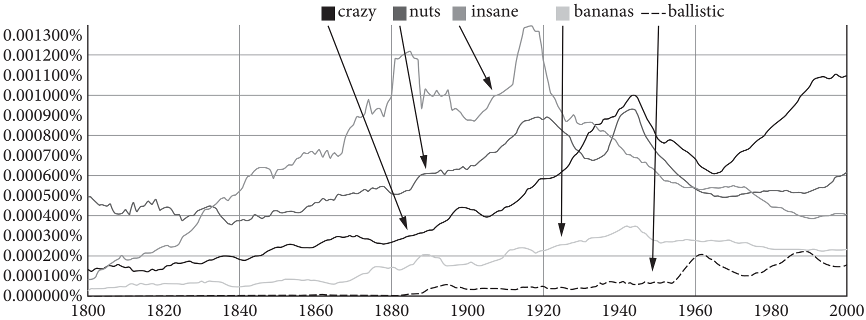


FIGURE 2.5. Frequency of occurrence of adjectives meaning 'crazy' from Google N-gram viewer

The other constructions that we have considered include additional specifications beyond the basic structures, and are therefore more marked. These specifications are of two types. In the case of the verb–particle construction, the position of the particle is restricted, and it is debatable whether these constructions can be derived entirely from general principles external to English. In the case of the *V one's Y off* and *V one's way* constructions, there are additional restrictions on form, and the correspondence with CS is further restricted. On the other hand, the sound+motion construction is entirely regular in form, but supplies additional meaning in the presence of a PP that denotes a path.

What we are measuring when we look at coverage is degree of idiosyncrasy, which is often referred to in the literature as 'idiomaticity.' Isolated form–meaning correspondences are highly idiomatic, to the extent that no words of similar meaning can be substituted for the words in the correspondence. With less idiomaticity, the construction becomes more compositional, that is, its meaning is more understandable in terms of the meanings of the parts. In other words, the meaning of the construction can be formulated in terms of the more general features of the words, rather than in terms of the specific forms.[24]

[24] For an investigation into how idiomaticity is measured, and how it influences the distribution of forms in corpora, see Wulff (2008, 2009). An earlier study by Nunberg et al. (1994) provides an important perspective on idiomaticity and compositionality.

2.3.3 Strain

Jackendoff's principle (19) suggests not only that deviation from the regular pattern should be costly, but that degree of deviation from the regular pattern should correspond to relative cost. I sketch here a preliminary approach to degree of deviation in terms of 'strain.'

Strain measures the extent to which a construction deviates from the canonical structure of the language, as formulated in terms of category membership or constituent order. If there is a correspondence that specifies that a particular category precedes another category and the interpretation of this sequence, it is an additional cost to exempt a lexical item from this correspondence. I assume that grammars are organized to minimize strain, other things being equal, and that as strain increases, there is increased pressure to reorganize the grammar.

This does not mean that strain is always reduced—there are certainly instances in which strain is increased in a defined area of the grammar because of changes elsewhere in the language, or the reduction of strain is inhibited by other factors. Hence the use of "other things being equal," which is always a crucial proviso when considering the way in which the grammar of a language reacts to specific areas of complexity.

Here is an example of strain. Culicover and Jackendoff (2005) discuss some "words that go in the wrong place," as illustrated in (49). To this list I add a few more English cases in (50).

(49) a. two years $\left\{\begin{array}{l}\text{hence}\\ \text{ago}\end{array}\right\}$ ~ $*\left\{\begin{array}{l}\text{hence}\\ \text{ago}\end{array}\right\}$ two years

b. popcorn $\left\{\begin{array}{l}\text{galore}\\ \text{aplenty}\end{array}\right\}$ ~ $*\left\{\begin{array}{l}\text{galore}\\ \text{aplenty}\end{array}\right\}$ popcorn

c. the official responsible ~ the responsible official

d. your objections notwithstanding ~ notwithstanding your objections

e. tall enough ~ *enough tall

(50) a. Ask not what you can do for your country.

b. Be that as it may.

c. Come what may.

d. Easy come, easy go.

e. Come Monday morning, you are not going to be looking so pleased with yourself.

f. One swallow does not a summer make.

g. can't seem [= 'doesn't seem to be able']

Saying that a word goes in the "wrong place" with respect to its head, either optionally or obligatorily, sets it aside from the canonical case in which nothing special needs to be said. The category of *galore* is arguably Adj, and typically the order in English is Adj-N.[25] In order to avoid **galore popcorn*, the lexical entry for *galore* must stipulate its post-head position.

In order to get the desired result in terms of the measurement of complexity, we must guarantee that the cost of a lexicon in which there is a general term in a construction is significantly lower than that of a lexicon in which there are many individual items with the same non-canonical property. Since by assumption the individual items and the construction are in the lexicon, the measurement of strain is more specifically the measurement of lexical strain.

The description of the idiosyncratic properties of these elements is essentially the same as that of the general construction, which makes the comparison relatively straightforward. The lexical entry for *galore* is given in (51a), while for comparison that of the normal adjective *pretty* is given in (51b). I use the symbol ⇔ to indicate the form–meaning correspondences. I use subscripts to show how the meaning of *galore* corresponds to its position.

(51) a. pretty: Adj ⇔ PRETTY
b. galore_1: N Adj_1 ⇔ MANY_1

The corresponding construction in a language that allows both pre- and post-nominal adjectives would be (52), where Adj′ is the interpretation of Adj.

(52) $\text{N Adj}_1 \Leftrightarrow \text{Adj}'_1$

We can measure the strain of this formulation in a rough way by dividing the number of adjectives that follow N by the total number of adjectives in the language. The number in English is so small that the strain is extremely small. So a few isolated post-nominal adjectives contribute relatively little to the complexity of the grammar.

But suppose that the number of post-nominal adjectives is significant. At some point, say 10 percent or perhaps even less, the exceptions are no longer odd exceptions, but are candidates for a subregularity. Listing each adjective individually as an exception adds considerably to the strain, since they constitute 10 percent of the population of adjectives. Hence we might expect

[25] It is not entirely certain that *galore* and *aplenty* are adjectives. They could be quantifiers, and the argument would not be affected, since quantifiers precede the noun in English, too. *Galore* means 'abundant,' so it could be an adjective that denotes quantity. But unlike *abundant*, it cannot be compared: *more abundant*, **more galore*. On the other hand, unlike quantifiers like *many*, *much*, and *a lot*, *galore* cannot appear without a head noun: *I have many/much/a lot*, **I have galore*. Similarly for *aplenty*, although *I have plenty* is fine.

a reorganization of the grammar to reduce the strain (again, other things being equal), through the bifurcation of the class of adjectives that modify nouns into two sets of adjectives, each of which is a term in a distinct construction, on the basis of some meaning difference. This is what appears to be the case in languages such as French, Spanish, and Italian that show both orders in NP (see Laenzinger 2005 and works cited there). A further respecialization of the two constructions would be for each one to have a general interpretation associated with it, e.g. attributive for Adj-N and predicative for N-Adj. At this point, an adjective could appear in either construction, and the constructions would then be indexed to the meaning difference and not indexed to particular subcategories of adjectives.

The reduction in strain is a special case of increased coverage, where a set of words is replaced by a general characterization of the set, as discussed in §2.3.2. I hypothesize that simply having an unrelated set of words listed as a term in a construction is less costly than the strain of having a large set of unrelated words in the "wrong place." This is a special case of the notion that explicit mention in a construction of a term in a canonical position is less costly than specifying a non-canonical order.

Another source of strain in a construction is the use of a word of one category as though it is a member of another category. This is exemplified by (50a,c,e), where the main verb behaves as though it is a member of the category V_{AUX}. These special constructions are likely remnants of the stage of English when all verbs functioned as the head of S=IP with respect to negation and inversion (see Chapter 7).

This last example is particularly instructive, because it suggests one plausible causal mechanism for the creation and modification of constructions in the course of language change. The level of strain is a measure of complexity; the reduction of strain is reduction in complexity. The reduction is accomplished by reinterpreting what is apparently unprincipled variation in the syntactic structure as reflecting a meaning difference. The intuition that variation is reinterpreted in this way is in fact a long-standing one in both formal and informal linguistic theorizing.[26] Chapter 7 considers several cases of grammatical change from this perspective in some detail.

[26] An extreme version of this metric is the Uniqueness principle of Wexler (1978). As with other general principles and constraints of mainstream generative grammar, I take such a principle to be properly a measure of relative complexity, and not an absolute prohibition against certain types of correspondences.

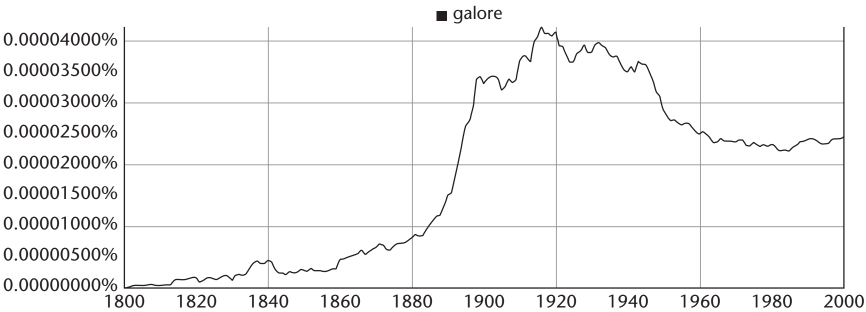


FIGURE 2.6. Frequency of *galore* from Google N-gram viewer

The reduction of strain is exemplified by the elimination of special properties of words like *galore* that put them into non-canonical positions. Whether this will actually happen with a given lexical item in the course of language change is dependent on a number of factors, however, besides the degree of strain. For instance, *galore* always occurs post-nominally and is a relatively rare word, so that there is a high degree of correlation (apparently 100 percent[27]) between its appearance in the language and its appearance in this position. Such a high correlation works against regularization, since N-*galore* is a construction in itself—it is *sui generis*, and in fact one might argue that the syntactic category of *galore* is defined uniquely by its participation in this construction (see footnote 25). To the extent that N-*galore* has a special rhetorical character, as well, it is likely to persist in the language. Figure 2.6, generated by Google N-gram viewer, and Figure 2.7, generated from the 150 billion word Google Books database (Davies 2011), show that its use has been growing since 1800, although it appears to have peaked in the 1910s and has stabilized at a lower level since then.

In contrast, a word with idiosyncratic ordering properties that shares certain semantic properties with regular words, such as *enough*, might begin to shift to a regular distribution more readily. A Google search finds isolated instances of such expressions as *enough tall*,[28] *enough thin*,[29] *enough pretty*,[30]

[27] All of the instances of *galore* in the Corpus of Contemporary American English appear in the construction N *galore*.

[28] "Sometimes the content is enough tall to do that, but sometimes the content is too short and the DIV won't touch the bottom of the screen" <http://stackoverflow.com/questions/2363455/making-div-height-at-least-as-tall-as-the-page-in-css>.

[29] "energy stream density can still be obtained inside the shell if the shell is enough thin" <arxiv.org/pdf/1006.1036>.

[30] "His sister's black/But she is sure enough pretty" ("Living just enough for the city," by Stevie Wonder).

CLICK ON A BAR TO SEE MATCHING STRINGS FROM THAT DECADE

DECADE	1810s	1820s	1830s	1840s	1850s	1860s	1870s	1880s	1890s	1900s	1910s	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s
SIZE (MIL)	378	655	1,437	1,938	2,953	2,953	2,844	4,408	5,632	7,520	10,087	7,089	5,795	6,167	8,104	13,192	14,011	15,511	19,816	26,882
TOKENS	1	0	6	6	12	18	37	86	282	633	818	492	529	541	496	742	621	728	1,136	1,795
PER MIL	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.05	0.08	0.08	0.07	0.09	0.09	0.06	0.06	0.04	0.05	0.06	0.07

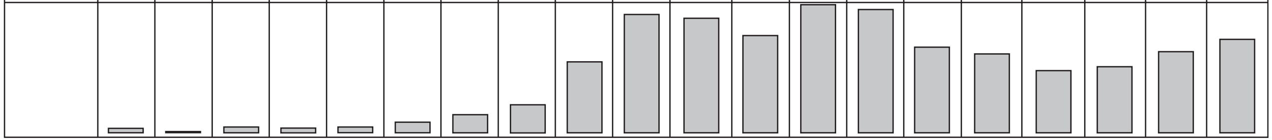


FIGURE 2.7. Frequency of N-*galore* in Google Books, 1810–2000

DECADE	1810s	1820s	1830s	1840s	1850s	1860s	1870s	1880s	1890s	1900s	1910s	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s
SIZE (MIL)	378	655	1,437	1,938	2,953	2,953	2,844	4,408	5,632	7,520	10,087	7,089	5,795	6,167	8,104	13,192	14,011	15,511	19,816	26,882
TOKENS	254	361	985	1,413	2,956	2,676	3,760	7,615	11,149	18,344	30,021	22,800	20,179	26,251	32,133	51,084	53,997	56,490	71,976	105,080
PER MIL	0.67	0.55	0.69	0.73	1.00	1.14	1.32	1.73	1.98	2.44	2.98	3.22	3.48	4.26	3.97	3.87	3.85	3.64	3.63	3.91

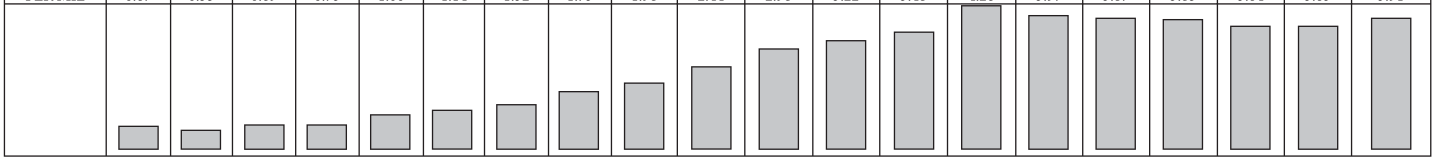


FIGURE 2.8. Occurrence of *enough Adj* in Google Books database, 1810–2000s

enough complicated,[31] used predicatively, as in *it is enough tall* instead of *it is tall enough*. Figure 2.8 shows the occurrence of expressions of the form *enough Adj* in Google Books. While the numbers are not large, they suggest a persistent but ultimately not yet successful tendency to regularize the position of *enough*. A Google search also turns up several instances of *an enough tall N*, *an enough rich N*, and *enough smart N*, for example; it is not clear to what extent these have been produced by native speakers.

[31] "Simply because life is already enough complicated and messed up the way it is" <http://www.facebook.com/group.php?gid=2253735512>.

PART II

English constructions

This page intentionally left blank

3

Isolating constructional complexity: two case studies

This chapter analyzes in some detail two cases of constructional complexity in English. §3.1 investigates the details of English relative constructions, including relative clauses and reduced relatives. This section builds on the analysis of English relative clauses of Sag (1997). While classical generative grammar takes the relative clause to be a prime instance of 'core grammar,' Sag's analysis shows that there are substantial idiosyncrasies in the English relative clauses surrounding a common general structure.

The analysis developed in §3.1 simplifies and extends Sag's treatment. I show that relative clauses and other relatives share an interpretation that each particular syntactic configuration corresponds to in a systematic way. This correspondence simplifies and unifies the description of the set of constructions. As a consequence of establishing this constructional uniformity, it is possible to isolate precisely the respects in which particular special cases are idiosyncratic. This exercise lends support to the view that constructional grammars are less complex than arbitrary collections of logically possible constructions.

A second class of constructions, taken up in §3.2, demonstrates that idiosyncrasies may also override the canonical structure of the language. These are English focus inversion constructions, exemplified by sentences such as *Cindy runs much faster than does any of her brothers.* Here the subject NP *any of her brothers* is in final position in the VP headed by *does.* Focus inversion is exceptional with respect to the canonical structure of English, which requires that the subject precede the inflected verb (except in cases of subject aux inversion, which is different from focus inversion).

Furthermore, I argue that it is no accident that these constructions have this property—VP-final position is a typical position for focus in English, and the focus inversion constructions take advantage of this fact. The fact that this correspondence violates the canonical syntactic requirements of the language again supports the constructional approach.

3.1 English relatives[1]

3.1.1 *Inheritance and complexity*

An important insight of accounts of linguistic structure that are formulated in terms of constructions is that inheritance of properties can be used to catalogue the relatedness among them. The general idea is that some construction B is an instance of a more general construction A, such that B inherits the properties of A and specifies some of its own in addition.

It is easy to see that the complexity of a set of constructions can be reduced to the extent that properties of constructions are inherited. This is a central point of Jackendoff (1975), as discussed in Chapter 2, and it goes back to the earliest work in generative grammar on the use of notational conventions to capture generalizations (Chomsky 1965: 42–5; Chomsky and Halle 1968: chapter 3). In particular, if A and B share properties for some principled reason, then the description should explicitly take account of this fact. The description then distinguishes the 'natural' situation where there are shared properties from the situation in which A and C (for example) share no properties. The metric for computing complexity is formulated so that the sharing of properties between A and B counts as more economical than does the non-sharing case of A and C. In current terms, the more properties B inherits from A, the less complex the description of the set of constructions.

Sag (1997) formulates inheritance relations between relative clauses in the notation of HPSG. HPSG is well suited to illuminating inheritance relations in syntax, because it permits the enumeration of types that have particular featural characterizations. For instance, on Sag's account, all relative clauses must satisfy the following constraint (p. 44).

(1) *rel-cl*

$$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{MC} & - \\ \text{INV} & - \\ \text{MOD} & \begin{bmatrix} \text{HEAD } \textit{noun} \end{bmatrix} \end{bmatrix} \\ \text{CONTENT} & \textit{proposition} \end{bmatrix}$$

[1] This section was originally published as Culicover (2011). It has been slightly revised for inclusion here. I am grateful to William Schuler, Detmar Meurers, Bob Levine, Thomas Hoffmann, Wolfgang Sternefeld, and audiences at the University of Tübingen for very useful advice and comments. Two anonymous reviewers offered numerous comments and questions that have led to significant improvements, and I thank them for their efforts. I am responsible for any and all remaining errors.

This constraint says that the relative clause cannot be a main clause (notated as [MC –]), that it does not show inversion (notated as [INV –]), and that it modifies a noun (notated as [MOD [HEAD *noun*]]).[2]

On Sag's analysis, all relative clauses satisfy the *rel-cl* constraint, whether they are finite or non-finite, whether the relative proform is a subject or a non-subject, or whether or not there is a relative proform. All that needs to be specified in the description of the infinitival relative, for example, are those properties that distinguish it from all other relative clauses, i.e. those that it does not inherit in virtue of being a relative clause, as well as those that it inherits in virtue of being infinitival.

Inheritance of constructional properties thus captures a generalization, in precisely the way envisioned by markedness theory (Chomsky 1964, 1965).[3] If a particular type of relative clause required inversion, for example, then it would be necessary to stipulate this fact explicitly, since otherwise this special type of relative clause would inherit [INV –] from *rel-cl*.

A particularly important aspect of the constructional approach is that such an idiosyncrasy is not impossible, but simply costly in terms of the complexity metric. While it can be accommodated in the description of the synchronic grammar of the language, it would be reasonable to expect that it would be rare, at least in the dialects of the language and in other languages where [INV –] is a feature of *rel-cl*. Where such an idiosyncratic property does appear, it might be expected that it would coexist with an otherwise identical construction that does obey [INV –], and that the more idiosyncratic construction would have a special interpretation that the regular one does not share. As far as I know, such costly idiosyncrasies do not exist, although they are logically possible; all of the cases of idiosyncrasy discussed in this chapter are of the type where a property of a particular construction supplements those inherited from more general constructions.

3.1.2 Relatives as constructions—Sag (1997)

With the general relationship between inheritance and complexity in mind, let us look at the set of relative constructions in English described by Sag (1997). I focus, as Sag does, on the restrictive relatives that are embedded in noun phrases.

[2] It is actually possible to have inversion in a relative clause when there is an initial negative topic. E.g.,

(i) a. a candidate who under no circumstances would I vote for
 b. these candidates, none of whom would I vote for,...

Inversion in these cases is not a property of the relative clause, but of the negative topic, although it is the head that is inverted. See Culicover (1991) for discussion of such examples.

[3] For more extensive discussion of this point, see Hoffmann (2011: chapter 6).

The common properties of relative clauses in Sag's account are given in (1). With the exception of [INV –] and [MC –], and that they are clauses, there are no properties that all relative clauses share. For example, not all relative clauses show an overt A′ chain, that is, a constituent in a non-argument position and a corresponding gap. In fact, not all relative clauses have an overt constituent in A′ position, and not all relative clauses have a gap.[4] And, furthermore, not all relatives are clauses.

In finite clauses, subjects, objects, oblique arguments, and adjuncts may be relativized. There may be a relative proform in initial position (2), or *that* (3), or zero (4). Zero is excluded when what is relativized is the highest subject (4a). A relativized subject of an embedded clause is possible in a zero relative (4d). *That* and zero are also possible in relatives modifying the heads *time*, *place*, and *reason* (5).

(2) a. the book which is on the table (is expensive) [Subject]
b. the book which you put __ on the table (is expensive) [Object]
c. the table on which you put the book __ (is expensive) [Oblique object]
d. the time when you put the book on the table __ (is unknown) [Adjunct]

(3) a. the book that __ is on the table (is expensive) [Subject]
b. the book that you put __ on the table (is expensive) [Object]
c. *the table that you put the book __ (is flat) [Oblique object]

(4) a. *the book __ is on the table (is expensive) [Subject]
b. the book you put __ on the table (is expensive) [Object]
c. *the table you put the book __ (is flat) [Oblique object]
d. the book [you said [__ is on the table]] (is expensive) [Embedded subject]

(5) a. The time (that) I left was a bit after midnight.
b. The place (that) I live is accessible to work.
c. The reason (that) I called is not important

In infinitival relatives, there can be a relative proform only in an initial PP. Hence there is no proform when the subject is relativized (6) or when a non-subject NP is relativized (7)–(8).

[4] In Chomsky (1977) it is simply assumed that all relative clauses contain A′ chains, and that they are present in the structure even if there is no overt evidence for them. As a consequence, it is necessary to analyze some relative clauses in terms of invisible movement of invisible operators that create invisible gaps. For general arguments against such an approach to syntax, see Culicover and Jackendoff (2005: chapters 2 and 3).

(6) the man $\left\{\begin{array}{l}\text{who should}\\ \text{*who to}\\ \text{to}\end{array}\right\}$ clean the kitchen

(7) a. the book (*which) to put __ on the table
b. the table (*which) to put the book on __
c. the table on which to put the book

(8) the time $\left\{\begin{array}{l}\text{*when}\\ \text{at which}\\ \varnothing\end{array}\right\}$ to put the book on the table __

Finally, there are 'reduced' relatives. These are present participial predicates (9) or passive participial predicates (10), adjective phrases (11), and prepositional phrases (12). The reduced relatives lack a relative proform and always relativize the "subject,"[5] hence they lack an apparent gap.

(9) the book (*which) sitting on the table

(10) the book (*which) written by Tolstoy

(11) a book (*which) so interesting (would be fun to read)

(12) a book (*which) about syntactic theory[6]

I refer to these constructions as "relatives," because there is no prima facie reason to call them "clauses." Relative clauses are also relatives, and they are clauses.

As can be seen from this brief summary, there are a number of morphosyntactic properties that can be used to characterize the different relatives. Some of these are correlated, and others appear to be independent of one another.

(13) a. finite | infinitival | participial
b. relativized subject | relativized non-subject
c. relativized NP | relativized non-NP
d. A′ constituent | no A′ constituent
e. overt gap | no overt gap
f. single word A′ constituent | phrasal A′ constituent
g. NP A′ constituent | non-NP A′ constituent

[5] The scare quotes around "subject" are intended to suggest that what is relativized is the argument that would be the subject if the relative was a finite clause.

[6] It may be that in *a book about syntactic theory*, the PP is an argument. Note, though, that *a book about syntactic theory* can be paraphrased by *a book which is about syntactic theory*, and *The book is about syntactic theory* is quite unexceptional. Such facts suggest that *about syntactic theory* can also function as a predicate.

On a classical approach to syntactic constructions, such as Chomsky (1977), we would assume that there is a uniform structure for all relatives, and treat the various idiosyncrasies of the different types as potentially part of the periphery. At the same time, we would seek principled accounts of as many idiosyncrasies as possible.

The constructional approach taken by Sag, on the other hand, is to accept the various superficially distinct construction types as they are, and to formulate their various properties in terms of an inheritance hierarchy. The type *wh-subj-rel-cl* is a special case of *wh-rel-cl*, which in turn is a special case of *rel-cl*, and so it inherits the properties of (1). Sag accounts for the fact that a finite relative can have a relative pronoun that is a subject, but an infinitival relative cannot, by stipulating that the type *wh-subject-rel-cl* is an instance of the type *wh-rel-cl* and of the type *fin-hd-subj-ph*. Thus it inherits all properties of relatives with *wh*-phrases in initial position, and it is a finite clause with a subject (Sag 1997: 452). Crucially, this property of infinitival relatives does not follow from anything.

The inheritance hierarchy can be represented explicitly by linking the various types. Sag's complete hierarchy for *wh*-relatives is reproduced in Figure 3.1.

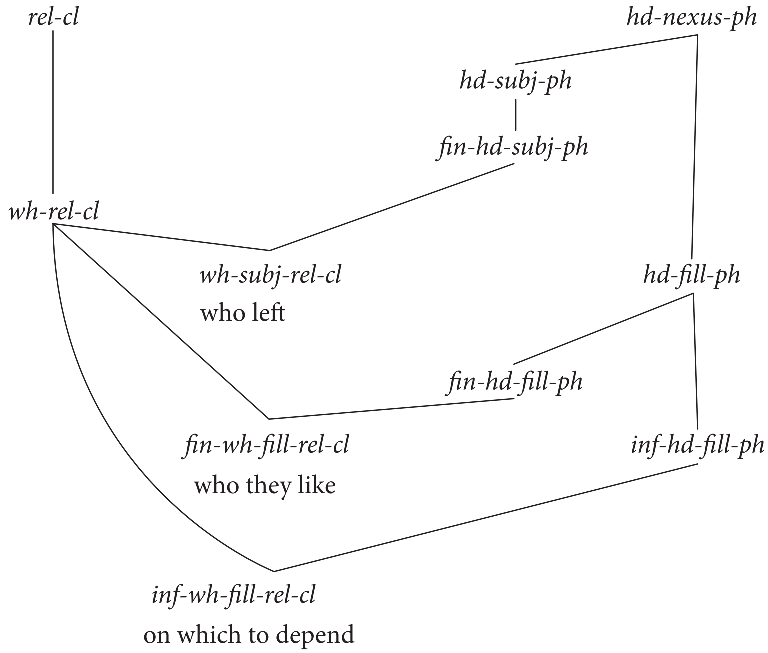


FIGURE 3.1. Inheritance hierarchy for *wh*-relatives (Sag 1997: 464)

There are a number of respects in which the subject of the infinitival relative does not pattern like the subject of the finite relative. While finite relatives systematically have overt subjects, an infinitival relative may have an overt subject (and the complementizer *for*) only if there is no overt *wh*-form.

(14) a. the woman for you to talk to __
b. the woman to whom to talk __
c. *the woman to whom for you to talk __

This constraint is shared with infinitival questions.

(15) a. I wonder who to talk to __
b. *I wonder who for you to talk to __[7]

Sag accounts for this idiosyncrasy of the infinitival relative by stipulating, through a constraint, that when the infinitival has a filler, it lacks an overt subject—see Figure 3.2.

The constraints define the type *inf-hd-fill-ph*, that is, an infinitival phrase with a filler (an A' constituent). The constraint HEAD specifies that such a type has an infinitival verb form, and the constraint HD-DTR specifies that it has a null subject. The ISA constraint specifies that this type of phrase is a phrase that has a filler, and hence inherits all properties of such a phrase that are not specified here.

Moreover, if there is a filler in an infinitival relative, it must be a PP.

[7] A Web search prompted by a reader's comment turns up a few cases of *the person who to VP*, e.g.,

(i) a. where do i find the person who to talk to about the quest? <http://answers.yahoo.com/question/index?qid=20100727153550AAPs5ed>
b. In this case, I'll refer to the radio/club DJ as being the person who to target. <http://independentmusicstartup.com/440/how-to-get-a-tastemaker-to-take-your-music-to-the-next-level/>
c. All requests for aid should include: . . . 3. The name and identity of the requesting person or the *person who to* contact upon arrival. <www.jdcap.org/SiteCollectionDocuments/EmergencyPlanExternal.pdf>
d. if you ae [sic] from out of Auckland and interested I can give you the number and person who to contact. <www.electricalforum.co.nz/index.php?action=more_details>
e. Please provide a phone number and name of a person who to contact to resolve this. <blog.klm.com/"the-ocean-of-the-air-connects-us-all"/1189/>

A few other examples, not cited here, are clearly written by non-native speakers. In any case, it should not come as a surprise if the regular pattern in infinitival *wh*-interrogatives has found its way into infinitival relatives, through first and second language learners. *The person who to contact* strikes me as significantly more acceptable than other infinitival relatives, and appears in approximately a dozen valid hits, including those cited in (i). But I find infinitival relatives with other *wh*-forms and other heads and VPs to be quite impossible.

TYPE	CONSTRAINTS	ISA
inf-hd-fill-ph	$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{comp} & \\ \text{VFORM} & \textit{inf} \end{bmatrix} \\ \text{HD-DTR} & \begin{bmatrix} \text{SUBJ} & \langle X \rangle \end{bmatrix} \end{bmatrix}$	*hd-fill-ph*

FIGURE 3.2. Constraint that rules out subjects in infinitival questions and relatives with overt fillers (Sag 1997: 461)

TYPE	CONSTRAINTS	ISA
inf-wh-fill-rel-cl	[NON-HD-DTRS <PP>]	*wh-rel-cl* & *inf-hd-fill-ph*

FIGURE 3.3. Constraint that allows only PP fillers in infinitival relatives (Sag 1997: 480)

(16) a. the man to whom to talk __
b. *the man who to talk to __

In this respect infinitival relatives contrast with infinitival questions, which permit an NP filler in initial position; see (15a). Sag accounts for this property of infinitival relatives with the constraint in Figure 3.3. Since this type of relative clause is *inf-hd-fill-ph*, it must have a filler. And since it is *wh-rel-cl*, it must have a *wh*-filler. The constraint [NON-HD-DTRS <PP>] specifies that this filler must be a PP.

While there are many other details that can be noted about various relative clause types, this brief summary should convey the flavor of the approach. It is instructive to note, for example, that the constraint that rules out a *wh*-subject NP in an infinitival relative (**the woman who to go*), that *wh-subject-rel-cl* is an instance of the type *fin-hd-subj-ph*, is different from the one that rules out a *wh*-non-subject NP in an infinitival relative (**the woman who to talk to*), Figure 3.1. These constraints, while observationally accurate, do not distinguish the idiosyncrasies from the regularities, because they are all couched in a uniform representation and have equal weight. Crucially, a comparable description would apply to an alternative English in which infinitival relatives were fully regular and the idiosyncrasies regarding subjects and PP fillers applied to finite relatives, for example.

There are various ways in which to rearrange the inheritance hierarchy so that the properties of the relative clause types are accurately captured. Simply arranging types does not convey a picture of what is natural and systematic and what is idiosyncratic, however. In the end, the question is, what are the true regularities, and what are the true idiosyncrasies? An optimal

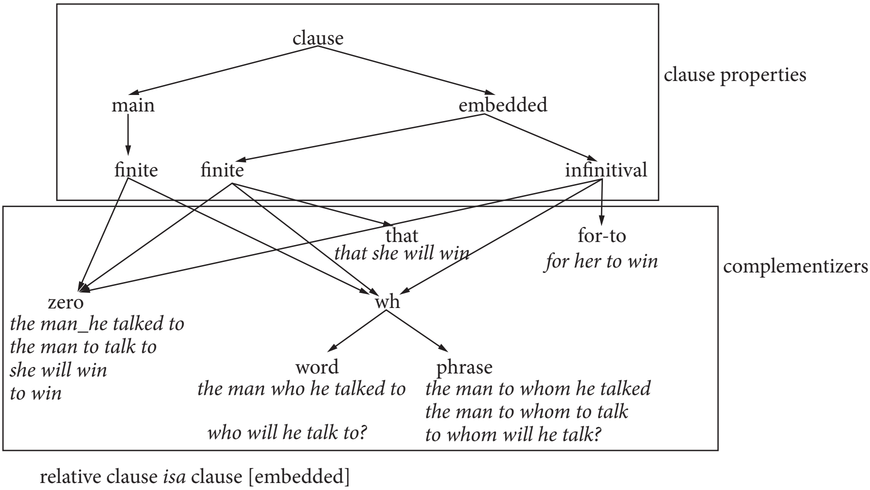


FIGURE 3.4. Summary of clause properties

characterization of the possible forms of English relative clauses would correlate generality with simplicity of description. In the next section, I reformulate Sag's inheritance hierarchy using a different inventory of relevant properties, and I argue that doing so yields a clearer and somewhat simpler picture of how this group of constructions is organized.

3.1.3 Simplifying the description

Let us consider the morphosyntactic properties of relative clauses in (13) that may enter into the formulation of constructions. There are two types of clauses, main and embedded. A relative clause must be embedded, that is, must not be a main clause. A clause in English may have a zero complementizer or initial *wh*-phrase, but only embedded clauses may have *for-to* or *that*. A *wh*-expression may be a word or a phrase. These properties of clauses are organized in Figure 3.4.[8]

Since a relative clause is [embedded], it will have all of the properties of [embedded] associated with [clause], and all of the properties of [clause], unless otherwise specified. Thus, if it is an embedded clause, it can be [finite] or [non-finite], if it is [finite] it can be [zero], [that], or [wh], and so on.

Figure 3.4 distinguishes eleven different types of relative clauses in terms of these properties—e.g. following the path from [clause] through [main] to

[8] I exclude free relatives from this characterization. They are not main clauses, but do not seem to have many properties of relative clauses beyond their A′ chains. I take up the question of how they are to be handled under a constructionalist approach in §3.1.5.

[phrase] defines 'finite tensed main clause with initial *wh*-phrase' and so on. There are no explicitly relational properties, e.g., whether there is an overt subject, whether the subject is relativized, or whether there is a filler-gap relation.

The problem now is to decide how to formulate constructions that express the properties that define canonical and exceptional syntactic structures, and their corresponding interpretations. Recall that our goal is to find a characterization of the various constructions so that their predictable properties are properly accounted for by inheritance. I take the primitive components of visible syntactic structure themselves to be the properties that define individual constructions, that is, just those given in Figure 3.4. So, for example, a finite clause is a constituent of category S with the feature [finite], and a finite relative clause is the syntactic configuration in which the finite clause is an adjunct to an N, as in (17).

(17) [$_{\text{NP}}$... N S[FINITE]]

Descriptions such as these, with their correspondences to conceptual structure, define the constructions of the language.

If a clause is [embedded], then it will inherit all of the properties of [embedded] and [clause] unless otherwise specified. Crucially, the way that the alternatives are framed in Figure 3.4 does not allow for the possibility that an embedded clause can be both [for-to] and [wh]. Since these are mutually exclusive primitive syntactic properties of embedded clauses, and infinitival clauses are embedded clauses, they are mutually exclusive properties of infinitival relatives (and infinitival questions, which are not shown in Figure 3.4). Thus cases such as **to whom for you to talk* are directly ruled out—there is no need to stipulate that a *wh*-phrase and a subject cannot coexist in clause-initial position of a relative clause.

Let us focus on the part of the inheritance hierarchy dealing with relative clauses. Figure 3.5 notes just those types that do not follow from [embedded] or [clause].

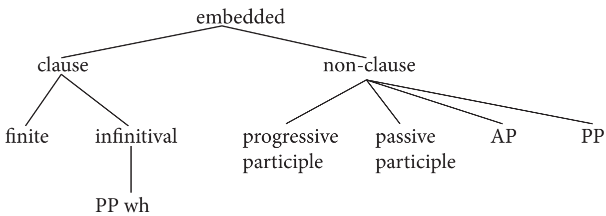


FIGURE 3.5. Type hierarchy for English relatives

By convention, mention of the type *PP wh* excludes those possibilities not mentioned, so there are no 'non-PP *wh*-infinitival' relative clauses. If no alternatives are mentioned, as in the case of [finite], all possibilities for [clause] are allowed.

Note that the participial, AP, and PP adjuncts are [embedded] but not [clause]. Making this distinction allows us to sidestep the [zero/wh/that/for-to] alternatives of Figure 3.4, which are applicable only to clauses. While Sag calls these relative clauses, and they are often referred to as reduced relative clauses, they differ from true relative clauses in four respects: (i) they lack tense inflection, which is a sign of clauses, (ii) they relativize only the subject argument, which is not a property of English relative clauses, (iii) they lack both the head and tail of a chain, and (iv) they can be complements of *be*, which is also not a property of English relative clauses. They are in fact predicates, functioning as relatives, and not clauses, and should be designated as such.

This last observation suggests a further simplification. It is reasonable and conventional to view a relative clause as denoting a property, in the sense that it contains an open argument that must be bound externally. If this is correct, then it is reasonable to say that anything can be a relative (clause or otherwise) if it can correspond to a property in CS, that is, if it has the semantic representation $\lambda x.\mathbf{F}(x)$. As discussed below, introduction of the lambda notation makes it possible to account for the relation between the head of the relative clause and the relativized position within the relative clause without requiring a chain in the syntactic representation. If we add the lambda notation to our description of the relative constructions, what follows is that the only stipulations that we have to make are the following:

- relatives are interpreted as properties;
- English relatives are adjoined to the right of the head in NP (as in Sag's *rel-cl* (1));
- if there is a relative proform, it must be in or dominated by the constituent in clause-initial position (in English);
- an infinitival relative clause must be [PP wh] (in English).[9]

The only idiosyncratic stipulation, then, is the last. It is clear that this is the correct result, because the same condition does not hold for *wh*-questions—cf. *wonder who to talk to* vs **the woman who to talk to.*

[9] Such relatives are in fact completely impossible in Danish, Swedish, Norwegian, and German as well (Sabel 2006).

To sum up, then, there are several English relative and relative clause constructions. Their properties are fully predictable from general properties of English syntax, except for one. All the others are inherited from one of the following generalizations: relatives are interpreted as properties, relative clauses are clauses, relative clauses are embedded. What is not entirely predictable, however, is what form a constituent corresponding to a property may take in English, how an English relative clause gets to be interpreted as a property, and where the embedding takes place.

Given the foregoing observations, the properties of relatives can be described in terms of a set of correspondences between syntactic form and meaning, along the lines of the Parallel Architecture of Jackendoff (1997; 2002b). These are taken up in the next section.

3.1.4 Relative correspondences

3.1.4.1 A construction type Relatives as a group are constructions where a constituent corresponds to a property. This correspondence defines a general construction type *relative*, as follows. The property is represented as $\lambda x.\mathbf{F}(x)$, and is applied as a modifier to the representation corresponding to head of the NP.[10] The syntax-CS correspondence is shown in (18).

(18) RELATIVE

SYNTAX	XP
	$\updownarrow$
GF	
	$\updownarrow$
CS	$\lambda x.\mathbf{F}(x)$

I assume, following Culicover and Jackendoff (2005: chapter 9), that there is no syntactic movement, but that a relative clause contains a gap that is linked in CS to a variable. If there is an A′ constituent in non-canonical position, it corresponds to an antecedent in CS that binds the variable. If there is no A′ constituent in non-canonical position, the construction supplies the antecedent for the binding relation. Thus the various types of relative clauses are not

[10] A representation along these lines includes the types of information given in the attribute-value matrices of HPSG, such as Sag's (1997) *rel-cl*, displayed in a different and, I believe, more transparent way.

directly related to one another through a common syntactic derivation, but through their correspondences with the same core conceptual structure representation.[11] I take up the details of how the interpretations are computed immediately below.

3.1.4.2 Embedded relatives One general correspondence concerns the configuration in which English relatives are embedded, namely $[_{NP}$ N XP$]$. XP is interpreted as a modifier of N. Hence it must be a relative (not necessarily a relative clause) whose variable is bound by N′, the CS representation corresponding to N. In the correspondence in (19), α is used to index the binding relation between the head and the variable and XP′ represents the CS representation corresponding to XP, which is independently given by (18), and not repeated here. MOD stands for 'modifier.'[12]

(19) *N MODIFICATION (ENGLISH)*

SYNTAX $[_{NP}$ N XP$]$

GF

CS $N'^{\alpha}[\text{MOD:XP}'](\alpha)$

N MODIFICATION is a construction. It specifies the correspondence between a form and a meaning. What is important to observe about it is that the binding relation between N'^{α} and α is specified by the construction itself. If there is a gap in the relative clause, it does not correspond directly to α, but to the variable in the lambda expression. The variable α, or an expression containing the variable α, is substituted for x by lambda reduction, so that the antecedent comes to bind it in the corresponding position in **F**.

Since N $\leftrightarrow$ N′ is independently given as a lexical correspondence, the contribution made by RELATIVE (18) and N MODIFICATION (19) is that in the English NP, (i) XP follows the head, (ii) XP is a modifier (corresponding

[11] It is natural to extend this approach to all A′ constructions, including interrogatives and topicalization, and in fact I do so in Chapter 4. How the lambda expression fits into the overall interpretation has to be specified in each construction. It is through the common semantics that we can account for many of the similarities between A′ constructions observed by Chomsky (1977). However, the behavior of A′ constructions with respect to the extraction constraints of Ross (1967) and others must be explained in other terms, perhaps the processing complexity of long-distance dependencies (see Chapter 5).

[12] I am glossing over many technical matters that are far from trivial, but not central to the main points of the analysis. Many details hinge on the formal representation of N′ and the function of MOD.

to [MOD [HEAD *noun*]] in (1)), and (iii) there must be a variable α in XP′ corresponding to XP that is bound by N'^{α}.

How this last condition is satisfied depends on the syntax of XP. In the case of a relative clause there are several possibilities, as we have seen in our description in §3.1.3. Consider first the *wh*-relative, as in (20).

(20) $[_{NP}$ N $[_{S}$ wh-N ... [e] ...]]
[e.g. *the woman who I saw* [*e*]]

I assume that the correspondences for the relative pronouns are lexical. For example, the lexical entries for *who, which,* and *where* are as in (21).

(21) *RELATIVE PROFORMS*

$$\begin{array}{ccc} \text{who} & \text{which} & \text{where} \\ \updownarrow & \updownarrow & \updownarrow \\ \alpha_{\text{person}} & \alpha_{\text{thing}} & \alpha_{\text{location}} \end{array}$$

When there is an overt relative pronoun, the variable α in the CS representation corresponds to and is thus licensed by the pronoun, while the remainder of the representation corresponds to the relative clause and the head of the NP. In this case, [*I saw* [*e*]] corresponds to λx.**SEE**(EXP:ME, THEME:x) and *woman* to WOMAN$^{\alpha}$. The interpretation is then

(22) WOMAN$^{\alpha}$[MOD: λx.**SEE**(EXP:ME, THEME:x)](α)

Pied-piping is a bit more complex—I discuss it in §3.1.4.5 below.

An interesting consequence of the constructional formulation is that *that*- and zero-relatives must be distinguished from *wh*-relatives. *Wh*-relatives are licensed by the lexical relative pronoun constructions, as well as by N MODIFICATION. In the case of *that*- and zero-relatives, there is no *wh*-form, so the variable α must be licensed by the N MODIFICATION construction itself. This is the correct result, and is consistent with Hoffmann's (2011) argument that, contra Sag (1997), *that*-relatives are to be treated on a par with zero-relatives and not with *wh*-relatives.[13]

[13] Hoffmann (2011: chapter 6) argues that relative clauses should be organized into subtypes that take into account correlations between register and the pied-piping/preposition-stranding distinction. Preposition-stranding appears to be associated with a more informal register than pied-piping. Since the question of register is orthogonal to the formal description of the constructional possibilities, I do not address it in my analysis here.

3.1.4.3 Connectivity Following a proposal in Culicover and Jackendoff (2005: chapter 7), I take an A′ construction to be a special case of connectivity. An A′ construction is characterized by a constituent in a non-argument position (usually but not always in clause-initial position), and a gap in the position corresponding to the syntactic function of the A′ constituent. In connectivity, a constituent is interpreted as the argument of an open expression, and the variables in the constituent are linked to the operators in the open expression through lambda reduction (Sternefeld 2001). The general correspondence is given in (23).

(23) *Connectivity*

SYNTAX $[_S\ X\ \ Y]$

GF

CS $Y'(X')$

Condition: X c-commands Y

What this correspondence says is that even if X is not linked to a gap in Y in an apparent syntactic chain configuration, if X c-commands Y then Y′ can be interpreted as a function applied to X′. Connectivity arises from a range of syntactic configurations where the interpretation licenses the functional application shown in (23). For instance, in a pseudo-cleft we have the *wh*-clause applied to the focus, e.g.

(24) [What Mary ate [e]] was a doughnut.

Here, *what Mary ate* corresponds to λx.**EAT**(MARY,x), *doughnut* translates to DOUGHNUT, and the correspondence for the copula ultimately produces the interpretation (25)—

(25) [λx.**EAT**(MARY,x)](DOUGHNUT)

—which is equivalent to **EAT**(MARY,DOUGHNUT) after lambda reduction. (For formal accounts of connectivity, see Jacobson 1994 and Sternefeld 2001.)

In an A′ construction, the A′ constituent (that is, X in (23)) is interpreted as the argument of the function corresponding to the remainder of the sentence, which contains the gap. Thus, the composition of *who* and [*I saw* [*e*]] of (20) yields (26).

(26) [λx.**SEE**(EXP:ME,THEME:x)](α_{person})

This reduces to (27).

(27) **SEE**(EXP:ME,THEME:α_{person})

The correspondence is illustrated schematically in (28).

(28)

SYNTAX	[$_{NP}$ woman [$_{S}$ [$_{NP}$ who] I I^{0}[PAST] [$_{VP}$ see [e]]]]
GF	Subj Obj
CS	WOMAN$^{\alpha}$[MOD:λx.**SEE**(EXP:ME, THEME:x)](α_{person})

The crucial part of this correspondence is that the *wh*-phrase corresponds to the variable in CS. Since this variable also corresponds in this case to the gap, by lambda reduction the antecedent and the gap form a chain that is mediated by the links between them and their CS representations.

Recall that the syntax-CS correspondence in (19) is constructional. It does not follow from the meaning of a noun and an adjoined phrase that the phrase will be interpreted as a property of whatever the noun denotes in CS, unless we build the interpretation into the grammar as a language-particular correspondence or a universal principle. At the same time, the construction is a very general one, since it applies to any XP. What we see in (28) is how an S with a gap may participate in this construction, in virtue of the fact that the relative clause corresponds to a property, as specified in (18).

3.1.4.4 Finite relatives Construction (18) defines as *relative* the class of phrases that may have an interpretation of the form λx.**F**(x) and function as modifiers. A relative clause is a clause that has an interpretation of this form, while a reduced relative is a non-clause that has such an interpretation. What must be independently specified, then, are the correspondences for various relative constructions so that they fall under this general description. The correspondence for the simple *wh*-relative is fairly straightforward, as we have seen, but there are other constructions and properties that must be considered.

Consider finite and infinitival *wh*-relatives. A superficial description of the former is given in (29).[14]

[14] Here and throughout I omit consideration of the "fine structure" of the left periphery. Questions regarding the branching structure, the status of Spec, and the presence or absence of invisible heads do not appear to be germane to a constructional analysis, at least not at this preliminary stage of the analysis.

(29) $[_{S} [_{XP} \ldots \text{wh-N} \ldots]\ (\text{NP})\ \text{V}[\textsc{finite}] \ldots]$

(29) is a description of an S that has a phrase in initial position that counts as a *wh*-relative, either because it is a *wh*-word or because it contains one in the appropriate configuration, not specified here. This phrase is optionally followed by an NP (the subject) and then a finite verb.

We can define a *wh*-relative in this way even though the clause-initial XP may have any one of a number of grammatical functions, including subject. If XP is not an NP (or even if it is an NP but not the subject) and the sentence lacks a subject, the relative is ill-formed because of the independent constraint in English that a finite clause must have an overt subject. So a phrase like **the woman to whom gave the book* is ruled out.

What allows an XP to count as the clause-initial constituent of a *wh*-relative clause is a complex question that space limitations will not allow me to explore in detail here. Suffice it to say that it can be a relative proform, an NP with *whose* in its specifier, a PP such as *from whom*, or a more complex expression, such as *sitting next to a picture of whom* (see Ross's 1967 discussion of pied-piping).[15]

3.1.4.5 Pied-piping As we have seen, the *wh*-relative clause corresponds to a CS representation that contains a variable. But because of pied-piping, the variable does not necessarily correspond directly to the gap in the relative clause. In the correspondence, the *wh*-phrase corresponds directly to the CS variable, while the phrase in A′ position that contains it corresponds to whatever CS property the gap corresponds to. (30) illustrates.[16]

[15] What remains to be accounted for, among other things, is why a relative pronoun cannot be embedded more deeply in the clause-initial constituent, as in

(i) a. *(This is) the woman$_k$ [from the lawyer [who sued whom$_k$]]$_i$ Sandy received a letter t$_i$

b. *(This is) the woman$_k$ [that Sandy was rude to whom$_k$]$_i$ everyone knew t$_i$

The question is why *from the lawyer who sued whom* and *that Sandy was rude to whom* do not inherit the property of satisfying the relative clause construction, while phrases such as *from whose lawyer* and *sitting on next to a picture of whom* do. In fact it appears that the constraints of Ross (1967) apply to the linking of the head N (in this case *woman*) with the relative proform in situ. If there is no movement in this construction, it follows that such constraints are not constraints on extraction or even constraints on A′ chains in the strict sense. An alternative account in terms of processing complexity may prove to be revealing in such a case; see Hofmeister et al. (submitted) and references cited there.

[16] In the original formulation I had [e] corresponding to [GOAL:x] and *to whom* corresponding to α. The correspondences have been exchanged in the current formulation, which I think is more accurate.

(30) SYNTAX [$_{NP}$ woman [$_{S}$ [$_{PP}$ to whom] I I^0[PAST] [$_{VP}$ speak [e]]]]]

GF Subj

CS WOMAN$^{\alpha}$[MOD:λx.**SPEAK**(AGENT:ME,x)](GOAL:α)

A similar analysis applies to pied-piping with *whose* in the specifier of an NP, as in *the woman whose picture I saw*. The *wh*-phrase corresponds to a variable in CS, and the constituent in A′ position is the argument of the property corresponding to the rest of the sentence. So the representation for *whose picture I saw* is (31).

(31) SYNTAX [$_{NP}$ woman [$_{S}$ [$_{NP}$ whose picture] I I^0[PAST] [$_{VP}$ see [e]]]]]

GF Subj Obj

CS WOMAN$^{\alpha}$[MOD:λx.**SEE**(EXP:ME,THEME:x)](PICTURE[POSS:α])

Lambda reduction substitutes the representation PICTURE[POSS:α] for the variable *x*, which produces the CS representation in (32), where WOMAN$^{\alpha}$ binds the possessor of PICTURE.

(32) WOMAN$^{\alpha}$[MOD:**SEE**(EXP:ME,THEME:PICTURE[POSS:α])]

Lambda reduction may thus be seen as 'reconstruction' in the sense in which it was envisioned in transformational analyses such as (Chomsky 1977) (see Sternefeld 2001).

Next, consider the case in which the relative proform is a subject, as in *who saw them*. Here, the *wh*-phrase corresponds to the variable but there is no gap. The correspondence is shown in (33).

(33) SYNTAX [$_{NP}$ woman [$_{S}$ [$_{NP}$ who] I^0[PAST] [$_{VP}$ see them]]]

GF Subj Obj

CS WOMAN$^{\alpha}$[MOD:λx.**SEE**(EXP:x, THEME:THEM)](α)

The key here is that *who* corresponds to *x*, because it is the subject, and Subj corresponds to the EXP role of **SEE** by default; *who* also corresponds to α, because of the lexical entry of *who*. So by lambda reduction, WOMAN$^{\alpha}$ binds EXP:α. Note that it is not necessary to make any special stipulation for the case in which the *wh*-phrase is the subject. And because *wh*-, *that*, and zero are alternative clause-initial configurations, the presence of *who* excludes the other possibilities.[17]

The correspondences for the *that*-relative and zero relative are similar to the ones for the *wh*-relatives. The difference is that the variable α is licensed by the N MODIFICATION correspondence (19) itself, since there is no lexical form that corresponds to it directly. While *that* and zero appear to be in free variation in general, there are certain special properties of zero relatives that are arguably attributable to processing factors. For example, in standard English is not possible to have a zero relative when the highest subject is relativized, e.g.

(34) You shouldn't have a dog *(that) dislikes other dogs.

Chomsky and Lasnik (1977) attribute this restriction to a grammaticalized constraint that avoids the garden path effect associated with the absence of *that*. There are well-formed zero relatives of this type in non-standard English presentational sentences, e.g.,

(35) There's a guy outside says he's seen your brother.

For discussion of the contexts in which these relatives occur, see Harris and Vincent (1980) and Tagliamonte et al. (2005).

In addition, it is generally less acceptable to extrapose a zero relative (36b) and to have topicalization in a zero relative (37b), compared with a *that*-relative.

(36) a. A movie just came out that everyone wants to see.
b. ?A movie just came out everyone wants to see.

(37) a. He's the man that at the party, I gave your phone number to.
b. *He's the man at the party, I gave your phone number to. [i.e. I gave your phone number to the man at the party]

[17] The incompatibility of *wh* and *that* is a property of English. Other languages permit both to appear in the same clause; see e.g. Bayer (2002). Hence their constructions will have to be formulated somewhat differently than the English ones.

Again, it is possible that these variations in acceptability are not a matter of grammar per se, but arise from the interaction between the grammar and the interpretive process (Jaeger 2010).

3.1.4.6 Infinitival relatives Let us turn now to infinitival relatives. In general, an infinitive may have a *for*-phrase, a *wh*-phrase, or zero in initial position. The *for*-phrase marks the subject. If there is no *for*-phrase, the subject grammatical function is not realized overtly. There are four possibilities, shown in (38)–(42).

In the first correspondence, the *for*-phrase corresponds to subject, as specified by the correspondence for infinites in general, the gap corresponds to *x*, and the variable is licensed by N MODIFICATION. GF marks the grammatical functions Subj and Obj that mediate between the syntactic structure and the CS.

(38) SYNTAX book [$_S$ [$_{PP}$ for you] to [$_{VP}$ read [e]]]

GF Subj Obj

CS BOOK$^\alpha$[MOD:λx.**READ**(AGENT:YOU, THEME:x)](α)

In the second correspondence, the relative proform corresponds to the variable, its role is specified by the preposition, and the gap corresponds to *x*. According to the correspondence for relatives with an A′ constituent in clause-initial position, the corresponding representation for this constituent is the argument of the lambda expression. I represent *with which* as INSTR(UMENT):α.[18]

(39) SYNTAX pencil [$_S$ [$_{PP}$ with which] to [$_{VP}$ draw [e]]]

GF Subj

CS PENCIL$^\alpha$[MOD:λx.**DRAW**(AGENT:y,x)](INSTR:α)

[18] I gloss over a number of details in this representation that are orthogonal to the main points of this discussion, such as whether *with which* is an argument or an adjunct, and whether the semantic representation of **DRAW** should include an event variable.

Lambda reduction allows the variable to be bound by the head of the NP. The variable *y* is interpreted as the arbitrary controlled subject of the infinitive, which does not correspond to an overt syntactic constituent.[19]

As noted earlier, a clause-initial phrase in an infinitival relative must be a PP. That restriction must be specified in a correspondence that explicitly says that the VP is infinitival, as in (40). To the extent that the restriction does not follow from any general principle or pattern, but is idiosyncratic to the construction, it counts as a genuine constructional cost.

(40) *Infinitival wh-relative*

SYNTAX	N[$_S$[$_{PP}$ ...wh-N ...] to [$_{VP}$... [e] ...]]
GF	Subj
CS	N′$^\alpha$[MOD:λx.**F**(y,x)](PP′)

The rest of the structure of the PP does not have to be given explicitly in the correspondence, since it is independently defined for the language. Since PP contains a *wh*-phrase, PP′ contains α.

In the third schema, there is no overt subject, so the subject GF corresponds to a variable that must be bound by control, or assigned an arbitrary interpretation.

(41)

SYNTAX	book [$_S$ to [$_{VP}$ read[e]]]
GF	Subj Obj
CS	BOOK$^\alpha$[MOD:λx.**READ**(AGENT:y,THEME:x)](α)

The gap corresponds to *x*, and therefore to the variable after lambda reduction. Because there is no relative proform to specify the variable, the variable is again introduced by N MODIFICATION.

Finally, in the fourth schema, there is no gap. But because the clause is infinitival and lacks a subject, the subject GF corresponds to a free argument in CS. In this particular case, the variable is not treated as arbitrary control,

[19] Here I am assuming the analysis of control of Culicover and Jackendoff (2005), inspired by the HPSG analysis of Pollard and Sag (1994), where there is no syntactic PRO, but a linking between a GF and the appropriate CS argument. On the question of PRO, see also Culicover and Wilkins (1986).

because it is bound by the head noun of the NP, through lambda reduction, and N MODIFICATION is satisfied, as in (42).

(42) SYNTAX woman[$_S$ to [$_{VP}$ solve [the problem]]]

GF Subj Obj

CS WOMAN$^{\alpha}$[MOD:λx.**SOLVE**(AGENT:x,THEME:PROBLEM)](α)

Lambda reduction yields (43).

(43) WOMAN$^{\alpha}$[MOD:**SOLVE**(AGENT:α,THEME:PROBLEM)]

In this way we explain straightforwardly why it is possible to have infinitival relatives like those in (44).

(44) a. the man [to see [e]]
b. the man [to do the job]

In (44a), there is no overt subject. If Subj corresponds to *x*, as in λx.**SEE**(x,y), then the second argument *y* will not be bound, which leads to semantic ill-formedness, since English does not have implicit non-subjects. (45) illustrates.

(45) MAN$^{\alpha}$[MOD:λx.**SEE**(x,y)](α) ⇒ *MAN$^{\alpha}$[MOD:**SEE**(α,y)]

So the choice of variables must be λx.**SEE**(y,x), where *y* corresponds to Subj and is interpreted as an arbitrarily controlled argument.

In (44b), there is only one variable in the interpretation, so it must correspond to Subj, and the interpretation is well-formed.

(46) MAN$^{\alpha}$[MOD:λx.**DO**(x,JOB)](α) ⇒ MAN$^{\alpha}$[MOD:**DO**(α,JOB)]

This explanation for the possibility of infinitival relatives raises the question of why not all languages with infinitives permit infinitival relatives.[20] This question is of course not specific to the current analysis—in fact, it arises under any analysis of any phenomenon which is arguably possible in a language but does not occur. It arises in particular for any language in which an infinitive is treated on a par with sentences, either in terms of the syntax, the semantics, or both.

There are two cases, *wh*-infinitival relatives and zero-infinitival relatives. With respect to the former, Sabel (2006) proposes that *wh*-infinitives are

[20] Thanks to an anonymous referee for highlighting this question.

possible only if the left periphery of an infinitive may contain "a base-generated phonetically realized element," clearly a constructional stipulation. With respect to the latter, it may be that bare infinitives in languages that lack infinitival relatives do not correspond to properties, in contrast to infinitives in English, which would again be a constructional stipulation. However, this is only a speculation, and a definitive solution would require a significantly broader investigation than is possible here.

Another point that should be noted regarding infinitival relatives is that the CS representations here are systematically incomplete. What is missing is the deontic interpretation. The relative clause *for you to see* means 'who/that you should see,' the relative clause *to see* means 'that one should see,' *to do the job* means 'who/that should do the job,' and so on. This semantic idiosyncrasy is further evidence that we are dealing with distinct (but related) constructions in this domain, albeit constructions that are very closely linked to one another in an inheritance hierarchy.

3.1.4.7 Non-clausal relatives I conclude by considering those relatives that are not clausal, and do not appear superficially to contain a syntactic gap. Recall that by "relative" I mean any XP that can be interpreted as a property, that is, as an expression of the form $\lambda x.\mathbf{F}(x)$. These XPs are in fact the class of predicates. In the lexical entry for the head of a predicate that can serve as a relative there must be an external argument, so that it can be predicated of something. If the XP is *angry at Sandy*, for example, the CS interpretation is [λx.**ANGRY**(x,SANDY)], if it is *eating cookies*, the CS interpretation is [λx.**EAT**(x,COOKIES)], if it is *on the table*, the CS representation is [λx.**ON**(x,TABLE)], and so on. Notice that there is a variable in the interpretation bound by lambda even though there is no syntactic gap.[21]

If one of these predicates is the right sister to N in NP, then N MODIFICATION licenses the interpretation in which *x* is bound by the head, e.g.

[21] It is of course possible to stipulate that there cannot be a variable in the interpretation unless there is a gap in the syntactic representation. Doing so would reduce all relatives to relative clauses, with invisible structure for those that are not superficially clausal. Given that it is straightforward to state the correspondence for a non-clausal predicate without making this assumption, it would be necessary to find some motivation for the stipulation beyond the fact that the relatives can all be interpreted as properties. For extensive discussion of the consequences of imposing syntactic uniformity for constructions where there is semantic uniformity, see Culicover and Jackendoff (2005: chapters 2 and 3).

(47) a. a man angry at Sandy
$\text{MAN}^{\alpha}[\text{MOD: } \lambda x.\mathbf{ANGRY}(x,\text{SANDY})](\alpha) = \text{MAN}^{\alpha}[\text{MOD:}\mathbf{ANGRY}(\alpha,\text{SANDY})]$
b. a dog eating cookies
$\text{DOG}^{\alpha}[\text{MOD: } \lambda x.\mathbf{EAT}(x,\text{COOKIES})](\alpha) = \text{DOG}^{\alpha}[\text{MOD:}\mathbf{EAT}(\alpha,\text{COOKIES})]$
c. cake on the table
$\text{CAKE}^{\alpha}[\text{MOD: } \lambda x.\mathbf{ON}(x,\text{TABLE})](\alpha) = \text{CAKE}^{\alpha}[\text{MOD:}\mathbf{ON}(\alpha,\text{TABLE})]$

That these predicates can function as relatives follows directly from N MODIFICATION with no additional stipulation. It is sufficient that the RELATIVE correspondence for the predicates licenses $\lambda x.\mathbf{F}(x)$ and that N MODIFICATION licenses the external argument and the variable.[22]

3.1.5 *Appositive and free relatives*

There are two other constructions in English that are typically referred to as "relatives," appositive relatives (48a) and free relatives (48b).

(48) a. It started to rain, which I really resented.
b. Sam will eat what(ever) you serve him.

Both are called "relatives" because they have certain formal properties that resemble those of restrictive relatives—they have *wh*-forms in A′ position, and they have gaps when the *wh*-form is not a subject.

Are these "relatives" to be incorporated in the analysis above? On the face of it, the answer is "no." An appositive relative is not necessarily embedded in an NP, doesn't necessarily have a nominal antecedent, and is not interpreted as a modifier. A free relative is not embedded in any phrase, although it is not a main clause, lacks an antecedent, and is interpreted referentially.

In fact, what these "relatives" have in common with relatives is part of their interpretation. The interpretation of the restrictive relative associates the interpretation $\lambda x.\mathbf{F}(x)$ with the relative, and takes this to be a MOD of the head. The same $\lambda x.\mathbf{F}(x)$ is in the interpretation of the appositive and free relative, as defined by the respective constructions. In the case of the appositive relative, the interpretation is coordinate, with the antecedent of the

[22] Interestingly, a predicate nominal cannot function as a relative.

(i) *a woman a doctor [cf. a woman who is a doctor]

This would follow directly if the interpretation of a predicate nominal, unlike the other predicates, lacks an external argument in its CS representation. However, the formal semantics literature treats predicate nominals on a par with predicate adjective phrases (see for example Partee 1992: 113–15). I am not able to pursue this question here.

relative proform incorporated into the interpretation of the appositive, which corresponds to the lambda notation. So the interpretation of (48a), for example, is 'It started to rain and I really resented it,' where the second *it* refers to 'It started to rain.' The general correspondence is given in (49). $\mathbf{F}_1$ is the interpretation corresponding to the expression that contains the antecedent XP, which may be XP itself, and $\mathbf{F}_2$ is the interpretation of the appositive relative exclusive of the relative proform.

(49) *Appositive*

SYNTAX $[_{ZP} \ldots \text{XP} \ldots]$ $[_{S}\ \text{YP[WH]} \ldots]$

CS $\mathbf{F_1}(\text{XP}'^{\alpha})\ \&\ \lambda x.\mathbf{F_2}(x)(\alpha)$

The construction for the free relative is similar to that of the appositive, in that the lambda expression forms part of the interpretation. The free relative denotes a set of entities as specified by the *wh*-phrase, and a property of these entities. For example, in (48b) the free relative means 'THING/STUFF α such that you serve him α.' The general form of the correspondence is given in (50).

(50) *Free relative*

SYNTAX $[_{S}\ \text{X[WH]} \ldots]$

CS $\text{X}'^{\alpha}[\text{MOD}:\lambda x.\mathbf{F}(x)](\alpha)$

3.1.6 Summary

We have seen that it is straightforward to describe distinct relative constructions by specifying their syntactic properties and the meanings that correspond to them. The properties that follow from more general constructions of the language or from universals do not need to be specified for each construction. Those properties that do not follow from anything constitute costs for the constructions with respect to the grammar as a whole. Those that follow from universals are cost-free except for what is specific to the language.

The situation for relatives then turns out to be quite transparent, and somewhat simpler than Sag's original analysis suggests. The major observations are these:

- Relative clauses are clauses, and inherit certain properties in virtue of this. Clauses, and therefore finite relative clauses may be zero, *that*-, or *wh*-, but not a combination. Infinitival clauses, and therefore infinitival relative clauses, are zero, *for-to*, or *wh*-, but not a combination.
- The interpretation of relatives is inherited from general principles. One interpretation principle covers zero and *that*-relative clauses. The interpretation of overt A′ constituents follows from the interpretation principle for connectivity.
- It is necessary to stipulate that the *wh*-phrase in an infinitival relative must be PP.

What does this tell us about constructions, and how we should describe them? It suggests that we should seek analyses that minimize the complexity of descriptions as much as possible. It may not be possible to account for everything in terms of fully general rules and principles—there may be true idiosyncrasies. This does not mean, however, that there are no genuine generalizations. The analysis of English relatives offered here shows that in fact relatives in English are very regular, and that it is possible to precisely isolate the properties that they inherit in virtue of being English expressions that express properties, and distinguish them from those properties that are truly idiosyncratic.

3.2 Focus inversion[23]

The preceding section discusses a relative clause construction in English in which an idiosyncrasy rules out a possibility that should otherwise be attested, given the general patterns of the language. In this section I discuss a case in which a construction is permitted that should not be possible, given the general patterns of the language. The construction is focus inversion of the type found in English comparatives. I show that this type of inversion has the subject of the sentence "in the wrong place." This is possible because putting the subject in this position allows it to function robustly and unambiguously as a contrastive focus for the purpose of comparison. Inclusion of this construction in the grammar of English contributes to complexity; however, as in the case of other constructions (for some examples see Chapter 4), the idiosyncrasy that gives rise to this complexity is not completely arbitrary, in that its explanation takes full advantage of general properties of the language.

[23] Portions of this section are adapted from Culicover and Winkler (2008), hence the frequent use of "we." I have reproduced data, argumentation, and discussion from the original paper only where it bears on the main argument of this chapter.

3.2.1 *Inversion*

In addition to the canonical NP-(AUX)-VP structure, English has a number of constructions in which the subject follows the inflected verb. These are called "inversion constructions." Some examples are given in (51) and (52).

(51) a. $\left\{\begin{array}{l}\text{what are you}\\ {}^{*}\text{what you are}\end{array}\right\}$ looking at? [SAI]

b. $\left\{\begin{array}{l}\text{Are you}\\ \text{You are}\end{array}\right\}$ going to just stand there and stare? [SAI]

(52) a. Into the room oozed the syrup. [Locative inversion]
b. There is on the table a diverse selection of imported delicacies. [Stylistic *there* insertion]
c. Sandy is much smarter than $\left\{\begin{array}{l}\text{is the professor}\\ \text{the professor is}\end{array}\right\}$. [Comparative inversion]

As noted previously, a central goal of contemporary derivational theories in generative grammar has been to provide accounts of such apparently non-canonical constructions within the framework of general principles. For instance, SAI in (51) is typically analyzed as raising of I^0, the head of IP, to C^0, the head of CP, across the subject, which is the Spec of IP. Such movement is local, structure-preserving, conforms to X′ theory, and can presumably be licensed through a mechanism such as obligatory feature discharge.[24]

Hence SAI can be seen as fitting into a minimalist core syntax, although not entirely without stipulations. What has to be stipulated about SAI in English, on such an approach, is that I^0 and C^0 have a feature that must be discharged by movement of I^0 to C^0. Not all languages have this property. But it is a simple matter to assume that the feature in question is universal but parametric, in the sense that a language may but need not have it.[25]

The fact that such an analysis can be formulated supports a research program in which the properties of particular grammatical constructions are a consequence of the feature matrices of functional heads which trigger movement. Such an approach contrasts with previous work in the *Principles and Parameters* framework (Chomsky 1986a; Chomsky and Lasnik 1993) and is in fact closer to the more traditional approach in which constructions are assumed to have some idiosyncratic properties that cannot be completely

[24] See for example Pesetsky and Torrego (2000).

[25] This is simply a restatement of the facts, of course, and it could be argued that the property in question is not a particularly convincing universal (see for example Culicover 1999). I do not pursue the point here, but pursue an alternative analysis in Chapter 7.

reduced to general principles. The idiosyncrasies are isolated in the features on the heads (Borer 1984).

In contrast, the inversion constructions in (52), loosely referred to as "stylistic inversion" (SI), present a greater challenge for the minimalist theory of syntactic derivation. In some of these cases, the subject follows the main verb V, not simply a single auxiliary verb. There is no natural way to account for SI in terms of movement of V given the independently motivated and widely accepted analysis of the English verbal cluster in which a non-auxiliary V never leaves VP;[26] the only V that moves to the position before the subject, C^0, is an auxiliary.

Nevertheless, it often held that the syntax of these constructions, even the more exotic ones, is relatively straightforward. For example, Culicover and Levine (2001) argue that in the case of the particular type of SI exemplified in (52a), the subject is VP-internal, and does not move to Spec of IP. Rather, the PP does. The ordering V-NP is either an underlying order, or V moves to the left of the subject NP but not into I^0. Stylistic *there* insertion in (52b) is analyzed as an instance of subject postposing by Ward et al. (2002). Merchant (2003) argues that the case of SI exemplified in (52c), like SAI, involves movement of the verb to the left.[27]

SI constructions are different from SAI in three crucial respects:

- they appear in general to be optional, while SAI is obligatory;
- they do not fall straightforwardly under raising of I^0 to C^0 (although we might analyze them as movements of the verb to the left, as noted);
- they are tied to focus and other discourse properties; see Birner (1996); Birner and Ward (1998); Rochemont (1978, 1986); Rochemont and Culicover (1990); Winkler (2005).

While the idea that the various types of inversion have particular discourse properties is not at issue, it is not trivial to give a precise accounting of what these properties are; see for example Chen (2003). Accounting for these properties in a principled way is part of the challenge of understanding what their properties are. Moreover, it is not immediately clear why standard SAI of the sort found in questions does not have the same kind of focus and discourse properties as SI does, in spite of their superficial similarities.

[26] We include in this category analyses in which V raises to an empty *v*-head that is lower than I^0; see e.g., Johnson (1991).

[27] Note also that there have also been proposals in the literature that SI, at least in cases such as (52c), involves not movement of the verb to the left but positioning of the subject NP to the right; see Rochemont and Culicover (1990) and Chen (2003).

On this last point, it should be noted that Goldberg and Del Giudice (2005) have argued that "utterances with SAI are non-declarative and non-positive" (419), and that this property of SAI "motivates" its formal properties.[28] (See also Goldberg 2006, 2009.) As pointed out by Borsley and Newmeyer (2009), this is a fairly weak claim, particularly since on Goldberg's account, constructions such as those that use SAI that are formally related do not necessarily share all aspects of meaning. So, for example, SAI in negative inversion is non-positive but declarative—

(53) Never would I agree to do something like that.

—while SAI with *so* is positive (and arguably declarative) but dependent for its full interpretation on prior discourse, and hence not prototypically declarative and positive.

(54) a. So was I.
b. *So was I worried.

Exclamatives, argue Goldberg and Del Giudice, while declarative and positive, are historically descended from questions, which apparently exempts them from the generalization.

Goldberg and Del Giudice are correct that SAI constructions are not used to express prototypical positive assertions. However, it is not clear that they have anything in common except this lack of prototypicality. The idea that there is a particular meaning linked to SAI was addressed in Culicover (1971, 1973). I argued there that SAI is a formal property of the grammar of English that cannot be reduced to a uniform interpretation. This fact does not undermine the motivation for the constructional analysis, but supports it. The complexity of the grammar is reduced by the fact that the same formal resources are used to express a variety of different discourse properties.

One important virtue of Goldberg and Del Giudice's analysis is that it draws our attention to the question of why inversion constructions exist in English at all. In Chapter 7 I analyze SAI as a residue of the Germanic verb-second construction (V2). This is a case in which loss of the full generality of a construction has led to its redefinition for a number of special-purpose constructions, including most of those discussed by Goldberg and Del Giudice. However, one construction, comparative inversion (52c), turns out *not* to be an instance of SAI. I focus primarily on comparative inversion (CI) in the remainder of this chapter.

[28] The proposals of Goldberg and Del Giudice were not addressed in Culicover and Winkler (2008).

CI has a number of properties that make it appear at first to be a special case of SAI (Merchant 2003). Yet it has very special discourse properties, and its syntactic structure is constrained in ways that normal SAI is not. To show that CI is a special construction, we demonstrate that:

- in CI, the canonical subject position (Spec,IP) is not filled. Hence CI contrasts dramatically with SAI, and with other cases of SI.
- CI has specialized focus properties, which helps to explain why it exists as a construction in English.

In §3.2.2 we review Merchant's (2003) analysis of CI and show that it does not account for the full range of attested CI possibilities. In particular, the subject can follow more than one verbal element, as in *than **would have** any of the men who*..., which is not expected if CI is a variety of SAI. We consider a number of alternative syntactic analyses of CI, and conclude that it is indeed characterized by a particular syntactic property: the canonical subject position in Spec,IP is empty and the subject appears to remain in its initial position. We explore several alternatives for accounting for this fact, and suggest that the most straightforward approach is to assume that the subject is in fact in situ. For convenience, we characterize this syntactic feature of CI as "suspension of the EPP," by which we mean that the subject is not in its canonical position in Spec,IP, as is the case in canonical comparatives (CC).[29] But the subject-in-situ property of CI and related constructions raises some challenging theoretical issues that may not be best captured in terms of the EPP per se as a formal device.

The analysis of CI requires us to be explicit about the relationship between constituent order, intonation, and focus structure and how this construction differs from CC. This issue is taken up in some detail in Culicover and Winkler (2008). We argue there that CI is licensed as a consequence of the general English strategy of right-aligning a constituent in focus at the right edge of an intonational phrase. What is special about CI (in contrast to CC), as discussed in §3.2.3, is that the accented subject occurs immediately after the auxiliaries and is interpreted as contrastive focus (CF). The canonical subject position remains unfilled. Beyond this characteristic, CI is completely unexceptional in the way that it articulates with focus and givenness-marking. §3.2.4 shows how to characterize focus inversion in constructional terms.

In §3.2.5 we situate the analysis of CI within the broader context of specialized stylistic inversions. Analysis of the broader set of inversion

[29] Chomsky (1995: 55): "The Extended Projection Principle (EPP) states that [Spec, IP] is obligatory, perhaps as a morphological property of I or by virtue of the predicational character of VP (Williams 1980, Rothstein 1983)."

constructions shows that although each has somewhat different syntactic and semantic properties, they all show the same characteristic interaction: the constraint of placing a focused subject VP-finally overrides the requirement of realizing the subject in the canonical subject position.

3.2.2 Comparative inversion

The first question to consider is whether comparative inversion (CI) as exemplified in (52c) is an instance of SAI, as has been generally assumed. If it is not, then we have a reasonable basis for arguing that it is a distinct construction whose properties have to be specified in the grammar of English. The argument is further strengthened by the demonstration that the structure of CI is not derived by moving constituents from one position to another; rather, the postverbal subject is in situ as a property of the construction.

SAI is conventionally analyzed as raising of the verb in I^0 (or some appropriate functional head) to C^0. This movement puts the inflected auxiliary immediately before the subject. If CI cannot be analyzed in this way, the obvious alternatives are leaving the subject in situ in Spec,vP or moving the subject to the right, or the equivalent, depending on assumptions about the underlying structure.

Merchant (2003) crucially relies on an SAI analysis of CI in order to account for the fact that the main verb cannot appear when the auxiliary verb appears to the left of the subject NP.

(55) a. Sandy will run faster than Kim will (run).
b. Sandy will run faster than will Kim (*run).

Merchant's observation is that in (55b), where there is inversion, VP ellipsis appears to be obligatory. This is a puzzle, since in general VP ellipsis is not required when there is inversion; cf. the yes-no question *Will Kim run?* However, the data would have a very different status if (55b) was not a case of SAI. If it was not, we would still have to explain why the main verb cannot appear. But we would not be faced with the puzzle of why *Will Kim run?* is grammatical when it lacks a VP.

The evidence that CI is not a special case of SAI is that the subject NP can be preceded by a verbal sequence. That is, the actual structure of these constructions is roughly of the form in (56a), not (56b), which is the SAI structure. V_{AUX} stands for Modal, *have* and *be*, and V* for any number of auxiliary verbs (including none) that may follow the tensed form.

(56) a. X V_{AUX}[TENSE] V* NP…
b. X V_{AUX}[TENSE]$_i$ NP t_i…

We leave open for now the question of whether there is any movement in the derivation of (56a), so we don't show any traces or bracketing.

A Web search for sequences of the form *than-*V_{AUX}*-V*-NP* found numerous legitimate examples, a small sample of which are reproduced here. There are two categories of examples. One category shows VP ellipsis; these have sequences of the form *than-Modal-have-NP*. The other category shows predicate ellipsis; these have sequences of the form *than-have-been-NP*, *than-Modal-be-NP*, and *than-Modal-have-been-NP*.[30] Both categories are distinguished by the fact that they contain verbal sequences that precede the subject NP. (In these examples the CI construction is marked by square brackets and the accented material is capitalized.)

A. than-Modal-have-NP

(57) a. Interesting that his positioning, facing out of the frame but looking back into it, seems to make the portrait at once more candid and more dignified [than would have any of the alternative "CONVENTIONAL" compositions].
b. They argue that they produced more readable and better researched reviews and editorials [than could have the academics under whose NAMES the papers appeared].

B. predicate ellipsis
i. than-have-been-NP

(58) a. But Mokotów was much further from the City Center [than had been the OLD town] and the evacuation under the German lines all the more perilous.
b. Some ranchers still permitted the military to use their property, though the vehicles were far more destructive to the land [than had been the soldiers on HORSEBACK].

ii than-Modal-be-NP
ii. a. than-might-be-NP

(59) a. Today's air travel is unquestionably a much worse experience [than might be the proverbial ROOT canal].

[30] These searches were formulated in terms of particular verbal sequences followed by the definite or indefinite article, e.g. *than might be the/a/any*. Searches with other determiners produced no useful hits, nor did searches with proper names (e.g. *George Bush*).

b. Do you think your polemical stance is more effective in convincing people of your ideas [than might be a more MAINSTREAM and POLITE approach]?

ii. b. than-would-be-NP

(60) a. It strikes me that might be a more appropriate way to think about what we are up against here, [than would be any OTHER major conflict].
b. Who was responsible for keeping the records would be a more reliable witness as to their accuracy as a whole [than would be any of the ORIGINAL MAKERS].
c. All copies were expensive and the author/publisher, having produced one copy of the work, was no better situated to make another copy of the work [than would be ANY HOLDER of the work].

iii. than-Modal-have-been-NP
iii. a. than-would-have-been-NP

(61) a. The Prussian army would have been very much surprised to see their king without his hat; but no more so [than would have been the PEOPLE of those days to find Mrs. Madison without her TURBAN].
b. In other words, once a storm has destroyed a residential area, the repair costs are higher [than would have been the cost of preventing the release of GREENHOUSE GASES, which would have prevented the storm from HAPPENING].

iii. b. than-might-have-been-NP

(62) a. As a pageant, "Cammina Cammina" is far less colorful and even provocative [than might have been the PAGEANT put on by the PEASANTS seen in the precredit sequence].
b. To her, thinking, as she ever was thinking, about Johnny Eames, Siph was much more agreeable [than might have been a YOUNGER man who would have endeavoured to make her think about HIMSELF].

Examples such as these constitute prima facie counterevidence to the claim that CI is a special case of SAI. An SAI account would require a sequence of verbal head positions in the left periphery to which the verbal elements could move. Or, the verbal sequence would have to be moved to the C-position as a whole (similar to what was proposed by Rochemont and Culicover 1990: 95 for SI). The least stipulative alternative would be an account without movement of the auxiliaries. We take up this approach in the next section and provide arguments for the hypothesis that the subject also does not move.

3.2.3 The subject is 'low'

Consider now how to derive (56a). Logically there are four possibilities:

1. the subject is in canonical subject position (e.g. Spec,IP) and all of the verbs move to the left;
2. the subject is in canonical subject position and moves to the right;
3. the subject is in canonical subject position, and everything in I′ moves to the left of it;
4. the subject is in situ in Spec,vP, and remains in situ.

We rule out without further discussion option 1—although technically workable, it would require far too many ad hoc stipulations to be credible. Option 2 is more plausible, but involves movement of the subject to the right and adjunction to IP, which is ruled out on standard assumptions. Moreover, local movement is a less attractive option than placement in situ, if syntactic alternations can be accommodated naturally without movement. Option 3 conforms more closely to standard assumptions about the direction of movement, although it stipulates an ad hoc movement of I′ (or some XP that contains some empty material along with I′) and other complexities.

Option 4 is the least complex in terms of derivation and stipulations (although it is not without problems, as shown in Culicover and Winkler 2008). Crucially, it differs from the others in that the subject is 'low' in this construction, in the sense that it is in Spec,vP. The essential property of this alternative is that the subject is in situ, a condition that could be satisfied in other ways as well.

In this section and the next we present some empirical evidence that suggests that subject in situ is the correct analysis. This allows simplification of the discussion, in that we do not have to weigh the pros and cons of options 1–4. Since our main concern here is primarily to establish and explain the properties of CI, we do not pursue in any detail here the technical issues that might arise in a derivational theory from keeping the subject in situ in this construction.

In order to show that the subject is in situ, we need to show that it does not pattern with a subject that is in Spec,IP, or with heavy NP shift applied to a subject in Spec,IP. Since CI is a very rigid construction, there can be no extraction from the subject, or extraction of the subject in the comparative clause, which makes it difficult to test the configuration. But the configuration can be inferred from the behavior of the subject with respect to parasitic gaps and multiple *wh*-questions, and from its ordering relative to other constituents of VP in somewhat complex but acceptable pseudogapping constructions.

3.2.3.1 Parasitic gaps On option 4, the subject is a constituent of VP, on the surface in a position similar to that of the subject in *there*-insertion. On the other options, it is either in Spec,IP, or moved out of Spec,IP by a rule similar to heavy NP shift (HNPS). The behavior of CI with respect to parasitic gaps suggests that it groups with direct objects in situ. It does not behave like a heavy-shifted object, nor does it behave like an inverted subject in SAI.

Consider first the fact that it is possible to have a parasitic gap in the subject of a comparative, with CI (63a) or without CI (63b).[31]

(63) a. ?a person that Sandy gave more money to *t* than would have [even good friends of *pg*][32]
b. a person that Sandy gave more money to *t* than [even good friends of *pg*] would have

The following examples show that parasitic gaps in direct objects or objects of prepositions are not problematic.

(64) a. a person that Sandy gave money to *t* after talking to [some good friends of *pg*]
b. a person that Sandy gave money to *t* after introducing [some good friends of *pg*] to Otto

Example (65) shows that parasitic gaps do not appear in heavy-shifted NPs, which argues against the movement analysis of CI. Compare this with the grammatical (64b).

(65) *a person that Sandy gave money to *t* after introducing to Otto [$_{NP}$ some good friends of *pg*]

A similar contrast can be found for regular *there* and stylistic *there*, the latter being the counterpart of heavy NP shift for subjects in English. It is possible to have a true gap in an NP in the regular *there* construction.

[31] There is no question that examples with parasitic gaps are overall more complex than examples with simple extractions or no extractions. The analysis here is based on my own judgments that some cases of parasitic gaps are relatively unexceptional, which I do not mark, some are possible but complex, which I mark with '?,' while others are even worse, which I mark with '*.' The judgments are difficult to make, are somewhat unstable, and not everyone will share them.

[32] In Culicover and Winkler (2008) this example was not marked with '?,' but should have been.

(66) a person that there was [a picture of *t*] on the table

But a true gap cannot appear in the NP in the stylistic *there* construction.

(67) *a person that there was on the table [a picture of *t*]

This contrast is found for parasitic gaps as well. There can be a parasitic gap in the VP-internal NP (68a), but not in the heavy NP-shifted NP (68b).

(68) a. ?a person that I didn't recognize *t* even though there was [a picture of *pg*] on the table
b. *a person that I didn't recognize *t* even though there was on the table [a picture of *pg*]

The generalization appears to be that parasitic gaps are possible with VP-internal NPs, regardless of their grammatical function. Gaps in rightward-shifted NPs, however, are not possible.

At the same time, there is evidence that a parasitic gap is incompatible with a subject when there is SAI. Example (69a) shows a parasitic gap in an NP in canonical subject position. Example (69b), on the other hand, shows the same NP with a parasitic gap in inverted subject position. Example (69c) shows that SAI in the subordinate clause without a parasitic gap, while somewhat complex, is more acceptable than the SAI sentence with the parasitic gap.

(69) a. ?a person that Sandy criticized *t* because [some good friends of *pg*] refused to contribute to the proceedings.
b. *a person that Sandy criticized *t* because to virtually none of the proceedings did [any good friends of *pg*] contribute posters.
c. a person that Sandy criticized *t* because to virtually none of the proceedings did [any good friends of Robin] contribute posters.

The behavior of CI with respect to parasitic gaps then shows that CI is not the same as HNPS. The subject is not a postposed heavy NP. The subject may be in situ in Spec,vP, or it may be in its canonical subject position. However, the evidence that we have already considered demonstrating that the subject in CI may follow a sequence of auxiliary verbs shows that this latter possibility is not a viable one. The remaining possibility is the one that we argue for, namely that the subject is in situ in Spec,vP.

3.2.3.2 Multiple wh-*questions* We provide evidence that shows that in multiple *wh*-questions involving comparatives, the subject in CI behaves like a direct object, and not like a subject in CC. We take this to be indirect evidence that the subject CI is in Spec,vP.

First, note that it is possible to have a multiple *wh*-question in a comparative when the comparative lacks a verb, as in (70).

(70) a. Which (of the) girls run faster than which (of the) boys?
b. Which (of the) girls did Sandy praise more than which (of the) boys?
c. Who runs faster than who?
d. Who did Sandy praise more than who(m)?

Next, note that a subject in a canonical comparative clause may participate in a multiple *wh*-question, as seen in (71b). (We show the same structures with quantifiers to rule out the possibility that the ungrammaticality in (72)–(74) below has something to do with quantifier scoping into a comparative.)

(71) a. Some (of the) girls praised Leslie more than some (of the) boys did.
b. Which (of the) girls praised Leslie more than which (of the) boys did?
c. Who praised Leslie more than who did?

Examples (72b,c) show however that the direct object in a canonical comparative clause cannot participate in a multiple *wh*-construction.

(72) a. Sandy praised some (of the) girls more than she did some (of the) boys.
b. *Which (of the) girls did Sandy praise more than she did which (of the) boys?
c. *Who did Sandy praise more than she did who?

And (73b,c) show the same for the object of a preposition.

(73) a. Sandy gave more money to some (of the) girls than she did to some (of the) boys.
b. *Which (of the) girls did Sandy give more money to than she did to which (of the) boys?
c. *Who did Sandy give more money to than she did to whom?

Finally, consider the following examples, which show that multiple *wh*-questions where the CI subject is a *wh*-phrase are ungrammatical.

(74) a. Which (of the) girls run faster than which (of the) boys do?
b. *Which (of the) girls run faster than do which (of the) boys?
c. *Who runs faster than does who?

The subject in CI patterns with the direct object *wh*-phrase in (72b,c) and (73b,c), and not with the preverbal subject *wh*-phrase in the canonical comparative. This data is consistent with the hypothesis that the subject *wh*-phrase *which (of the) boys* in (74b), where there is CI, is in Spec,vP.

3.2.3.3 Pseudogapping In some cases, the postverbal subject may precede other VP material but not the verb, but it may never follow such material.

(75) a. Sandy made more money in 2001 than did $\left\{\begin{array}{l}\text{Leslie}\\ \text{any of the other students}\end{array}\right\}$ in 2002.

b. *Sandy made more money in 2001 than did in 2002 $\left\{\begin{array}{l}\text{Leslie}\\ \text{any of the other students}\end{array}\right\}$.

c. *In 2001 Sandy made more money than did in 2002 $\left\{\begin{array}{l}\text{Leslie}\\ \text{any of the other students}\end{array}\right\}$.

(76) a. Sandy ate more cookies at the party than did $\left\{\begin{array}{l}\text{Leslie}\\ \text{any of the other students}\end{array}\right\}$, slices of cake.

b. *Sandy ate more cookies at the party than did slices of cake, $\left\{\begin{array}{l}\text{Leslie}\\ \text{any of the other students}\end{array}\right\}$.

(77) When my son was younger and played soccer, I found that the parents got more pleasure from watching the games than did the kids from playing them, or the coaches from coaching them.

This constituent ordering follows if the subject is assumed to be at the left edge of the verb phrase, following the auxiliaries and tense and preceding the other constituents of VP. In fact, the observed ordering is what we would expect if the verb were simply omitted, with or without other constituents of VP, although this idea runs afoul of standard assumptions about how ellipsis works. The data are also compatible with the view that the two focus constituents undergo various movements to the left edge of the verb phrase, leaving behind a VP-shell that can be deleted; for such an analysis of pseudogapping see Lasnik (1999) and for a critique, Culicover and Jackendoff (2005).

The conclusion that the subject in CI remains in situ is problematic for derivational syntactic theories that require a subject argument to be licensed configurationally. This issue is not central to our current concerns; for detailed discussion, see Culicover and Winkler (2008).

The analysis of Culicover and Winkler (2008) demonstrates how the placement of the subject of CI in situ in VP has the result that the CI must be heavily focus-accented. The conclusion is that there is an interaction between the requirement that a focused constituent be right-aligned in its intonational phrase and the requirement that a subject must appear in Spec, IP, that is, the EPP. In derivational terms, the right-alignment requirement prevents movement of the contrastive subject to Spec,IP; in the constructional terms used here, it is more specific than the general rule for subject placement, and therefore preempts the general rule. (This is an instance of the general Elsewhere Condition; Kiparsky 1973).

In contrast, in CC, where contrast is marked not by position but simply by focus accent, the EPP is stronger than the focus alignment requirement and the subject appears in its canonical position (see Culicover and Winkler 2008 for further arguments).

Contrastive focus marking allows an additional focus following the subject; this produces the pseudogapping pattern seen in §3.2.3.3. (It may be that this focus marking is linked to a particular syntactic configuration, e.g. locating the focus constituents in the Spec of a functional head, but the question is a complex one and I will not pursue it here.)

Consider next the pseudogapping cases in (78a,b). (78a) answers the question *who made more money in which year?* and (78b) answers the question *who ate more of what at the party?*. Here, the parentheses marked with subscript 'ip' indicate an intonational phrase.

(78) a. SANDY made more money in 2001 (than did LESLIE)$_{ip}$, (in 2002)$_{ip}$
b. SANDY ate more COOKIES at the party (than did LESLIE)$_{ip}$, (pieces of CAKE)$_{ip}$.

In (78a,b) the contrastively focused subject precedes other vP-internal material, such as a prepositional adverb in (78a) or the object of a transitive verb in (78b). The information-structural requirement of two-variable comparisons is that both variables must be contrastively focused. The syntax–phonology interface conditions (Culicover and Winkler 2008) predict that in constructions with two CFs, as in CI, the CFs are right-aligned in their own intonational phrase.

The examples in (79) below show that the requirement of CF marking of the subject and a second remnant (here the object of a ditransitive verb) is sensitive to the discourse context. Contrastive Focus is realized on the element which is not givenness-marked by the comparative context (see also Rooth 1992; Schwarzschild 1999).

(79) a. SANDY sent her relatives more electronic BUSINESS CARDS (than did MANNY)$_{ip}$, (electronic CHRISTMAS CARDS)$_{ip}$.
b. SANDY sent her relatives more ELECTRONIC business cards (than did any OTHER family member)$_{ip}$, (PRINTED ones)$_{ip}$.

The discourse context of (79b) is more complex than the one for (79a). While (79a) answers a multiple *wh*-question *who sent her relatives more of what kind of electronic cards*, (79b) answers a question comparing senders out of the set of family members and objects taken out of the set of business cards. Note that in (79b) the alignment of the focus accent is not violated because each focused element occurs at the right edge of ip modulo the backgrounded and deaccented NPs *family member* and the one-anaphora *ones* in the *than*-clause.

The contrastiveness requirement on the remnants in CI cannot be loosened, as seen by the ungrammatical continuations in (80a) and (80b).

(80) SANDY sent her relatives more electronic BUSINESS cards

a. *(than did she)$_{ip}$ (electronic CHRISTMAS cards)$_{ip}$
b. *(than did MANNY)$_{ip}$ (them)$_{ip}$

The final phonological factor which must be controlled for in pseudogapping examples is a phenomenon known as 'stress clash.'

(81) a. BILL studied MATH a lot longer than did JOHN GERMAN.
b. BILL studied MATH a lot longer than did any of his FRIENDS classical CHINESE.

If two contrastive accents are realized in two adjacent constituents, such as *John* and *German* in (81a), there is a stress clash. In (81b) however, the two CF exponents are separated by the intervening pre-nominal adjective, which avoids a stress clash on the two adjacent nouns. A stress clash can also be remedied by a longer pause between the adjacent foci, or by repeating the complete VP as in (82).

(82) BILL reads LATIN much better (than does JOHN)$_{ip}$ (read GERMAN)$_{ip}$.

Examples like (82) are discussed at greater length in the context of VP ellipsis in CI in Culicover and Winkler (2008).

3.2.4 Focus inversion as a construction

The Subject construction given in Chapter 2 (repeated here as (83)) states that a Subj in English corresponds to the XP left sister of the inflected verb—this is the constructional equivalent of the EPP.

(83) *Subject*

SYNTAX	$[_S$ XP V[TENSE]]
GF	Subj

In focus inversion, the subject does not appear in this position. It is therefore necessary to state the correspondence in such a way that it can take precedence over (83). As discussed in connection with relative constructions, a more specified correspondence takes precedence over a less specified one. A first approximation of the focus inversion construction is given in (84).

(84) *Focus inversion*

SYNTAX	$[_S$ {than|as} V[TENSE] ... $[_{VP}$ ... NP XP]]
GF	Subj
IS	FOCUS

This correspondence deviates from the canonical correspondence rule in (83) in that it specifies that the subject follows the verbal auxiliary and is to the right in VP.

The question now arises as to why this non-canonical option is available in English. As discussed in more detail in Chapter 7, the ordering of a subject after the verb and even after other constituents of VP was an option in earlier forms of English, where arguments were marked with morphological case. The conditioning factors appear to have been driven by articulation with information structure; more specifically, in older varieties of Germanic, constituents at the right edge of VP are interpreted as focus by default. In fact, there is some evidence that direct objects and other constituents could follow V in Old English, an otherwise V-final language (Pintzuk and Kroch 1989).

It is reasonable to hypothesize that CI (and other focus inversion constructions) are fossils that have acquired highly specialized discourse meanings that they inherited from the earlier stage of the language, where ordering was relatively freer. For evidence for a postverbal focus position in older stages of Germanic, see among others Gergel (2008a, 2008b); Bies (1996); Hinterhölzl and Petrova (2005); Petrova (2006); Sapp (2005, 2007), and Schallert (2007).

3.2.5 More focus inversion constructions

The preceding suggests that there is a specific CI construction that functions as an alternative to the canonical structure for subjects. To broaden the empirical basis for such a conclusion, we review here the main characteristics of a number of other focus inversion constructions. These demonstrate a range of complexities that are related to, but not identical to, those of CI. Although there are different idiosyncratic properties that must be stipulated for each construction, in all of them the subject occurs at the right edge of the intonation phrase and must carry a focus accent.

3.2.5.1 So-goes One construction in English that shows optional inversion with obligatory focus on the low subject is *so-goes*. To our knowledge it has not been discussed previously in the literature.

(85) As IOWA goes, so goes the NATION.

Inversion is preferred in the second (the *so*-) clause, but is not necessary. However, when there is no inversion the construction lacks its special rhetorical force,[33] and also the requirement that the subjects be in focus.

(86) a. As Iowa goes, so the nation goes.
b. As Iowa goes on MONDAYS, so it goes on TUESDAYS.
c. *As Iowa goes on MONDAYS, so goes it on TUESDAYS.

Note also that there can be inversion in the first (the *as*-) clause.

(87) As goes IOWA, so goes the NATION.

But then there must be inversion in the *so*-clause. That is, the *so-goes* construction is subject to a parallelism constraint requiring the word orders of the antecedent and the *so-goes* clause to match, and only differ with respect to the focused NPs.

(88) *As goes IOWA, so the NATION goes.

The *so-goes* construction is well represented in naturally occurring text.

(89) a. As goes the statue, so goes the war.[34]
b. "As goes the SBC, so goes the larger church in the USA and, as goes the church, so goes the nation." —E. Ray Moore, Jr, USAR RET CH (LT.COL.)[35]
c. As goes Greece, so go we all.[36]

[33] Krifka (2007: 1) proposes that "rhetorical structuring partly belongs to IS."
[34] <http://findarticles.com/p/articles/mi_m6836/is_3_49/ai_n25118494/>.
[35] <http://www.exodusmandate.org/art_20060425-resolution-for-sbc-annual-meeting.htm>.
[36] <http://blog.beliefnet.com/roddreher/2010/02/as-goes-greece-so-go-we-all.html>.

It is most natural to say that the canonical form in (86a) and the inverted realization in (87) are forms of a special *so-goes* construction. At first sight it appears that this construction is of the form as in (90) with the focus accent realized on X in the antecedent clause and on Y, the low subject, in the *so*-clause.

(90) $\left\{\begin{array}{l}\text{As X goes}\\ \text{As goes X}\end{array}\right\}$, so goes Y (FOCUS).

However, we see from the following examples that there is some flexibility. First, it is possible to use verbs other than *go(es)*.

(91) As runs the FOX, so flies the BIRD.[37]

It is also possible to have an auxiliary verb preceding *go*. Note *how goes* in (92a) as an alternative to *as goes*.

(92) a. But he is watching the case with interest. "How goes 'Oprah,' so will go the LAW," he said, adding that 14 more states are considering similar bills.[38]
b. As Maine goes against global warming, so might go the NATION.[39]
c. As Massachusetts health "reform" goes, so could go Obamacare.[40]

As in CI, the information-structural requirement of the *so-goes* construction is that the subject in the *so-goes* clause receives a strong pitch accent and is interpreted as focus. The interpretation of the focus, however, differs. The semantics of the *so-goes* construction requires a strong focus which is interpreted as an addition to the set and not as an exclusion, as in the CI cases (see Krifka 1999). The data suggest that the construction has roughly the following superficial description.

(93) $(\left\{\begin{array}{l}\text{As X goes}\\ \text{As goes X}\end{array}\right\},)$ so $\left\{\begin{array}{l}\text{V}_{\text{AUX}}\text{[TENSE] GO}\\ \text{GO[TENSE]}\end{array}\right\}$ NP[Focus]

The interpretation of the accented constituent as additive focus depends on the semantics of *so* in relation to the discourse context and happens at the interpretive component.

[37] <http://www.tuaw.com/2006/12/20/as-runs-the-fox-so-flies-the-bird/>.
[38] <http://www.mad-cow.org/Constitution.html>.
[39] <http://www.earthplatform.com/global/heating>.
[40] <http://www.washingtonpost.com/wp-dyn/content/article/2010/07/18/AR2010071802733.html>.

3.2.5.2 Inversion after so *So-goes* is not completely *sui generis*. Inversion after *so* and after *as* are independently possible in English, although they do not have the same interpretation as they do in *so-goes*. Consider *so* first.

(94) a. I was there, and so was SANDY.
b. Leslie eats pretzels, and so does SANDY.
c. The Yankees will make a bid for Beckham, and so will the RED SOX.
d. Sandy was there, and so was I.

When *so* is used in this way, meaning 'also,' inversion is obligatory.

(95) a. *I was there, and so Sandy was.
b. *Leslie eats pretzels, and so Sandy does.
c. *The Yankees will make a bid for Beckham, and so the Red Sox will.
d. *Sandy was there, and so I was.

When there is no inversion, *so* is possible, but then it means 'therefore,' not 'also.' And in such cases, focus may be on the subject or a non-subject.

(96) a. I was there, and so SANDY was, too.
b. I was there, and so I was not HERE.

It is marginal to have more than one verb in the *so*-clause when there is inversion, but when there is only one, inversion appears to be SAI. But when the subject is "heavy," inversion around the entire verbal cluster appears to be possible as well, and perhaps even preferable.[41]

(97) a. Leslie had been there, and so $\left\{\begin{array}{l}\text{had I}\\ \text{*had been I}\\ \text{*had I been}\end{array}\right\}$.

b. Leslie had been there, and so $\left\{\begin{array}{l}\text{had Sandy}\\ \text{had been Sandy}\\ \text{*had Sandy been}\end{array}\right\}$.

c. Leslie had been there, and so
$\left\{\begin{array}{l}\text{had those demonstrators that you told me about}\\ \text{had been those demonstrators that you told me about}\\ \text{*had those demonstrators that you told me about been}\end{array}\right\}$.

[41] Examples such as these contradict the claim of Goldberg and Del Giudice (2005) that *so*-inversion is a variety of SAI.

So-inversion is typically possible with a pronominal subject, but as (97) suggests, as the verbal cluster becomes more complex, a pronominal subject is less acceptable.[42]

Examples of *so*-inversion with complex verbal clusters can be found through a Web search. (We have added the focus accent in the following.)

(98) a. As the pyramid rose, the working space would have diminished, of course, and [so would have the number of TEAMS that could simultaneously work ATOP it . . .][43]
b. [Yet so would have been the chances of entrapping and destroying large Hezbollah forces on the GROUND . . .][44]
c. His hair was light, and [so would have been his COMPLEXION, had it not been burned red by exposure to the hot sun of the TROPICS . . .]
d. . . . would have taken his place—the chain of command—and that person's tasks would have been the same and [so would have been his positional LIMITATIONS . . .][45]

3.2.5.3 Inversion after as Turning next to *as*, simple constructed examples appear to be quite well-formed, with pronominal and non-pronominal subjects.

(99) a. Sandy is very forgetful, as is LESLIE.
b. Sandy likes baseball, as do I̱.
c. Sandy would have refused to do that, as would ISABELLA.

But when the verbal cluster is more complex, a heavier subject NP is preferred.[46]

[42] The differences in acceptability here may have to do with the relative weight of the pronoun and the auxiliary verbs. It in fact appears to be possible to account for a range of sequences by assuming that the weight of what precedes the pronoun must be less than that of the inverted subject NP. Since the pronoun is relatively light, very little can precede it before unacceptability emerges. A heavier NP can appear after a longer sequence of verbs, and only the heaviest of NPs can appear after the longest sequence of verbs, e.g. *would have been.*

[43] <http://www.americanscientist.org/my_amsci/restricted.aspx?act=pdf&id=12805316473962>.

[44] <http://www.nysun.com/opinion/what-sharon-would-do/46280/>.

[45] <http://www.j-bradford-delong.net/movable_type/2003_archives/002216.html>.

[46] It appears, moreover, that reducing *have* to *'ve* may render the examples with pronominal subjects somewhat more acceptable.

(i) Sandy would have been very angry, as would've been HE.

Reduction also appears to improve the examples with *would have NP* in (165b,c), which sound quite awkward without it.

(100) a. Sandy has been very angry, as $\left\{\begin{array}{l}\text{has} \left\{\begin{array}{l}\text{Leslie}\\ \text{HE}\end{array}\right\}\\ \text{has been} \left\{\begin{array}{l}\text{Leslie}\\ \text{*HE}\end{array}\right\}\end{array}\right\}$.

b. Sandy would have been very angry, as $\left\{\begin{array}{l}\text{would} \left\{\begin{array}{l}\text{Leslie}\\ \text{HE}\end{array}\right\}\\ \text{would have} \left\{\begin{array}{l}\text{Leslie}\\ \text{*HE}\end{array}\right\}\\ \text{would have been} \left\{\begin{array}{l}\text{Leslie}\\ \text{*HE}\end{array}\right\}\end{array}\right\}$.

c. Sandy would have been very angry, as $\left\{\begin{array}{l}\text{would}\\ \text{would have}\\ \text{would have been}\end{array}\right\}$ all of the people who invested in the project.

As in the case of the other SI constructions, if there is no inversion, focus is not restricted to the subject.

(101) a. Sandy has been very angry, as LESLIE has (been).
b. Sandy has been very angry, as she USUALLY is.

These examples of *so-* and *as*-inversion are characterized by the fact that they involve inversion around auxiliary verbs, including *be*. The *so-goes* construction shows that there can be inversion around the verb *go* and other semantically related verbs. Examples found through a Web search show that other verbs may also show *so-* and *as*-inversion.

(102) a. I got lost in the maze of narrow, twisting alleys of the Bazaar quarter—Smyrna's only district to have conserved a picturesque air, as had noted the WRITER . . . [47]
b. As had noted the Chairman of Board of Tsesnabank JSC DAUREN ZHAKSYBEK, the initial sum of the deal was planning in size of $25 m., but was increased . . . [48]
c. Dante has it, i.e. of the armorial device of the Visconti, as runs the note to the passage by Shelley HIMSELF.

3.2.6 Conclusions

I showed in this section that there are a number of constructions in English that are characterized by the fact that the subject is in situ following the finite inflection and other auxiliaries. On mainstream assumptions the subject is in vP, and fails to move to Spec,IP in order to satisfy the EPP. On non-

[47] <http://www.helleniccomserve.com/smyrnapostblaze.html>.
[48] <http://www.cbonds.info/cis/eng/news/index.phtml/params/id/357892>.

mainstream assumptions, e.g. *Simpler Syntax*, the subject is in VP, and is there in virtue of the specifications of the construction.

The explanation for why it is natural that a subject in English can appear in VP after the inflected verb has to do with the interaction of constituent order, prosody, and focus interpretation. The right edge of a VP is the canonical locus of the accent. The inversion constructions recruit this general property of English, specifying that the subject appears in this position. In virtue of the positioning of the subject, it acquires the accent, and is thus directly interpretable as focus. It is reasonable to explore extending this general approach to stylistic inversion (e.g. *Into the room walked Robin*), stylistic *there* (*There walked into the room an impossibly large elephant*), and possibly even non-stylistic *there*.

The basic idiosyncratic stipulation that a focus inversion construction requires, then, is that the subject appears after the auxiliary material. Everything else follows from the general properties of the language, another case where complexity is maintained in the grammar, but restricted. The situation is essentially the same as that discussed in the case of relative clauses in §3.1. The individual construction is distinguished from the canonical form of the language in a particular respect; in other respects its properties follow from those that are more general. In the case of relatives, the idiosyncrasy is an exception: there cannot be a *wh*-NP in initial position in an infinitival relative. In the case of focus inversion, the idiosyncrasy is a deviation: the subject must be in situ in VP.

As we have seen, there is no possibility in the focus inversion constructions that the focus can fall on anything other than the subject. Whether this constraint has to be part of the specification of the construction, or whether it follows from general mechanisms concerning how focus is assigned, is an open question. The latter option lends support to the view that a construction need not have a special meaning associated with it in the constructional lexicon; the constraint on infinitival relatives is another piece of support for this view.

4

Constructions and the notion 'possible human language'

A central goal of linguistic theory is to characterize the notion 'possible language.' In principles and parameters theory (PPT—Chomsky and Lasnik 1993), this goal was addressed in part by defining 'possible language' in terms of the possible configurations of parameter settings. On this view, different combinations of parameter values define different languages. As noted in Chapter 1, in the earliest formulations, parameters concerned the configuration of the 'core,' while the so-called 'periphery' was left relatively unexplored (Chomsky 1981b). The core parameters are those that have to do with head–complement order, the possibility of (overt) *wh*-movement, and so on.

The PPT approach has proven to be less than fully successful in fully explicating the notion of 'possible language,' in part because it is not concerned with the so-called 'periphery,' which actually turns out to constitute a sizeable portion of linguistic knowledge. Maximally general, high-level macro-parameters account at best for only a small portion of the variance among languages.[1] Such parameters do describe some central and salient differences among languages, but there is a vast amount of variation that falls outside of their scope. Focusing attention on 'core grammar' as defined by these parameters fails to take into account the considerable systematicity and variability that is found in the periphery, as well as the variability that is found in the core. On the other hand, low-level micro-parameters (Black and Motapanyane 1996) can in principle record all of the variance, but are nothing more than descriptive devices for annotating the idiosyncrasies that the macro-parameters fail to account for (Culicover 1999).

Even in a constructional theory in which any arbitrary form–meaning correspondence is in principle possible, there needs to be some account of what is natural, in order to begin to approach an understanding of 'possible

[1] Moreover, it is open to question whether there are any core phenomena that are actually accounted for in terms of substantive parameters—see Culicover (1999) and Newmeyer (2005) for discussion.

language.' In this chapter I show that in a diverse set of constructions identified in *Syntactic Nuts* and elsewhere, there is significant systematicity accompanied by particular aspects of idiosyncrasy. But the idiosyncrasy is not entirely arbitrary. While idiosyncrasy, and thus formal complexity, exists in natural languages and must be accounted for in grammars, a constructional approach allows us to see how the idiosyncrasy is constrained by its relationship to the more general.

The notion of naturalness is nicely illustrated by the English sluice-stranding construction, discussed in §4.1. §§4.2–4.5 review several constructions discussed in *Syntactic Nuts* and shows how their behavior follows from the more general structures of the language.

4.1 Sluice-stranding

Sluice-stranding is characterized by sluicing in which the *wh*-phrase precedes a preposition, as in (1).

(1) What about?; What with?; Who with?; What for?; etc.

As Merchant (2001) has emphasized, it is entirely natural that English would have such a construction, given that English also has preposition-stranding, as in ***What** were you talking **about**?*. It would be very surprising if a language that lacks preposition-stranding, such as German, had sluice-stranding.[2] On the other hand, it might not be terribly surprising to find a language that has preposition-stranding but not sluice-stranding; Swedish and Danish appear to be such languages.

In mainstream approaches, the connection between the general pattern and the special construction is a derivational one—the special construction is derived transformationally from the general one, e.g. by deletion in the case of sluicing and sluice-stranding (see Merchant 2001). I argue in this section that such an account is not viable because it makes the wrong predictions about what is and is not a possible sluice-stranding expression. Moreover, the derivational analysis is a puzzle for a language that has sluicing (and it appears that in fact all do) and preposition-stranding, but not sluice-stranding.

[2] German has fixed expressions such as *worüber* 'where about,' and permits *ich weiss nicht worüber* 'I don't know what about.' This is clearly not derived by preposition-stranding, which is not permitted in German. As in the case of sluice-stranding in English, *worüber* (as well as other expressions of the form *wo*+P) are lexical. All of the combinations are possible: *woran*, *worauf*, *woraus*, *worin*, *wobei*, *womit*, *wovor*, *wovon*, *wogegen*, *worunter*, *wonach*, *wozu*. In some cases the meanings are transparent, and in others, idiomatic.

4.1.1 Correspondences for questions

Sluice-stranding has to be related to three other syntactic phenomena in English: (i) *wh*-questions, (ii) sluicing, and (iii) preposition-stranding. Looking first at *wh*-questions, it is useful to recall the lexical specifications for the relative proforms given in Chapter 3. These are repeated in (2).

(2) *Relative proforms*

who	which	where
$\updownarrow$	$\updownarrow$	$\updownarrow$
α_{person}	α_{thing}	$\alpha_{location}$

Just as the relative proforms correspond to variables, so do the interrogative proforms. The variable corresponding to the relative proform is bound by the head of the NP that contains the relative. The variable corresponding to the interrogative proform is bound by an interrogative operator Q^{α} that is introduced by the *wh*-question construction (3). The entire *wh*-phrase is the argument of the property corresponding to the rest of the sentence, which contains a gap.

(3) *Wh-question*

SYNTAX $[_S\ [_{XP} \ldots X^0[\text{WH}] \ldots]\ YP]$

GF

CS $\mathbf{Q}\ [YP'\ ](\ldots\alpha\ldots)$

In this construction, the position of the XP in the structure specifies the scope of the operator at CS, while the interrogative itself corresponds to a variable in CS. The location of this variable in the CS representation depends, in turn, on the syntax of the clause that contains the *wh*-phrase.

For example, if the sentence is *Who did you see?*, the *wh*-phrase *who* corresponds to the variable α, and specifies the scope of the coindexed operator. The object is a gap, and therefore corresponds to *x* in the argument structure, as shown in (4).

(4) SYNTAX $[_S$ who did you see [e]]

GF Subj Obj

CS $\mathbf{Q}\ [\lambda x.\mathbf{SEE}(\text{EXP:YOU,THEME:}x)](\alpha_{person})$

By lambda reduction, this CS representation reduces to (5).

(5) $\mathbf{Q}^{\boldsymbol{\alpha}}$[**SEE**(EXP:YOU, THEME:$\alpha$)]

On the other hand, if the sentence is *Whose book did they read?*, *whose* is a possessor of *book*, therefore the POSS argument of BOOK must be a variable. The correspondence before reduction is shown in (6).

(6) SYNTAX $[_S\ [_{NP}$ whose book] did you read [e]]

GF Subj Obj

CS **Q** [λx.**READ** (AGENT:YOU,THEME:x)](BOOK[POSS:α_{person}])

After reduction the CS is

(7) $\mathbf{Q}^{\boldsymbol{\alpha}}$[**READ**(AGENT:YOU,THEME:BOOK[POSS:α_{person}])]

The correspondence here follows the same pattern as the relative clause pied-piping correspondence in Chapter 3.

These correspondences highlight an aspect of the analysis of *wh*-chains in *Simpler Syntax* that distinguishes it in a very fundamental way from classical analyses, either transformational or monostratal. On classical views, the constituent in A′ position is linked syntactically to the gap. Either it originates in the position of the gap and is moved to the A′ position, or it forms a syntactic chain with the gap, or it satisfies an unsaturated selection feature of some lexical head.

But in *Simpler Syntax*, constituents in A′ position form a chain with the gap only through the CS representation. These A′ constituents are 'orphans,' in the sense that they lack inherent syntactic function and corresponding direct thematic interpretation in the sentence. In the case of the interrogative, for example, the chain consists of the correspondence between the A′ constituent and the operator $\mathbf{Q}^{\boldsymbol{\alpha}}$, which results in the binding of the variable α by the operator. There is an independent correspondence between the constituent that contains the variable and the gap. Lambda reduction produces the result that the operator binds the variable in CS. In the case of a clause-initial *wh*-word, the chain is straightforward and the A′ constituent syntactically binds the gap in the usual way, through c-command. But in the case of pied-piping, the interrogative does not bind the gap, because it does not c-command it. However, in both cases, the operator binds the variable after lambda reduction.

4.1.2 Interpreting sluicing

With the foregoing as background, let us consider how we would interpret sluicing, as in *(Someone called, but) Sandy forgot who.* The verb *forget* selects an interrogative complement, and the interpretation of *who* is accounted for by the WH-QUESTION correspondence in (3). Hence we get the following correspondence.

(8) SYNTAX $[_S$ Sandy $[_{VP}$ forgot $[_S$ who $]]]$

GF Subj

CS **FORGET**(EXP:SANDY,THEME:**Q** (α_{person}))

As discussed at some length in *Simpler Syntax* (see also Culicover and Jackendoff 2012), sluicing is an ellipsis construction. Thus the content of the Theme has to be computed on the basis of the discourse. One aspect of the computation involves identifying the property that is applied to the variable α. For example, if the discourse is "Lee was talking to someone," reconstruction of the ellipsis must construct the following representation.

(9) λx.**TALK**(AGENT:LEE,GOAL:x)

And this representation is combined with that in (8) to yield (10).

(10) **FORGET**(EXP:SANDY, THEME:$\mathbf{Q}^{\boldsymbol{\alpha}}$[λx.**TALK**(AGENT:LEE,GOAL:x)] (α))

This reduces to (11)—

(11) **FORGET**(EXP:SANDY, THEME:$\mathbf{Q}^{\boldsymbol{\alpha}}$[**TALK**(AGENT:LEE,GOAL:α)])

—which is also the CS of *Sandy forgot who Lee was talking to.*

It follows, then, that nothing special has to be said about sluicing beyond the mechanism by which the ellipsis is reconstructed and licensing conditions for the sluiced phrase.[3] As argued in *Simpler Syntax* and Culicover and Jackendoff (2012), this mechanism, bare argument ellipsis,

[3] For example, in a language with case-marking, the sluiced phrase must have a case that is compatible with the case of the corresponding phrase in the antecedent. See Culicover and Jackendoff (2012) for discussion.

is very general in English and applies without special stipulations to sluicing. Therefore, given English bare argument ellipsis, sluicing in English is essentially cost-free.[4]

Sluice-stranding, exemplified by *who to*, is the same as sluicing, with an additional cost. It is necessary to say which *wh*-expressions participate in the construction, and which prepositions participate in combinations with particular *wh*-expressions. But the ordering follows from the interpretation of the *wh*-word, which gives the sluice-stranding construction its interrogative interpretation. And the interpretation of the *wh*-word as the complement of the preposition follows from the fact that the preposition lacks an overt complement, which means that the gap following the preposition corresponds to a variable in CS, just as it does when there is normal preposition-stranding. (12) illustrates.

(12) SYNTAX $[_S$ who $[_{PP}$ to [e]]]

GF

CS **Q** $[\lambda x.\text{GOAL:x}](\alpha_{\text{person}})$

By reduction, this becomes $\mathbf{Q}^{\boldsymbol{\alpha}}[\text{GOAL:}\alpha]$. Reconstruction of the ellipsis supplies the remaining part of the CS.

The restrictions on sluice-stranding are discussed in some detail in *Syntactic Nuts*. Only single *wh*-words may appear in this construction—*who with* but **which person with*. The *wh*-words may appear in restricted combinations with prepositions—*what about* but **who about* —, and some prepositions do not participate at all—*what on* but **what after*. Sluice-stranding is more restricted than sluicing and pied-piping. Every case of sluice-stranding corresponds to a well-formed case of sluicing, and there is no preposition that permits stranding in sluice-stranding but not in full sentences. As argued in *Syntactic Nuts,* a constructional account is needed in order to capture the idiosyncrasies.

The syntactic component of the correspondence in (12) captures the stranding property of the construction by treating P as the head of PP with

[4] For the simple cases, the antecedent supplies the interpretation of the open proposition into which the fragment is integrated. But as Beecher (2006) argues, there are some cases where pragmatic inference is at play. A similar argument is made for the interpretation of bare argument ellipsis in *Simpler Syntax*.

a gap. If we assume that the ability of a preposition to strand in English is independently characterized in terms of this property in the lexicon, the possibility of the structure in (12) is not problematic.

It is important to observe how the existence of a general construction, in this case preposition-stranding, licenses the existence of a more specialized construction without requiring that the latter is in some sense derived from the first. Inheritance in cases like these in fact captures exactly the same generalization that derivation does, but without particular syntactic mechanisms and ad hoc stipulations.

In the remaining sections, I review some "assorted nuts" that were discussed in *Syntactic Nuts*. My goal in each case is to illustrate how they are related to the more regular constructions of the language, and to pinpoint their idiosyncrasies. The main point, as in the preceding account of sluice-stranding, is that specialized constructions are not completely unique and *sui generis*, but rest solidly on the platform of the constructions of the grammar of the language of which they are a part.

4.2 Comparative correlatives

One case, the comparative correlative (CC), has generated substantial discussion in the literature as to whether it is in fact a syntactic nut. I review the debate briefly and show how to account for the compositional aspects of the construction while isolating and accounting for its idiosyncratic properties.

CC is superficially composed of two comparative expressions, as illustrated in (13).

(13) The more Sandy eats, the hungrier he gets.

Each expression is of the form

(14) $[\text{the}]_i \begin{Bmatrix} \text{more XP} \\ \text{XP -er} \end{Bmatrix} [_{\text{S}} \ldots\ t_i \ldots]$

and the interpretation is roughly that the quantity denoted by the second comparative is correlated with the quantity denoted by the first comparative. A paraphrase of (13) is (15).

(15) As Sandy eats more (and more), he gets hungrier (and hungrier).

Culicover and Jackendoff (1999) (see also Culicover and Jackendoff 2005) argue that the CC construction exemplified in (13), while showing many regularities characteristic of English, is idiosyncratic in two respects:

- it is necessary to stipulate in CC that it is possible to have a clause of the form in (15), and
- it is necessary to stipulate the semantic relationship between the two clauses of CC.

The claim of Culicover and Jackendoff (1999) is that the relationship between the two clauses does not follow compositionally from the interpretation of meaningful elements in some particular syntactic configuration. Such a situation is constructional—the meaning is associated with the structure, and cannot be determined on the basis of the meaning of its parts and how they are put together syntactically.

Since the original Culicover–Jackendoff proposal, there have been arguments to the effect that the comparative correlative is fully regular and that its properties follow from UG (Den Dikken 2005). And there have been arguments to the effect that the comparative correlative cannot be fully accounted for on the basis of UG because its properties vary from language to language (Abeillé and Borsley 2008).

Adopting the general perspective of the preceding chapters, a natural approach to resolving this issue would be to first formulate the minimal syntactic description of this construction that is sufficient to account for its formal properties, and then to formulate the minimal account of the mapping between this form and the meaning of the construction. On this basis, it would be possible to see to what extent the form conforms to the general pattern of the language, to what extent it is a projection of independently required semantic structure, and to what extent it is necessary to stipulate aspects of the form/meaning correspondence.

The most salient properties of the construction are the following.

- *the*: The definite article has an idiosyncratic use in this construction, but it is a word of English.
- comparative: The comparative is fully regular, and follows the constraints on the distribution of *-er* vs *more*.
- The comparative correlative expresses a relation between a difference in the value of a property on one scale (*the more Sandy eats*) and a difference in the value of a property on another scale (*the hungrier he*

gets). *The*...*the*... says that the differences in the values on the two scales correlate, with the second dependent on the first.[5]

To isolate what part of the construction is responsible for what part of the meaning, consider sentence (16).

(16) Sandy runs fast, Leslie runs fast.

This sentence can have the interpretation 'Sandy runs fast and Leslie runs fast.' Or it can have the interpretation 'Sandy runs fast and subsequently Leslie runs fast.' Or it can have the interpretation 'Sandy runs fast and as a consequence Leslie runs fast.' Similar interpretations are associated with two conjoined clauses (Culicover and Jackendoff 1997).

The minimal assumption about the connection between the clauses is that it is a pragmatically licensed option. That is, there is no element in either clause that means 'subsequently,' 'consequently,' or 'therefore'—it is an inference from the juxtaposition of the two propositions and their individual interpretations. Similar paratactic effects occur throughout language, and across a range of categories; e.g., *no shoes, no shirt, no service*; *easy come, easy go*; *like father, like son* (see Culicover 2010 for a range of examples).

However, there are some cases where the semantic relation between the two constituents becomes conventionalized. That is, it becomes part of the constructional meaning. I suggest that this is the case in the CC construction. As in an example like (13), the two constituents are apparently of the same category—they are both *the*-clauses, whatever category that is. The interpretation is, roughly, that the quantity denoted by the second is somehow correlated with and dependent on the quantity denoted by the first. There is nothing in the construction that has this meaning, and so the simplest assumption is that it is a pragmatically licensed interpretation assigned to the parataxis. But there is no alternative neutral coordinate interpretation. Therefore, it is reasonable to assume that the correlative interpretation is constructional.

Not every property of the construction is arbitrary, however. We can isolate a meaning for *the* in this case, roughly 'the extent to which X is more.'[6] So the

[5] Hence the term "correlative conditional" is sometimes used to describe this construction.

[6] Den Dikken (2005) proposes that *the* has a specific interpretation, but claims that the entire interpretation of the construction can be read off of the syntax. Culicover and Jackendoff (2005) argue that this is possible only if the missing parts of the interpretation are encoded as invisible constituents in the syntax, a step that is motivated only by the goal of making the syntactic representation sufficient for the interpretation. There is no independent evidence for such a richer syntactic structure.

interpretation of (14) will include representations for 'the extent to which John eats more' (call it EXTENT(EAT+)) and 'the extent to which John gets hungrier' (call it EXTENT(HUNGRY+)). But crucially, there are two parts of the meaning that have to be stipulated: (i) EXTENT(EAT+) and EXTENT (HUNGRY+) are correlated, and (ii) EXTENT(HUNGRY+) depends on EXTENT(EAT+).

This example of a construction is particularly important because it helps to define a category of constructions, namely, those that are conventionalizations of pragmatically licensed meanings. In order for this conventionalization to be possible, there has to be some particular formal characteristic that can become the signature of the construction. In this case, it is clause-initial *the more*. The correlative interpretation is supported even when there is one *the*-clause, which is interpreted as the independent quantity; cf. (17).

(17) a. Sandy gets hungrier (and hungrier), the more he eats.
b. The more Sandy eats, he gets hungrier (and hungrier).

It is possible to explicitly denote the correlative connection between two clauses. This produces part of the reading as the comparative correlative, but without the implication that there is a dependency. A word that clearly conveys a weak correlation is *as*:

(18) As Sandy slept, a fog rolled across the lawn.

What this means is that the time of the fog rolling across the lawn overlaps with a part of the time of Sandy's sleeping. However, there is no explicit causal connection between the two.

But a connection may be inferred if it is plausible, as in (19).

(19) As Sandy slept, his dogs destroyed the furniture.

The connection in this case is not causal in the sense that Sandy being asleep made the dogs destroy the furniture. But one might reasonably infer that the dogs were allowed to destroy the furniture because Sandy was asleep. Again, this is a matter of plausibility. If the main clause denotes a change of state that can plausibly be understood as causally dependent on the subordinate clause, a stronger connection is difficult to suppress, as seen in (20).

(20) As Sandy stood up, the money fell out.

However, in none of these cases is the connection between the events explicit.

If the two clauses denote compared quantities, the correlative meaning clearly implicates a dependency. E.g.,

(21) As Sandy got angry, the dogs got more nervous.

Example (22) clearly implicates that the dogs getting more nervous was connected to Sandy getting angrier, although an independent interpretation is possible. When we explicitly state that the change involves ongoing increases in two distinct quantities, the correlation even more strongly implicates the dependency, as in (22)–(24).

(22) a. As Sandy eats more, he gets hungrier.
b. Sandy gets hungrier, as he eats more.

(23) a. As Sandy eats and eats, he gets hungrier and hungrier.
b. Sandy gets hungrier and hungrier, as he eats and eats.

(24) Sandy gets increasingly hungry, as the amount he eats grows.

In each case, the *as*-clause contains an expression denoting increase and the generic form of the verb. The parataxis invites the inference that the differences on one measure at various times correlate with differences in the other measure, and moreover that the measure in the main clause is dependent on the measure in the subordinate clause.

To sum up, it is possible to understand why the CC construction works this way by seeing it as the limiting case of a range of ways of expressing the connection between two properties or events. CC is special, however, in that it specifies the precise form in which the connection must be expressed, and also does not allow for the two clauses to be interpreted as correlated but independent. In this sense, it is a genuine construction.

To isolate the idiosyncrasy, we must state the correspondence between the form and the meaning both of (14) and of the construction where there are two such expressions. Let us begin with the correspondence for the comparative, which is independently required (adapted from Beck 1997).

(25) *Comparative*

SYNTAX $\left\{\begin{matrix}\text{more XP} \\ \text{XP - er}\end{matrix}\right\}$

$\updownarrow$

CS $\lambda d.XP'[DEG:d]$

Note that the CS representation of *more/-er* does not include the relation '>' and what the comparison base is. I assume that these meaning components are provided by *than*, either explicitly or implicitly. *Than* is not present in the CC construction.[7]

[7] It is therefore not possible to have **The faster Sandy swims than Leslie, the more impressed I am.*

The simple correspondence in (25) allows us to take advantage of an existing general correspondence that can be incorporated into the special correspondence for the construction. The contribution of comparative *the* is given in (26).

(26) *THE (COMPARATIVE)*

$$\begin{array}{ll} \text{SYNTAX} & \text{the} \\ & \updownarrow \\ \text{CS} & \forall x.\lambda P.P(x) \end{array}$$

This interpretation is constructional and in fact lexical, in that it is associated with the item *the*. *The* applies to properties in such a way that it introduces the universally quantified variable x as the argument of the property P.

When we combine *the* and the comparative, we get the interpretation in (27).

(27) *the more XP*

$$\begin{array}{lll} \text{SYNTAX} & \text{the} & \left\{\begin{array}{l}\text{more XP} \\ \text{XP - er}\end{array}\right\} \\ & \updownarrow & \updownarrow \\ \text{CS} & \forall x.\lambda P.P(x) & \lambda d.XP'[DEG:d]) \end{array}$$

As in the case of other constructions, how the meanings of the parts are composed is stipulated in the correspondence. In this case, *the more/-er XP* has the interpretation that the meaning of *more/-er XP* is an argument of the meaning of *the*. The CS representation in (27) then reduces as shown in (28).

(28) $\forall x.\lambda P.P(x)(\lambda d.XP'[DEG:d])$
$\Rightarrow \forall x.(\lambda d.XP'[DEG:d])(x)$
$\Rightarrow \forall x.(XP'[DEG:x])$

This means 'for any x such that x is the quantity of XP′, ...'

Next we have the clause in which *the* is initial. As with other A′ constructions, the constituent containing the gap, e.g. *(that) Sandy runs*, has the function of a modifier of the expression to which it is adjoined. As in the case of other A′ constructions, the main expression binds a variable that is an argument to the property denoted by the modifier. (29) provides an illustration.

(29) *CC* CLAUSE

SYNTAX the $\boxed{\left\{\begin{array}{l}\text{more XP}\\ \text{XP - er}\end{array}\right\}}$ $[_S \ldots [e] \ldots]$

CS $[\forall x.\lambda P.P(x)(\lambda d.XP'[DEG{:}d])]^{\alpha}[\lambda y.\mathbf{F}(y)](\alpha)$

The reduction of this CS given in (31).

(30) $[\forall x.\lambda P.P(x)(\lambda d.XP'[DEG{:}d])]^{\alpha}[\lambda y.\mathbf{F}(y)](\alpha) \Rightarrow$
$[\forall x.\lambda P.P(x)(\lambda d.XP'[DEG{:}d])]^{\alpha}[\mathbf{F}(\alpha)] \Rightarrow$
$[\forall x.(\lambda d.XP'[DEG{:}d])(x)]^{\alpha}[\mathbf{F}(\alpha)] \Rightarrow$
$[\forall x.XP'[DEG{:}x]]^{\alpha}[\mathbf{F}(\alpha)]$

For example, if the expression is *the more Sandy eats*, the interpretation is

(31) $[\forall x.STUFF[DEG{:}x]]^{\alpha}[\mathbf{EAT}(AGENT{:}SANDY,PATIENT{:}\alpha)]$

that is, 'for any x, x the amount of stuff that Sandy eats.'

Now we come to the CC construction itself. I use the relation **CORREL**(X, Y) to denote the correlation between X and Y, where Y depends on X. The two expressions both correspond to representations such as (29). Since each denotes a quantity, the correlation is between the two quantities, as given in (32).

(32) $\forall x.XP_1'[DEG{:}x]^{\alpha}[\mathbf{F_1}(\alpha)]$, $\mathbf{CORREL}(x, y \mid \forall y.XP_2'[DEG{:}y]^{\beta}[\mathbf{F_2}(\beta)])$

So, for a sentence such as *the more Sandy eats the hungrier he gets*, the CS represent7ation is (33).

(33) $\forall x.STUFF[DEG{:}x]^{\alpha}[\mathbf{EAT}(AGENT{:}SANDY,PATIENT{:}\alpha)]$,
$\mathbf{CORREL}(x, y \mid \forall y.\mathbf{HUNGRY}[DEG{:}y]^{\beta}[\mathbf{BECOME}(SANDY,\beta)$

This means 'for any amount x of stuff α such that Sandy eats α, the amount y correlates with x, for any y amount of hungriness β such that Sandy becomes β.' This seems to be intuitively correct as the core interpretation of the CC construction, although it leaves out a number of details (see Beck 1997 for an extensive discussion of the semantics of CC).

What we have done here, then, is to formulate the interpretation of CC so that the parts that are independently required contribute to the meaning as a whole. As we have seen, what has to be stipulated about CC is the interpretation of *the* and the correlative interpretation of the parataxis, just as Culicover and Jackendoff originally claimed.

4.3 Concessives

English has a number of constructions that are described as 'concessives'—they grant the truth of a proposition, the existence of an object, or the value of a variable, as background to performing some other speech act, such as an assertion or request. Some examples are given in (34).

(34) a. Even if it's raining, you need to go outside.
b. (Even) though you're not tired, sit down.
c. Obama claims "success" in isolating Iran, although China and others still resist sanctions.
d. Levels of the main greenhouse gas in the atmosphere have risen to new highs in 2010 despite an economic slowdown in many nations that braked industrial output.

The concessives in (34a–c) concede the truth of some proposition, and the one in (34d) concedes the existence of something. Another common concessive is *no matter*, which concedes an arbitrary value to some variable, as exemplified in (35).

(35) a. No matter what the weather is like, you need to go outside.
b. No matter how tired you are, sit down.
c. Obama claims "success" in isolating Iran, no matter what China and others do.
d. Levels of the main greenhouse gas in the atmosphere have risen to new highs in 2010, no matter how much the economy in various nations has slowed down.

A curious property of *no matter* is that it can lack a copula, but nevertheless expresses predication, as discussed in *Syntactic Nuts*. Some typical examples are given in (36). The *no matter* phrase in each case is of the form *no matter wh-XP NP*, where XP is typically an adjective denoting a scale, and NP is definite, and a reasonable paraphrase of the missing copula is 'may be.'

(36) a. You need to go outside, no matter what the weather (may be).
b. No matter how tired your feet (may be), sit down.
c. Obama claims "success" in isolating Iran, no matter how negative the positions of other nations (may be).
d. Levels of the main greenhouse gas in the atmosphere have risen to new highs in 2010, no matter how slow the economy in various nations (may be).

No matter what NP can be paraphrased by *irrespective of NP* (37). And *no matter* itself can be paraphrased by *irrespective of*, but then *may be* is required (38).

(37) You need to go outside, irrespective of the weather.

(38) a. You need to go outside, irrespective of what the weather *(may be).
b. Irrespective of how tired your feet *(may be), sit down.
c. Obama claims "success" in isolating Iran, irrespective of how negative the positions of other nations *(may be).
d. Levels of the main greenhouse gas in the atmosphere have risen to new highs in 2010, irrespective of how slow the economy in various nations *[may be].

Other concessives are *regardless*, *irregardless*, *notwithstanding*, *disregarding*, and *irrespective*, all of which are somewhat marginal without the copula.[8]

(39) $? \left\{\begin{array}{l}\text{Regardless of}\\ \text{Irregardless of}\\ \text{Notwithstanding}\\ \text{Disregarding}\\ \text{Irrespective of}\end{array}\right\}$ how bad the weather, you need to go outside.

The examples suggest that *no matter* is a specialized construction with particular syntactic properties that must be explicitly marked. While the concessive interpretation can perhaps be isolated in the phrase *no matter*, it must be specified that the form can be as illustrated here.

So we may conclude (as I do in *Syntactic Nuts*) that there is a construction *no matter* that has the selectional property in (41), with the interpretation that XP is predicated of NP.

(40) *no matter*: ____ XP[WH] NP[DEF]

This construction exists along with the other selectional option, which is interrogative CP[WH].

If this was all there was to say about concessives in English, it would constitute evidence that it is necessary to stipulate certain syntactic properties of constructions in the constructional lexicon. But the situation turns out to be somewhat more interesting. Example (41) shows that the English comparative correlative, discussed in the preceding section, also allows the copula to be omitted when the predicate is adjectival or nominal.

(41) The bigger the car, the smaller the driver.

[8] I judge these cases to be marginal. They are certainly attested, but rare. E.g.,

(i) a. *Regardless of how rich* or how conspicuous, only uncivil nuts like to drive this type of cars. <http://www.insideline.com/mercedes-benz/cls-class/2012/2012-mercedes-benz-cls63-amg-2010-los-angeles-auto-show.comments?sort=>.
b. The exclusive artefact to go is to do your best, *disregarding how rich*, how poor, where you live, who you are, where you work. <http://ourfamilymagazine.com/page/183>.

And even more strikingly, concessives with *-ever*, which is a paraphrase of *no matter*, also show this property.[9]

(42) a. $\left\{\begin{array}{l}\text{No matter what}\\ \text{Whatever}\end{array}\right\}$ the cost, I always buy two tickets.

b. I always buy two tickets, $\left\{\begin{array}{l}\text{no matter what}\\ \text{whatever}\end{array}\right\}$ the cost.

What we see, then, is that there are three distinct concessive constructions, *no matter*, *-ever*, and CC, that allow for predication without a copula, where the predicate precedes the subject. In addition to (40), we have (43a) for *-ever* and (43b) for CC.

(43) a. XP[WH]-ever NP
b. the XP NP

Since these are all concessives, it does not seem to be an accident that they all share this constructional property. Intuitively, the metric for constructional complexity should assign a lower cost to constructions to the extent that they share properties. The problem in the case of concessives, and in many other cases, is how to formulate the properties so that this result is possible.

What we need to do is isolate the properties that the concessives share with one another. We can then show how the constructional correspondences for the various concessives follow the same pattern. While the construction is somewhat idiosyncratic, it is no accident that the concessives (and, as far as I know, only the concessives) have this idiosyncrasy. So, while the phenomenon is not general from the perspective of all of English or UG, it is locally general for English concessives.

An insight about how to proceed comes from the analysis of the interpretation of *-ever* in Dayal (2004). Dayal takes this interpretation as equivalent to that of 'free choice' *any* and represents it as a universal quantifier (see also Haspelmath 1997). As discussed in Chapter 3, a *wh*-word corresponds to a variable. Thus, *however big* corresponds to (44).

(44) $\forall$x.**BIG**[DEG:x]

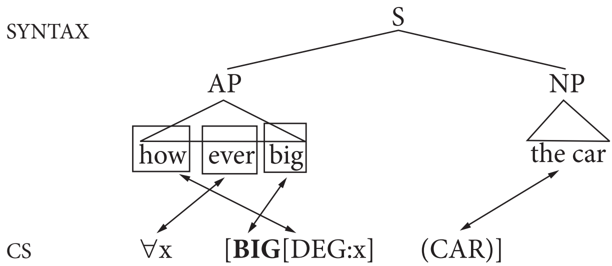
For *however big the car*, this representation yields (45), where we specify that BIG is predicated of CAR.

(45) $\forall$x[**BIG**[DEG:x](CAR)]

[9] The preferred *-ever* forms appear to be *whatever* and *wherever.*

This meaning in (45) corresponds transparently to the superficial morpho-syntactic structure, as shown in (46).

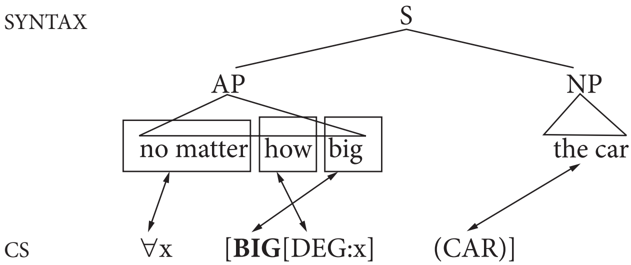
(46)



This means ‘for all quantities x, such that x is the size of a car.’[10]

Similarly, we can state the correspondence between the structure of *no matter* and the meaning roughly as follows.[11] Here, *no matter* corresponds to ∀x—it has the same meaning as -*ever.*

(47)



Interestingly, the free-choice interpretation is also expressed by the construction *it doesn’t matter.* The predicate expresses free choice and is a variant of *no matter,* so we should not be surprised to find that the complement may lack the copula. While not common, naturally occurring examples can be found.

[10] It is probably no accident that -*ever,* which is related to *every,* corresponds to the universal quantifier.

[11] In addition to *no matter what NP,* English appears to permit *no matter NP* and *It doesn’t matter NP,* as in

(i) a. We will do it, no matter (what) the cost.
b. It doesn’t matter (what) the cost, we will do it.
c. It doesn’t matter the color of skin, it doesn’t matter the nationality, it doesn’t matter the weight or height, it doesn’t matter the money someone has when you are choosing a partner. <http://hubpages.com/hub/I-would-love-to-hear-how-you-feel-about-dating-soemone-who-is-overweight>, accessed June 18, 2012.

(48) a. Love to go out and hang with the friends and ***it really doesn't matter what the*** **occasion or the event** as long as it's with friends.[12]

b. Second, it doesn't matter what the circumstances, advertising and/or promoting content from another site is against the rules.[13]

c. It doesn't matter what the cost/I have succeeded and I'm proud.[14]

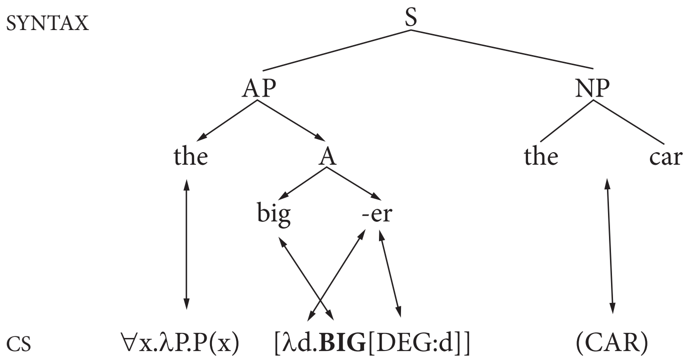
Remarkably, the construction can even be found with *give a damn.*

(49) They don't ***give a damn what the cost,*** they don't want to carry that garbage to the front.[15]

At this point let us return to the CC construction. Taking the correspondences in (46) and (47) as representative, note that part of the interpretation of CC is the free-choice universal quantifier as well. In CC there are two clauses with *the more*, and each contains a universal quantifier, as shown in (32). And both allow for the special copulaless construction under discussion.

A reasonable hypothesis, then, is that the omission of the copula and placement of the predicate before the subject is connected constructionally to the free-choice universal interpretation. The correspondence for *the bigger the car* is given in (50).

(50)



The CS reduces to

(51) $\forall x.[\mathbf{BIG}[DEG{:}x](CAR)]$

which is precisely the interpretation assigned to *no matter* and *-ever*. When we pair two *the*-clauses in a correlative parataxis, as in (41), the interpretation is (52)—

[12] <http://www.doulike.com/Iowa_Online_Dating/photos/359298.html?c=28>, accessed June 18, 2012.

[13] <http://www.bungie.net/Forums/posts.aspx?postID=66299234>, accessed June 18, 2012.

[14] From *The Blues of a Businessman*, Lyrics by Joel Sattler, based on Claude Dubois.<http://www.wattpad.com/2449537-the-blues-of-a-businessman-song-lyrics-by-joel>

[15] <http://newspaperarchive.com/the-delta-democrat-times/1968-02-07/>, accessed June 18, 2012.

(52) $\forall$x.[**BIG**[DEG:x](CAR)], **CORREL**(x, y | $\forall$y.[**SMALL**[DEG:y](DRIVER)])

—if we apply the correspondence uniformly to *the*... *the*... independently of the content of the complements of *the*.

Summing up, we have seen that these concessives share a correspondence of the form $\forall$x.**F**(x), where $\forall$x expresses the meaning of 'free-choice *any*' and **F** is quantifiable. Under these conditions, the special syntactic configuration exemplified in (50) is possible. This idiosyncratic property is therefore not totally arbitrary, but appears to be inherited from the general category of concessives. While not all concessives permit this construction, the fact that a good number of them do means that the grammar is less complex than if each concessive had different idiosyncratic properties, not only of form but of interpretation.

Examples like (41) inherit the correlative interpretation from the CC construction, and omission of the copula from the concessive interpretation. Thus, the inheritance hierarchy is as given in Figure 4.1.

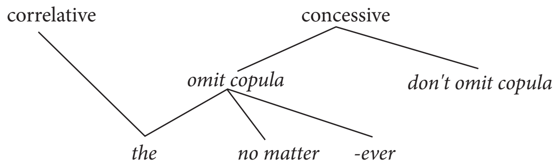


FIGURE 4.1. Inheritance hierarchy for concessives

4.4 Imperatives

Another important type of construction is the imperative. Traditionally, imperatives are one of the three fundamental sentence types in natural language, along with declaratives and interrogatives. While the syntactic properties of these latter two are the main concern of core grammar, mainstream approaches to syntax typically ignore imperatives.

The reason for this, one might suspect, is that imperatives tend to have idiosyncrasies that make it difficult to fit them into a picture in which superficial syntactic properties are reflections of universal principles and parameter settings. To take just one example, Government Binding theory incorporated principles and parameters to account for the distribution of phonologically null arguments. One parameter concerns whether a language allows empty pronominals, the so-called 'pro-drop' parameter. English is not a a pro-drop language, while Spanish is. However, the subject in the English imperative is typically empty, and must refer to the addressee. While it is possible to stipulate this property of the imperative, and perhaps even explain it in pragmatic terms (Bach and Harnish 1979), it clearly does not fit into the general framework of parameters.

An early study of the English imperative in generative grammar was Culicover (1971). I argued there that certain superficial forms are signatures of the imperative construction. For example, (53) shows that the imperative may lack a subject in the main clause, and the verb is uninflected, (54) shows that *don't* marks a negative imperative, (55) shows that the imperative can have an overt subject that refers to the addressee, (56) shows that an overt subject inverts with *don't* in the negative, (57) shows that an emphatic imperative may have *do* preceding the verb, and (58) shows that overt subjects do not co-occur with *do*.

(53) a. Be quiet.
b.*Is quiet.

(54) a. Don't be such a fool!
b. *Be not such a fool!

(55) $\left\{\begin{array}{l}\text{Everyone}\\ \text{You}\end{array}\right\}$ be seated.

(56) a. Don't you be such a fool.
b. Don't everyone talk at once.
c. *Do you be so helpful.
d. *Do everyone sit down.

(57) a. Do be quiet.
b. Do sit down.

(58) a. *You don't be quiet.
b. *You do be quiet.

Lasnik (1999) shows that it is possible to encode some of these properties in an abstract Imperative functional head that alternates with Tense, an instantiation of I^0. Unlike Tense, Imperative does not allow an auxiliary verb to function as AUX and attach to it, so we get (54a) with *do* support and not (54b), which contrasts sharply with *He is* $\left\{\begin{array}{l}\textit{not}\\ \textit{n't}\end{array}\right\}$ *a fool*. And, presumably Imperative has an ad hoc feature that allows the subject to be null when it refers to 'you,' whereas tensed verbs do not.[16]

[16] Culicover (1971) argues that Subjunctive is a variant of Tense. As seen in (i), it does not permit a null subject.

(i) It is imperative that *(you) be there on time.

Lasnik suggests that the presence of *not/n't* can trigger *do*-support as it does with Tense, "however that is to be captured in the theory" (114). This is not as simple as it may first appear, because as we will see, if we take the proposal seriously, Imperative really doesn't function like Tense.

The goal, of course, is to subsume the appearance of *do* in the imperative under a general account of *do*-support. *Do*-support is triggered in classical analyses by the need to attach Tense to a verb. Taking Imperative to be a variant of Tense for this purpose, we have the following analysis in (59) of (53a) and (54a). In (59b), the attachment of Imperative to *be* is blocked by *not*, so *do*-support applies.

(59) a. Imperative be quiet ⇒ be+Imperative quiet
b. Imperative not be such a fool ⇒ do+Imperative not be such a fool

Now, in order to derive the inversion cases in a way that is parallel to inversion in interrogatives, we must assume that there is a functional head, call it F^0, that attracts Imperative, as in (60). F^0 is the counterpart of the interrogative head C^0[WH]. If *not* contracts to *n't*, it attaches to Imperative, and thereby inverts when Imperative raises to F^0.

(60) F^0 you Imperative not be such a fool
⇒ F^0 you Imperative+n't be such a fool [contraction]
⇒ Imperative+n't+F^0 you be such a fool [inversion]
⇒ do+Imperative+n't+F^0 you be such a fool [*do* support]

(I am obviously glossing over some peripheral details here, such as exactly when *do* gets inserted and how Imperative and *n't* get attached to one another and to *do*.)

We can order *do*-support before contraction and get the same result.

(61) F^0 you Imperative not be such a fool
⇒ F^0 you do+Imperative not be such a fool [*do*-support]
⇒ F^0 you do+Imperative+n't be such a fool [contraction]
⇒ do+Imperative+n't+F^0 you be such a fool [inversion]

In the emphatics, we also get *do*. If this is a consequence of *do*-support, then there must be something that blocks attachment of Imperative to the verb. Call this Emph (as Klima did in the earliest analyses in the 1960s).

(62) Imperative Emph be quiet ⇒ do+Imperative Emph be quiet

But now we have a problem when we introduce F^0.

(63) F^0 you Imperative Emph be quiet
⇒ Imperative F^0 Emph you be quiet [inversion]

⇒ do+Imperative F^0 Emph you be quiet [*do*-support]
'*Do you be quiet'

We could put a feature on Imperative that blocks it from raising to F^0 when it is adjacent to Emph, but then we would also have to block *do*-support because of the ungrammaticality of (64) as an imperative.

(64) F^0 you Imperative Emph be quiet ⇒ F^0 you do+Imperative Emph be quiet
'*You do be quiet.'

As we see, then, there are problems of overgeneration when we try to subsume the imperative under the syntactic analysis that works for tensed sentences. This is because of the idiosyncrasies of the imperative.

The point that *Syntactic Nuts* makes about the imperative is not that an analysis in derivational terms is impossible. It is possible, technically, but it requires various stipulations, not just about the observed lexical items, but about hypothetical invisible functional heads such as Imperative and F^0, in order that just the observed forms are derived. Not only is it necessary to assume that these heads are present in the syntax, but it is necessary to constrain them in order to avoid overgeneration.

Syntactic Nuts argues that when we see a construction with apparently idiosyncratic properties, a plausible analysis is that it actually is a construction with idiosyncratic properties. The properties that the construction shares with the language as a whole, or with other more general constructions, are inherited and do not need to be specified in the description of the construction itself. But the idiosyncrasies do need to be.

In the case of the imperative, it appears that the simplest characterization is the following set of constructions, essentially what was proposed in Culicover (1971).

(65) a. Imperative: (NP) VP[BARE]
b. Negative imperative: *don't* VP
c. Emphatic imperative: *do* VP
d. Emphatic negative imperative: *don't* NP VP
e. Nothing else is a possible imperative

For pragmatic reasons, the NP in (65a) and (65d) must refer to the addressee or to a set containing the addressee, e.g.,

(66) Don't $\left\{\begin{array}{l}\text{you}\\ \text{everyone}\\ \text{anyone}\\ \text{any of you clowns}\\ \text{you or Sandy}\end{array}\right\}$ touch that red button marked "Armageddon"!

The next question to consider is, what precisely are the idiosyncrasies and what follows from inheritance?

- One clearly idiosyncratic property is that the verb appears in the bare form. There can be no modal auxiliaries, which occur only in tensed S's.
- Everything in an English imperative is where it is expected to be, given the properties of English. The verb is initial in VP, and the subject, when it appears, precedes the VP. So this does not have to be stipulated, although the SUBJECT construction does have to be modified to allow for subjects of non-finite, but inflected root VPs.
- When the imperative is negative, the form that is used to mark this is *don't*. The form of the auxiliary in this case is the bare form, *do*, not *does*. This is clearly not an accident, even though it is not straightforward to account for this form by the same mechanism that derives *do*-support in tensed S's, because there is no tense in imperatives (see the first point in this list).
- When the imperative is emphatic, the form used to mark this is *do*. Again, this is no accident. *Do* is used emphatically in declaratives as well.
- Overt subjects appear only with emphatic positive imperatives and negative imperatives.
- There is inversion in negative imperatives with overt subjects. While it is not clear that this possibility can be predicted, the syntax of inversion appears to be completely unexceptional from the perspective of English syntax in general.

It appears, therefore, that there are just three things that have to be stipulated about the English imperative. First, the verb has the bare form; second, inversion applies in a negative imperative; third, the subject may and sometimes must be omitted.[17] Everything else follows directly from general properties of English grammar, but only if they are formulated constructionally. The use of *don't* and *do* follows from the general English constructions informally characterized in (67)–(69).

(67) DO: when there is no auxiliary, and the construction requires an auxiliary (e.g. because of negation), use *do*.
(68) NEGATION: to express sentential negation, put *not* after the auxiliary or use an auxiliary with *n't* attached.
(69) EMPH: to express emphasis (on truth value), accent the auxiliary.

And Inversion is defined informally for English as follows.

[17] It is conceivable that omission of the second-person subject in the imperative is a pragmatic universal.

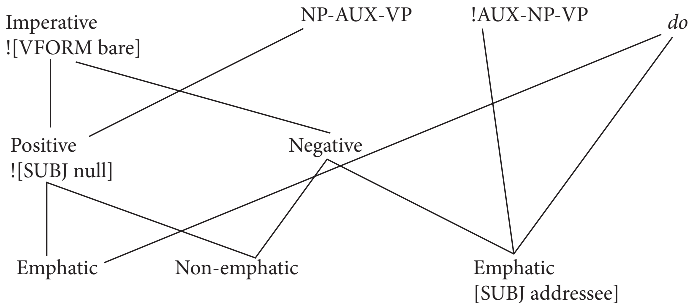


FIGURE 4.2. Inheritance hierarchy for imperatives

(70) SAI: the AUX position immediately precedes the subject.

Given this construction, the negative imperative has the following description, which is an idiosyncratic stipulation.

(71) NEGATIVE IMPERATIVE: *don't* appears in the (inverted) AUX position.

If there is a subject in the imperative, *don't* precedes the subject. If there is no subject, then *don't* precedes the VP.

The inheritance relations for the English imperatives are summarized in Figure 4.2. The stipulated properties are marked with '!'.

As seen in this hierarchy, while there are indeed idiosyncratic properties of the imperative constructions that need to be stipulated, many of them follow from general properties of the imperative, or are special instances of other constructions in the language, e.g. inversion and *do*-support. So, again, the situation is somewhat more natural than complete arbitrariness, which is precisely what we expect to find in a set of constructions that are closely related in function.

4.5 *Not*-topics (*not in my car (you won't)*)

I conclude this discussion of assorted nuts with discussion of a construction all of whose properties follow by inheritance. Demonstrating that such a case exists is an important confirmation of the general approach argued for in this book. We expect, in fact, that the language will maximize the extent to which constructions inherit general properties of the language. Those that do not completely do so are specialized constructions in the narrower sense, because they have idiosyncratic properties that have to be explicitly associated with them. But what might appear to be an idiosyncratic construction may turn out not to be, in this sense, if we get the general properties and the inheritance relations right.

In *Syntactic Nuts* I suggested that the construction exemplified by (72)—

(72) Not in my car, you won't.

—is a syntactic nut, that is, an idiosyncratic construction. I show here that further investigation of ellipsis, especially in Culicover and Jackendoff (2005) and Culicover and Jackendoff (2012), provides evidence that this is not a special construction, but comes by its properties in a systematic and predictable way. First I show that *not*+XP is a case of bare argument ellipsis where the XP is a focus and *not* corresponds to wide-scope negation. A sentence such as (72) is then analyzed as the concatenation of such a bare argument with VP ellipsis. The full interpretation is supplied by interpreting these two types of ellipsis with respect to some antecedent.

The construction in its simplest form is elliptical. It is a negative response to a prior question or statement, but also serves to constrain the scope of the response. E.g.,

(73) A: Are you interested in going swimming today?
B: Not today.

(74) A: It looks like the market is up.
B: Not today.

In terms of its syntax, this construction appears to be a variant of bare argument ellipsis, where a phrase stands alone as a statement and is interpreted on the basis of its relationship to the preceding discourse. Example (73) shows that the interpretation of *not*+XP can be contrastive, based on a match with the antecedent, while example (74) shows that the interpretation can be additive or 'sprouting,' in the sense of Chung et al. (1995). Both cases are instances of focus.

What is interesting in either case is that *not* is understood as taking scope over the proposition expressed by the antecedent, and is associated with the focus. Hence (73) has the interpretation in (75)—

(75) It isn't today that I am interested in going swimming.

—while (74) has the interpretation in (76).

(76) It isn't today that the market looks like it is up.

One reason why this construction appears to be special is that the combination *not*+XP does not occur freely in arguments or adjuncts in full English sentences, as shown by the following.

(77) a. *I am interested in going swimming not today.
b. *It looks like the market is up not today.

When wide-scope negation is used in a full sentence it appears in the auxiliary, and is attracted to the focus in the interpretation.

(78) a. I'm not interested in going swimming TODAY. [= It is [not today] that I am interested in going swimming.]
b. It doesn't look like the market is up TODAY. [= It is [not today] that the market looks like it is up.]

But if the antecedent sentence itself is negative, then *not*+XP can extend or restrict the scope of negation, as seen in (79)–(80).

(79) a. I'm not interested in going swimming, not today.
b. It doesn't look like the market is up, not today.

(80) a. I'm not interested in going swimming this week, not even today.
b. It doesn't look like the market is up, not even today.

These examples suggest that *not*+XP is a negated focus construction. It appears to be restricted to contexts where XP is a fragment. Hence, not surprisingly, this use of *not*+XP can also be seen when it is a bare argument, as in (73), (74), and (81).

(81) A: It doesn't look like the market is up.
B: (No,) not today.
B′: (No,) not even today.

The *not*+XP phrase in these cases confirms the negation of the antecedent, in the same way as a negative assertion can be confirmed by another negative assertion, e.g.,

(82) A: It doesn't look like the market is up today.
B: No, it doesn't (look like the market is up today).

But, as we have seen in other examples, *not*+XP can also be used contrastively with a positively asserted antecedent.

However, the following examples show that when *not*+XP is adjoined to a sentence it must confirm a negative antecedent in the sentence; that is, it cannot scope over the antecedent, in contrast to what we saw in the BAE dialogues in (73) and (74).[18]

[18] These cases are acceptable with *but*, which changes the construction from one in which *not* +XP is interpreted as an adjunct of the antecedent to a bare argument ellipsis.

(i) a. I'm interested in going swimming, but not today.
b. It looks like the market is up, but not today.

(83) a. *I'm interested in going swimming, not today.
[$\neq$ I'm not interested in going swimming today.]
b. *It looks like the market is up, not today.
[$\neq$ It doesn't look like the market is up today.]

The reason why *not*+XP cannot be adjoined to a positive clause is that it can only be interpreted as restricting or extending the negation in clause. (The cases of contrast are seen in dialogues such as (73) and (74).) The examples in (83) are unacceptable for much the same reason as the starred cases in (84) are: they are self-contradictory.

(84) a. Sandy is a fabulous chef, he really is(*n't).
b. They said he would win, and he $\begin{Bmatrix}\text{WILL} \\ \text{*WON'T}\end{Bmatrix}$.
c. They said he would win, but he $\begin{Bmatrix}\text{*WILL} \\ \text{WON'T}\end{Bmatrix}$.

Consider now the fact that the antecedent of *not*+XP may be repeated in dialogue as an elliptical negative sentence, to which the *not*+XP is attached. This produces the following types of examples.

(85) A: Can I go swimming today?
B: You can't, not today.
B′: Not today, you can't.

(86) A: It looks like the market is up.
B: It isn't, not today.
B′: Not today, it isn't.

The B′ cases in (85)/(86) are of course instances of the construction that we are concerned with. In fact, we see the same alternatives when the elliptical sentence is positive.

(87) A: Can I go swimming today?
B: You can't, today.
B′: Today, you can't.

(88) A: It looks like the market is up.
B: It isn't, today.
B′: Today, it isn't.

It thus appears that, quite generally, B′ is in fact a reordered variant of B, and that *Not in my car, you won't* is simply the combination of *not*+XP and an elliptical sentence. Thus, if it is possible to have (89a–e) or the equivalent, then it is possible to have (90a–e).

(89) Can I go swimming?
   a. Yes, you can go swimming today.
   b. Yes, you can today.
   c. Yes, today.
   d. Yes, today you can.
   e. Yes, today you can go swimming.

(90) Can I go swimming?
   a. No, you can't go swimming today.
   b. No, you can't today.
   c. No, not today.
   d. No, not today you can't.
   e. No, not today you can't go swimming.

Since this distribution is not restricted to *not*+XP, there is no need to have a special construction for *Not in my car, you won't*. It is a fully predictable combination of topicalization, focus, ellipsis, and negation.

Of course, a full account of this construction requires independently motivated accounts of how VP ellipsis and bare argument ellipsis work, as well as an account of how a parenthetical expression like *today* or *not today* is interpreted with respect to a sentence that it is adjoined to. Both of these topics are complex, and I cannot begin to do justice to them here. For proposals regarding the first, see Culicover and Jackendoff (2012). For the second, see Dehé and Kavalova (2007).

4.6 Summary

In this chapter I discussed a number of specialized constructions, or syntactic nuts, in English, and showed how it is possible to distinguish the respects in which they are regular and idiosyncratic.

Recall that our concern is to provide some substance to the notion 'possible language' within the framework of constructions. Since in principle constructions can express arbitrary correspondences between forms and meanings, it is essential to formulate some notion of how constructions cluster in a language around canonical patterns, and how the idiosyncrasies of constructions in a given language are constrained by their relationships to the more general constructions. This perspective does not promise to distinguish sharply between 'possible' and 'impossible' languages. However, it does suggest to what extent patterns of constructions are natural and likely to be encountered cross-linguistically.[19] Along the spectrum of naturalness, the more natural constructions contribute less complexity to the grammar in ways discussed in the preceding chapters, while the less natural ones contribute more complexity.

In this sense we have come back full circle to the notions of 'core' and 'periphery,' to view them in a different light. These notions are not about what is universal and what is specific to particular languages, but what patterns are more or less likely to be encountered cross-linguistically. We saw why languages have sluicing (§4.1.1), how sluice-stranding is linked to preposition-stranding (§4.1.2), how the same constructional components appear in the various concessives (§4.2–§4.3), and where the idiosyncrasies are in the English imperative (§4.4).

What emerges from these studies, I suggest, is that if there is any sense to the notion 'core,' it is the part of the grammar of the language that comprises its most general and widely used constructions. These are the constructions that express correspondences between syntactic form and aspects of conceptual structure that are central to the encoding and communication of thought, e.g., the marking of thematic roles, reference, predication, and attribution, declarative, interrogative and imperative, negation, and so on. The periphery then consists of more specialized constructions that play off of this 'core,' that is, idioms and constructions that express conceptual structures that are less than central, such as concessives and the comparative correlative.

[19] For analyses of some possible but nevertheless quite rare phenomena, and proposals about how they might come about as a consequence of linguistic change, see Booij (2008).

PART III

Processing complexity and grammar

This page intentionally left blank

5

Reflexes of processing complexity[1]

A persistent theme in linguistic theory is that there are constraints that govern the types of syntactic configurations that can exist in human language. These constraints for the most part prohibit syntactic chains where the tail of the chain, typically a gap, is within a particular syntactic configuration, while the head of the chain, typically a constituent in an A′ position, is external to the configuration. For this reason, these constraints have been called 'island constraints,' where the configuration containing the gap is an 'island' (Ross 1967).

The conventional view is that island constraints are part of the grammar. On this view, a sentence that violates an island constraint is ungrammatical in the same way that a sentence whose constituents are in the wrong order is ungrammatical—there is a violation of grammatical form.

Island constraints are presumed to be universal, part of the cognitive mechanism that embodies the human capacity for language acquisition. As with other features of Universal Grammar, constraints are incorporated into the grammars of individual languages in virtue of their role in restricting the form that such grammars may take.

This chapter explores an alternative view, which is that at least some phenomena attributed to island constraints are properly the consequence of processing complexity, not UG. Drawing heavily from the recent literature, I argue that the constraints are descriptions of syntactic configurations, and perhaps also form–meaning correspondences, that lead to processing

[1] I am very grateful to the Synners discussion group at OSU for giving me the opportunity to present some of the ideas in this chapter in a very preliminary form. Their questions and suggestions have led to substantial improvements and modification. Thanks go in particular to Yusuke Kubota, Chris Worth, Mike White, Elizabeth Smith, and Judith Tonhauser. Elaine Francis read an earlier version of this chapter and gave me very useful comments. I also thank Susanne Winkler for her comments and much valuable discussion, and Johannes Heim for his critical review. Many additional problems and opportunities for improvements were identified during presentations at the University of Paris VII and the University of Maryland. I am responsible for any errors.

complexity. Thus the corresponding unacceptability judgments are not reflections of what form the grammar should take, but reflect the interactions between particular syntactic configurations and the mechanisms that map these configurations into interpretations.

5.1 Universals

A prototypical example of an island is the constraint against extraction from a relative clause, illustrated in (1b). Here, the relative pronoun *which* has been extracted from the relative clause *who has just come out with **which**.*

(1) a. Sandy knows an author [$_S$ who has just come out with a very interesting monograph on island conditions].
  b. *This is the very interesting monograph on island conditions which$_i$ Sandy knows an author [$_S$ who has just come out with t_i]

Because these constraints are prohibitions against certain syntactic configurations, it might be expected that the ungrammatical configurations never occur in actual discourse. It is usually held that the constraints cannot be learned, especially given the plausible assumption that there could be no evidence for the learner that bears on the ungrammaticality of an example such as (1b).[2] But, given that they are not learned, the argument goes, the constraints must somehow be known by the learner prior to exposure to language. Thus, they are either specific a priori knowledge about the structure of language, or they are instantiations in language of more general a priori cognitive constraints. In the absence of evidence for more general constraints, the standard position has been that they are specific to language.

Such constraints must then be linguistic universals, the argument continues, since they concern possible linguistic structures in the absence of evidence from any particular language. They are thus part of the language faculty, the device in the mind that is devoted to acquiring language. This device embodies a definition of what constitutes a human language, by imposing constraints on all grammars. The search for universals is thus an

[2] For the role of positive evidence in learning grammar, see Wexler and Culicover (1980), Lasnik (1989), as well as recent discussions of Bayesian learning as applied to language acquisition, e.g. Chater and Vitányi (2007); Lidz and Waxman (2004); Regier and Gahl (2004); Tenenbaum et al. 2006. Early thinking about constraints (e.g. Chomsky 1973) did not take into account the possibility that the absence of a construction that should otherwise occur could be used as evidence that it cannot occur.

inquiry into the nature of this language acquisition device (Wexler and Culicover 1980).[3]

The perspective taken in this book is a somewhat more multifaceted and broader interpretation of this view. I assume here that there are a number of factors at play that determine the form that a language may take.

1. Constraints on meaning, and on the possible correspondences between meaning and form, are candidates for universals.
2. The pressure to reduce complexity will affect the form that language takes, whether the complexity exists in acquisition, in the representation of grammatical knowledge, in the computation of meaning in real time, or in the management of information density in communication.
3. If there are constraints on form per se, these are candidates for universals, to the extent that they cannot be reduced to extragrammatical factors 1 and 2.

The logic behind Factor 1 is that the most ubiquitous properties of meanings are likely to be reflected in the forms that express them. For example, reference to objects and their properties is a necessary part of communication, so it should not be surprising that in the grammar of every language there are formal categories that correspond, albeit imperfectly, to this semantic category. Due to the autonomy of syntax, the formal categories are not completely reducible to the semantic categories, but the influence of the latter is seen in the former. One example of this relationship is the widespread use of A′ chains to express *wh*-interrogatives, which arguably reflects the universal operator-variable relation in conceptual structure. While not every language uses every universally available device, those devices that do occur are plausibly understood as reflections of the form–meaning correspondence.

Factor 1 also bears on phenomena such as binding theory. It is plausible that the binding conditions (Chomsky 1981b) that have been argued to regulate the syntactic relationship between a pronoun and its antecedent reflect CS constraints on binding (Reinhart and Reuland 1991, 1993; *Simpler Syntax*).

Next, consider Factor 2. Since complexity is a measure that is independent of any particular language, it is plausible that there are universal properties of language that can be explained in terms of complexity. For example, if a particular construction is difficult for humans to learn or to process in real

[3] These paragraphs outline a version of the 'poverty of the stimulus' argument laid out by Chomsky (1986b). It is a topic with an extensive literature—see Pullum and Scholz (2002) and references cited there.

time, we would expect it to be rare. If it is simply too complex for humans to learn or to process in real time, it should not be found in any human language. There is no need, on this view, to assume that a formal constraint against such a construction is part of the human linguistic capacity.

Factor 3 recognizes the possibility that there may be formal linguistic universals that are not reducible to universals of meaning or processing constraints.

The types of universals that I focus on in this chapter have to do with grammatical form, but are not explainable in terms of the form–meaning correspondence per se or universals of meaning. I argue, building on the work of Hawkins, Kluender, Hofmeister, and Sag et al. (see references below) and extending it in certain respects, that these constraints do not reside in the language faculty, but are a consequence of the pressure in the system that processes language to reduce complexity in the computation of the form–meaning correspondence.

5.2 Where do universals live?

Before I consider particular cases, there is a fundamental methodological point to be addressed. In a recent review, Phillips (2011) summarizes three distinct positions that one can take regarding the relationship between processing and grammatical constraints. On one view, which I call the "grammatical constraint" view, a grammatical constraint is part of the universal capacity for language and is responsible for judgments of ill-formedness (typically characterized as "ungrammaticality") of sentences with particular properties. An example would be the Complex NP Constraint (CNPC) of Ross (1967), which is taken to be responsible for the ill-formedness of a sentence like (1b). On the grammatical constraint view, (1b) is ungrammatical because the relative pronoun *which* is linked to a gap inside of a relative clause, which is an island, in violation of the CNPC.

The second view, which I call "processing complexity," has been advocated by Kluender (1998) and Hofmeister and Sag (2010). On this view, example (1b) is not ungrammatical. It satisfies all formal requirements of English for which we have concrete evidence, e.g. the relative pronoun *which* is in clause-initial position, the gap is in the complement position of the verb, the relative clause follows the head noun, and so on. The processing complexity view is that as the cost of computing the interpretation grows, a threshold of acceptability is crossed, at which point a sentence is judged as "ungrammatical." What is problematic about (1b), on this view, or at least challenging at present, is that it appeals to an account in which the demand for computational resources is so great that it exceeds this acceptability threshold. It is essential on such an account to identify precisely what these computational resources are, to say what the threshold is, and to show that manipulation of the computational costs correlate with differences in acceptability judgments.

A related view is the "decathlon model" (see Featherston 2005b; Keller 2000 and work cited there). On this view, the violation of a constraint has a cost associated with it, which in itself may not be sufficient to render a given sentence unacceptable. Constraint violations are summed over a sentence, and if the total cost exceeds a threshold, the sentence is unacceptable. These costs are related to "computational workload" (Featherston 2005b: 7).

A third view, which Phillips (2011) calls "grounded," is a hybrid of the first two. This view is due most notably to Fodor (1978) and Hawkins (1994). On this view, there are genuine grammatical constraints, but they reflect or are "grounded" in processing complexity. On this view, a structure such as the one exemplified by (1b) starts out, so to speak, as grammatical in principle but is avoided by speakers due to its complexity. In the limit, this avoidance becomes encoded (somehow) in the grammar as a constraint. Thus the knowledge that (1b) is ill-formed derives not from the fact that interpreting it exceeds some threshold of acceptability, but from the related grammatical constraint that rules it out. As Phillips points out, it is not clear that there is an empirical test of groundedness that can distinguish it from the other two views.

I assume here a fourth view that is most closely related to groundedness, which I call "embodiment," building on an approach proposed in Culicover (1996) and Culicover and Nowak (2003). On this view, the mental representation of a grammatical structure is not static and symbolic, but part of a dynamical system whose function is to map between form and meaning. A grammar is an idealized description of the input and output of this function, but is not a description of the internal configuration of the dynamical system. The internal configuration not only incorporates grammatical knowledge, but also reflects how the language is used, including frequency and social factors.

An important point about this representation of language is that there is a correlation between how the system is configured with respect to a particular structure and judgments of acceptability. Very frequent structures are generally judged to be unexceptional (Bybee 2006). A structure that does not conform to some rule of the language—that is, a grammatical violation—is generally judged to be strongly unacceptable—that is, "ungrammatical." Such a case would be the wrong word order in English, for example, **book the.* Another would be an agreement violation, for example, **He are nice.*

Beyond this, it has been argued that even relatively small differences in the predictability of syntactic structures correlate with processing complexity, realized as surprisal; see Levy (2005) and Jaeger and Tily (2010). Culicover and Nowak (2003) employ a physical metaphor to capture the intuition that greater frequency of a construction corresponds to greater acceptability, other things being equal. The metaphor is that of a *flow* of related *trajectories* that has a certain strength and direction. The strength is measured by the width of the flow, corresponding to coverage, and its depth, corresponding to frequency of experience. A sentence is represented as a trajectory from points in various regions corresponding to words and phrases. Unexceptional structures are those represented by strong flows—these are the agglomeration of many individual trajectories that share the same direction and move through the same regions. The processing of an unexceptional example involves traversing a path that fits entirely within such a flow; no energy is required to move with the flow, and there is no resistance. On the other hand, an ungrammatical example requires movement at some point against the flow (in the case of word order violations) or movement into regions where there is no flow (in the case of agreement violations).

The middle ground is occupied by structures that may be partly well-formed in terms of their component parts, but are put together in such a way that they are difficult to process. In the physical metaphor, the trajectories move through areas where there are no flows, and hence require some energy to traverse. Other things being equal, greater processing complexity is reflected in lower frequency of occurrence, and thus weaker representation in the dynamical system. Hence the ill-formedness of the structure is not explicitly represented in the grammar, but it is encoded in the system in which grammatical knowledge is incorporated, in virtue of it being outside of the flow. The relatively low frequency of a given structure may be sufficient to produce unacceptability, through surprisal. Or it may be the case that low frequency combines with real time processing factors to render a given example highly unacceptable. In either case, the unacceptability of an example like (1b) would not be due to a grammatical constraint that explicitly rules it out, but to the fact that it is weakly represented in the system that processes language, if at all.

There may, however, be facilitating factors that make instances of a complex structure more acceptable. For example, it could be the case that a particular complex form is relatively frequent in a language or dialect. In such a case, it would be judged to be somewhat more acceptable than a comparable form in another language.

The combination of low frequency with real time processing factors may be responsible for a range of observations in the literature that show that violating a grammatical constraint does not always lead to ill-formedness (see Hofmeister and Sag 2010 for a review). For example, it appears that the CNPC does not appear to hold strongly when the NP is indefinite, the sentence is presentational, and the relative clause is infinitival. Thus, sentences like (2) are often judged to be less severely unacceptable than those like (1b).

(2) This is the kind of sink that there really aren't a lot of plumbers around to fix.

In fact, it has been argued that in Swedish it is systematically possible to violate the CNPC. That is, examples such as the following are acceptable.

(3) a.

Jag	känner	till	en	bok	som	en	flicka	gav	en	pojke.
I	know	of	a	book	which	a	girl	gave	a	boy

'I know of a book which a girl gave a boy.'

b.

En	flicka	Känner	jag	till	en	bok	som	gav	en	pojke
A	girl	know	I	of	a	book	which	gave	a	boy

c.

En	pojke	känner	jag	till	en	bok	som	en	flicka	gav.
a	boy	know	I	of	a	book	which	a	girl	gave

(Allwood 1982)

At the same time, Allwood observes that not all cases of extraction from complex NPs are equally acceptable, and that the acceptability depends to some extent on the choice of verb and the grammatical function of the constituent that is relativized. Moreover, examples in English that are comparable to acceptable examples in Swedish are not acceptable.

Facts such as these suggest that extraction from a relative clause contributes a certain amount of complexity, which contributes to unacceptability, and that factors such as definiteness contribute the rest. Such an analysis would be consistent with the fact that extraction from a definite is problematic even in the absence of a relative clause (Chomsky 1973).

(4) *This is the sink which I postponed the plumber's repair of (until Monday).

Moreover, the threshold for unacceptability appears to be higher in Swedish than in English.[4]

To sum up, embodiment is the view that processing complexity may produce unacceptability judgments in two ways. Such judgments may arise from limitations on resources in real time computation. Or they may reflect the representation of the low frequency of particular configurations in the system that computes the form–meaning correspondences.

4 It may be that the location of the threshold correlates with the frequency of the construction in the two languages, but the full explanation for the difference between the two languages is a question for future research.

The rest of this chapter is organized as follows. §5.3 reviews briefly a range of complexity factors that appear to interact to produce graded judgments of acceptability in island configurations. §5.4 uses evidence from parasitic gaps to argue that island effects can in fact be reduced to the processing complexity associated with these factors. §5.5 shows how complexity may produce the appearance of an island violation when none exists, lending further support to the view that what we are dealing with in the case of islands is not grammar, but processing.

5.3 Islands

5.3.1 Complexity factors

Let us consider next some of the factors that have been argued to incur processing costs. Among the factors that have been identified in the literature through experimental methods are the following.

(5) a. Constructing an A′ chain (i.e. holding an A′ constituent for subsequent linkage to a gap) (Gibson 1998: Alexopolou and Keller 2003: Ariel 1999).
b. Identifying the gap in an A′ chain (see Featherston 2001 for a review).
c. Processing referring expressions (Gibson 1998; Kluender 1998). Degrees of referentiality produce differential processing load (Ariel 2004).
d. Processing embedded tensed Ss (as contrasted with non-finite Ss) (Gibson 1998; Kluender 1998).
e. Processing inside of subjects (Kluender 2004).
f. Processing inside of clauses (Kluender 1998).
g. Crossing dependencies (Frank 2004; Frazier 2008; Kobele 2007; Levy et al. 2012; Stabler 2004).
h. Self-embedding (Miller and Chomsky 1963; Gibson 1998; Pickering and Barry 1991).
i. Hypothesizing the position of a gap that turns out to be filled already (the filled gap effect), thereby causing a type of garden path (Clifton and Frazier 1989).
j. Repairing garden paths (Fodor and Inoue 1994; 2000; Lewis 1998).
k. Pragmatic factors (Kroch 1989; Goldberg 2006; Ambridge and Goldberg 2008).
l. Frequency-based prototypes (Dąbrowska 2008) and surprisal (Levy 2005; Jaeger 2010)

Kluender in particular proposes that to a considerable extent the unacceptability of island violations can be accounted for in terms of the sum of the complexity contributed individually by several factors during processing (see also Featherston 2005b). One reason to believe that this approach is on the right track is that for every constraint that has been proposed in the literature for English there are cases that native speakers find quite acceptable (Hofmeister 2007; Sag et al. 2008; Hofmeister and Sag 2010). In some cases, there are non-obvious factors that make apparently identical structures differ in acceptability, but overall the evidence suggests that the judgment of unacceptability does not stem solely from the configuration, contrary to what we would expect if the source of the unacceptability was a matter of grammaticality, that is, a well-formedness condition. (As suggested in the discussion of embodiment in §5.2, the non-occurrence of a syntactic configuration may itself be responsible for some of the processing complexity.)

The factors that I focus on below are those identified by Gibson (1998) in his study of the complexity of relative clause chains. These are:

- the number of referential elements that have to be processed between the filler and the gap;
- the location of the gap within a finite complement vs a non-finite complement.

5.3.2 Grammatical constraint violations

Let us consider some particular cases. The following examples show a range of judgments for each of the most commonly studied constraints, and the contribution of some of the complexity factors in (5). Since all of the examples involve A′ chains, I do not mark (5b) and I mark (5a) only for multiple chains. The judgments '*,' '??,' and '?' are not intended to be precise, but represent my judgment of the relative acceptability of various configurations.[5] The absence of a mark does not necessarily mean that an example is completely unproblematic, since even the best of these examples are quite complex.

Complex NP Constraint (no chain into a relative clause (RC))

(6) a. *The campaign$_i$ that [$_{NP}$ someone [$_{RC}$ to spearhead t_i]]$_j$ was finally thought of t_j is now in total disarray. [(5e,f)]

b. *The campaign$_i$ that [$_{NP}$ someone$_j$ [$_{RC}$ who$_j$ could actually spearhead t_i]]$_j$ was finally thought of t_j is now in total disarray. [(5d,e,f)]

c. ?That's the campaign$_i$ [$_{NP}$ that I finally thought of someone [$_{RC}$ to spearhead t_i]]. (Kluender 2004) [(5f)]

[5] In some of these cases my judgments are consistent with those of the original literature, e.g. Kluender (2004). In other cases, my judgments are confirmed by the experimental study of Kurtzman and Crawford (1991).

d. ??That's the campaign$_i$ [$_{NP}$ that I finally thought of someone [$_{RC}$ who$_j$ could actually spearhead t_i]]. [(5d,f)]

(7) a. *That book$_i$ that I finally met [$_{NP}$ the first reviewer [$_{RC}$ to actually rave about t_i]] at a party is finally in paperback. [(5d,e,f)]
b. ??That's the book$_i$ that I finally met [$_{NP}$ the first reviewer [$_{RC}$ to actually rave about t_i]] at a party. [(5d,f)]
c. ?That's the most recent book$_i$ to completely mislead [$_{NP}$ the first reviewer [$_{RC}$ to actually rave about t_i]]. [(5f)]

Subject island (no chain into a subject)

(8) a. *Who$_i$ does [that she can bake ginger cookies for t_i] give her great pleasure? (Kluender 2004) [(5c,d,e,f)]
b. ??Who$_i$ does [us baking ginger cookies for t_i] give her great pleasure? (Kluender 2004) [(5c,e,f)]
c. ?Who$_i$ does [baking ginger cookies for t_i] give her great pleasure? [(5e, f); subject of *baking* is bound by *her*]
d. Who$_i$ does [being able to bake ginger cookies for t_i] give her great pleasure? (Kluender 2004) [(5e,f); subject of *baking* is bound by *her*]

(9) a. *a person who$_i$ [that Robin attacked t_i so viciously] didn't surprise me [(5c,d,e,f)]
[cf a person who$_i$ it didn't surprise me that Robin attacked t_i so viciously]
b. *a person who$_i$ [the fact that Robin attacked t_i so viciously] didn't surprise me [(5c,d,e,f)]
c. ??a person who$_i$ [for Robin to attack t_i so viciously] wouldn't surprise me [(5c,e,f)]
[cf a person who$_i$ it wouldn't surprise me [for Robin to attack t_i so viciously]]
d. ?a person who$_i$ [Robin's vicious attack on t_i] didn't surprise me [(5c,e)]
e. ??a person who$_i$ [Robin attacking viciously t_i] would really bother me [(5c,e,f)]
f. ?a person who$_i$ [attacking viciously t_i] would frankly give me great pleasure [(5e,f)]

CED (no chain into a subordinate clause)

(10) a. *a person who$_i$ Susan got real drunk [after she thought about t_i] [(5c,d,f)]
b. a person who$_i$ Susan got real drunk [after thinking about t_i] [(5f)]

(11) a. *?a person who$_i$ Susan went to Washington [so that her daughter could work with t_i] [(5c,d,e,f)]

b. ?a person who$_i$ Susan went to Washington (in order to) for her daughter to work with t_i [(5c,f)]

c. a person who$_i$ Susan went to Washington (in order) to work with t_i [(5f)]

Wh-island (no chain into a wh-question)

(12) a. a person who$_i$ I wonder [what$_j$ I should give t_j to t_i for Christmas] [(5a,c,d,f)]

b. a person who$_i$ I'm wondering what$_j$ to give t_j to t_i for Christmas [(5a,f)]

c. *a person to whom$_i$ I'm wondering what$_j$ I should give t_j t_i [(5a,c,d,f)]

d. ??a person to whom$_i$ I'm wondering what$_j$ to give t_j t_i [(5a,f)]

(13) a. *a present$_i$ that I wonder who$_j$ I should give t_i to t_j for Christmas [(5a,c,d,f,g)]

b. *a present$_i$ that I wonder who$_j$ to give t_i to t_j for Christmas [(5a,f,g)]

(14) a. ?a car$_i$ that I wonder how$_j$ you think you are going to fix t_i t_j [(5a,c,d,f,g)]

b. a car$_i$ that I can't figure out how$_j$ to fix t_i t_j [(5a,f)]

(15) a. ?a person who$_i$ I'm wondering whether I should talk to t_i [(5c,d,f)]

b. a person who$_i$ I'm wondering whether to talk to t_i [(5f)]

Subject island and wh-island

(16) a. *She is a person who$_i$ [whether the President should talk to t_i (or not)] is unclear. [(5c,d,e,f)]

b. ?She is a person who$_i$ [whether to talk to t_i (or not)] is quite unclear. [(5e,f)]

Most of these cases are familiar from almost a half-century of research on island phenomena. What they show is that it is actually possible under certain circumstances to form a chain into a relative clause, a subject, a subordinate clause, or an embedded question. But the acceptability of such a chain is also typically determined by one or more of the other complexity factors, such as the processing of a referring NP within the chain, crossing dependencies, a gap in a tensed S, and the repair of a garden path.

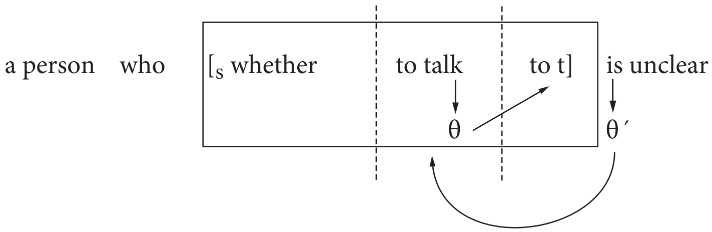
In many cases it is possible to construct minimal pairs, where the effect of a particular factor or a small number of factors can be isolated. For example, pairs such as (8a,b) show the effect of finiteness, while (8b,c) show the effect of an overt subject. However, it is important to stress that these judgments are very subtle, and they may be affected by factors other than those summarized here, as shown by the difference between (8c,d). Therefore it would be premature to attach to each postulated complexity factor a specific cost and to try to predict where judgments will fall based on the total cost incurred in an example with several factors added together.

5.3.3 Computing correspondences

It is important to highlight the fact that on the approach being pursued here, the factors summarized in (5) are seen as active during the building of structure and thus contribute to the judgment of unacceptability in general, not just in the case of the traditional island configurations. Consider, for example, the extraction illustrated in (16b). The sentential subject, while its grammatical function might be identified by default, precedes the verb.[6] Consequently it is not possible to link this Subject GF with the corresponding thematic role in CS at the point at which the chain <*who, t*> is formed. Until it assigned a role, the sentential subject has to be stored in temporary memory as a syntactic construct.

We thus have the situation illustrated in (17).

(17)



In this case, *who* is not interpreted as an argument of the main predicate of the relative clause, although it is in a position where such an interpretation is possible; cf. *a person who I like, a person who saw me*, etc. So the phrase *whether to talk to* immediately has the status of ruling out a Subject GF for *who*. One alternative is that *whether to talk to* is a sentential subject of the relative clause introduced by *who*.[7] This sentential subject lacks a thematic role until the predicate is processed. The chain <*who,t*> can be formed, but the trace is in an incompletely processed constituent. It is plausible that the sum of the processing load for maintaining the chain <*who,t*> and the processing load for maintaining the uninterpreted subject exceeds the acceptability threshold. (For specific proposals about how storage requirements and chain formation interact in processing of cases such as this, see Lewis 1998, 2000.)

The head of the chain in this case is *who* in A′ position, which clearly signals the necessity of constructing a chain. Interestingly, and not surprisingly from the processing perspective, when there is weak or no indication that there is a

[6] Alternatively, the complex phrase could be a topicalized non-subject, e.g.,

(i) ??a person who [whether or not to talk to *t*], nobody actually knows

In either case, the phrase lacks a thematic interpretation at the point at which the chain is being formed.

[7] The same problem arises if the *wh*-clause is a topicalized non-subject, as discussed in footnote 6.

chain in this case, acceptability is much lower.[8] This situation occurs in a *that*- or zero-relative clause, where the head of the chain is not readily accessible, as in (18a,b). In comparison, when the subject is extraposed, as in (19), the predicate *unclear* assigns a thematic role to the subject well before the gap is identified.

(18) a. *a person [$_S \left\{\begin{matrix}\varnothing \\ \text{that}\end{matrix}\right\}$ [$_S$[whether to talk to *t*] is entirely unclear]]

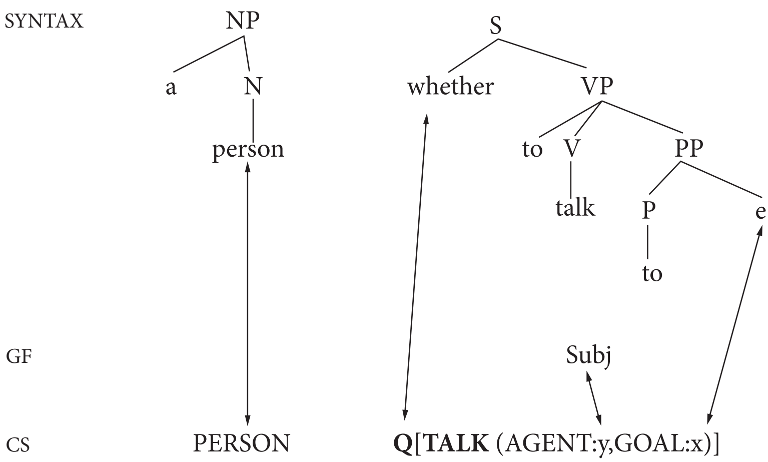
b. ?(?)a person [$_S$ who [$_S$ whether to talk to *t*] is entirely unclear]]

(19) a. a person [$_S \left\{\begin{matrix}?\varnothing \\ \text{that}\end{matrix}\right\}$ [it is unclear whether to talk to *t*]]

b. a person [who it is unclear whether to talk to *t*]

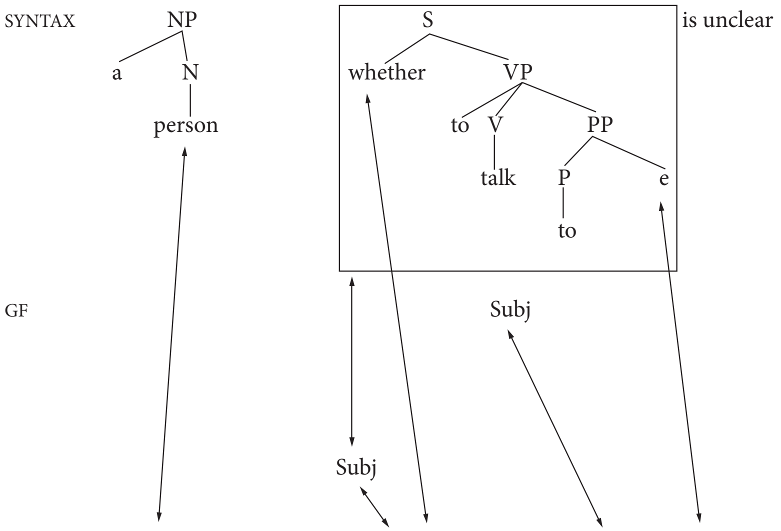
On standard mainstream analyses, these sentences and (16) have essentially the same syntactic representation, where [Spec,CP] contains a relative operator, *who*, or an empty operator *OP*, and the head position C^0 is phonetically empty or expressed as *that*. But in *Simpler Syntax*, the chain in the case of (16) is headed by *who*, which is linked to the head of the NP by agreement, while in the case of (18), the chain is headed by *person*.

The chain in *Simpler Syntax* is mediated by the correspondence between syntax and conceptual structure, so that in a simple case where there is no relative pronoun, there is a link from the head noun *person* to the CS element PERSON. The structure at the point at which the sentential subject is processed is essentially (20). Note that PERSON lacks an index, since at this point, given the absence of a relative marker, there is no evidence that there is a relative clause. Consequently the MOD:λx in the relative clause correspondence does not appear in the early form of the CS representation.

(20)


[8] For discussion of the role of overt relative marking in processing and its reflex in frequency of use, see Tagliamonte et al. (2005).

Since *whether to talk to* lacks an identifiable grammatical function, it cannot be incorporated into the phrase headed by *person.* It is only when we get to *is (entirely) unclear* that the processor recognizes that *whether to talk to* is the subject, and that there is a clause following *person* that is functioning as a relative clause. At this point, it is necessary to go back and mark PERSON with the index α, have this index bind the variable α, slot this variable into the position marked by *x* in CS, and assign to the *whether* clause the Subj GF of the relative clause, in order to derive (21). This processing has the character of a garden path repair—while it is not grammatically impossible, it requires considerable resources to hold the various incompletely processed constituents in memory and thus produces the subjective experience of strong unacceptability, typical of many garden paths.

(21)


CS $\text{PERSON}^{\alpha}[\text{MOD:}\lambda x.\textbf{UNCLEAR}(\textbf{Q}[\textbf{TALK}\ (\text{AGENT:}y,\text{GOAL:}x)])](\alpha) \Rightarrow$
$\text{PERSON}^{\alpha}[\text{MOD:}\textbf{UNCLEAR}(\textbf{Q}[\textbf{TALK}\ (\text{AGENT:}y,\text{GOAL:}\alpha)])]$

Note that when there is a grammatical marker of the relative clause, as in (18b), the effect is somewhat weaker, although extraction from the embedded interrogative is still problematic, for independent reasons.

In the next section I look at a number of other factors that contribute to the complexity of computing chain dependencies, in particular, specified NPs and finite tense. The picture that emerges is that the specificity of reference of a constituent facilitates the computation of dependencies in which it is the antecedent, but inhibits the computation of dependencies in which it intervenes between the antecedent and the dependent.

5.4 Parasitic gaps

5.4.1 Parasitic gap constructions

The island phenomena that we have been considering here have to do with chains. As noted, there are basically two options available to explain these phenomena. Either the unacceptability is due to ungrammaticality, that is, violation of one or more well-formedness conditions on the chains, or it is due to processing complexity. The fact that the judgments are graded in acceptability suggests that the latter is a factor. (Of course, both factors could be operative.)

How do we go about choosing between these alternatives, besides demonstrating that there are graded judgments for all of these island violations? One way to proceed is to show that the same graded judgments occur when there is no extraction, but there is nevertheless a chain. I argue here that such a case occurs in parasitic gap constructions. If the parasitic gap is not directly linked to the antecedent, but is interpreted as dependent on the antecedent in virtue of the true gap, it would be reasonable to conclude that unacceptability of the parasitic gap is due to the complexity of linking the gap to its antecedent, and not a strictly grammatical property, such as might be invoked in the case of an A′ construction such as *wh*-movement. Such an account presumes, of course, that there is no movement in the case of the parasitic gap, a point that I return to momentarily. And it presumes that the constraint is specific to extraction.

Once it has been established that there are complexity effects in the linking of a parasitic gap with its antecedent, the same logic can be used to argue that this is all that is going on in the linking of a true gap with its antecedent. That is, we can argue from parasitic gaps to true gaps, with the ultimate goal of eliminating grammatical constraints on chains.

A parasitic gap (notated *pg*) is exemplified in (22).

(22) a class that I failed *t* without ever signing up for *pg*

It is well known that parasitic gaps are typically more acceptable than their true gap counterparts, when the latter are island violations.

(23) a. ?a class that I passed Spanish without ever signing up for *t*
b. ?(?)a class that I paid my tuition without ever signing up for *t*

Thus, we cannot say that the relationship between the parasitic gap and its (apparent) antecedent is subject to the same grammatical constraints as is the corresponding true gap and its antecedent. If they were the same, the parasitic chain and the true chain should be equally acceptable.

But if what is going on is processing complexity, then we would expect to find that the parasitic gap shows the same acceptability as the true gap, incremented by the amelioration that the true gap contributes to the parasitic gap. Not all languages have parasitic gaps, and most of those that do lack the range of parasitic gap possibilities found in English (Culicover 2001). Since in principle a gap may function as a parasitic gap (as it does in English), I hypothesize that parasitic gaps are constructional in the sense that I have been using the term in this book. That is, they are defined form–meaning correspondences that are licensed in a specific language. Whether a given language permits a parasitic gap in a given environment is connected to whether it permits parasitic gaps in a range of other environments that are more general, in a sense to be made more precise.

In §5.4.2 I consider the variability that exists in the acceptability of a range of English parasitic gaps. As suggested by the variability seen in the case of the island phenomena discussed in §5.3, it is plausible to attribute the variability in judgments here directly to the dynamical construction of the chain, that is, to the identification of the antecedent at the point at which the gap, true or parasitic, is encountered. This approach allows us to account for the fact that parasitic gaps behave exactly like extraction gaps when they themselves appear in islands, as noted originally by Kayne (1983). The picture that ultimately emerges is that there is no extraction per se, no specifically grammatical constraints on chains, and the acceptability of any antecedent gap chain is a function of the complexity factors that are at play in the course of computing the dependency.

In the classic examples that compare parasitic gaps and true gaps, the true gap is unacceptable while the parasitic gap is acceptable. My interpretation of this is that it is the existence of the parasitic gap construction in the language that licenses the parasitic gap chain. If the parasitic gap chain does not itself contain too many complexity factors, the resulting judgment is that it is acceptable. But if the parasitic gap chain itself is complex, then the licensing of the configuration by the construction is not sufficient to bring the complexity down to a point where the chain is judged to be acceptable. As I show below in §5.4.3, in contexts where there cannot be a true gap, even with ameliorating factors, a parasitic gap is not acceptable. The conclusion is that in general, the acceptability of a chain is determined by the complexity factors that are determined by processing, as Kluender (1998) originally proposed.

There are two main possibilities for accounting for the possibility of parasitic gaps: chain composition and across-the-board (ATB) extraction. I argue here for a variant of the latter. The first possibility, chain composition, assumes that the parasitic gap is the trace of the extraction of an invisible operator (Chomsky 1986a). By assumption, this extraction does not in itself involve an island, but takes place entirely within the island, so there is no violation. In (24), the parasitic gap chain is $<OP_j, pg_j>$

(24) a class OP_i that I failed t_i [OP_j without ever signing up for pg_j]

The linking of the parasitic gap with the antecedent is accomplished by composing the true gap chain and the parasitic gap chain into a single chain.

This approach has the virtue that extraction of the empty operator may violate island constraints. But it requires a stipulation, namely that the two chains become composed into one, which does not explain anything about parasitic gaps but simply encodes the fact that they exist. To the extent that chain composition is supposed to be a general principle of linguistic theory, it fails to capture the fact that many languages lack parasitic gaps, and others have a more restricted inventory of parasitic gaps than do English and Swedish. Stipulations still have to be made about where parasitic gap constructions are possible.

Furthermore, there is no independent motivation for movement to the left periphery of an adjunct along the lines of the *without*-clause in (24). Nothing else can appear in this position in English, a fact that lends a decidedly ad hoc character to the analysis. So while the chain composition analysis in some form may well capture the facts (but see Levine and Sag 2003 for some criticisms), it does not constitute a particularly transparent account of why the parasitic gap construction behaves the way it does.

The second approach, ATB extraction, was proposed by Sag (1983) and has been adopted by others in a variety of theoretical approaches to extraction. On this view, sentences with parasitic gaps are treated as extensions of coordinate structures, where the property of containing a gap is shared by the phrase containing the true gap and the one containing the parasitic gap. This approach captures the fact that when the parasitic gap is inside of an island, there is an island violation (Kayne 1983).

(25) * a class $that_i$ I failed t_i without talking to [the professor who taught pg_i]
[cf *this is the class $that_i$ I talked to [the professor who taught t_i]]

Since there is an ATB chain from the antecedent to both the true gap and the parasitic gap, the island violation follows directly.

If we assume that a particular construction is derived by standard extraction mechanisms, then we would expect it to have all of the properties of standard extractions, not just some of them. In particular, if parasitic gaps are derived by ATB extraction, they should have the properties of ATB extractions in conjunctions. But languages that have ATB extraction may lack parasitic gaps. For example, it is perfectly possible to have ATB extraction in German, e.g., (26).

(26) Welche Zeitung hat Hans aufgeräumt und Maria gelesen?
which paper has Hans filed and Maria read?
'Which paper did Hans file and Maria read?'

But there is a broad consensus that German lacks parasitic gaps (Kathol 1997; Haider and Rosengren 2003), and in fact the parasitic gap construction corresponding to (26) is not possible.

(27) *Welche Zeitung hat Hans aufgeräumt, bevor Maria las
Which paper has Hans filed before Mary read
(Reich 2007)

Moreover, even languages that have parasitic gaps and ATB extraction do not have the variety of parasitic gap constructions that English and Swedish have (Culicover 2001). For example, English allows parasitic gaps in subject relatives but most languages lack such parasitic gaps.

(28) Sandy is a person who$_i$ everyone that meets pg_i likes t_i instantly.

Therefore, while we could characterize parasitic gaps as ATB, as in Sag (1983), it is still necessary to treat parasitic gaps as a specialized variety of ATB extraction, with idiosyncrasies that are not found in coordinate ATB. (For discussion and criticism, see Postal 1993.) That is, we have to specify in what non-coordinate configurations generalized ATB extraction is possible.[9]

It is important to note in this regard that if a language permits parasitic gaps in only one position, that position will be an argument of a non-finite *without*-adjunct, as in (22), repeated here (Culicover 2001).

(22) a class that I failed t without ever signing up for pg

This construction is the minimal extension of a coordinate with ATB extraction. In fact its interpretation is coordinate, meaning roughly 'and not.' Munn (2001) shows that parasitic gap constructions and coordinates share many properties beyond containing gaps in the conjuncts.

I suggest, therefore, that the parasitic gap construction is a language-specific extension of coordinate ATB extraction in which the set of licensed

[9] A classical empirical problem for the ATB approach with respect to English parasitic gaps is that the case of the gaps need not be the same, although presumably they must both match the case of the antecedent. For example,

(i) a. a class that$_i$ the department said t_i was required for graduation without allowing me to take pg_i

b. Robin is someone who$_i$ even good friends of pg_i believe t_i likes power entirely too much. (Levine and Sag 2003)

However, the same problem arises for ATB extraction from conjunctions.

(ii) a class that$_i$ the department said t_i was required for graduation and then didn't allow me to take t_i

Thus, case mismatch cannot be used as an argument against an ATB analysis.

environments is explicitly given. The extension is motivated originally by semantic coordination, but can be further generalized to non-coordinate cases. In the course of processing, the chain between the antecedent and the gap is formed dynamically, and what has to be accounted for is the accessibility of the antecedent at the point at which the parasitic gap is encountered. Thus we expect to find that the conditions that inhibit or facilitate interpretation of the parasitic gap will play a similar role in the interpretation of ATB extraction, as observed by Munn (2001), and as discussed further in §§5.2–5.3.

Extensions of ATB to non-coordinate configurations are not restricted to cases where one of the gaps is inside of an island. A language like English, which is very liberal in regard to where a parasitic gap may appear, should license parasitic gaps in non-coordinate non-islands. The examples in (29) show that it does.

(29) a. a friend who I never tell stories about *t* to *t*
b. a movie that I starred in *t* with the director of *t*
c. the member who we mentioned to *t* that we were going to kick out *t*
d. a man who you should never show such pictures of *t* to good friends of *t*

Finally, recall that on the ATB analysis the parasitic chain may violate island constraints. Consequently, if this analysis is correct, the island constraints cannot be part of UG, that is, prohibitions against certain syntactic configurations. Rather, they are reflections of processing complexity, and play a role the grammar of individual languages in virtue of being "grounded" or "embodied," as discussed in §5.2.

5.4.2 Complexity factors in parasitic gaps

As suggested in the preceding section, parasitic gap chains are just like true gap chains, except that the true gap chain licenses the parasitic gap chain in a configuration where it would otherwise be ruled out. When the parasitic gap chain itself is sufficiently complex, the improvement contributed by the true gap chain is not sufficient to make the parasitic gap acceptable.

In this section I consider the relevance of the complexity factors discussed in §5.3.2 to the acceptability of parasitic gaps. Recall that these factors have been identified as affecting the complexity of chains in general, independently of parasitic gaps. My judgments of the examples that illustrate the effects in adjuncts of an overt subject vs PRO subject, and of finite vs non-finite tense, are consistent with the experimental results reported on in Kurtzman and Crawford (1991).

Example (30a) shows that a chain into a subject is marginal. (30b) shows that addition of a true gap improves the first chain so that the sentence is judged (relatively) acceptable, a classical parasitic gap effect.

(30) a. *What did your attempt to repair *t* ultimately damage your reputation.
b. ?What did the attempt to repair *pg* ultimately damage *t*? (Phillips 2006: 796)[10]

But the true gap itself is not sufficient for acceptability if we introduce other complexity factors into the parasitic gap chain. (31a,b) show the effects of a definite NP intervening in the chain, while (31c) throws in the additional factor of finite tense.

(31) a. ?Which car did the request for you to repair *pg* ultimately devalue *t*?
b. ??Which car did Robin's attempt to repair *pg* ultimately damage *t*?
c. *Which car did the request that you should repair *pg* ultimately devalue *t*?

Examples involving extraction from a subject based on those in (8)–(9) above appear to produce similar judgments when the presence or absence of a subject within the subordinate clause is manipulated. The examples with overt subjects (32a) and (32a–e) are somewhat less acceptable than those that lack overt subjects.

(32) a. *Who does [us baking ginger cookies for *pg*] give great pleasure to *t*?
b. *?Who does [baking ginger cookies for *pg*] give great pleasure to *t*?
c. ?Who does [being able to bake ginger cookies for *pg*] give great pleasure to *t*?

(33) a. *a person who that Robin attacked *pg* so viciously didn't bother *t*
b. *?a person who the fact that Robin attacked *pg* so viciously didn't bother *t*
c. ??a person who for Robin to attack *pg* so viciously wouldn't bother *t*
d. ??a person who Robin's vicious attack on *pg* didn't bother *t*
e. ?a person who Robin attacking viciously *pg* would really bother *t*
f. a person who attacking viciously *pg* would really bother *t*
g. a person who vicious attacks on *pg* would really bother *t*

And we find that locating the parasitic gap after the true gap (that is, not in a subject), as illustrated by the examples in (34) and (35), reduces the unacceptability in many cases.

[10] The '?' judgment here is mine.

(34) a. Who did you issue an invitation to *t* without [us baking ginger cookies for *pg*]?
b. Who did you issue an invitation to *t* without [baking ginger cookies for *pg*]?
c. Who did you issue an invitation to *t* without [being able to bake ginger cookies for *pg*]?

(35) a. ?a person who it didn't bother *t* after Robin attacked *pg* so viciously
b. ??a person who I failed to inform *t* about the fact that Robin attacked *pg* so viciously
c. a person who it wouldn't bother *t* for Robin to attack *pg* so viciously
d. ?a person who I failed to inform *t* about Robin's vicious attack on *pg*
e. (?)a person who I failed to inform *t* about Robin attacking viciously *pg*
f. a person who I failed to inform *t* about attacking viciously *pg*
g. a person who I failed to inform *t* about vicious attacks on *pg*

The parasitic chain into a finite adjunct (36b) is less acceptable than the one into a non-finite complement (36a).

(36) a. a person who I talked about *t* without anyone recognizing *pg*
b. ?a person who I talked about *t* although no one recognized *pg*

A final set of examples that illustrate the effects of complexity factors on the acceptability of parasitic gaps are those in which the head of the chain is not an NP. It is generally believed that a parasitic gap must have an NP antecedent, so that an example such as (37) with a PP antecedent is severely unacceptable.

(37) *To whom did the idea to talk *pg* ultimately appeal *t*?
[cf. The idea to talk to Otto ultimately appealed to him.]

However, the examples in (38), due to Levine et al. (2001), show that non-NP antecedents are possible. Notice that all of these examples but one use *without*, semantically the least complex subordinator, equivalent to 'and not.' Other subordinators impose additional meanings, such as temporal sequence.

(38) a. How harshly do you think we can treat them *t* without in turn being treated *pg* ourselves? [adverbial P-gap][11]
[cf ??How harshly do you think we can treat them *t* after having been treated *pg* ourselves?]

[11] Elaine Francis (p.c.) suggests the following example which is even better than (38a), and also contains a parasitic gap.

(i) How harshly do you think we can treat them *t* without them in turn treating us *pg*?

b. That's the kind of table on which it would be wrong to put expensive silverware *t* without also putting *pg* a fancy centerpiece. [PP P-gap] [cf *That's the kind of table on which it would be wrong to put expensive silverware *t* before also putting *pg* a fancy centerpiece.]

c. I wonder just how nasty you can pretend to be *t* without actually becoming *pg*. [AP P-gap]
[Cf *I wonder just how nasty you can pretend to be *t* before actually becoming *pg*.]

d. That drunk, it would be impossible for me to get *t* without Robin getting *pg* as well. [AP P-gap]
[Cf *That drunk, it would be impossible for me to get *t* after Robin gets *pg* as well.]

e. How drunk can you appear (to be) *t* without actually becoming *pg*?/ How drunk can you appear (to be) *t* without actually being *pg*? [AP P-gap]
[Cf *How drunk can you appear (to be) *t* before actually becoming *pg*?]

f. That's exactly how strange we got them to act *t* (after, of course, acting *pg* ourselves). [Adverbial P-gap]

g. A doctor, you could spend your whole life trying to be *t* without ever becoming *pg*. [non-referential NP P-gap]
[Cf *A doctor, you could spend your whole life trying to be *t* before actually becoming *pg*.]

h. That Robin is a spy would naturally be difficult to refute *t* without (someone) having first conjectured *pg*. [sentential P-gap]
[Cf *That Robin is a spy would naturally be difficult to refute *t* after (someone) having first conjectured *pg*.]

Crucially, all of these parasitic gaps are in non-finite, subjectless adjuncts that follow the true gap. It appears that, whatever the reason, a non-NP antecedent of a parasitic gap is by itself highly problematic, although the examples demonstrate that it is not completely impossible, as long as there are no other complicating factors. So, we would predict that any additional complexity factor would render a case with a non-NP antecedent impossible, even while nothing else about the configuration is changed. And in fact this is what we find.

(39) a. *How harshly do you think we can treat THEM *t* without them in turn being treated *pg*? [overt subject][12]

[12] But note that while *them* adds just enough complexity to make (39a) quite unacceptable, it does not have this effect in the example in footnote 11, which is presumably simpler in other respects (e.g. the subordinate clause is in the active, not passive).

b. *That's exactly how strange we got them to act *t* (after, of course, we acted *pg* OURSELVES). [tensed S and overt subject]
c. *Without actually BECOMING *pg*, I wonder just how nasty you can PRETEND to be *t*. [parasitic gap precedes true gap]
d. *How harshly can we ask them *t* [whether they acted *pg*]? [tensed S and overt subject; *wh*-island]
e. *That Robin is a spy we had almost refuted *t* before we had learned *pg*. [tensed S and overt subject]
f. *How nasty did the people who became *pg* originally expect to be *t*? [tensed S, overt subject, parasitic gap in a subject]

Let us return now to a point made earlier. The examples in (30) and (34)–(35) suggest that a chain whose formation causes an unacceptability judgment, whether it is into an island or not, may be improved by the presence of another chain. The following examples develop this idea further. In (40)/(41) we see additional examples showing that a parasitic chain into a subject NP and into a relative clause in a subject is more acceptable than a true chain into the same domains.

(40) a. a man who even friends of *pg* dislike *t*
b. ?a man who even friends of *t* dislike that movie

(41) a. a man who everyone who knows *pg* dislikes *t*
b. *a man who everyone who knows *t* dislikes that movie

Examples (40b)/(41b) are familiar subject condition/CNPC violations.

Example (42a) shows that, surprisingly, two chains, both of which are true chains into a subject NP, are more acceptable than one chain into either subject NP.[13] (I mark all of the gaps as *t*, since it is not clear which are true gaps and which, if any, are parasitic.)

(42) a. a man who friends of *t* think that enemies of *t* are everywhere
b. ??a man who friends of *t* think that goblins are everywhere
c. ??a man who friends of mine think that enemies of *t* are everywhere

Since both chains are subject condition violations, as shown by (42b,c), and there is no true gap that satisfies the island constraints, (42a) should be ungrammatical.

The examples in (43) show the same thing for chains into relative clauses in subjects. (43a) violates the CNPC with respect to both subjects.

[13] Examples such as these were developed in discussions with Robert Levine.

(43) a. a man who everyone who knows *t* thinks that everyone who dislikes *t* is misguided
b. *a man who everyone who knows me thinks that everyone who dislikes *t* is misguided
c. *a man who everyone who knows *t* thinks that everyone who dislikes me is misguided

Finally, the examples in (44)–(45) show that it is possible for a parasitic chain into a subject to license a parasitic chain into an adjunct. Again, there is no chain in (44a)–(45a) that satisfies all island constraints.

(44) a. a man who several friends of *t* tried to get in touch with me before I contacted *t*
b. ??a man who Sandy tried to get in touch with me before I contacted *t*

(45) a. a man who several friends of *t* tried to get in touch with me before any enemies of *t* could (get in touch with me)
b. ??a man who Sandy tried to get in touch with me before any enemies of *t* could (get in touch with me)

The data reviewed in this section support the constructional analysis of parasitic gaps. The parasitic gap construction in English licenses a gap in a sentence that contains another gap that does not c-command it. The chain of both gaps is headed by the same antecedent, as in ATB extraction. A parasitic gap is less acceptable when there are complexity factors in its chain, most notably an overt subject or finite tense.

5.4.3 Some impossible parasitic gaps

I turn now to the final claim made about parasitic gaps in the introduction to this section. The idea that parasitic gaps are licensed only if they are in a context that does not exceed some complexity threshold predicts that there will be some configurations in which parasitic gaps are impossible. In these cases there is so much complexity that the ameliorating effect of the second gap does not boost the parasitic gap over the threshold.

One set of cases is due to Kayne (1984), who showed that parasitic gaps show island effects. E.g.,

(46) *a person who I talked to *t* without recognizing the people who insulted *pg*

This is precisely what we expect, since the parasitic gap is inside of a tensed relative clause with an overt subject. As expected, if we eliminate the two complexity factors, the parasitic gap becomes more acceptable, if not perfect in this configuration. (I have changed the lexical items in order to make the infinitival relative semantically plausible.)

(47) a car that I made an offer on *t* without having any money to pay for *pg* [cf. *a car that I made an offer on *t* without having any money that would pay for *pg*][14]

Similar examples can be constructed with other island configurations.

(48) a. *a person who I talked to *t* without admitting that the fact that I disliked *pg* didn't bother me [sentential subject]
b. ?a person who I talked to *t* without ever admitting that disliking *pg* didn't bother me

(49) a. *a person who I talked to *t* until I got nervous because I didn't recognize *pg*
b. ?a person who I talked to *t* before getting nervous because of not recognizing *pg*

Next, consider *that-t* and other complementizer (C^0)-*t* effects.

(50) a. a candidate that I believe (*that) *t* will win
b. *a candidate that I wonder whether *t* will win

As noted in Culicover (1993a), these gaps are not ameliorated by a true gap when they appear as parasitic gaps.

(51) a. I told supporters of Hillary last week [(that) she might actually win]
b. the candidate that I told supporters of *t* last week [(*that) *pg* might actually win]

(52) a. I asked supporters of Hillary last week [whether she might actually win]
b. *the candidate that I asked supporters of *t* last week [whether *pg* might actually win]

But *that-t* effects are ameliorated by an intervening adverb (Culicover 1993a). This is the 'adverb effect.'

(53) a person who I believe that $\left\{\begin{array}{l}\text{most likely}\\ \text{certainly}\end{array}\right\}$ *t* will win

And the adverb effect is seen in some but not all C^0-*pg* cases.

(54) a. the candidate that I told supporters of *t* last week [that under certain circumstances *pg* might actually win]

[14] Thanks to Elaine Francis for suggesting these examples to me.

b. ??the candidate that I asked supporters of *t* last week [whether under certain circumstances *pg* might actually win]

The adverb effect and the impossibility of ameliorating a *that-pg* with a true gap suggest that the *that-t* effect is in fact a local processing phenomenon, as suggested by Culicover and Levine (2001). They suggest that it is a garden path effect that interferes with the identification of the gap of the chain in the immediate context of *that*. If this is correct, we would not expect the true gap to improve the parasitic gap itself.

More generally, we should expect that problems that arise in the identification of gaps will not be ameliorated by true gaps, since the processing complexity is a local phenomenon having to do with the identification of the gap itself. The lack of amelioration in such cases is in fact seen in the following examples, where the immediate postverbal indirect object is the gap.

(55) a. the customer who I sold a defective iPod to *t*
b. ?the customer who I sold *t* a defective iPod

Jackendoff and Culicover (1972) attribute the marginal character of (55b) to a type of garden path.[15]

As predicted, the parasitic gap in this construction is no more acceptable than is the true gap.

(56) a. the customer who I sold a defective iPod to *pg* without even providing *t* with a receipt
b. ?the customer who I sold *pg* a defective iPod without even providing *t* with a receipt.

(57) a. the customer who I robbed *t* after selling a defective iPod to *pg*
b. ?the customer who I robbed *t* after selling *pg* a defective iPod

In this case, the difficulty of identifying the gap immediately after the verb is not ameliorated by the true gap, although the possibility of having a gap within the adjunct (the clause headed by *after*) is licensed by the parasitic gap construction. In other words, the construction says that a gap within a particular configuration is possible, but does not specify how to find the gap if there are other factors that make gap identification problematic.

5.4.4 Summary: processing gaps

This section has presented evidence that shows that the construction of parasitic chains is subject to all of the complexity factors that apply to

[15] The unacceptability judgment of A′ extraction in double object constructions is not shared by all speakers.

extraction chains, because it is an extraction chain. These complexity factors produce the appearance of island violations in the case of extraction chains, when their total complexity exceeds threshold. Comparable parasitic chains are somewhat less complex, because of the effect of the second gap, which is the characteristic property of these constructions. But these chains otherwise show all of the same complexity effects.

The implication of these data is that a certain level of complexity inhibits access to a gap in the course of constructing a chain. Suppose that we have the configuration in (58). There is one *wh*-phrase in A′ position, and two gaps, t_1 and t_2.

(58) wh-NP [X [t_1]] . . . [Y [t_2] . . .]

In order for a well-formed interpretation to be formed, the *wh*-phrase must bind a gap. At the same time, both gaps must be bound. If the first gap is not bound because the intervening X over which the chain must be formed is too complex, then at the point at which the *wh*-phrase binds the second gap, the first gap is unbound. This is a type of garden path effect, which can only be repaired by going back to the *wh*-phrase and the first gap and forming a chain.

That this is in fact what happens is suggested by an experimental study by Phillips (2006). Phillips shows that some gaps in subjects are accessible to the processor—the processor knows that such gaps are potential parasitic gaps that might be licensed by a true gap. Other gaps in subjects do not appear to be accessible to the processor. Here is a minimal pair.

(59) a. What did the attempt to repair *pg* ultimately damage *t*? (Phillips 2006: 796)
b. *What did the reporter that criticized *pg* eventually praise *t*?

Phillips shows that not only is (59b) judged unacceptable, but there is no evidence for activation of a gap inside of the relative clause in online processing. The processor doesn't even try to bind the parasitic gap. In contrast, a pronoun is readily available.

(60) What did the reporter that criticized it eventually praise *t*?

A plausible interpretation of the difference in the processing of the examples in (59) is that chains are not built into forbidden territory when the complexity level is high enough (see Hofmeister et al. forthcoming). The failure of a parasitic gap in (59b) means that at the close of the relative clause, there is an unresolved gap, which produces a garden path effect.

Observe, finally, that the processor does not know for sure that the first gap is a parasitic gap. It can only determine that it is if it has in fact formed a chain to the first gap, and then encounters the second gap. This sequence of events suggests that the same processes are at play in the construction of parasitic chains and true chains. But if this is the case, there is no reason to treat the two types of chains any differently from the perspective of the grammar. What is relevant is how much complexity is involved in construction of the chain. If there is too much complexity, the chain will not be formed. If there is some complexity, the chain can be formed, with some degree of unacceptability.

5.5 Complex ellipsis

If the approach to complexity and acceptability outlined above is on the right track, we should find other cases of demonstrable complexity that produce the subjective experience of unacceptability. An interesting example noticed by Lasnik (2001) and discussed by Merchant (2008) is given in (61).

(61) *They want to hire someone who speaks a Balkan language, but I don't know which$_i$ (Balkan language) they do ([‘want to hire [$_{NP}$ someone [$_S$ who speaks t_i]]’).
=‘*They want to hire someone who speaks a Balkan language, but I don't know which (Balkan language) they do.’

This example shows ellipsis of a VP that contains a relative clause, and a chain formed from *which* and a trace that is (presumably) inside of this missing relative clause. The bracketing illustrates that the material in the ellipsis contains a gap inside of an island. Since the relative clause is an island, it is natural to attribute the ungrammaticality here to a violation of the CNPC.

As Lasnik says, ‘The island is gone at PF, erased by VP Ellipsis, but the example is still seriously degraded.’ The fact that this sentence is degraded is bit of a puzzle, given that in general, ellipsis constructions such as sluicing in which there is a gap inside of the invisible island appear not to be subject to island constraints, as originally noted by Ross (1969).[16] E.g.,

(62) They want to hire someone who speaks a Balkan language, but I don't know which$_i$ (Balkan language) (‘they [$_{VP}$ want to hire [$_{NP}$ someone [$_S$ who speaks t_i]]]’).
‘They want to hire someone who speaks a Balkan language, but I don't know which (Balkan language).’

On the face of it there is nothing grammatically wrong with (61); in fact, it is syntactically very simple, in virtue of the VP ellipsis. The difference between

[16] However, it should be noted that Ross marked cases where an island is sluiced with ‘?’, suggesting that for him they were not completely acceptable.

(61) and sluicing in (62) appears to be that in the case of ellipsis the antecedent VP (or its meaning) has to be reconstructed, while in the case of ellipsis, the entire antecedent clause (or its meaning) has to be reconstructed.

Merchant's solution is to posit there are in fact differences between the two constructions in terms of invisible syntactic structure. In the case of sluicing, the entire chain between the *wh*-form and its trace is absent at PF, but in the case of ellipsis, there is a remnant of the chain in the form of an invisible intermediate trace that is attached to the VP and present in the derived structure. This trace is inside of the island while its antecedent is outside, and therefore is sufficient to produce the violation.

The alternative that I pursue here is based on the approach of *Simpler Syntax*. According to *Simpler Syntax*, there is no invisible structure, and no movement. Hence on this view there is no syntactic island in (62). The question, then, is why there appears to be an island violation in (61). I say "appears" because crucially, all that we actually know about (61) is that it is unacceptable. The challenge is to find an alternative source of the unacceptability that does not rely on an island violation.

We can find such a source in the process by which a fragment is assigned an interpretation. Interpretation requires matching of the fragment to a constituent or constituents of the antecedent. This match is used to license the fragment syntactically, and to identify the portion of the meaning of the antecedent that is to be composed with the meaning of the fragment to produce full interpretation for the fragment. For a detailed discussion of how this interpretation proceeds under *Simpler Syntax* assumptions, see Culicover and Jackendoff (2012)).

In the case of sluicing, it is sufficient to find a match in the antecedent directly, and substitute the interpretation of the fragment for the interpretation of the match. The interpretation of (62) for *which* is given in (63). As discussed in Chapter 3, the variable α corresponds to the interrogative *which*, and $\mathbf{Q}^{\boldsymbol{\alpha}}$ is the interrogative operator that binds it.

(63) *Antecedent interpretation*

$\mathbf{WANT}(\mathrm{THEY},\mathbf{HIRE}(\mathrm{Y}^{\beta}[\mathrm{INDEF};\mathrm{MOD}{:}\mathbf{SPEAK}(\beta,\mathrm{BALKAN\text{-}LANG})]))$

Fragment interpretation

$\mathbf{NOT}(\mathbf{KNOW}(\mathrm{ME},\mathbf{Q}^{\boldsymbol{\alpha}}(\alpha_{\mathrm{thing}})))$

Interpretation of the fragment requires constructing a lambda expression from the antecedent interpretation, by identifying the indefinite NP in the antecedent and doing lambda abstraction on it. This yields (64).

(64) $\lambda x.\mathbf{WANT}(\mathrm{THEY},\mathbf{HIRE}(\mathrm{Y}^{\beta}[\mathrm{INDEF};\ \mathrm{MOD}{:}\mathbf{SPEAK}(\beta,x)]))$

Applying this interpretation to the interpretation of the fragment produces a full CS representation for the fragment.

(65) **NOT**(**KNOW**(ME,$\mathbf{Q}^{\alpha}$(λx.**WANT**(THEY,**HIRE**(Y^{β}[INDEF; MOD: **SPEAK**(β,x)]) (α_{thing}))))))

This reduces to

(66) **NOT**(**KNOW**(ME,$\mathbf{Q}^{\alpha}$(**WANT**(THEY,**HIRE**(Y^{β}[INDEF; MOD:**SPEAK** (β,α_{thing})]))))))

which expresses the meaning 'I don't know the thing such that they want to hire someone who speaks it.'

In the case of VP ellipsis, the match is different. In the simple case, the target of the match consists of the parts of the antecedent that correspond to the parts of the elliptical sentence. E.g., in

(67) Brian speaks a Balkan language, but I don't.

The matches are given in (68). The Target is the match in the antecedent.

(68)

	Target	Fragment	
	Brian	I	[matching subjects]
	[*assertion*]	don't	[matching polarity]

On the basis of identifying the matches, which are determined by the syntactic structure, we can then identify the overt VP in the antecedent that provides the interpretation of the missing VP in the fragment, in this case *speak a Balkan language*. Again, this can be done by constructing an intermediate λ-expression and applying it to the matching constituents (Lappin 1996).

As Lasnik (2001) shows, there are even cases where VP ellipsis is unacceptable, but there is no island violation. The c-examples in (69)–(70) are less acceptable than the b-examples (the '*'s are Lasnik's).

(69) a. They said they heard about a Balkan language, but I don't know which Balkan language they said they heard about.
b. They said they heard about a Balkan language, but I don't know which Balkan language.
c. *They said they heard about a Balkan language, but I don't know which Balkan language they did.

(70) a. They studied a Balkan language, but I don't know which Balkan language they studied.
b. They studied a Balkan language, but I don't know which Balkan language.

c. *They studied a Balkan language, but I don't know which Balkan language they did.

Fox and Lasnik (2003) explore a derivational approach to these phenomena. The analysis is complex, so I will not try to replicate it here in its entirety. The central idea is that that sluicing is derived by long movement, that is, movement that does not leave intermediate traces. Without deletion, such a derivation would produce an island violation, because it would not be successive cyclic and would cross too many barriers, in the sense of Chomsky (1986a). (In a footnote Fox and Lasnik assume that all maximal projections are barriers, including VP.) But on the assumption that deletion repairs islands, sluicing is grammatical even though it involves long movement; the deletion in sluicing deletes the offending islands.

On the other hand, on their account VPE does not delete the offending island violation (the one caused by long movement) and thus long extractions with VPE do not improve acceptability.

This solution appears to be the closest counterpart in movement theory to the *Simpler Syntax* account. The two solutions have the following components.

	Movement account	Simpler Syntax *account*
1. Initial *wh*-	long movement of *wh*-	*wh*- in initial position in construction
2. Sluicing	delete IP under identity with antecedent IP	interpret *wh*- as scoping over interpretation corresponding to antecedent IP
3. No island violations in sluicing	deletion eliminates island violations	no island violations because no movement and hence no chain
4. Island-type violations in VPE (70)	deletion does not eliminate an island that arises from long movement out of CP	processing complexity in reconstructing the trace in the missing VP in order to bind the initial *wh*- and assign an interpretation

The key difference between the two approaches lies in points 3 and 4. It is not necessary to eliminate island configurations in *Simpler Syntax*, because they do not exist where there is no syntactic chain. The movement account relies on a syntactic constraint against long movement across barriers, while an account that follows *Simpler Syntax* relies on a processing account.

In a sense, chain complexity is the counterpart of the barrier to extraction that the movement account assumes, but the processing account fits in more naturally with approaches such as those cited in this chapter (e.g. Kluender 1998) that attribute a range of unacceptability judgments, including those arising from center embedding and syntactic islands, to processing complexity. As expected, such judgments are gradient as processing complexity varies. The account does not require stipulations such as "barriers" which have no independent motivation. They permit a simpler notion of grammaticality—it is simply well-formedness with respect to the linear and hierarchical arrangement of syntactic categories and morphosyntactic agreement.

Crucially, the two accounts can be tested empirically by constructing cases in which there must be long movement (on the movement account) but where VPE does not produce unacceptability. Across-the-board movement is such a case.

(71) Students at various levels are studying a Balkan language, but I don't know which Balkan language$_i$ the graduate students are studying t_i and the undergraduates aren't (['studying t_i']).

The syntactic structure in the two conjuncts is the same, and they share the same *wh*-phrase as head of the chain. In fact, when we reverse the conjuncts, the ill-formedness returns.

(72) ??Students at various levels are studying a Balkan language, but I don't which Balkan language$_i$ the undergraduates aren't (['studying t_i']) and the graduate students are studying t_i.

And if both conjuncts show VPE, then as expected the sentence is ill-formed in the same way that (61) is.

(73) *Students at various levels are studying a Balkan language, but I don't know which Balkan language$_i$ the graduate students are (['studying t_i']) and the undergraduates aren't (['studying t_i']).

Since it is simply the order in which the conjuncts are processed and not the syntactic configuration that determines the judgments, it appears that the unacceptability of the VPE cases such as (61) is not the consequence of an island violation.

The alternative to assuming that deletion is sometimes sensitive to syntactic complexity in the form of islands is to assume that the complexity of processing the antecedent in order to interpret the ellipsis is responsible for the appearance of an island violation. What we are experiencing, as suggested throughout this chapter, is not the island violation, but the complexity of building the chain. In this case, however, the chain does not connect parts of an actual sentence, but a reconstruction of part of an antecedent.

Let us consider how this works. When there is an A′ chain into an elliptical VP, we have to reconstruct the VP in memory on the basis of the antecedent, find a position for the gap in this virtual VP, construct the A′ chain, and construct the interpretation. The complexity of this process is apparent in examples that are far simpler than (61) and do not involve islands, as Lasnik (2001) and Fox and Lasnik (2003: 148) observe; cf. the starred examples in (69)–(70) and the even simpler cases in (74)–(77).

(74) a. Mary was talking to someone, but I forget who.
b. ?Mary was talking to someone, but I forget who she was ('[$_{\mathrm{VP}}$ talking to *t*]').

(75) a. Mary talked to someone, but I forget who.
b. ?Mary talked to someone, but I forget who she did ('[$_{\mathrm{VP}}$ talk to *t*]').

(76) a. Mary was talking to someone, but who?
b. *Mary was talking to someone, but who was she ('[$_{\mathrm{VP}}$ talking to *t*]')?

(77) a. Mary wants to talk to someone, but who?
b. *Mary wants to talk to someone, but who will she ('[$_{\mathrm{VP}}$ talk to *t*]')?

In the case of (61) basically the same reconstruction process has to take place. In order to reconstruct the missing VP, it is necessary to identify the part of the antecedent VP that corresponds to the gap that forms a chain with *which*, construct the chain, and interpret it. The VP in this case is (78).

(78) want to hire someone who speaks a Balkan language

In this case, the complexity factors that produce unacceptability in forming a chain into a relative clause are at play, rendering (61), the first example that we looked at, even more unacceptable than the examples (74)–(77).

On the other hand, a gap that causes local complexity as a consequence of a garden path should not be more problematic than these simple cases. This is because the sequential processing that explains the garden path effect is not at play when the missing VP is reconstructed on the basis of the antecedent VP.

(79) a. ?Who did Mary give *t* the money?
b. ?Mary gave somebody the money, but I forget who she gave *t* the money.
c. ?Mary gave somebody the money, but I forget who she did ('[give *t* the money]').

(80) a. *Who does Mary think that *t* will call?
b. ?Mary thinks that someone will call, but I forget who she does ('[think that *t* will call]').

(81) a. Who does Mary think that most likely *t* will call?
b. ?Mary thinks that most likely someone will call, but I forget who she does ('[think that most likely *t* will call]')

As far as I can tell, this prediction is correct: the unidentifiable invisible (and, I claim, non-existent) gaps in the absent VPs in (79)–(81) are no worse than the unproblematic (and equally non-existent) gaps in the absent VPs in (74)–(77). And the mitigating factor of the adverb effect in (81) does not seem to be active when the VP is elliptical.

Again, this result suggests that the unacceptability of the cases with overt gaps is due to online processing complexity, while the unacceptability of the elliptical cases is due to the complexity of reconstructing the VP, identifying the gap position, and linking it to the antecedent for the purpose of constructing the interpretation.

Consider, finally, how this approach deals with the coordinate examples discussed in connection with Fox and Lasnik's (2003) account. On a processing account, the presence of the verb in the first conjunct of (71) allows for construction of the CS representation of the question, where the trace corresponds to a variable bound by the interrogative operator. I.e. for *which Balkan languages graduate students are studying* we have

(82) $\mathbf{Q}^{\boldsymbol{\alpha}}(\lambda x.\mathbf{STUDY}(\text{GRADUATE-STUDENTS},x))(\text{BALKAN-LANGUAGE}(\alpha)) \Rightarrow$
$\mathbf{Q}^{\boldsymbol{\alpha}}(\mathbf{STUDY}(\text{GRADUATE-STUDENTS},\text{BALKAN-LANGUAGE}(\alpha)))$

Hence when *undergraduates aren't* is encountered in the processing of the second conjunct, it is possible to reconstruct the missing VP as the non-contrasting part of (82), namely

(83) $\mathbf{Q}^{\boldsymbol{\alpha}}(\mathbf{STUDY}(\text{X},[\text{BALKAN-LANGUAGE}(\alpha)))$

On the other hand, processing of (70) requires that *undergraduates aren't* be interpreted as (83). But the antecedent interpretation is that of *studying a Balkan language*, that is,

(84) $\mathbf{STUDY}(\text{X},\text{BALKAN-LANGUAGE})$

The processing necessary to get from (84) to (83) involves a number of steps, most notably transforming BALKAN-LANGUAGE into $\mathbf{Q}^{\boldsymbol{\alpha}}(\ldots, \text{BALKAN-LANGUAGE}(\alpha))$ and merging that with $\mathbf{STUDY}(\text{X},\text{Y})$. I attribute the unacceptability of examples such as the b-examples in (74)–(77) to the relative complexity of this reconstruction process.

5.6 Summary

The picture that I have presented in this chapter is one in which there is a sharp distinction between grammaticality and acceptability. Grammaticality is conformity to syntactic and morphosyntactic well-formedness conditions, formulated in terms of the primitives of grammatical theory. I hypothesize that the *Simpler Syntax* approach is sufficient for this purpose (see Culicover 2009 for a summary). Acceptability is tied to processing complexity: if processing is too complex, unacceptability arises. Ungrammaticality also produces unacceptability, hence it is difficult to tell the two causes apart on the basis of subjective judgments alone.

This view raises two further questions. First, why do languages differ in what is judged to be acceptable (cf. the difference between English and Swedish CNPC violations)? We might think that all languages should be the same, since the complexity of processing the same structure should be equally complex regardless of the language (other things being equal). Second, if the unacceptability of certain constructions is due to complexity, why do they not get better with practice in these cases over time?

The answer to both questions is essentially the same. Processing complexity is not something that we have control over, such that practice itself will make certain constructions easier to process over time. It is generally assumed that the processor is impenetrable to external influence (see e.g. Fodor 1983). But what can vary through experience is the threshold at which a certain level of complexity is felt to be unacceptable for a given construction. I hypothesize that the threshold is affected by frequency of use. If I am correct that complexity is "embodied" in the representation of knowledge of language, complex and infrequent configurations will produce very high surprisal, and hence strong unacceptability judgments, while equally complex but more frequent configurations will produce lower surprisal and hence weaker unacceptability judgments. So if a particular complex construction is used sufficiently often for a particular purpose in a group of speakers, it can achieve the status of acceptability while retaining its complexity. In this way we might begin to understand the results of work such as Featherston (2005a), Fanselow et al. (2005), and Fanselow and Frisch (2006), which suggest that complexity is present and apparently invariant across languages even when subjective acceptability judgments vary.

PART IV

Acquisition, change, and variation

This page intentionally left blank

6

Explaining complexity: the learner in the network

I turn in this chapter to the question of why complexity exists at all in language, given that there are strong pressures to reduce and ultimately eliminate idiosyncrasy in grammars. The answer given in this chapter is that complexity can arise as a consequence of language acquisition and can be sustained against the pressure to eliminate it by the structure of the social network in which language is situated.

As before, the constructional framework proves to be central to the explanation. I assume that learners are predisposed to acquire constructions and that they generalize them minimally beyond the available evidence. Competing constructions inhabit the social network, and even coexist in individuals. In competition, a less complex construction will drive out a more complex construction that has the same function, unless there is sufficient independent support for the more complex one. This support can take the form of high frequency of occurrence, limited contact between individuals with different constructions, or geographical isolation of the more complex variety. Similarly, instances of processing complexity can be acceptable in a given language in cases where they are strongly supported by high frequency of occurrence and lack of competition from less complex alternatives.

6.1 The paradox of complexity

To this point I have argued that the formal complexity of grammars and processing complexity both play a role in determining the form that language takes. The pressure to limit formal complexity means that specialized constructions cluster around the general constructional patterns of a language, and idiosyncrasies are by and large minimal deviations from the general. Avoidance of processing complexity has the consequence that the constructions that are most commonly used are the ones that minimize processing resources, and more complex constructions (and combinations of constructions) are relatively rare.

The question that naturally arises, then, is why formal and processing complexity persist at all. This is a question that has always bedeviled explanations for language change that invoke the reduction of complexity as the engine of change. Whether the pressure to generalize and to reduce complexity resides in the language learner, in the mechanisms that govern linguistic communication, or somewhere else, it would seem that over time the rough edges would be smoothed out by the pressure towards simplicity and generality.

In fact, it appears that the opposite is the case. Complexity is maintained in grammars, as evidenced by the review of constructional complexity in Parts 1 and 2 of this book. And a case can even be made for the growth of complexity in grammars. For instance, the ancestors of English and most of its relatives lack preposition, or P-stranding, as in *What are you looking at*. P-stranding is quite rare in the world's languages, so it is not unreasonable to assume that it adds to the complexity of the grammar. It would thus appear that English has grown in complexity, at least in this respect.

So there is a paradox: while we can point to cases of grammars changing in the direction of reduced complexity, complexity nevertheless continues to exist, and in some cases might be argued to grow. We have to address the question of where complexity comes from even as there is pressure to reduce it.

A familiar response to this paradox is that all languages are equally complex. In those cases where certain aspects of a grammar are simpler, other aspects of the grammar are more complex in compensation. For example, loss of inflectional morphology is conventionally held to be a simplification of grammar. McWhorter has in fact argued that creoles, which are extreme cases of loss of inflection, are "the world's simplest languages" (McWhorter 2001). At the same time, languages with no inflectional morphology regularly use word order to mark grammatical functions, while languages with rich morphology appear to have more flexible word order. It is conceivable that fixed word order introduces a degree of complexity that is not found with free word order. A language with fixed word order requires additional constructions to mark focus and topic, while languages with free word order can use a subset of the independently available orders for this purpose, since the morphology marks grammatical functions. So it might be argued that inflectional morphology and fixed word order bring equal complexity to the grammar.

Along the same lines, it can be argued that languages that have relatively simple syntax and morphology are correspondingly more complex in the representation of the form–meaning mapping. Assuming that meaning is fixed, a simpler form requires a more complex interpretative mapping in order to express the same meaning. For example, a language without subordinating conjunctions like *although* or *because* would have to have inference rules to convey these relations, given just concatenated clauses. On the other hand, if *although* or *because* are explicitly mentioned, no inference is necessary.

Unfortunately, comparisons such as the foregoing are inevitably impressionistic, since they do not invoke explicit measures of complexity. As noted by a number of the contributors to Sampson et al. (2009), the claim that all languages are equally complex is not one that can be effectively evaluated on the basis of the facts. There are so many components of grammar that can be compared, and so little agreement on what constitutes relative complexity across components, that virtually any story that one wants to tell can be told. For example, to say that inflection and fixed word order are equally complex assumes that we have a way of measuring the complexity of each one independently, but we don't. Similarly, it is not clear how to justify the claim that an explicit subordinating conjunction like *although* or *because*, and the associated syntactic structure, is as complex as inference rules that infer 'although' and 'because' on the basis of experience and belief.

Because of this imprecision I will focus in this chapter on two more specific questions. First, how in principle could a more complex construction be introduced into a grammar, other things being equal? Second, given that a grammar contains a certain degree of complexity, what factors would sustain it in the face of pressure to eliminate complexity?

6.2 Growing complexity

In this section I consider some circumstances under which simplification in one aspect of a grammar can at the same time produce additional complexity. I focus on changes that are arguably internal to the language and do not involve contact. For discussion of a range of cases in which contact between dialects leads to the creation of new dialects that are in certain respects more than their sources, see Trudgill (1989, 1992, 1995, 2004a, 2004b, 2009, 2011).

6.2.1 Infinitival relatives

Consider once again the analysis of the English relative constructions in Chapter 3. The goal of that analysis is to isolate the idiosyncrasies in these constructions against the backdrop of syntactic and semantic generalizations. The analysis locates idiosyncrasy precisely in the exclusion of infinitival relatives with non-PP constituents in initial position. The exclusion of these clause-initial relative markers in this case is a departure from complete uniformity.

The question is, how did this idiosyncrasy arise? A first step towards an answer is to observe that Old English infinitival relatives lacked relative pronouns entirely (Kemenade 1987: 151); they used either zero or the equivalent of *that*. When what was relativized was in a PP, the relative was a zero-relative and the preposition was stranded, e.g.,

(1)	Drihten,	þu	þe	gecure	þæt	fæt	on	to	eardienne.
	Lord,	you	yourself	chose	that	vessel	in	to	live
	(Blick 157/ Kemenade 1987: 151)								

It is therefore plausible to assume that infinitival *wh*-relatives are an innovation. The innovation is in the direction of increasing uniformity, by extending *wh*-relatives from the tensed to the infinitival cases. Interestingly, the OE tensed relatives did not permit P-stranding with a clause-initial *wh*-; only clause-initial pied-piped relative PP is possible (van Kemenade 1987: 152–3). So the generalization can be seen to be based on the earlier pattern. However, with this first extension of the pattern from finite to infinitival relatives, the generalization of clause-initial relative NP has not occurred, uniformity is not yet achieved, and so strain, in the sense of Chapter 2, is introduced into the grammar.

If such a story is correct, it might lead us to assume that there are two forces, one working to increase uniformity and the other to reduce complexity (in this case, strain) by eliminating an idiosyncrasy. Furthermore, in this particular case the strength of the pressure to increase uniformity is greater than that working to eliminate strain. We know this because of the success of the change.

But it is in fact entirely natural to conceive of the competition in terms of a single force in the direction of reducing strain, but operating at several levels at the same time. In this particular case, for example, generalization of pied-pieded relative PP to all relatives reduces the exceptionality of infinitival relative clauses. Another generalization, not yet realized in English, is the full extension of the clause-initial relative forms to infinitivals, which would permit *a man who to talk to* and the attested *the person who to contact*.

Is it then possible to quantify the extent of the individual pressures, at least in relative terms? It is in fact possible to show that the change that is apparently happening subsumes the change that is not happening, that is, the existing idiosyncrasy. Consider that we have two variables, one ranging over the type of the clause (Tensed/Infinitival Relative), and one ranging over the clause-initial syntax (zero/PP/NP). The distribution of the alternatives at the OE stage is given in Figure 6.1. A shaded cell indicates that the property is possible. The diagonal striping in one cell marks the change to come.

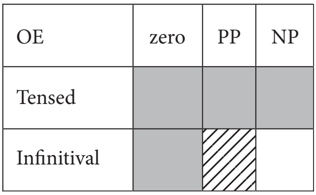


FIGURE 6.1. Extraction possibilities in OE

It is clear that extending the infinitival case to PP increases uniformity, by filling a previously empty cell. It makes the zero infinitival relative less idiosyncratic. But it is also clear that this extension now isolates the NP cell as an idiosyncrasy.

Thus with the extension to PP we have greater uniformity from the perspective of 'relative clauses,' but less uniformity from the perspective of 'infinitival relative clauses.' The situation is unstable, because of the strain. In principle it could resolve in one of two directions to achieve full uniformity. The most likely possibility, given the possibility of wh-NP in initial-position infinitival questions (i.e. *wonder who to talk to*), is that the last cell is filled in over time. But a logical possibility would be that wh-PP infinitivals and zero infinitivals are lost, which would move English back into the group of languages that lack infinitival relatives entirely.

The widespread occurrence of infinitival relatives and infinitival questions in English counts as support for the infinitival construction in general, and acts as a barrier against the change back to the entire absence of infinitival relatives. Whether the extension to NP actually will occur and if it does, how soon it will occur, is difficult to say without knowing how much resistance there is against eliminating the gap in the paradigm marked by the empty cell in Figure 6.1. At this point I am unable to say how this might be measured, although at this point I see no reason why the elimination of the gap should not ultimately occur.[1]

6.2.2 Preposition-stranding

As noted above, P(reposition)-stranding is another candidate for the emergence of complexity, as contrasted with the reduction of complexity. Hawkins (2004: 204–5) explains the rarity of P-stranding in processing terms as follows. In left-to-right processing of the A′ chain, the processor seeks to resolve the

[1] At the same time, it is entirely possible that extension to infinitival relatives has been followed by loss of infinitival relatives that occurred in other languages. The most likely trajectory is one where the zero infinitival relatives emerged briefly, and then were lost. Whether there is documentary evidence for such a sequence of changes is a question for future research.

A′ constituent as quickly as possible. Thus it posits a trace wherever it can. In the case of P-stranding, such a position exists immediately after the verb in examples such as (2).

(2) [Which student]$_i$ did you ask ([t_i]) John about [t_i].

However, linking the A′ constituent to the first (potential) gap is a mistake, which has to be repaired subsequently so that the A′ constituent can be properly linked to the gap in the PP. On the other hand, if there is no P-stranding but rather pied-piping, this problem does not exist, on Hawkins's account—there is only one position for the gap corresponding to the PP.

(3) [About which student]$_i$ did you ask John [t_i].

Explanations of this sort have a decidedly "just so" character, since they fail to consider all of the possibly countervailing forces that could just as easily be invoked to account for alternatives that do not happen to exist. For example:

- There are many cases in which the verb is intransitive, e.g., *Which student$_i$ are you talking about t_i?* Why did English not evolve P-stranding for just these cases, which do not suffer the processing problem of (2)?
- In V-final languages, e.g. Old English as illustrated in (1) above, there is no opportunity for the processor to misparse the A′ constituent based on the verb, since the preposition and the gap precede the verb. While it is possible that an NP in A′ position triggers the search for a gap immediately, without the verb the chain cannot be interpreted. One could argue that there is a better chance of interpreting the chain when the preposition is encountered, because it provides an interpretation for the A′ before the verb is processed. So one might expect P-stranding to be more common in V-final languages than in V-initial languages, which is not the case.

Newmeyer (2007: 113–16) argues that while P-stranding is rare, it is simpler than pied-piping. His argument turns on the observation that in order to have P-stranding it is necessary to allow P to govern the trace of extraction, thus grouping P with the other lexical categories. This is more general than restricting trace of extraction so that it occurs only in the context of N,V, and A, but not P.

The difficulty with this argument is that N and A themselves do not license extraction. It is possible to extract out of NP and AP only by stranding a preposition (*what did you read a book *(about)*; *what are you angry *(about)*). The generalization then is actually between extracting NP arguments of V, or extracting NP arguments of V and P. The latter may be somewhat more general if we consider V and P to be members of a more general category (Chomsky

1972; Jackendoff 1977). While they are similar in some respects (e.g. they express relations), they are very different in others (e.g. verbs are inflected but prepositions are not), and it is not clear on what basis there would be a generalization of V to P. So we cannot rule out the possibility that there are ways in which P-stranding introduces complexity.

Nevertheless, it is clear that P-stranding allows for a generalization of A′ chains whose gaps occur in NP positions. In order to explain why English allows for P-stranding but other languages do not, we must first understand what it is about English that makes this generalization possible. A hint comes from the history of P-stranding. In Old English, as observed above, P-stranding was possible first in finite relative clauses introduced by *þe* or *þæt* 'that' (Allen 1980a: 224–36). Somewhat later, in the twelfth century, A′ chains could be headed by *wh*-pronouns; in late Middle English, P-stranding was extended to passives; and in Modern English it is fully generalized. Moreover, in Chaucer the only preposition that is stranded in A′ chains headed by *wh*-pronouns is *inne* 'in.'

The question, then, is why P-stranding occurs at all in OE. According to Bennis (1986: 275ff.), OE allowed postpositional constructions with a particular set of pronouns, which could be extracted from the phrase and be situated leftmost in an embedded clause. Here are some examples:

(4) a. . . . ðæt hie ðær mehten betst frið binnan habban
that they there might best security within have
'that they might have the best security in there'
(Oros. p.116.6/Allen 1980b: 102)

b. . . . gif ðær gebedo æfter fylgeað
If there prayers after follow
'if prayers follow after that'
(CP. 399.33/Vat 1978: 35)

Bennis suggests (p. 280) that stranding in relative clauses results from movement and deletion of one of these pronouns. We can recast this phenomenon in constructional terms, as in (5) ('PRO' is a CS argument whose reference is determined contextually).

(5) SYNTAX $[_S$ ðær ... $[_{PP}$P]]

CS λx. ... **P**′(x)(PRO)

The emergence of P-stranding in relative clauses can then be seen as a generalization of this construction to the case in which the antecedent of the gap is not a pronoun, but *þe* coindexed with the head of the relative clause, as shown in (6).

(6) SYNTAX $[_S\ \{\text{ðær}|\text{Þe}\}\ \ldots\ [_{PP}\text{P}]]$

CS $\lambda\text{x}.\ \ldots\ \mathbf{P}'(\text{x})(\{\text{PRO}|\alpha\})$

The correspondence for *þe* is essentially the correspondence for zero relatives in Modern English given in Chapter 3.

The emergence of P-stranding occurs in a V-final context, which as noted is not subject to the processing considerations proposed by Hawkins. In fact, Bennis suggests that the postposition construction from which P-stranding originally developed is characteristic of a V-final language, e.g. Dutch. But it is not certain that P-stranding can emerge only in a V-final language, and in fact there are many V-final languages in which it is impossible (Bayer 1995). It occurs in Swedish, Danish, and Norwegian, and in French dialects (King and Roberge 1990; Abels 2003). Koopman (1984: 49–56) claims that Vata and Gbadi may have P-stranding, and neither language is V-final. P-stranding is possible in main clauses in Jacaltec (Craig 1977), which is VSO and strictly prepositional.

The other side of this issue concerns the prominence of pied-piping. Since pied-piping preserves the phrase, its essential grammatical properties are available as soon as it is encountered in processing, and its thematic properties can be hypothesized. For example, the PP *to whom* can be tentatively identified as denoting a Goal simply on the basis of *to*, although this guess might have to be revised in light of subsequent lexical information indicating that the phrase is part of an idiom, e.g. *the person to whom I took a liking*. In contrast, a bare NP or *that* in initial position in a relative clause conveys no information about the grammatical function of the chain that is to be formed.

In the case of P-stranding, then, we appear to have a competition between interpretation of A′ chains on the one hand, and generality of the chain construction that licenses gaps on the other. In such a competition, the ultimate outcome may depend not on the linguistic content, but on how each is situated in the social network. As I suggest in §6.3.2 below, and in Chapter 8, an innovation may receive support from the configuration of the social network even in contexts where there are forces working against it. Whether this is what happened in the case of English P-stranding is difficult to demonstrate conclusively, but the scenario sketched out here is a plausible one.

There is an additional competition here that may be relevant as well. Working within the *Conditions* framework of Chomsky (1973), van Riemsdijk

(1978) observed that as a maximal projection, PP should block extraction, because of the subjacency condition. Simplifying somewhat, subjacency rules out a chain that crosses a maximal projection. Van Riemsdijk proposed that extraction from PP is permitted just in case there is an 'escape hatch' in the phrase, typically a specifier position such as is found in NP and S in English. (7) illustrates. While the configuration in (7a) is ruled out by Subjacency, the one in (7b) is not.

(7) a. $^*[\mathrm{NP_i} \ldots [_{\mathrm{PP}}\ \mathrm{P}\ t_i]]$
b. $[\mathrm{NP_i} \ldots [_{\mathrm{PP}}\ [_{\mathrm{Spec}}\ t_i\ ]\ \mathrm{P}\ t_i]]$

Van Riemsdijk's proposal is an extension of Chomsky's approach to markedness (discussed in §1.3.1). Again simplifying somewhat, extraction from a maximal projection is not possible unless explicitly permitted in the grammar. Requiring the presence of a specifier position in all 'long' extractions is a way of notating this possibility using the syntactic structure itself.

If this general scenario is on the right track, it is plausible to treat reduction in complexity in the case of P-stranding as not a matter of extending an A′ construction such as Wh-question or Relative to a broader set of syntactic environments. Rather it is a matter of first extending the set of environments that may contain a gap, and then extending the set of environments that may contain a lexical head that lacks an overt complement. The latter permits a characterization of constructions such as those illustrated in (8), without requiring appeal to movement. The generalization of P-stranding in English is by now virtually complete, so that it comprises not only zero relatives (8a), but *wh*-relatives (8b), *wh*-questions (8c), pseudo-passives (8d), tough-movement (8e,f), comparatives (8g), nominalizations (8h), *worth* (8i), and *take* (8j).

(8) a. the painting I was looking at
b. the painting which I was looking at
c. What were you looking at?
d. This painting has rarely been looked at.
e. This painting is fun to look at.
f. Terry is a tough person to look at.
g. We have more paintings than we could ever look at.
h. This painting can't stand that much looking at.
i. This painting isn't worth looking at.
j. This painting will take a microscope to look at.

This state of affairs can be understood in constructional terms as the dissociation of the context in which the gap may appear from the construction in which the gap appears. That is, $[_{\mathrm{PP}}\ \mathrm{P}]$ is possible even when there is no A′

chain. (I assume, following *Simpler Syntax*, that there is no gap in the passive, hence no chain.)

It is plausible that any construction that licenses an A′ chain headed by NP will extend to NP complements of P, other things being equal. Assuming that this is the case, the complexity of P-stranding depends on the cost of locating the gap of the chain in PP. There is ample evidence that chains that extend into maximal projections are more complex than those that do not. Kluender (1998) found additional processing complexity across an S boundary. There are many languages, including varieties of German and Russian, that do not permit long-distance extraction. The familiar island constraints of Ross (1967), Chomsky (1973), Chomsky (1986a), etc. involve chains into maximal projections. One way to understand the emergence of P-stranding is to take the prohibition against chains into maximal projections to be the grammaticalization of processing complexity. In the absence of exceptions, it is an absolute prohibition. The pressure to generalize NP-headed A′ chains pushes in favor of reduced formal complexity, while the constraint against chains into maximal projections pushes in favor of reduced processing complexity. If the language changes in such a way as to permit chains into certain maximal projections, e.g. S, then the prohibition is weakened, and the possibility exists that it can be further weakened to allow chains into NP, AP, and finally PP.

6.3 Maintaining complexity

Let us turn now to the question of how complexity is maintained even when there is pressure to eliminate it. One instance of such a situation is a construction that is arguably ruled out in one language because it exceeds some threshold of complexity, but is judged to be acceptable in another language. One example already discussed in Chapter 5 is that certain sentences that violate the Complex NP constraint in English are judged to be grammatical in Swedish (Allwood 1982; Engdahl 1997). Another is the occurrence of considerable morphological complexity in the face of fully regular paradigms, found in many languages.

I propose here that there are countervailing forces in the social environment of a language that may prevent a complex construction from being eliminated in favor of a less complex construction. The essential assumption that underlies the account is the following: Learners are capable of acquiring a construction of arbitrary complexity, within the bounds permitted by such resources as memory. I argue for this assumption in §6.3.1.

Whether the construction is maintained is then a function of three factors: how complex it is, the extent to which alternative constructions are preferred, and the amount of evidence in the linguistic environment that supports it.

I illustrate the dynamics of construction distribution through a population using a computational simulation in §6.3.2.

6.3.1 Learning constructions

The classical view of language acquisition in generative grammar is that much of what is known by a learner is already present in the learner prior to exposure to linguistic input. In Principles and Parameters theory it was assumed that certain contingent knowledge is a matter of setting parameters which fix certain variable properties of language, such as the linear order of the head of a phrase and its complements. On this view a very common construction such as a *wh*-question with a phrase in A′ position is a component of UG and does not have to be learned. What does have to be learned is whether such a construction is part of the grammar of a given language, since not all languages mark *wh*-questions with A′ chains. One might imagine that this is a simple task, where the setting or the parameter is triggered by exposure to actual instances of *wh*-questions with *wh*-phrases in A′ position.

But parameter setting is inadequate as a realistic account of language acquisition for several reasons. First, it fails to account for the actual richness of variation among languages, unless one assumes arbitrarily fine-grained "micro-parameters," a step that reduces the theory of parameters to vacuity. Second, it does not appear to be able to provide an explanatory account of the course of acquisition. For instance, there is no evidence that learners erroneously set the value of a parameter even though in many if not most languages, there is evidence for both the correct and incorrect parameter values in the actual linguistic input. Third, even the most fundamental parameter settings do not appear to actually account for the grammatical state of affairs in many languages. And fourth, there is a paradox: it is not clear how a learner can assign a structure to a triggering input, e.g. an example of a *wh*-question, and identify it as evidence for a particular parameter setting, without already having acquired the construction so that it can be recognized for what it is. For more detailed discussion of all of these points, see Culicover (1999) and work cited there.

The natural alternative, also argued for on conceptual grounds in Culicover (1999), is a constructional one. An account of how language acquisition proceeds in terms of constructions is proposed by Tomasello (2000, 2003). Tomasello proposes that young children's early grammars are composed of highly idiosyncratic constructions formed around single lexical items or restricted sets of lexical items. Only later are these constructions generalized in terms of broader syntactic categories, so that they begin to look like what are conventionally thought of as "rules" of grammar.

On this view, then, the learner proceeds without any prior knowledge of what the grammatical constructions or rules of the language are beyond whatever constraints might be inherent to the form–meaning correspondence. Each one is constructed by the learner on the basis of the actual input. Even very regular patterns that might be characterized as "rules" in classical grammatical theory are bundles of more specific constructions, defined in terms of sets of lexical items. An open question on this approach is how, if at all, the learner manages to generalize beyond the specific input to which it has been exposed. Another is how learners get from local constructions to constructions that license unbounded dependencies. But even if we lack a clear and compelling answer to these questions, it is still plausible that early learning proceeds on the basis of the formation of highly specific constructions.

A computational simulation of constructional learning along these lines is described in Culicover and Nowak (2003) and Culicover et al. (2008). The simulation illustrates how complex constructions can be readily acquired as long as the linguistic environment does not provide evidence for more general constructions that carry out the same function. I summarize here the main points of this simulation.

The objective of this simulation is to explore the interactions between three main factors in the theory of language acquisition and the effects of these factors on the course and content of acquisition: (i) the computational capacities of the learner and its prior knowledge about language and grammar, (ii) the input to the learner, and (iii) the target of learning. The outcome of the simulation is an explicit representation of the knowledge of the simulated language learner that can be examined in detail, and compared with what human learners are presumed to know.

We make minimal assumptions in our simulation about each of (i)–(iii). In the case of (i), prior knowledge, we assume that the learner has no knowledge of particular grammatical categories, linguistic structure, or grammatical principles. The learner has only the capacity to extract correspondences between form and meaning based on the statistical properties of the linguistic input, to form categories based on similarity of distribution, and to form limited generalizations. As suggested in Chapter 5, just as some syntactic universals might emerge from the form–meaning correspondence, it is plausible that certain aspects of syntactic structure in the learner may be reflections of corresponding structure in CS. There are, however, aspects of syntactic structure that cannot be reduced to the correspondence, and must be learned on the basis of the linguistic input.

In the case of (ii), the input, we assume that the learner is presented only with pairs consisting of forms (including phrases and sentences) and their corresponding meanings. In the case of (iii), the target of learning, we assume that the target of learning is not a grammar in the sense of Mainstream Generative Grammar (in any of its manifestations), but a set of constructions that is compatible with what a native speaker knows about the language, but not necessarily one that is compatible with any particular linguistic theory.

Clearly, these assumptions are in many respects too strong, and a realistic account of how language is acquired will have to elaborate (i)–(iii) in many ways. The objective of the simulation is not to demonstrate that the strongest form of these assumptions is correct, but to determine in exactly which ways they are too strong. Moreover, it is of some interest to discover how far a minimalist language learner constructed along these lines can get, even given these very minimal assumptions.

Our simulation is called ***CAMiLLe***, for Conservative Attentive Minimalist Language Learner. ***CAMiLLe*** is exposed to sets of form–meaning pairs, e.g.,

(9) `house = HOUSE($TYPE:BLDG)`
`see the house ? = YNQ(SEE($EXP:YOU,$THEME:HOUSE ($TYPE:BLDG)))`

I leave open the question of how a learner knows what meaning a given expression is to be paired with, and simply assume that this question has a resolution in developmental cognitive psychology; see for example Bloom (2000).

On the basis of each form–meaning pair, ***CAMiLLe*** attempts to formulate one or more correspondence rules. It is important to point out that ***CAMiLLe*** proceeds from the assumption that strings of words and their corresponding meanings are organized according to heads and non-heads (dependents or adjuncts). Our simulation experiments have suggested that without this assumption, ***CAMiLLe*** cannot acquire even minimal linguistic knowledge. But even with this assumption, since there is no overt connection between the individual words and the individual meanings, ***CAMiLLe*** is prone to making many bad rules, as well as correct rules. For example, after having encountered the sentence/meaning pairs in (10)—

(10) `Ted is nice. = BE($THEME:TED, $PRED:NICE)`
`Ted is small. = BE($THEME:TED, $PRED:SMALL)`

—***CAMiLLe*** has enough information to guess that *Ted is* means `TED`, or `BE`, or `BE($THEME:TED)`. ***CAMiLLe*** keeps track of the weight of the evidence that supports each hypothesis, so that after enough experience, the diversity of exemplified correspondences continues to support the first hypothesis, but not the other two. This experience provides evidence that *Ted* corresponds to `TED` and *is* corresponds to `BE`. The evidence is purely statistical; the rules that are not supported remain in ***CAMiLLe*** but gradually get pushed out by rules that are more strongly supported by the evidence.

If ***CAMiLLe*** finds that two rules have a similar form, then to the extent possible it forms a cluster (i.e. a mini-category). For example, if ***CAMiLLe*** has strong evidence for the following two correspondences—

(11) `ted is <=> BE($THEME:TED)`
`sally is <=> BE($THEME:SALLY)`

Then ***CAMiLLe*** will form a correspondence rule of the form

(12) `[ ted;sally] is <=> BE($THEME:[ TED;SALLY] )`

Again, like the assumption of headedness, the assumption that there are larger categories does not come for free—it is explicitly built into the computational apparatus of ***CAMiLLe***.

Clearly, correspondences such as the ones illustrated here are not equivalent to rules of grammar in the traditional sense. For one thing, they are much too specific—they do not mention broad categories but simply clusters of individual elements. For another, they provide information only about the linear order of elements, not structure.

However, these correspondences are constructions in the sense that I have been using the term in this book, in that they describe form–meaning correspondences. They do not differ from more general constructions except in degree of specificity. While the more general constructions can and arguably should be characterized in terms of classes of elements and categories, the early constructions formulated by ***CAMiLLe*** are simply sets of words that appear to have similar distributions and functions.

It is at least plausible that what ***CAMiLLe*** comes up with is comparable in some important respects to what an early language learner comes up with, prior to the point at which generalization and the formation of large-scale categories and correspondences kicks in, not to mention the recognition of discontinuous dependencies. A hypothesis (suggested by Tomasello 2003 that we arrived at independently through our own preliminary experiments with ***CAMiLLe***) is that first there is a pre-grammatical stage, which is modeled by ***CAMiLLe***, followed by a grammatical stage. In the pre-grammatical stage we expect to see everything treated as though it is a idiomatic construction. In the grammatical stage, we expect to see those aspects of the language that are fully or almost fully regular to be reflected in significant generalizations that go beyond the learner's actual experience. Those aspects of the language that retain some significant idiosyncrasy, e.g. constructions of the sort noted in

```
house = HOUSE($TYPE:BLDG)
see the house ? = YNQ(SEE($EXP:YOU,$THEME:HOUSE($TYPE:BLDG)))
mary = MARY($TYPE:PERSON)
here's mary = $POINT($THEME:MARY($TYPE:PERSON))
see mary ? = YNQ(SEE($EXP:YOU,$THEME:MARY($TYPE:PERSON)))
john = JOHN($TYPE:PERSON)
see john ? = YNQ(SEE($EXP:YOU,$THEME:JOHN($TYPE:PERSON)))
here 's a flower = $POINT($THEME:FLOWER($TYPE:PLANT))
see the flower = $IMP(SEE($EXP:YOU,$THEME:FLOWER($TYPE:PLANT)))
here's a boy = $POINT($THEME:BOY($TYPE:PERSON))
see the boy = $IMP(SEE($EXP:YOU,$THEME:BOY($TYPE:PERSON)))
horsie = HORSE($TYPE:ANIMAL)
see horsie ? = YNQ(SEE($EXP:YOU,$THEME:HORSE($TYPE:ANIMAL)))
a baby ! = BABY($TYPE:PERSON)
see the little baby ? = YNQ(SEE($EXP:YOU,$THEME:BABY($TYPE:PERSON)))
nice baby ! = BABY($TYPE:PERSON)
talk to the baby = $IMP(TALK($AGENT:YOU,$GOAL:BABY($TYPE:PERSON)))
talk to mary = $IMP(TALK($AGENT:YOU;$GOAL:MARY($TYPE:PERSON)))
i see mary = SEE($EXP:ME,$THEME:MARY($TYPE:PERSON))
i am talk ~ing to may = TALK($AGENT:$ME;$GOAL:MARY)
do you see john ? = YNQ(SEE($EXP:$YOU,$THEME:JOHN($TYPE:PERSON)))
```

FIGURE 6.2. Sample input to *CAMiLLe*

earlier chapters, are retained in their pre-grammatical form, to the extent that the evidence to the learner is strong enough to maintain them (see §6.3.2).

Using constructed input we are able to test *CAMiLLe*'s ability to deal with a particular grammatical phenomenon. *CAMiLLe* requires a certain amount of exposure to a grammatical phenomenon in order to form a reasonably informed hypothesis about it. A file of naturally occurring, e.g. transcribed speech to children from the CHILDES database (MacWhinney 1995), in general does not provide enough instances of a given phenomenon,[2] and exposing *CAMiLLe* to composites of files, while potentially useful, does not allow us to focus on specific grammatical phenomena. So we constructed files by hand. An example of a constructed input file is given as Figure 6.2.

[2] It is of course an empirical question whether for any given grammatical phenomenon, the naturally occurring data taken as a whole provides sufficient evidence for a learner. If it does not, then this is an argument (from poverty of the stimulus) for innateness. The sorts of things that we are interested in here are those that are not universally found in languages of the world, and therefore we may presume that they are learned on the basis of evidence in the linguistic input to the learner. I think that this is the simplest way to understand such variation, but it is not possible to rule out other possibilities, such as universals that are not present unless triggered by experience, differential maturation, and so on.

The purpose of this particular file is to try to get ***CAMiLLe*** to correlate individual words with their meanings, and to correlate position in the string with semantic role. (The roles used here are Theme, Exp(eriencer), and Agent.) The output after processing ten sentences consists of 103 rules, many of them overlapping, and many of them highly idiosyncratic but low in weight. Lack of space precludes listing all of the rules here—I show a few "correct" rules in (13) and a few "incorrect" rules in (14).

(13) **5. [89]** `MARY <=>` mary
6. [82] `JOHN <=> john`
16. [13] `YNQ(*NULL* :SEE) <=> 1.see 3.?`
21. [10] `$IMP(*NULL* :SEE) <=> 1.see`
27. [9] `SEE($THEME:[ JOHN; MARY;] ) <=> see+1->[ john; mary;]`
95. [2] `$POINT($THEME:MARY) <=> 1.here' s 2.mary`

(14) **1. [172]** `SEE($EXP:YOU) <=> 1.see`
23. [10] `YNQ(*NULL* :SEE) <=> 1.see 2.the 3.?`
72. [2] `BOY <=> 3.boy | a+1->boy see | see X boy`
103. [2] `FLOWER <=> 1.here 2.' s 3.a 4.flower |1.see 2.the 3. flower`

The "correct" rules and the "incorrect" rules all reflect how ***CAMiLLE*** responds to the actual data. For instance, (14.1) reflects the fact that *see* in initial position correlates highly with the meaning `SEE($EXP:YOU)`. This is because there are a lot of sentences beginning with *see* (imperatives and questions) in which the subject is not expressed overtly. We may take this to be a very early stage of development, in which the learner has not yet determined that such sentences have a missing subject; such a determination can be made when the learner recognizes that all sentences of English have subjects.

More strikingly, (14.23) shows that in the limited input data, *see the* correlates highly with the interrogative of *see*. This is an artifact of the particular dataset, and is not an error on ***CAMiLLe***'s part, but a correct (but transitory) hypothesis under the circumstances. Similarly, ***CAMiLLe*** finds evidence to form correspondences between the meaning `BOY` and the word *boy* in third position, *a boy*, and *see... boy*. A more diverse set of experiences will disabuse ***CAMiLLe*** of these errors, or at least it should if we have designed the simulation correctly. And it is possible, although difficult to determine experimentally, that actual learners may form such incorrect, yet fleeting, correlations in the early stages of learning.[3]

3 An interesting case of such a fleeting construction is documented by Hamburger (1980).

Rule (13.16) reflects the fact that *see* is used as an interrogative (with '?' in third position in the sentence—an artifact of the input data). Similarly, (13.21) reflects the imperative case. (13.27) indicates that ***CAMiLLe*** has identified *John* and *Mary* as elements that have the same distribution (with respect to the Theme of *see*). This observation may, if we wish, form the basis for a generalization that *John* and *Mary* have the same distribution with respect to everything, although we will want to exercise caution in formulating the rule of generalization. Finally, (13.95) is a small construction, correlating *here's Mary* with pointing to Mary.

Although constructed data is easiest to control for experimental purposes, ***CAMiLLe*** was also presented with caretakers' speech to children taken from the CHILDES database.[4] A sample of the results of processing this input is given in (15).

(15) **1. [59152]** `BE($THEME:[ HE; IT; THAT; THIS; WHAT; WHERE; WHO;] ) <=>[ he; it; that; this; what; where; who;] +1->is`
6. [38487] `[BED; BOOK; BUG; BUNNY; CHAIR; COOKIE; CRAYON; DUCK; IT; PICTURE; THAT; THIS;] ($REF: [ $DEF; $INDEF;] ) <=> [ a; the;] [ ?; bed; book; bug; bunny; cookie; duck; picture;]`
15. [12036] `[KNOW; LIKE; SEE; THINK; WANT;] ($EXPERIENCER:YOU) <=> you[ know; like; see; think; want;]`
62. [1491] `IMP([ GET; GIVE; LIKE; PUT; SAY; SEE;] ) <=> 1.[ get; give; like; put; say; see;]`
155. [372] `[WHAT; WHERE;] ($REF:$WH) <=> 1.[ what; where;] is | [ what; where;] +1->is`
163. [353] `PLAY_WITH <=> play+1->with`
195. [266] `YNQ(*NULL*:WANT) <=> do X want`
228. [227] `WANT <=> want+1->to`
590. [43] `NEG(*NULL*:BE) <=> 2.is 3.not`

Rule (15.1) characterizes a construction in which there is a pronominal Theme of *be*. (15.6) shows that *a* and *the* correlate with the features [DEF] and [INDEF] on nominal concepts. This may ultimately form the basis for a more general rule NP → Det-N, although getting there requires additional

[4] ***CAMiLLe*** learns by determining correspondence rules mapping form and meaning. Since the sentences in the CHILDES database do not have meanings associated with them, in order to use those sentences as input to ***CAMiLLe*** it is necessary to provide them all with meanings. Doing so manually is prohibitively labor-intensive. What we did was parse the sentences with a fast (but inaccurate) parser (Mini-par), translate the output of the parser into rudimentary meanings, present the resulting sets of sentence/meaning pairs to ***CAMiLLe*** and ask ***CAMiLLe*** to figure out the correspondence rules. I am grateful to Anna Feldman for implementing the translation.

generalization. (15.15) correlates *you* before a set of verbs expressing knowledge and perception with the experiencer role of these verbs. And so on for the rest of the rules shown. Each one indicates that ***CAMiLLe*** has correctly extracted some correspondence, one that is specific to lexical items. This is how we characterize the most specific end of the constructional correspondence spectrum. Hence ***CAMiLLe*** appears to be capable, in principle at least, of carrying out the preliminary work of forming correspondence rules.

This simulation shows that if a learner takes every correspondence to be an individual construction, that construction will persist in the learner's grammar unless it is replaced by a construction that generalizes one or more of the terms. The constructions in (15) show generalization of the distributional properties of individual lexical items such as the pronouns (`[he; it; that; this; what; where; who;]`), verbs with Experiencer subjects (`[know; like; see; think; want;]`), nouns (`[bed; book; bug; bunny; cookie; duck; picture;]`) and so on. The formation of categories is the first step in generalizing constructions beyond pure idioms.

CAMiLLe was also able to determine that certain linear orderings correspond more generally to CS interpretations. Rule (15.155) identifies sentences beginning with *what/where is* as *wh*-questions, and (15.195) identifies inversion with *do* as the signature of a question with the verb *want*. These rules are the result of ***CAMiLLe*** first forming individual constructions for each sequence of forms, and then extracting regularities of the form–meaning correspondences. Additional details and limitations of such generalization are reported on more fully in Culicover et al. (2008).

To summarize, the simulation illustrates how constructions of arbitrary complexity can be acquired by learners, and how these constructions will persist in grammars unless they are preempted by evidence for greater generality. The question that I now turn to is how complexity can resist preemption when there is evidence in the linguistic input for a less complex construction that performs the same task.

6.3.2 Constructions in contact: viruses in the body linguistic

It is commonplace in contemporary theoretical linguistics to think of a language as the linguistic knowledge in the head of a native speaker, that is, the grammar. This is Chomsky's notion of I-language (Chomsky 1986b: 21–4).

Another useful way of thinking of a language is that it is the sum of all of the rules (or constructions) in the heads of all speakers in a language community. Thinking of a language in this way is somewhat more complicated because unlike the notion 'head of a native speaker,' the notion of a 'language community' is at first glance somewhat less well-defined—at least its boundaries are less clear. However, it has long been recognized (see e.g. Chomsky 1965: chapter 1) that the former notion is by no means straightforward either, and in order to work with it it is necessary to idealize considerably. The classical idealization abstracts away from the variation that is inevitable in the native speaker's environment due to the fact that s/he interacts with multiple individuals in the course of learning and communicating, while each speaker interacts with a different set of individuals.

Since each individual's knowledge of language is different from every other individual's knowledge of the same language, the linguistic experience of everyone will be somewhat different. The differences may have to do with the linguistic substance, with the social contexts in which various constructions are used, with the frequency with which various constructions are used, and so on.

In a constructional grammar, the constructions are individuated. A speaker may have one construction for expressing a particular meaning, or several. Learners acquire language by being exposed to correspondences that are covered by particular constructions, and by hypothesizing these constructions or variants of these constructions. So, intuitively, constructions spread through the lexicons of individuals, and they are transmitted from one speaker to another through contact, similar to the way viruses are transmitted through the population; in a sense, they are 'viruses' in the 'body linguistic,' understood both as the linguistic individual and the linguistic social group.

The epidemiological analogy is in fact a very instructive one (see Enfield 2008 for appeal to this analogy with respect to non-grammatical phenomena). Like a virus, a construction resides in an individual and its presence in that individual may or may not be expressed overtly. The construction may be latent, as a consequence of the individual's exposure to it, but not so strong that the individual will actively use it. "Expression" of a construction corresponds to its use in communication—in the analogy, this would correspond to the construction/virus reaching a certain threshold of activation or "infection," at which point the symptoms appear.

The analogy breaks down somewhat when the construction/virus is below the threshold. In the linguistic case, an individual may accept or recognize the construction, but might not produce it, and therefore not influence others. But in the case of the virus, the individual may transmit it to others without showing symptoms of the infection.

Finally, individuals may be naturally more or less resistant to constructional infection, due to individual psychological or social factors; this corresponds to susceptibility of being influenced by the language of others. By assumption, children are particularly susceptible to constructional infection. But adults may vary in their susceptibility and some may be quite resistant.

Just as it makes sense to study the state and growth of the virus in the individual (virology) or in the social group (viral epidemiology), so it makes sense to study the state and growth of the construction in the individual and in the social group. In both cases, it is impossible to fully understand how the individual reaches its actual state without understanding the dynamics of the group interactions. We appeal to the fact that constructions exist not just in individuals, but in groups of individuals in contact, as a way of accounting, at least in part, for the fact that their idiosyncratic properties are sustained even in the face of competing tendencies to simplify and generalize.

The constructional view is particularly well suited to taking an epidemiological perspective on the relationship between the state of the grammar in the individual and in the group. Related constructions may exist independently of one another, and compete with one another for primacy in the individual. Constructions may merge when they come into contact if they cover complementary phenomena. Constructions that are closely related and doing the same work may carve out distinct functions so that they do not compete. Constructions may divide and apply to different parts of the form–meaning correspondence. In contrast, it is difficult to see how to model such phenomena in lower-level terms such as features, or in higher-level terms, such as maximally general rules.

6.3.2.1 A computational simulation of contact As an aid to conceptualizing the role of contact in the status of complexity, it is useful first to consider the essential features of contact in a group through a computational simulation. By assumption, contact between individuals is neutral with respect to their linguistic competence. Thus, we consider contact between individuals within a speech community, contact between individuals from closely related speech communities (i.e. dialects), and contact between individuals from different languages to be the same from the perspective of the social network simulation.

The simulation involves a set of agents in an n*n array. There are several dimensions of variation, which we may call ‘features’ or ‘constructions’ or ‘parameters.’ Each agent has a value for each dimension. For the simplest case, we suppose that there are three dimensions of variation, each with two values. Thus there are $2^3=8$ combinations of feature values, that is, eight possible ‘languages.’

Assume that the three feature values are distributed randomly over the population. The initial state of affairs prior to contact is shown in Figure 6.3.

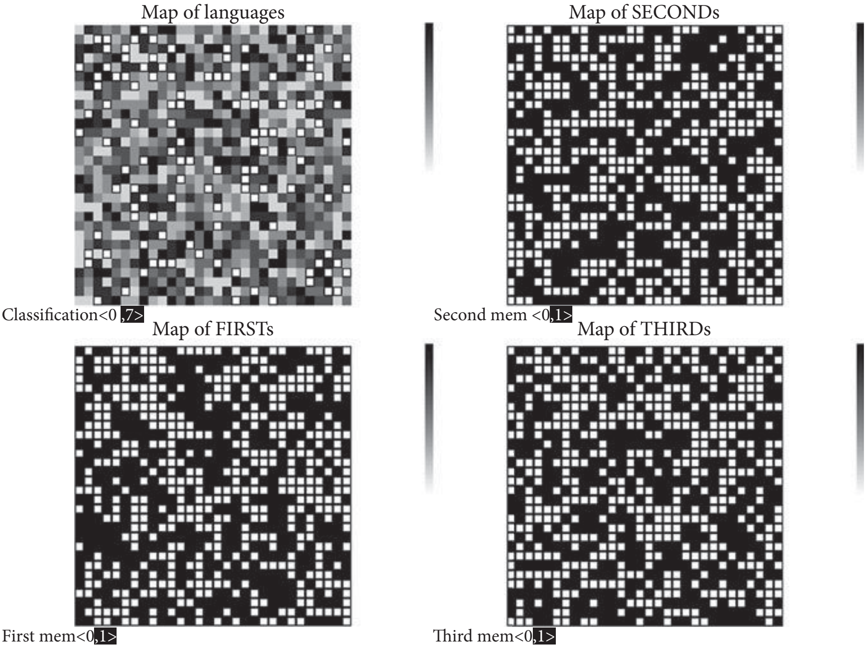


FIGURE 6.3. Feature values at initial state

The composite of the feature values for the three binary features is shown in the upper left. This simulation assumes that at the outset there is an even distribution of the possible combinations across the population, so that all eight languages are attested and have roughly the same share of the population.

At each step of the simulation, agents influence one another with respect to each feature of variation. An agent will change the value of a feature if more of its interaction partners have the other value (where what actually constitutes "more" can be made explicit and weighted by distance between interacting agents, number of agents who interact with one another, strength of agents, individual resistance to change, and other parameters of the simulation model).

Figure 6.4 illustrates the distribution of languages and features after sixty-nine steps of a single simulation in which the agents in Figure 6.3 influence one another and possibly change as a consequence of their interactions. This display shows the state of each agent for each feature at this point in the simulation. As a consequence of the changes from Figure 6.3, some combinations of feature values increase their distribution over the population, while others decrease. The histogram in Figure 6.5 shows the population levels of the eight languages at this point.

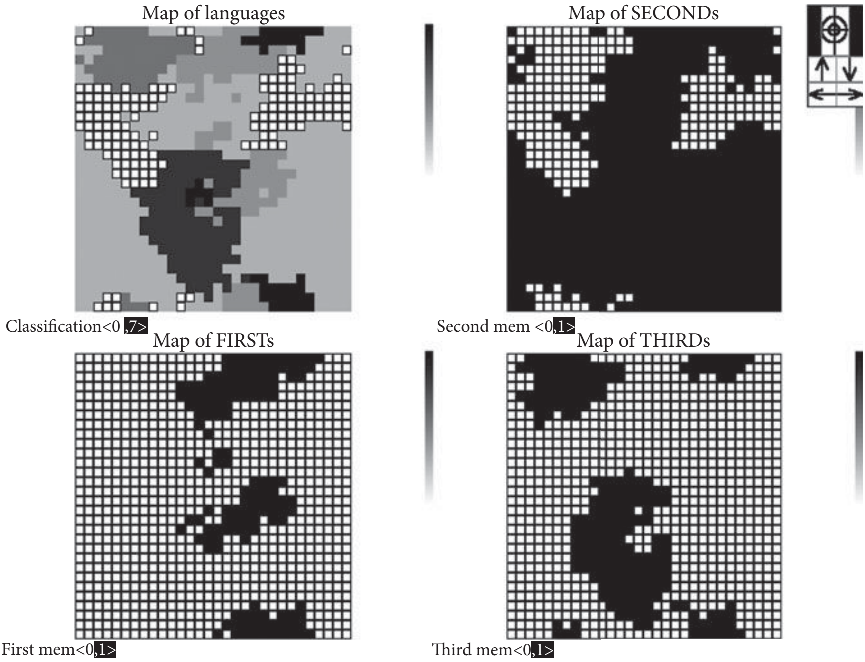


FIGURE 6.4. Feature values after clustering

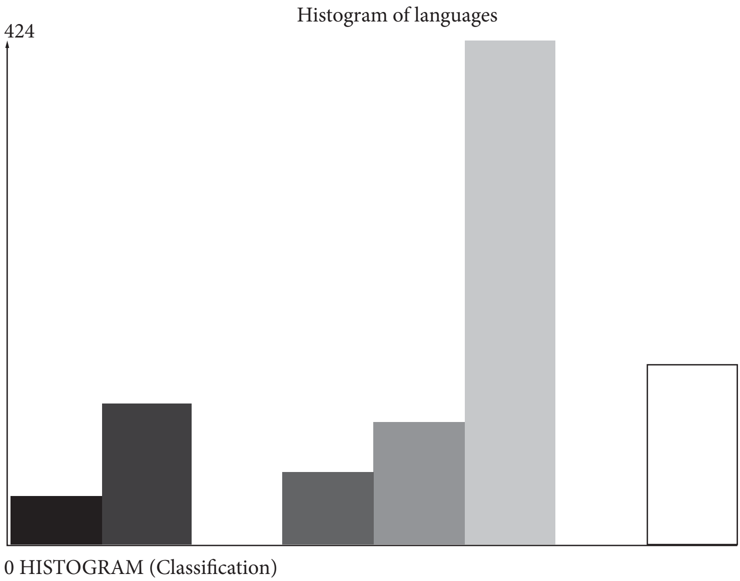


FIGURE 6.5. Number of languages of each type

Notice, crucially, that two of these languages are dead by this point, i.e. there are no agents with the relevant combination of feature values.

The loss of combinations of features illustrated in this particular instance of the simulation is not unique to the starting point shown in Figure 6.3. It is a consequence of the particular assumptions made in the simulation about how individuals interact in the network. Running the simulation again under the same parameters of interaction yields a different pattern of feature values and combinations each time, but the basic results are the same—some combinations survive, and some die out.

The appearance of gaps in the set of languages is an extreme case of a more general property of the simulation model, which is that feature values correlate. The value of a feature for an individual is determined by the interactions between that individual and its interaction partners. Over time all of the feature values of an individual tend to become more highly correlated with those of its interaction partners. The network evolves over time so that there is a tendency for one feature value to appear on an agent when another appears on the same agent. However, the correlation is not 100 percent, so that the opposite pattern may also occur, although less frequently. An illustration of the history of correlation in a particular simulation is given in Figure 6.6. As can be seen, features F2 and F3 correlate increasingly closely over the course of the simulation, while features F1–F3 and F1–F2 do not.

It is instructive to observe the kind of pattern that is associated with such correlation. In Figure 6.7 we see the space for F2 and F3 after 100 steps. The visual evidence reveals the tendency for F2 and F3 to have the same values, even through there are many exceptions.

The lessons offered by this simulation are several. The most important is that even when there is an initial random distribution of properties across the population, local differences in the densities of the options lead to the predominance of one option over the other. Going back to Figure 6.5, for example, the absence of Languages 3 and 7 reflects the predominance of Black over White in the SECOND property. So, beginning with a random distribution, we almost inevitably get clustering of properties, and with that, gaps in the space of combinations. Consequently, properties begin to correlate with one another. But the clustering, the gaps, and the correlations may have nothing to do with what the actual properties are—they are simply consequences of the interactions in the model, and of the fact that neighbors interact with neighbors, and stronger neighbors influence weaker neighbors.

Hence it is not possible to conclude anything of linguistic interest simply from clustering, gaps, and correlations. The distributional properties that we observe may have linguistic causes, or they may not. However, a theory of

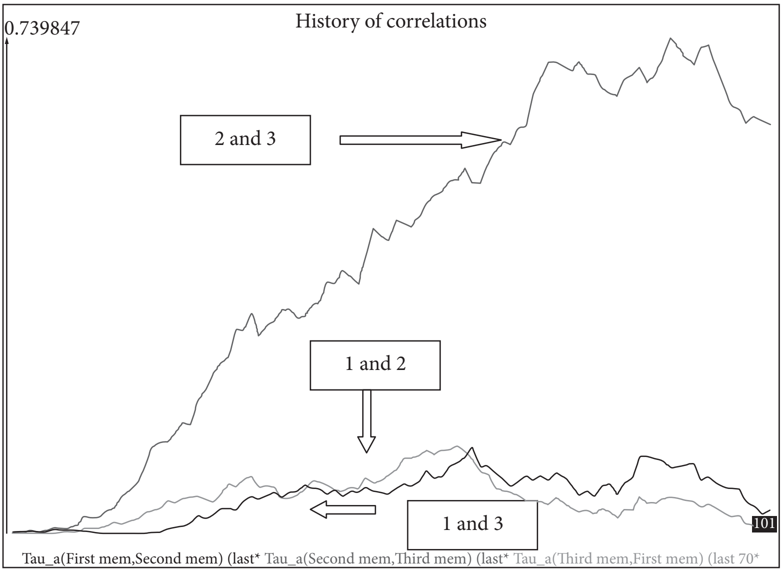


FIGURE 6.6. Feature clustering

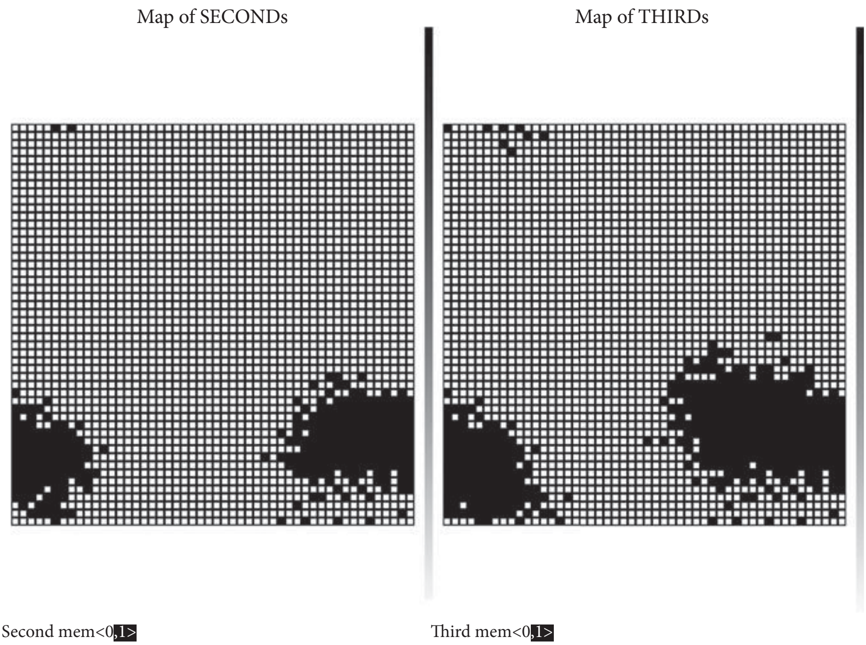


FIGURE 6.7. Features 2 and 3 after 100 steps

complexity combined with a theory of change through contact permits us to create scenarios in which certain constructions and combinations of constructions are sustained in the population, while others are rarely or not at all found. I take up this point in §6.3.2.2.

6.3.2.2 The role of bias The account of complexity maintenance that I outline here has two components. On one hand, the nature of interaction in the social network leads to clustering of properties, as illustrated in the preceding section. Clustering is the distribution of a property over a connected set of agents. On the other hand, complexity produces a bias that favors one option against another that is in competition with it. The simplest case of such a competition consists of two values of the same property, as in the simulation described above. The classical view of markedness of Trubetskoy (1939) and Jakobson (1968) is based on this type of bias in phonological feature values. Contemporary markedness theory (Kean 1980; Battistella 1996) develops a refined view of bias in which the bias for a particular feature value is not necessarily absolute, but may be contextual.

The effect of bias on the distribution of properties in the network depends in part on the structure of the network in which the bias exists. Properties that are shared by a cluster of agents tend to be resistant to competition from other properties. Thus a tight-knit or relatively isolated cluster may resist the effects of the bias. To see why this is, recall that the bias adds strength to a particular value, and therefore may induce an agent with another value to switch to it. But what also contributes to the strength of a feature value is the number of agents that have this feature value, compared with the number that have the other value.

Consider what happens when there is a cluster. An agent on the edge of the cluster may change its property value as a consequence of influence of agents outside the cluster. But there is strong support from inside the cluster for that agent to change back to the predominant value in the cluster. Moreover, the agents in the center of the cluster are more resistant to influence from outside, simply because they are more insulated from external influence—they interact with fewer agents who live outside the cluster. So the influence of the bias is less, to the extent that there are well-populated clusters that can resist it.

The physical distribution of agents is not the only factor, however. In the computational simulation, it is possible to change the interaction distance, which produces interesting results in combination with other parameters of the model. Putting it simply, the greater the interaction distance and the greater the number of individuals that an agent interacts with, the less stable are the clusters and the more likely it is that many small clusters will give way to fewer larger clusters. This result is, of course, a reduction in the possible combinations and more gaps in the class of possible languages.

On the other hand, bias makes the competition in the network between alternatives uneven. In the simulation described in the preceding section, each value of a property is of equal strength. So, for example, if agent A is interacting with two agents, and one has the value Black and one has White, then the influence of each on A is equal, and A is equally likely to choose Black as White.

Now we introduce a bias. For convenience, we set the parameters of the model so that A does not change from its current value, e.g. White, unless confronted with a greater number of Black than White. If A has White and interacts with two agents, one of which has Black and the other of which has White, A stays with White. But if the two agents both have Black, then A changes to Black. These states of affairs are illustrated in Figure 6.8.

Bias introduces an imbalance in the strength of feature values. Suppose that there is a bias in favor of Black. Now, if agent A has White, and interacts with one White and one Black, the bias in favor of Black means that Black now outranks White in the competition. So A will change to Black. Figure 6.9 illustrates.

If the distribution of Black and White in the population is even, and the bias in favor of Black is strong enough, there is a good chance that in the end, every agent will switch to Black. The actual outcome will also depend on the initial distribution, the interaction distance, and the number of agents that each agent interacts with.

I assume that both formal and processing complexity can result in bias. In the case of formal complexity, what is in competition in the social network are the more or less general variants of a construction that express the same meaning, i.e. those with and without arbitrary conditions and exceptions. In the case of processing complexity, what is in competition in the social network are correspondences that require greater or fewer resources to compute a given meaning. On the view outlined here, the more general will win out over the less

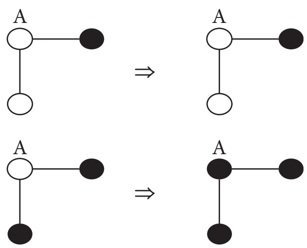


FIGURE 6.8. Interactions of agents in a network

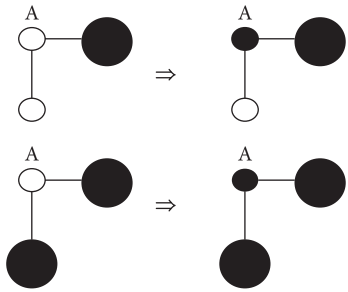


FIGURE 6.9. Interactions of agents in a network, bias on Black

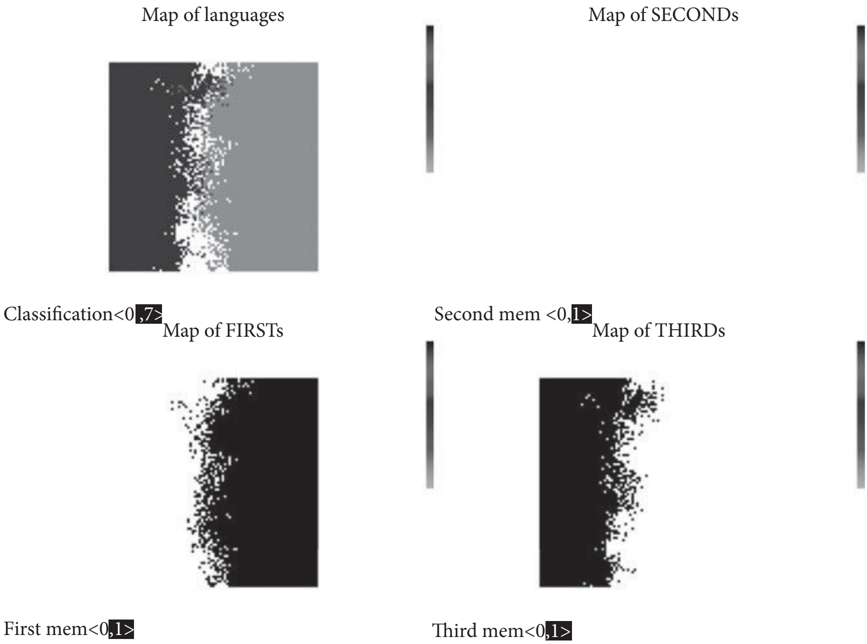


FIGURE 6.10. -1% bias of FIRST=BLACK. Step 450

general, other things being equal. And the less complex in processing terms will win out over the more complex, other things being equal.

But other things are not always equal. A cluster may resist external pressure to change, even when the property that is being maintained has a bias against it. The resistance to change may slow the loss of the dispreferred option. In a computational simulation, I assigned a slightly negative bias to one value (FIRST=Black). A picture of the state of the simulation after 450 steps is given in Figure 6.10.

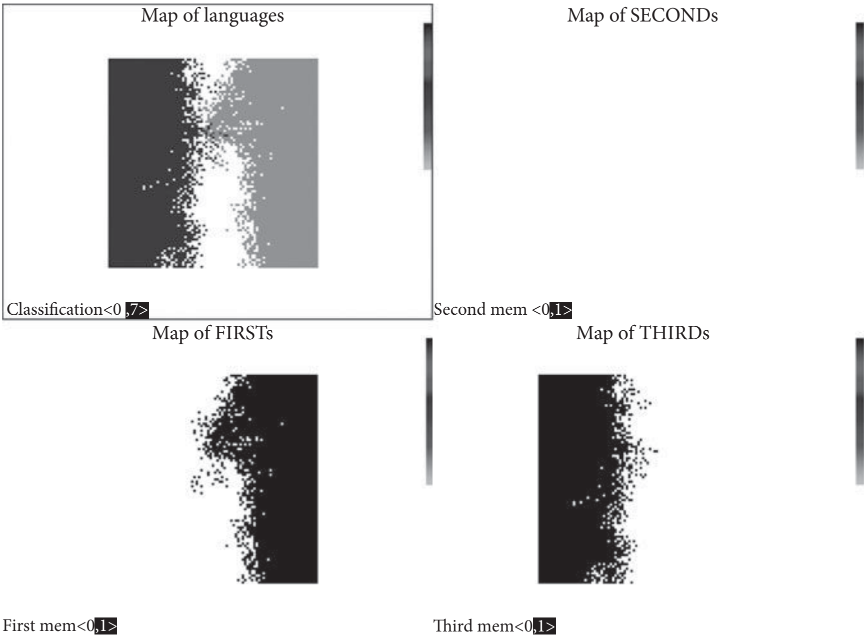


FIGURE 6.11. -1% bias of FIRST=BLACK. Step 2,209

After 2,200 steps the situation is still stable, in the sense that the area occupied by FIRST=Black (the right half of the area) is still fairly large. However it is smaller than it was before, as shown in Figure 6.11. At the same time, the distribution of the THIRD property is essentially unchanged from what it was at the start of the simulation.

It is only around step 3,500 that the bias ultimately causes the demise of FIRST=Black in the population (Fig. 6.12).

What this simulation illustrates is that it is possible for a less preferred option to remain in the population for a substantial amount of time, and in principle forever, as long as there are conditions that continue to support it. One condition is lack of contact with competitors; another is a strong and compact cluster. And something that is not reflected in the computational simulation but almost certainly has an effect in real social networks is the frequency of use of a dispreferred alternative. This last factor is a plausible explanation for the persistence of irregular verbal morphology, for example, in languages such as English or German. In the case of the German

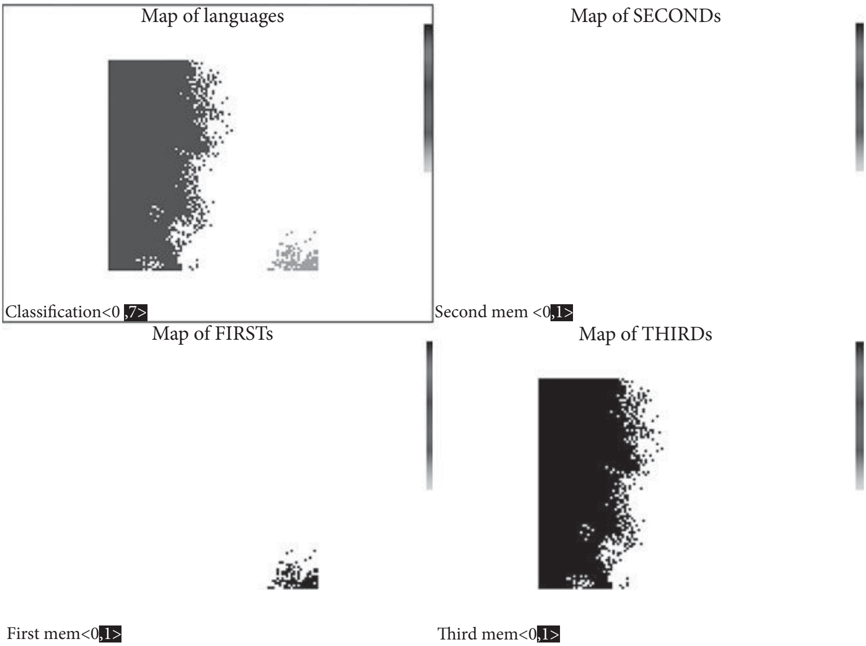


FIGURE 6.12. -1% bias of FIRST=BLACK. Step 3,460

morphology, there is in fact evidence that some of the less frequent strong verbs are moving to regular status (Nübling et al. 2006).[5]

Even when the linguistic alternatives are equal, arbitrary factors can lead to one alternative driving out the other in the competition. A possible example of the arbitrariness of constructional competition is a collocation like *more or less* or *up and down*.[6] The alternatives, *less or more* and *down and up* are equally meaningful, yet unacceptable as collocations. There is nothing wrong with the sequences *less or more* and *down and up* as long as they are used non-idiomatically, as in

(16) a. Do you want less or more coffee next time, or the same?
b. Terry moved the lever down and up.

[5] But there are also formal grammatical factors at play, which make the picture quite a bit more complicated: see Mailhammer 2006, 2007a, 2007b. Mailhammer shows that the OHG verbal system was more regular than the MHG system, in that ablaut variation was more predictable on the basis of the phonological properties of the verbal roots. Remnants of this regularity, he argues, are effective in preventing further loss of strong verb morphology. This is a nice case of "other things" not being equal, where complexity actually grows and is subsequently maintained.

[6] There are many other similar cases, e.g.: *come and go, in and out, back and forth, high and low.*

Imagine an initial state in which two syntactically symmetrical expressions such as these are equiprobable. Even in the absence of an initial bias in favor of *more or less* and *up and down*, a slight statistical advantage for these options shortly after the initial point could be translated into complete collocational victory in the long term. As one option becomes more frequently used with the particular rhetorical function, the other becomes less able to express it, to the point that it sounds odd, in contrast to the familiar collocation, which sounds entirely natural.

As for processing complexity, a hypothesis that appears to have some empirical support is that a relatively more complex configuration, such as extraction from a relative clause, will be used relatively rarely if there is a way to avoid it. The rarity of the configuration means that it is rare in the experience of the learner, which means that the learner will have no experience computing the particular correspondences that are required. As suggested in Chapter 5, the absence of the particular configuration in the network of English speakers will result in negative acceptability judgments. Greater frequency in Swedish means that the configuration will persist in the network of Swedish speakers, with correspondingly more robust judgments.

Consider another case, that of extraction from sentential complements. It is entirely possible that in terms of processing complexity alone, extraction from the complements of 'bridge' verbs, such as *believe*, is comparable to extraction from a relative clause. Familiar examples such as (17) show that such extraction is unproblematic in English, however.

(17) Where do you believe [$_S$ Sandy put the beer __]?

But English does not offer an alternative way of expressing the CS representation in which an argument of the complement is questioned or relativized. Hence this type of configuration is frequent in English, is well represented in the network, and yields strong acceptability judgments.

But extraction from complements is more complex than extraction from a main clause (Kluender 1998). While the complexity is not so great as to make the construction essentially non-existent (in contrast to extraction from relative clauses), it is reasonable to expect to find languages that use constructions that do not incur this degree of complexity to express the same CS correspondence. In fact there are many languages that have *wh* in situ, which completely avoids the complexity incurred by extraction. Others, such as Russian and dialects of German, have *wh*-movement but do not permit long extraction. These languages have developed alternative ways of expressing the relevant CS representations, including partial *wh*-movement in German. (For a review, see Sabel 2000.) Consider the case in Russian.

(18) ?* Kogo ty dumaeš, čto Ivan priglasil
who.ACC you think that Ivan invited
'Who do you think that Ivan invited?'

In order to express the CS representation, Russian has developed a partial *wh*-movement construction (as has German), exemplified in (19).

(19) Kak ty dumaeš, kogo Ivan priglasil
How you think who.ACC Ivan invited?
'Who do you think Ivan invited?'

The A′ chain (if there is one in the lower clause) is local, so this construction is less complex than the long extraction.

In the way contact and change are conceived of in this chapter, both long *wh*-movement and the alternative local construction are present in the 'body linguistic' for German and Russian. Because of the existence of the local construction as a way to express the appropriate CS representations, the long movement option either does not emerge or falls out of use. It is not ruled out in principle, of course, as seen in the fact that it is possible in other languages. The absence of positive evidence for the long movement means that it will be the unfamiliar option, which in turn has the consequence that it is judged highly unacceptable or "ungrammatical" by native speakers. But again, since it is quite possible in other languages, the unacceptability here is not a matter of ungrammaticality, it is a matter of low frequency and a preferred competitor living in the social network.[7]

6.4 Conclusions and perspectives

I have argued in this chapter that complexity interacts with the dynamical processes of language acquisition and language change in ways that shape the properties of grammars. Such an account offers the possibility of explaining where complexity comes from and why complexity persists in grammars in spite of the pressures to eliminate it.

The two central questions explore here were, Why does complexity arise? and Why is complexity maintained against the pressure to reduce it? My proposal is that in some cases, complexity arises locally when it is being reduced more globally in the grammar. One example involves English relative

[7] Of course I do not rule out the possibility that the syntactic structure of complementation is different in languages that permit long extraction and in languages that do not. However, I am not aware of any compelling evidence that independently motivates such a syntactic difference. Following the logic of *Syntactic Nuts*, it follows that we should seek an account of the difference where the possibility of not only describing the difference but explaining it appears stronger.

clauses, which are generalizing from finite to infinitival clauses, but the generalization is not complete: the zero and the wh-PP cases are permitted, but wh-NP case is not yet possible. This situation contrasts with the infinitival questions, where the full range of possibilities with initial *wh*-phrases is found.

Complexity can persist where the pressure to reduce it meets resistance. Resistance may take the form of high-frequency support for the more complex construction, isolation of the more complex construction from competition, or social cohesion of the speakers that have the more complex construction. As shown in a simulation, resistance to reduction may result in a situation where the more complex construction appears to be stable, but in the long run it may undergo reduction and even elimination as the pressure on it continues.

The next step is to consider actual cases in which complexity and social contact interact so as to yield attested paths of change, leading to attested grammars. This is the focus of Chapter 7, where I look at the history of the verbal sequence in English and analyze it as a series of changes driven by pressures to reduce complexity in a variety of ways.

7

Constructional complexity and change[1]

This chapter tracks the history of several related core constructions in English. I argue that the changes can be understood as the consequence of the pressure to reduce constructional complexity. I show how this approach can be applied to the shift from fully general V2 to more specialized subject aux inversion and to the rise of fully general *do*-support.

My central thesis is that certain changes in the history of English grammar can be understood as reducing constructional complexity in specific ways. I show how the pressure to reduce complexity gives rise to new constructions as well as the generalization of constructions into what are conventionally called "rules." I do not rule out the possibility that changes may also produce an increase in complexity, as discussed in Chapter 6. But I hypothesize that at least in the cases discussed here, such increases in complexity are the by-product of changes that reduce complexity elsewhere.

7.1 Change as explanation

In previous chapters I have shown how complexity plays a role in accounting for what is common in natural language, what is rare, and what does not appear at all. But complexity itself does not explain why a particular language has a particular property or set of properties. Why, for example, does English have *do*-support and subject aux inversion? Why does English have topicalization without V2 and German have topicalization with V2? What is the relationship between case-marking and scrambling? Where do syntactic nuts come from, and can we explain why the ones that appear in one language do not appear in another?

[1] This chapter is an adaptation and extension of Culicover (2008b). I am indebted to an anonymous reviewer of this manuscript for some very perceptive comments that have led to numerous clarifications and corrections. I have extensively reworked and simplified the discussion, clarified a number of points that were unnecessarily obscure in the earlier version, brought the notation in line with that used elsewhere in this book, and corrected a number of errors. Any remaining errors are my responsibility, of course.

The overarching idea is that the explanation for the form of a grammar at a particular point in time must take into account the history of the language. The account offered here incorporates complexity into a theory of language change. The following three principles form the basis of this theory; each has a long and distinguished history in linguistics.

- Over time, a grammar will tend to reduce local complexity, other things being equal.
- In the competition between constructions in a social network, the less complex will be preferred over the more complex, other things being equal.
- A grammar will attempt to maintain a one-to-one mapping between form and meaning to the extent possible.[2]

I show how in this chapter how an account of change that assumes these principles provides a reasonably promising account of the development of *do*-support and subject aux inversion in English.

Do-support is one of the most characteristic properties of English. It is found in the core constructions in (1).

(1) a. Sandy did not call. [negation]
    b. Did Sandy call? [interrogation]
    c. Robin didn't call but Sandy did. [ellipsis]

In fact, *do*-support occurs in English wherever an auxiliary verb is required but none is available. I reserve the term '*do*-support' for the English type of *do*-periphrasis. Other languages do not have full *do*-support, although many languages have periphrastic *do* of some type (van der Auwera 1999; Jäger 2006; Wichmann and Wohlgemuth 2008). My goal here is to give a constructional account that explains why periphrasitic *do* takes the form of *do*-support in English. Such an account contrasts with the familiar derivational approaches of mainstream generative grammar.

The historical developments and theoretical issues bearing on *do*-support and related constructions are quite complex and require considerably more exposition and analysis than is possible here. I therefore limit my discussion to the core cases typically addressed in formal grammatical studies and traditional descriptive grammars (such as Huddleston and Pullum 2002).

[2] I say "to the extent possible" because there are other factors than constructional synonymy that may contribute to complexity. For example, extraposition of relative clauses in English may simplify processing of complex NPs (see Hawkins 2004), and yet there does not appear to be any movement to identify the extraposed and non-extraposed construction with distinct meanings. This is a case in which processing complexity preempts formal complexity. An interesting topic for future research is precisely how preemption of this sort works.

I do not address a number of interesting phenomena, including the general question of why *do* specifically (and its counterparts in other languages) has a periphrastic function (see Jäger 2006 and references cited there), whether periphrastic *do* in English retains a residue of meaning (generative grammarians typically hold that it does not, but see Hirtle 1997 for an alternative view), and a variety of uses of periphrastic *do* in English that may not be reducible to one of the core cases.

The development of *do*-support involves a number of critical historical developments that have been addressed in the descriptive and theoretical literature. Old English (OE), an early ancestor of Modern English (ModE) spoken until about the mid-twelfth century, was a Germanic language with many of the properties of contemporary German.[3] Most importantly, OE was essentially a V2 language, verbs were fully inflected, and there was no distinct subclass of modal verbs comparable to what we find in ModE. The changes that are associated with the development of *do*-support include the following:

(2) a. growth of periphrastic *do*+V_{inf};
b. loss of full V2 reanalysis of V2 as 'residual' V2 in questions and other 'affective' environments;[4]
c. formation of a distinct subclass of modal verbs;
d. loss of case and establishment of 'positional licensing' of subjects;[5]
e. loss of preverbal *ne* and introduction of postverbal *not* as the marker of sentential negation.

Another change underway at approximately the same time was the shift from OV to VO; see Pintzuk and Taylor (2006). I assume that this alternation reflected in part non-grammatical ordering constraints in the English VP, which are still active in ModE (Hawkins 2004; Wasow 2002), and also fell under the requirement that the grammatical functions Subj and Obj be distinguished through word order, as overt case-marking was lost; see Kiparsky (1997).

Adopting the general perspective of Government Binding Theory (GB) and later Principles and Parameters Theory (PPT), a number of scholars have

[3] Although it appears that many of the changes seen in the development of ModE, such as the loss of V2, were already underway at this time. See for example Taylor (2005).

[4] Here I am following Rizzi (1996) in viewing English SAI as a historical remnant of V2. As outlined in §7.4, a constructional account allows us to draw a natural connection between the present grammar and the earlier form. In contrast, McWhorter (2007) writes that "V2 in English begins a decline in the fifteenth century, and is essentially dead by the seventeenth (Jacobsson 1951; Nevalainen 1997). The question obviously arises as to why." Not recognizing the relationship between SAI and V2, he offers an impressionistic account of contact between English and Scandinavian as the source of the change without actually answering the "why" question.

[5] The term is taken from Kiparsky (1997).

sought to detail aspects of the changes in word order patterns from OE through Early and Late Middle English (EME and LME) to ModE in terms of changes in the interactions between overt heads and complements, on the one hand, and abstract grammatical formatives, such as functional categories and grammatical features, on the other. This research has produced a number of proposals for changes not simply in superficial grammatical patterns, but in grammars. For some representative examples, see Kroch (1989), Kroch and Taylor (2000), Fischer et al. (2000), van Kemenade (1997), Kiparsky (1997).

This research, while it differs from author to author in terms of specific proposals, shares the characteristic that it makes crucial use of abstract syntactic structure and movement. For example, the V2 property of OE is typically (but not universally) accounted for by assuming that there is a functional head I^0 that follows the subject position and precedes VP. I^0 contains the finite inflection, which we denote here as [TENSE]. The highest V moves from the VP and adjoins to I^0, which produces the result that the main verb functions as the head of IP.

(3) $[_{IP}$ NP I^0[TENSE] $[_{VP}$ V...]] $\Rightarrow$
$[_{IP}$ NP V+I^0[TENSE] $[_{VP}$ __ ...]]

I refer to this phenomenon as V-to-I.[6]

In ModE, in contrast, only auxiliary verbs (including modals) are adjoined to I^0, on standard assumptions. An analysis along these lines captures the generalization that in a language such as French and German, which has V-to-I, the distribution of the finite verb with respect to inversion constructions, negation, and adverbs closely parallels the distribution of the finite auxiliary verb in English, as outlined in the influential proposal of Pollock (1989).

On this view, the emergence of *do*-support is associated with the loss of V-to-I. Through this loss, English changes from a language of the French or German type to a language with the characteristic features of ModE. In ModE, finite-tense marking must be licensed on a main verb through a mechanism other than V-to-I; this phenomenon is typically referred to as Affix Hopping (AH) (see Chomsky 1957 for the original analysis; for recent reappraisals, see Bobaljik 1995, Lasnik 1999, and Freidin 2004).[7] *Do*-support, on this general view, marks the position of I^0 when it is not immediately adjacent to the verb and hence cannot appear on it through AH.

6 Another common term is 'verb-raising.'

7 The term refers to the fact that the main verb is marked with tense even though it does not appear in I^0. Chomsky 1957 solved this problem by positing a rule that "hops" tense to the right onto the main verb.

While this approach to *do*-support captures a number of important generalizations, it is problematic in two respects. First, it is not entirely adequate as an account of the synchronic English grammar, a fact that is well known but has never been satisfactorily resolved. Second, it accounts (in the sense of "keeping an account") for the changes in the language but does not explain them.

Characterizing the changes in the history of English in such derivational terms, while appealing in some respects, does not constitute the most explanatory account. §7.2 argues that an analysis that makes crucial use of I^0, as the V-to-I account does, is not the best way to explain the English verbal sequence, and is problematic in a number of other respects as well. §7.3 argues that a more satisfactory approach makes crucial use of the notion of 'construction.' On this approach, the various word orders are realizations of constructions that grow, contract, merge, and split over time and eventually end up in the form that we see in ModE. §7.4 argues that it provides a better explanation for the actual course of events observed in the history of English.

7.2 Derivations

While there have been significant changes in syntactic theory during the more than fifty years of generative grammar, there is a common thread that runs through all of the analyses of the English verbal cluster: the tense inflection that is marked on the finite verb is syntactically isolatable as a constituent of the sentence, and its distribution determines whether or not there will be *do*-support. Where tense appears, and hence whether and where *do* appears, is a consequence of a syntactic derivation that crucially makes use of movement to create the configuration in which the elements that are responsible for the overt form are arranged with respect to one another.

In this section I outline the essential properties of these analyses, and point out the problematic aspects of the general approach. I review briefly how syntactic variation is accounted for in these terms, and conclude that positing I^0 leads to an overall loss of generalization.

7.2.1 Affix hopping

The earliest generative and derivational account of the English verb cluster is that of *Syntactic Structures* (Chomsky 1957). It is a classic analysis and widely known, so I review it only briefly in order to highlight its critical features.

Chomsky assumes that the verbal cluster excluding the main verb is generated as a complex unit, consisting of (4).

(4) $\underline{\text{Tense}}$ (Modal) (have +$\underline{\text{en}}$) (be +$\underline{\text{ing}}$)

Each of the underlined elements is an affix which is adjoined to the verbal element to the right of it by AH. Hence if there is a Modal, *have* or *be* (henceforth a V_{AUX}), Tense is marked on this verb. E.g.,

(5) $\underline{\text{Tense}}$ Modal V ⇒ Modal+$\underline{\text{Tense}}$ V
$\underline{\text{Tense}}$ have +$\underline{\text{en}}$ V ⇒ have+$\underline{\text{Tense}}$ V+$\underline{\text{en}}$

and so on. The sequence of terms in rule (4) accounts for the observed ordering, while the association of the affixes *+en* and *+ing* with their corresponding auxiliaries and AH captures the fact that *have* selects VPs with perfect inflection, and *be* selects VPs with progressive inflection.

The strict linear ordering of Modal, *have*, and *be* ensures that the verb is marked with Tense only if there is no intervening element, as seen in (6).

(6) $\underline{\text{Tense}}$ V ⇒ V+$\underline{\text{Tense}}$
E.g., the manager **Past** call ⇒ the manager call+**Past**.
('The manager called.')

But if Tense is not immediately adjacent to V, for example because of the presence of *not*, AH is blocked. In this case, *do* must be inserted, as seen in (7).

(7) $\underline{\text{Tense}}$ not [$_{VP}$ V...] ⇒ (AH blocked) ⇒ do+$\underline{\text{Tense}}$ not [$_{VP}$ V...]
E.g., the manager $\underline{\text{Past}}$ not call ⇒ (AH blocked) ⇒ the manager do+$\underline{\text{Past}}$ not call
('The manager did not call.')

Similarly, if Tense is moved away from V by SAI, Tense is not adjacent to V. AH is blocked, and *do* must be inserted, as seen in (8).

(8) NP $\underline{\text{Tense}}$ V ⇒ $\underline{\text{Tense}}$ NP V ⇒ (AH blocked) ⇒ do+$\underline{\text{Tense}}$ NP V
E.g. the manager $\underline{\text{Past}}$ call ⇒ $\underline{\text{Past}}$ the manager call ⇒ (AH blocked) ⇒ do+$\underline{\text{Past}}$ the manager call
('Did the manager call?')

AH is also blocked, and *do* must be inserted, if V is moved away from Tense, e.g. by VP topicalization (9a), or V is deleted, e.g. by VP ellipsis (9b) or pseudogapping (9c).

(9) a. ... and call the manager *did*.
b. ... and she *did*.
c. She likes beets no more than she *does* sweet potatoes.

The key insight of this analysis is that unaccompanied Tense functions just like Tense paired with a V_{AUX}. The sequence Tense-($\left\{\begin{array}{l}\text{Modal}\\ \textit{have}\\ \textit{be}\end{array}\right\}$) functions as a unit with respect to negation, movement, and deletion. However, on the classical analysis it is not a constituent. Thus the analysis fails to "capture a generalization" in the traditional sense.

Another missed generalization is the fact that *do* is necessarily inserted after AH in such a way that it ends up in the same configuration with Tense that it would have if it was inserted before AH, that is, *do*+Tense. Thus, we have two ways of getting Tense onto the verb, one for V_{AUX} and main verbs, and one for *do*.

Moreover, sequences like *would not have*, *would not be*, and *would not have been* show that *not* precedes *have* and *be*. So when there is no Modal, the order of elements is as given in (10).

(10) Tense not have +en V
Tense not be +ing V

Since Tense-*have* and Tense-*be* also function as units with respect to SAI, it is necessary to shift the leftmost *have* or *be* to Tense before SAI applies. This is another way in which the classical analysis misses a generalization about the distribution of Tense and *do*-support. The requirement that the rules SAI, shifting of $\left\{\begin{matrix} \textit{have} \\ \textit{be} \end{matrix}\right\}$ to Tense, and AH must be strictly ordered is a further loss of generality.

Finally, it turns out that only an intervening *not* causes *do*-support, and an intervening adverb does not, even one that conveys negation.

(11) a. Sandy $\left\{\begin{matrix} \text{will} \\ \text{did} \end{matrix}\right\}$ not call.

b. Sandy $\left\{\begin{matrix} \text{will} \\ \text{*did} \end{matrix}\right\}\left\{\begin{matrix} \text{certainly} \\ \text{never} \end{matrix}\right\}$ call.

This is a long-standing puzzle that has never been entirely resolved, and is rarely addressed (but see Pesetsky 1989).

7.2.2 The shift to structure: I^0

It is natural to expect that the problems noted in the preceding section can be fixed by taking Tense-V_{AUX} to be a constituent. In contemporary terms, this is the assumption that Tense(-Modal) is a realization of I^0, the head of IP(=S). The verbal auxiliary is thereby brought into the general X-bar framework, which allows for a more uniform theory of phrase structure.[8] In SAI, I^0 undergoes head-to-head movement to C^0, and is thus local and structure-preserving.

[8] For discussion of the role of uniformity in the development of syntactic theory, see Culicover and Jackendoff 2005: chapters 2 and 3.

This step is in fact the leading idea in the extension of X-bar theory to the functional categories and is the basis of many subsequent influential proposals in mainstream generative grammar (MGG), including uniform binary branching and the derivation of surface word order from branching structure (Kayne 1994), the localization of parametric variation (Borer 1984), the structure of the left periphery (Rizzi 1997), the DP hypothesis (Abney 1987), and so on.

In general, Principles and Parameters theory defines the apparent location of I^0 as the head of the projection IP, and more generally, the location of an X^0 as the head of XP. If an overt head is assumed to originate in some other place than where it appears on the surface, for theoretical reasons, then there must be a derivation that moves it to its observed position.[9] This is seen in the MGG analysis of inversion in French and German, where the main verb moves out of the VP into I^0, and from there to C^0 (Pollock 1989; Den Besten and Edmondson 1983). In English, on the other hand, only the auxiliary verbs (including modals) adjoin to I^0 and function as the head of IP. The main verb remains in situ as the head of VP. Hence we have *Will you eat the vegetables?* but **Eat you the vegetables?*

But the assumption of I^0 does not actually improve the analysis of the English verbal cluster, it makes it worse. I outline the reasons why this is and conclude that I^0 should be abandoned. To make the analysis concrete, assume the following. (The notion [TENSE] is intended to indicate that the tense-marking is a feature.)

(12) a. I^0 is the head of IP;
   b. I^0 precedes (*not*) VP;
   c. I^0 contains [TENSE] or Modal[TENSE];
   d. [TENSE] is discharged by locally c-commanded V[TENSE]; *do* is inserted if [TENSE] is not discharged;
   e. each head (except [TENSE]) selects its complement. In particular
      i. [Modal[TENSE]] selects VP[BARE], that is, uninflected VP;
      ii. *have* selects VP[PAST.PRT] and *be* selects VP[PRES.PRT];
      iii. the features of a projection are realized on its head;

[9] This characterization of derivational analyses in terms of movement of V to I^0 is a simplification. There are more complex options within the general derivational framework, depending on one's specific assumptions about the fine details of the syntactic structure. For example, one may assume that there are functional heads T(ense)^0, AgrO^0, and AgrS^0; see for example Bobaljik (1995). In that case, it is necessary to account for the movement through a number of Spec positions, and the raising of V through each of the functional heads. Since the main point of the argument is that a non-derivational analysis is able to accommodate the gradual growth of *do*-support and the loss of V2 in a way that is not readily available to a derivational analysis, I have not focused on the details of various alternatives within the derivational approach.

f. $\begin{Bmatrix} \textit{have} \\ \textit{be} \end{Bmatrix}$ raise to I^0 when I^0 contains only [TENSE].

On this analysis, I^0 is a constituent, precedes *not*, and undergoes SAI, which captures a generalization in configurational terms. The cost of such an analysis amounts to three stipulations: (i) that a tensed verb in situ is licensed by [TENSE] in I^0 but not by Modal[TENSE] in I^0; (ii) that $\begin{Bmatrix} \textit{have} \\ \textit{be} \end{Bmatrix}$ preclude *do*, so that we don't derive **doesn't have called*, **doesn't be calling*; and (iii) that *do* must be inserted under certain circumstances (e.g. as a "last resort" in the Minimalist Program (Chomsky 1995)).

With the shift to I^0, questions arise about how exactly to get *do*-support to apply. Assume that [TENSE] and Modal are in I^0. I^0 will consist either of [TENSE] alone, or Modal[TENSE]—they are in complementary distribution in I^0.

The following VP then has two possibilities. First, it may be a bare VP headed by a main V, with no tense-marking. Second, it may be a tensed VP, since in a simple declarative clause the main V is tensed and in situ. These two possibilities combine with the two possibilities for I^0 to give four combinations.

(13) a. [TENSE] VP
b. [TENSE] VP[TENSE]
c. Modal[TENSE] VP
d. Modal[TENSE] VP[TENSE]

Not all of these combinations correspond to well-formed derivations. (13a) is possible only if [TENSE] and VP are ultimately not adjacent and *do* is introduced. (13b) is possible only if I^0 and VP are adjacent so that the equivalent of AH can apply. (13c) always leads to a well-formed sentence, regardless of what happens to VP. And (13d) can never yield a well-formed sentence.

To capture these facts, we need to assume that when [TENSE] is alone in I^0, it licenses and is 'discharged' by a tensed main verb V[TENSE]. But when [TENSE] appears on Modal, only a bare VP is possible.

(14) a. Unattached and undischarged [TENSE] is ill-formed, but it is saved by introduction of *do*. This takes care of (13a).[10]
b. VP[TENSE] is licensed by a locally c-commanding [TENSE], which is thereby 'discharged.' This accounts for (13b).[11]
c. Modal[TENSE] either does not have to be licensed, or it is automatically licensed because it is in I^0. This accounts for (13c).

[10] If we assume as an alternative that *do* is an underlying Modal, then there must be a filter that rules out *do*[tense]-V when *do* is unstressed.

[11] If we assume as an alternative that *do* is an underlying Modal, then there must be a rule that deletes unstressed *do* in I^0 when it immediate precedes and c-commands VP[tense].

d. VP[TENSE] is not licensed by Modal[TENSE], and thus (13d) is ungrammatical.

This analysis compares with the classical analysis as follows:

- The discharge of [TENSE] in (14b) is the functional equivalent of AH, as in the classical analysis.
- The failure of discharge of [TENSE] (the functional equivalent of the blocking of AH) produces *do*-support, as stipulated in (14a), as in the classical analysis.
- Since this analysis assumes that each verb is the head of its own VP, we will still have to say something about the situation in which the VP is headed by an auxiliary V (*have* or *be*) and I^0 contains only [TENSE]. That is, the auxiliary verb has to raise into I^0 if there is only [TENSE] in I^0. This is the equivalent to the auxiliaries moving to the right of tense in the classical analysis.[12]

There are five main puzzles with this type of account of *do*-support.

- Why is AH, or the equivalent, needed? In this particular analysis, why does [TENSE] in I^0 license V^0[TENSE], while [TENSE] on a modal or auxiliary does not?
- Why do auxiliary verbs with [TENSE] have to be raised into I^0? And when they are raised, why do they end up having exactly the form that they would have if they were subject to AH (or the equivalent)?
- Why is *do* inserted, and why does it get exactly the form that it would have if it was an underlying Modal?
- Why does AH (or the functional equivalent) require strict adjacency between I^0 and V?
- Why does negation block AH, or otherwise trigger *do*-support, while adverbs do not?

The first three questions point to the same redundancy that we found in the classical analysis. The redundancy arises even when we formulate the analysis in terms consistent with X-bar theory, feature discharge, licensing, and movement. Effectively, there are four ways in which [TENSE] can be licensed on a verb. It is Merged as Modal[TENSE], it is Merged as V[TENSE] and licensed by I^0[TENSE], it is the result of the raising of $\left\{\begin{matrix} \textit{have} \\ \textit{be} \end{matrix}\right\}$ to I^0[TENSE], or it is the result of *do*-support.

[12] If we assume as an alternative that *do* is an underlying Modal, then *have/be* must substitute *do* in I^0 when I^0 immediately precedes and c-commands it.

We might try to achieve some simplification by assuming that [TENSE] is in I^0 alone, and that Modal, *have*, and *be* are each heads of their own VP projection. Then in order to function as the head of IP, a tensed Modal, *have*, and *be* must all raise to I^0. The two situations are then as follows:

(15) $[_{I^0}$ [TENSE]] $[_{VP}$ V_{AUX}[TENSE] ...] $\Rightarrow$ $[_{I^0}$ [TENSE] V_{AUX}[TENSE]] $[_{VP}$ ___ ...] $\Rightarrow$ $[_{I^0}$ [~~TENSE~~] V_{AUX}[TENSE]]$[_{VP}$ ___ ...]
$[_{I^0}$ [~~TENSE~~]] $[_{VP}$ V[TENSE] ...] $\Rightarrow$ $[_{I^0}$ [~~TENSE~~]] $[_{VP}$ V^0[TENSE] ...]

We assume that the V_{AUX} that is raised discharges the [TENSE] in I^0.[13] The peculiarity of this alternative is that the raising of V_{AUX} to I^0 is not blocked by *not*, while the discharge of [TENSE] by the main verb is blocked by *not*.[14]

As far as I can tell, this is the best that can be done given the key assumptions of the account in which I^0 is the head of the sentence, heads raise to other heads, and *do* is inserted as a 'last resort' when [TENSE] is not otherwise licensed. Regardless of the technical details, the English verbal cluster displays idiosyncrasies that do not succumb to a systematic treatment in terms of branching structure, functional heads, and movement.

Bobaljik (1995) proposes an adaptation of the classical analysis, in which *do*-support is triggered by the failure of Tense to adjoin to V. On his analysis, Tense must be attached to a V in order to satisfy a morphological requirement, e.g. it undergoes a form of AH. When attachment of Tense is blocked, *do*-support applies so that the features of Tense can be expressed, along the lines that we have sketched. Putting AH in the morphology and not the syntax makes the adjacency requirement more natural, although it does not avoid stipulations. Bobaljik (1995: 62) must stipulate that intervening adverbs do not interfere with the adjacency of Tense and V. As in the classical analysis, $\left\{\begin{array}{l}\text{Modal}\\ \textit{have}\\ \textit{be}\end{array}\right\}$ must raise in his analysis, again to allow Tense to attach to V. And there is a residual redundancy, in that the morphology is now doing work that replicates what would happen in the syntax, if *do* was a modal in the first place.

[13] Such an assumption can be elevated to a principle: the [TENSE] on the V_{AUX} is not 'strong' enough to discharge the [TENSE] in I^0, while the [TENSE] on the main verb is. This is another way of saying that main verbs have the functional equivalent of AH, and *have* and *be* must appear in I^0 if there is no Modal.

[14] In the case of auxiliaries there are additional complications, because *have* is also a main verb, in which case it displays *do*-support (in American English) in most cases (except when the direct object is an NPI, as in *Have you any idea what you just said?*). *Be* functions like an auxiliary whether or not it is a main verb.

These redundancies can be eliminated if we abandon the notion that I^0, consisting of the sequence [TENSE]-(Modal) is the head of the sentence and that more generally, empty functional heads are responsible for the derivation of surface order. Such a step has the immediate virtue of eliminating problems that I^0 introduces in the analysis of V-final languages such as German and Dutch; for discussion, see Kiparsky (1996), Ackema et al. (1993), and Sternefeld (2006: 507ff.). Since raising of V to final I^0 puts the verb in the highest position in the clause, an extraposed sentential complement must be adjoined even higher than I^0 in the clause, instead of in a position in VP following the verb.

Eliminating I^0 and other functional heads also simplifies the analysis of constituent ordering, both synchronically and diachronically. The possibility of raising V to I^0 and also of moving constituents higher in the tree to Specifier positions of various functional heads introduces too many degrees of freedom into the analytical arena.[15] For example, on the assumption that the canonical order in the VP is [V (NP) XP*], any order in which the NP and other constituents appear to the left of V must be derived by movements of the NP and XPs to the left, to the Spec positions of functional heads. If the V is assumed to move as well, e.g. to I^0, the proper analysis of any given word order is problematic at best, especially if one allows a rich set of invisible functional heads above V (as in, for example, Cinque 1999). The situation is further complicated if the possibility of rightward movement is also allowed.

The difficulties that such a derivational approach presents for the analysis of a V2/OV language are in fact highlighted by Kroch and Taylor (2000) in their discussion of verb object order in EME. They write

> we know that Old English allowed both leftward scrambling and rightward extraposition of complements and adjuncts (Kemenade 1987, Pintzuk and Kroch 1989) and these movements obscure underlying order even in the absence of verb movement (Pintzuk 1991).

These observations echo the conclusion arrived at by Rochemont and Culicover (1997) in a critique of rightward movement and uniform binary branching: in the absence of independent stipulations regarding the direction of movement, the underlying structure, and the triggers of movement, the surface structure significantly underdetermines the analysis. The surface order alone even more severely underdetermines the analysis.

The alternative that I therefore consider here is related to the one ruled out by Kroch (2003), who claims that it: "depends on a fragile assumption; namely, on the existence of directionally consistent drifts in usage over long

[15] For discussion of the problem that such a situation presents for the language learner, see Culicover (1993b).

periods of time that are unconnected to grammar change." On the view that I develop, language change is still grammar change. But the grammar consists of constructions, defined in terms of shared surface properties and systematic correspondences with meaning, and not in terms of parameter settings, e.g. whether or not the language has V-to-I. Such a formulation allows precisely the type of drift that Kroch excludes.

7.3 English constructions

7.3.1 Some core constructions

As background for an analysis of constructional change in English I summarize here the main constructions of English. These are what I call the core constructions of English, using the notion as introduced originally in Culicover (1971) and Culicover (1973). These core constructions are those that are characteristic of the language, they express the particular idiosyncrasies of the language at the most general constructional level, and together, they define what distinguishes the grammar of English from the grammars of other languages.

There are seven basic grammatical constructions that I take to constitute a significant part of the core of English.

(a) *SUBJECT* (from Chapter 2)
The constituent that corresponds to Subj precedes the inflection.[16] The inflected element may be V_{AUX} or lexical V.

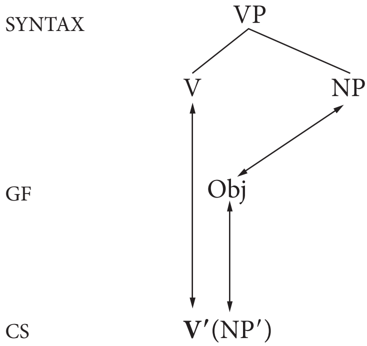
$$\begin{array}{lll} \text{SYNTAX} & [_S\ \text{XP} & \text{V[TENSE]}\ ] \\ & \updownarrow & \\ \text{GF} & \text{Subj} & \end{array}$$

(b) PREDICATE
An inflected auxiliary precedes the constituent that corresponds to the predicate (VP, AP, or NP).

$$\begin{array}{ll} \text{SYNTAX} & V_{AUX}\ \ X\ \ \text{Pred} \\ & \updownarrow\qquad\ \swarrow\!\!\nearrow \\ \text{CS} & \mathbf{V_{AUX}}'(\text{Pred}') \end{array}$$

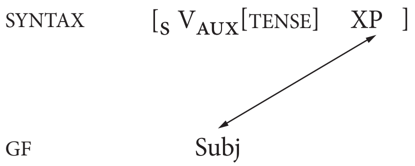
[16] The SUBJECT construction might be further generalized by including not only finite but non-finite (i.e. subjunctive, infinitival, and gerund) inflection.

(c) *Object* (from Chapter 2)
The direct object is the right sister of V.

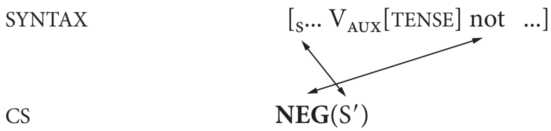


Combining (a), (b), and (c) gives us the correspondences for basic English S(Aux)VO sentences.

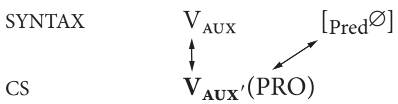
(d) *SAI*
In inversion, a finite auxiliary verb precedes the constituent that corresponds to Subj.



(e) *Negation*
Sentential negation *not* immediately follows a finite inflected auxiliary verb.[17]



(f) *Ellipsis*
A null predicate corresponds to a pro-predicate.



[17] For an auxiliary such as *won't* with contracted negation, NEG is part of its corresponding meaning.

(g) *DO*

do is the default finite auxiliary when there is no lexical auxiliary verb.

SYNTAX $[_{V_{AUX}}\text{do}[\textsc{tense}]]$

$\updownarrow$

CS **TENSE**

The fact that *do* itself lacks a correspondence with any CS function indicates that it is purely formal, that is, an expletive.[18] As the correspondence is stated, *do* may appear anywhere where an auxiliary verb can appear. How this restriction is implemented is discussed below. As we will see in §7.4.2, earlier variants of *do*-support were more restricted, in that the counterparts to (g) imposed different syntactic conditions on the distribution of expletive *do*.

Let us consider now how these constructions combine to license a sentence. Each construction states a set of conditions which a correspondence may satisfy in order to be considered grammatical. Lexical entries state form–meaning correspondences at the word and idiom level. For example, (16) states a correspondence for the verb *see*.

(16) *see*

SYNTAX $[_V \text{see}]$

GF Subj Obj

CS **SEE**(EXP:X,THEME:Y)

By default, the Subj GF links to the Experiencer or Agent role in CS, and the Obj links to Theme (Dowty 1991). If there is only one argument, by default Subj links to Theme. I refer to the correspondences between GFs and CS arguments collectively as thematic linking (TL).

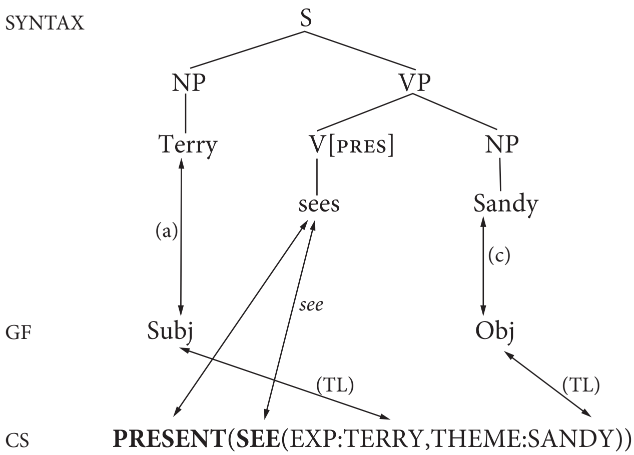
It is well known that the inflection appears on the main verb only if there is no auxiliary. This generalization is captured by constructions (a) and (b), since the subject immediately precedes the inflected element, and an inflected auxiliary precedes a predicate, including VP. This means that there is nothing corresponding to I^0 in a sentence with an inflected VP.

[18] It is interesting to note that unemphatic *do* and *did are* already referred to as expletives by Burn (1766) (quoted by Visser 1973: 1509): "John *Burn*, who in his A Practical Grammar of the English Language (*1766* p. 50) declares that 'Do and did, besides marking the time of the action, silently imply opposition and emphasis; otherwise Do, and Did, become mere *expletives*, and ought not to be used.'"

Nevertheless, the finite inflection on V has a CS function, just as it does when it appears on an auxiliary. So we must assume that V[TENSE] corresponds to CS representation **TENSE**(**V′**); thus, *sees* corresponds to **PRESENT** (**SEE**), and so on.

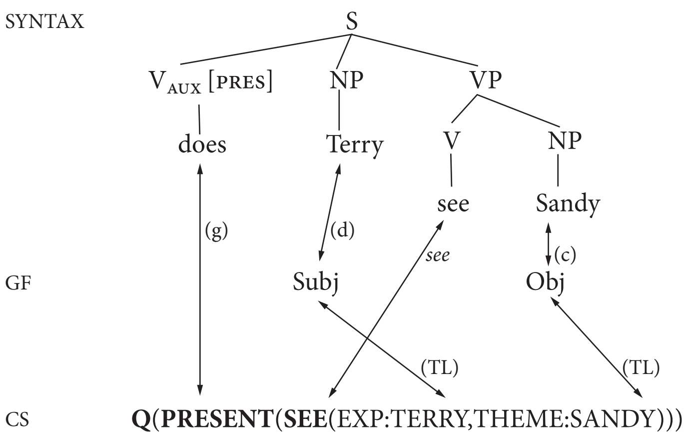
Take the sentence *Terry sees Sandy.* In (17) I mark each of the correspondences with the construction that licenses it.

(17)



The correspondence for *Does Terry see Sandy* is the same as the one for *Terry sees Sandy*, except that the Subj of the sentence is licensed by SAI (d) instead of SUBJECT (a). Note that the **Q** operator does not correspond to a particular element in the syntactic description, but is present in the CS in virtue of the SAI construction.

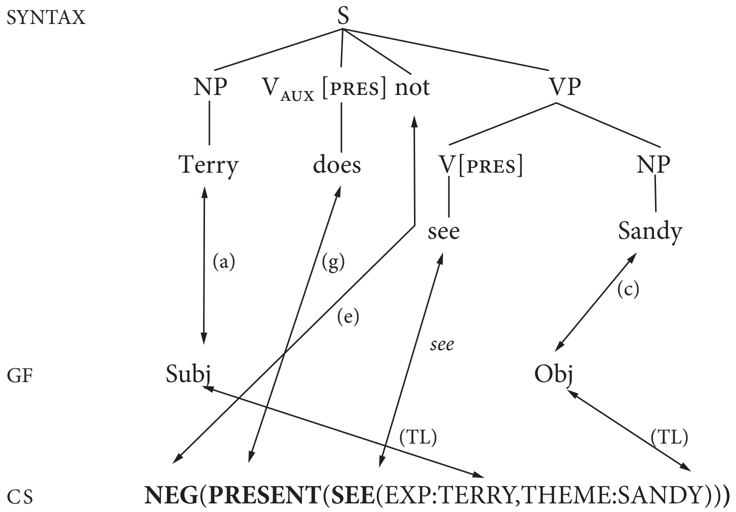
(18)



This representation presumes that the yes-no question construction incorporates SAI (d).

The correspondence for *Terry does not see Sandy* is given in (19).

(19)



On a constructional approach, a sentence is ungrammatical if there is an element that is not licensed by any construction. For example, if we place *not* after the tensed verb—**Terry sees not Sandy*—, *not* is not licensed, since the construction that licenses sentential negation in English is (e). *Not* must follow V_{AUX} and precede the VP in order to correspond to **NEG.**[19]

On the other hand, the sequence *do*[TENSE]-V is licensed, according to the constructions as we have formulated them thus far. But *do* is typically not used in simple affirmative sentences unless it expresses focus of some type. Hence sentences of the form (20a) are extremely rare, compared with sentences of the form (20b).

(20) a. Terry does see Sandy.
b. Terry sees Sandy.

Figure 7.1 shows the relative frequency of the two constructions as calculated by the Google N-gram viewer.

In fact, the generalization about *do*-support is that *do* appears only where V_{AUX} *must* appear. To understand why this is so, compare the correspondences for *do*[PRES]-V in (21) and V[PRES] in (22).

[19] *Not Sandy* is licensed by the construction *not X but Y*, however: *Terry sees not Sandy, but Leslie.*

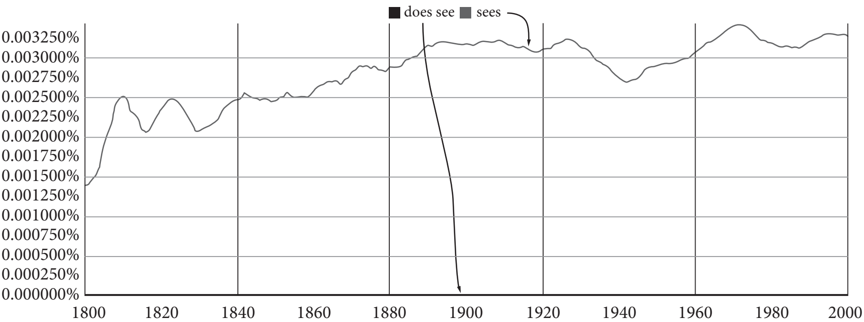


FIGURE 7.1. Frequency of *does see* and *sees* in Google N-gram viewer, 1800–2000

(21) SYNTAX do[PRES]] V

CS **PRESENT**(V′)

(22) SYNTAX V[PRES]

CS **PRESENT**(V′)

(21) is locally more complex than (22), because (21) states a more elaborate set of conditions for the same correspondence with CS than does (22). It is plausible, therefore, that other things being equal, correspondence (22) has driven out correspondence (21) over the course of time, along the lines discussed in Chapter 6 regarding competition between constructions in the social network (see also Kroch 1989). As discussed in greater detail in §7.4.2 below, the history of *do*-support shows that the two constructions were in fact in competition from the inception of *do* in the fifteenth century until the mid-eighteenth century. Construction (21) has been driven out by (22) with regards to expressing the correspondence with **PRESENT**(V′).

But the form in (21) is alive and well when there are other conditions added to it that do not require it to compete head-to-head with (22). For example, when *do* is strongly accented it is interpreted as *verum focus* or as contrast on the temporal reference of tense (23a). Weaker accent on *do* has a more subtle function (23b), one of affirmation. ((23b) should be read with an even accent on *does*, *see*, and *Sandy*.)

(23) a. Terry DOES see Sandy.
b. (Does Terry see Sandy?) Yes, Terry does see Sandy, but not very clearly.

I come back to the status of *do*-V in the discussion of the evolution of *do*-support in §7.4.2.

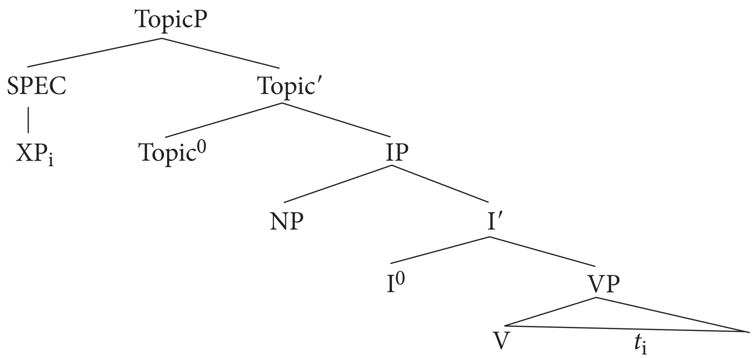
7.3.2 Ordering the correspondences

With these basic correspondences in place, we are now in a position to consider more complicated matters having to do with constituent ordering on the left periphery in English. The SUBJECT construction as stated licenses the Subj in the initial position of the clause. But it is possible for a constituent to precede the subject in A′ position. E.g.,

(24) An analysis like that, John could never subscribe to.

The classical method of accounting for this is to assume that there is a particular location in the syntactic structure into which the topic goes, along the lines of Rizzi (1997) and virtually all derivational analyses. (25) illustrates.

(25)



The ordering of topic and subject is then accomplished by ordering the topic position leftmost in the TopicP,[20] the subject position leftmost in IP, and letting IP be a right daughter of TopicP.

As Rizzi's (1997) analysis clearly shows, however, it is an oversimplification to say that the ordering is Topic-Subject. Although topics in English do precede the subject, they may precede the complementizer position in *wh*-questions, or follow the complementizer position in *that*-clauses.[21]

[20] This may be an oversimplification, since some sentences appear to contain more than one topic constituent. E.g.,

(i) a. Last week in Chicago, Sandy visited the Art Institute.
b. As for me, syntax, I'll never understand.

In both cases, the second topic is not leftmost. On further analysis, however, (i.a) might be analyzed as a composite topic, while *as for me* in (i.b) might be analyzed as extra-clausal. I leave these complications for future research.

[21] This sentence is a naturally occurring case of non-emphatic *do*-V.

(26) a. An analysis like that, how could anyone actually subscribe to?
b. It is abundantly obvious that an analysis like that, John would never subscribe to.

Therefore, on a structural analysis along the lines of (25) it is necessary to have (at least) two Topic^0 functional heads and two landing sites for a topicalized phrase. Such an account clearly misses a generalization, which is that the topic always precedes the part of the clause that provides the content; the complementizer *that* is simply a formal device that signals subordination.

It is interesting to observe that if we order the correspondences in terms of their application to a string of constituents, it is possible to reduce if not entirely eliminate such reference to structure. Assume that the topic expresses a well-defined correspondence with information structure (IS) called *Topic*; defining precisely what the precise function of topic is is a separate matter that I set aside here.[22] Reducing matters to their essentials, what we want to say is simply (27)—

(27) *Topic*: Topic is leftmost.

—where Topic is an XP that corresponds to a particular IS configuration. A representation that expresses this generalization is given in (28).

(28) *Topic*

$$\begin{array}{ll} \text{SYNTAX} & [_S\ \text{XP} \ldots] \\ & \quad\ \updownarrow \\ \text{IS} & \quad \text{Topic} \end{array}$$

As we have just seen, it is not possible to state the property 'clause-initial' straightforwardly in structural terms, because the initial part of one type of clause does not look like the initial part of another type of clause. But suppose that we consider the linear ordering of constituents, without considering the structure, e.g., (29), where XP is a topic and NP a Subj.

(29) XP NP V[TENSE] ...

Suppose further that in the CS/IS representations, XP corresponds to Topic and NP corresponds to Subj. Then the sequence of constituents in (29) satisfies (27), since XP counts as Topic.

Consider next the position of the subject. If there is no topic, then the subject is initial in the sequence. (There are of course constructions that locate the subject elsewhere.) If there is a topic, and we check off the topic as

[22] For a review, see Roberts (forthcoming).

clause-initial, then the subject is initial in the string of constituents following the topic. The simplest generalization about the position of the subject is the following, which paraphrases the Subject construction (a).

(30) *Subject*: An XP that corresponds to Subj is leftmost.

Now, the correspondences Topic and Subject conflict, since they both say that XP is leftmost. It is clear how to eliminate the conflict: order the correspondences so that Topic applies first to the string of constituents, and then Subject applies to the *remaining* string of constituents. For example, if the string is (29), XP is leftmost in the entire string, satisfying Topic, and the NP is leftmost in the remaining string, satisfying Subject.

This ‘dynamic’ approach to applying correspondences reveals an interesting relationship between structure and order. As Kayne (1994) demonstrated, it is possible to encode order in terms of structure, given certain assumptions. By the same token, it is possible to encode structure in terms of order, again given certain (but different) assumptions. In the latter case, we make it a part of the theory that a correspondence applies only to the portion of the string that is not already accounted for by a prior correspondence. Let us call this portion the ‘(linear) domain’ of the correspondence. By ordering the correspondences appropriately, it is possible to eliminate redundancies and exceptions. In particular, it is not necessary to mention the right context of a correspondence if the following holds:

- by default every correspondence applies from the left of its domain,
- the correspondences are (partially) ordered, and
- the linear domain of a correspondence excludes what is in the domain of an correspondence ordered before it.

What this means is that although Topic and Subject both mention ‘[$_{S}$ XP’ as a syntactic condition, (29) is fully licensed if we order Topic before Subject, as in (31).

(31) Topic > Subject

Continuing, the Predicate construction (b) is ordered after Subject, and VP constructions such as Object are ordered after Predicate. Crucially, the Predicate construction specifies the location of the tensed verbal element—it is leftmost following the subject (unless of course it is preempted by a construction based on SAI). The ordering of the basic constructions is given in (32).

(32) Topic > Subject > Negation > Predicate > Object (and other VP constructions)

Finally, we must specify that the complementizer C^0, e.g. *that*, if it is present, precedes the topic.

(33) *Complementizer*
C^0 is leftmost in its clause.

SYNTAX $[_S\ C^0 \ldots]$

This construction is ordered before Topic. The ordering of constructions then produces the canonical linear constituent order for English in (34).

(34) (C^0) (XP_{TOPIC}) $NP_{SUBJECT}$ V[TENSE] ...

There are of course alternative constituent orderings in English. These are licensed by other constructions, which have the property of relating particular constituent orderings to particular meanings, as we have seen.

7.3.3 Constructional domains

Consider next yes-no and *wh*-questions. The constructions are given in (35) and (36).

(35) *Yes-no question*

SYNTAX $[_S\ V_{AUX}$[TENSE] XP]

GF Subj

CS **Q**(S′)

(36) *Wh-question (Root)*

SYNTAX $[_S$ XP[WH] V_{AUX} [TENSE] XP ...]

GF Subj

CS **Q** (S′)(XP′[WH](α))

Both constructions incorporate SAI. WH-QUESTION makes use of a device introduced in Chapter 3 for interpreting relative clauses. The S corresponds to a lambda expression in CS, the *wh*-phrase in initial position corresponds to an argument of the lambda expression, and the *wh*-phrase contains a variable corresponding to the *wh*-word, which is bound by the operator, in this case, $\mathbf{Q}^{\boldsymbol{\alpha}}$.[23]

The following sentences show that a topic may precede a yes-no question and a *wh*-phrase in a *wh*-phrase.

(37) a. Tomorrow, where are we going to stay?
b. To Sandy, what do you plan to say

Hence the ordering of constructions applying to the periphery may be further elaborated as (38).

(38) $\text{COMPLEMENTIZER} > \text{TOPIC} > \left\{\begin{array}{l}\text{SUBJECT}\\ \text{YES - NO -Q}\\ \text{WH - Q}\end{array}\right\}$

Thus we arrive at the full ordering in (39).

(39) $\text{COMPLEMENTIZER} > \text{TOPIC} > \left\{\begin{array}{l}\text{SUBJECT}\\ \text{YES - NO -Q}\\ \text{WH - Q}\end{array}\right\} > \text{NEGATION} > \text{PREDICATE} > \text{OBJECT}$

This analysis assumes that the positions of complementizer *that* and clause-initial *wh*-phrases are distinct. The fact that they do not appear together is then a matter of interpretation, not syntax. *That* marks an embedded declarative, while a *wh*-phrase is interrogative, and each excludes the other.

7.3.4 do-*support*

Suppose now that we have an instance of a construction that incorporates SAI, and the only relation in CS is **F**, which corresponds to a main verb in the syntactic structure. In order for SAI to be licensed, [TENSE] in the interrogative must be realized on V_{AUX} to the left of the constituent that corresponds to Subj. A constraint that says that an element of a certain category must appear in a certain position is understood to specify that there must be something in the syntactic structure that participates in the correspondence. Thus, in order for SAI to be licensed there must be an expletive if there is no available lexical item. In this case, requiring that V_{AUX} must be leftmost in its domain means

[23] A correspondence where the S does not correspond to a lambda expression is ill-formed, because the *wh*-phrase cannot be assigned a function in CS. This is the explanation for sentences such as (i).

(i) *What did you eat the banana?

that in this case, **TENSE** will correspond to *do*[TENSE] if there is no operator in the CS that corresponds to V_{AUX}. Thus in English the default realization of V_{AUX}[TENSE] is *do*[TENSE], just as the default realization of Subj is *it*.

This analysis of *do*-support generalizes from cases of inversion to other cases where [TENSE] cannot be realized on a verb, but must be realized on V_{AUX}. This occurs when there is negation, ellipsis, or VP topicalization, all of which include V_{AUX} in their descriptions.

Since *do* occurs only if a construction explicitly mentions V_{AUX}[TENSE], the redundancies and adjacency problems raised in connection with the classical analysis and the contemporary reformulation do not arise. For example, in contrast with NEGATION, which specifies the position of *not* with respect to V_{AUX}, there is no construction that specifies the position of adverbs with respect to V_{AUX}. Rather, there is a construction that specifies that an adverb may be a left sister of a predicate. This construction licenses strings such as *will certainly call* and *probably called*. Therefore we do not get *do*-support when there is an adverb preceding the main verb.

There is no loss of generality in the location of [TENSE] in this case, since the correspondence that specifies the position of [TENSE] says quite generally that it appears on the leftmost verb in the sequence, following the subject. If this verb is a main verb, a modal, or an auxiliary verb, that is where [TENSE] will appear. And if a particular construction stipulates that there must be a V_{AUX}[TENSE] in a particular position, that is where *do*[TENSE] will appear.

The cost of such an analysis is that it requires that every construction in which there is *do*-support must explicitly mention the location of V_{AUX}. This includes not only SAI and NEGATION, but ELLIPSIS, gapping, pseudogapping and related constructions, and VP-topicalization.[24]

The history of *do*-support may now be viewed as the generalization of *do* from more highly specified constructions to the status of the default V_{AUX}[TENSE] that satisfies all correspondences in which the position of V_{AUX} is specified. The enabling conditions on this development are therefore the following:

- The emergence of the category V_{AUX}.
- The emergence of constructions that explicitly mention the position of V_{AUX} as a competitor to V.

In the next section, I review briefly the historical record that bears on this account. The emergence of *do* as an expletive V_{AUX} was preceded by its occurrence in a number of more restricted and specialized constructions.

[24] *Do*-support occurs also in pseudogapping. Ellipsis and VP-topicalization also occur in the environment of infinitival *to*.

The emergence of constructions that specify V_{AUX}, and in particular SAI, is the well-attested consequence of the increasing limitations on V2 in the course of the transition from OE and ModE.

7.4 Tracing the changes

I start with the premise that the best account of the change from OE to ModE is one that makes minimal assumptions about what changed, and imposes minimal stipulations on the various stages and the changes themselves.

The search for a minimal account is plagued by the fact that there are so many degrees of freedom in what counts as a possible solution.[25] Accounts in MGG try to limit the solution space by assuming an analysis of ModE in terms of I^0 and V-to-I, and formulating the changes in terms of the behavior of V with respect to I^0, and more generally in terms of movement, e.g. of constituents (such as DPs) to the Spec positions of functional heads (e.g. Agr^0). But the centrality of movement to the analysis, even with relatively specific constraints, poses significant challenges to attempts to narrow the domain of possible scenarios (see for example the quotation from Kroch and Taylor 2000 in §7.2.2).

Clearly, the details of an account of the change depend crucially on the characterization of the endpoint of the change. In the preceding section I outlined a grammatical account of ModE that does not assume movement. Rather, the relative positions of constituents of the sentence are accounted for in terms of correspondences between particular syntactic forms and CS representations. These constructions are ordered in such a way that they license certain constituent sequences through their correspondence with CS representations. There are two key components of this analysis:

- the realization of [TENSE] on the first V and default realization of *do* when there is no V_{AUX} in a construction that requires V_{AUX};
- the positioning of the subject.

These constitute the locus of the change. I believe that focusing on these aspects of the grammar permits a more faithful account of the actual facts seen in the course of time, and opens up some interesting possibilities for establishing causal relationships.

The historical developments that led to the current state of affairs in ModE are well documented. Here I review only those aspects that are central to the constructional story. In §7.4.2 I discuss the history of *do*-support itself, and

[25] For a similar observation in the domain of language acquisition, see Culicover (1993b).

review evidence that the occurrence of periphrastic *do* is a widespread phenomenon in languages of the world. §7.4.3 summarizes the transition from general V2 to SAI. This transition involves the emergence of the category V_{AUX}, the loss of V2 in cases of topicalization first when the subject is pronominal, and positional licensing of the subject.

7.4.1 Periphrastic do

As I pointed out in §7.2, the analysis of contemporary *do*-support is faced with two main problems: the redundancy of AH and the fact that adverbs do not enter into the analysis. These problems arise in a different form in the analysis of OE, since it must be assumed that before there was *do*-support, negation did not block AH (or its functional equivalent). Kroch (1989) suggests that in OE, as presumably in other languages, following the analysis of Zanuttini (1991), negation was a specifier and became a head over a period of time.

But on Zanuttini's analysis, in a number of languages negation is a head. These languages don't have *do*-support. So we have to assume that these languages don't have AH either. Thus the situation in English turns out to be a remarkable coincidence, in which a language that is apparently much like German in its syntax simultaneously manages to develop AH, negation as a head, and *do* as a dummy modal.

On the other hand, varieties of *do*-periphrasis are well attested in languages of the world. For example, Benincà and Poletto (2004) describe the Monnese dialect of Northern Italian in which *fa* 'do' is obligatorily used in yes-no and non-subject *wh*-questions, inverting with the subject. Some illustrative examples are given in (40).

(40) a. fa-l majà?
does-he eat?
'Does he eat?'

b. Ke fa-l majà?
what does-he eat?
'What does he eat?'

c. *maja-l?
*eats he?

d. *ke maia-l?
*what eats-he?

e. a-l majà?
has-he eaten?
'Has he eaten?'

f. ke a-l majà?
what has-he eaten?
'What has he eaten?'

g. *ke fa-l aver majà?
what does-he have eaten?

h. fa-l plöer?
does-it rain?
'Is it raining?'

i. a-l plöt?
has-it rained?
'Did it rain?'

j. *plöe-l?
*rains-it?

Crucially, *fa*-support in Monnese alternates with V-to-I. V-to-I occurs with negation, while *do*-support occurs in questions.

(41) a. l so mia
it I-know not
'I do not know it'

b. *fo mia savé-l
I-do not know.it

These examples are consistent with the proposal that Monnese has a form of *do*-support, limited to questions. But as Benincà and Poletto suggest, *fa*-support in Monnese can be seen as a type of light-verb construction, which has generalized to most of the lexicon and which is restricted to questions.[26]

The analysis of Monnese suggests that a similar development may have occurred in OE as a precursor to modern *do*-support. Jäger (2006) demonstrates that periphrastic *do* is very widespread in the world's languages. Its distribution does not appear to be determined by genetic relatedness or by geography. It does not appear to be correlated with particular grammatical properties of a language. It has many functions across languages, some of which are strictly formal while others are pragmatic.

[26] Their account seeks to explain such a state of affairs in more abstract terms, but this is the essence of their empirical observation.

There are, however, several distinct types of morphosyntactic environments in which *do*-periphrasis appears. According to Jäger (2006: 92), there are two types of *do*-periphrasis that apply to English:[27]

Type 1. The appearance of lexical or morphological material in the clause triggers verbal periphrasis, in most cases material that attaches to the lexical verb and thus prevents the realisation of regular verb morphology. This material usually belongs to a closed class and its function is similar to that of regular verb morphology, i.e. verbal categories, and/or adverbial modification.

Type 2. If a language has rigid or dominant word order, periphrasis is used to mark clause types that display a deviant or irregular word order or to maintain a close approximation of the regular word order in these, i.e. to keep the relative order of verb and object unchanged. Focalisation, topicalisation and interrogativity are the most common functions that can be associated with periphrasis in this context.

On this view, English *do*-support may originally have had the consequence of keeping certain verbs free of inflectional morphology, and of keeping the main verb and its complements adjacent. Both may be viewed as having to do with the minimization of complexity. The first reduces paradigmatic complexity, and the second minimizes the domain in which the dependency between the verb and its complements can be computed, in the sense of Hawkins (2004).[28] Of course, separation of the finite inflection from the main verb introduces its own complexity, as noted earlier.

Jäger's study shows that English is by no means unique in having *do*-periphrasis. This is to be expected, if the motivations for *do*-periphrasis in fact have to do with the reduction of complexity. What is characteristic of English is the spread of *do*-periphrasis from particular lexically constrained environments to specific syntactic environments, and its ultimate generalization to all environments where V_{AUX} is constructionally specified. However, there appears to be no reason in principle why similar developments could have not occurred in other languages. It would be reasonable to expect similar, but not identical, patterns to be found in a cross-linguistic study of the diachrony of *do*-periphrasis.

[27] See also van der Auwera (1999) and Wichmann and Wohlgemuth (2008).

[28] From this perspective it is no accident that restricted forms of *do*-support occur in non-standard German and German caretaker speech to children, e.g. *Er tut Bücher schreiben*. Not only is V in a predictable position close to its complements, but irregularities of verbal inflection are avoided, since the main verb always has the infinitival form.

7.4.2 Growth of English do*-support*

The historical development of *do*-support has been amply documented in the literature. Kroch (1989) has carefully and intensively analyzed the data in Ellegård (1953) to demonstrate that *do*-support developed in each of the contexts in which it currently appears in ModE, at the same rate, but at different stages. There are six contexts distinguished by Ellegård (1953) that Kroch tracks: negative declaratives (*he didn't go* ~ *he went not*), negative questions (*didn't he go* ~ *went he not*), affirmative transitive adverbial and yes/no questions (*did he eat the pizza* ~ *ate he the pizza*), affirmative intransitive adverbial and yes/no questions (*did he go* ~ *went he*), affirmative *wh*-object questions (*what did he eat* ~ *what ate he*), and 'contact' *do*, where *do* immediately precedes V (*he did eat the pizza* ~ *he ate the pizza*). Each of these grows in the percentage of cases that *do* is used vs the percentage of cases where it could be used, from the period 1400–25 to the period 1650–75, with the exception of 'contact' *do*. This use peaked in the mid-sixteenth century and then declined until it was virtually unattested by the end of the seventeenth century.[29]

Kroch argues that the rate of development of each use of *do* is constant, but that some uses are more advanced in time than others. This can be seen from the data in Figure 7.2 for the last period (1650–1700), where the percentages and total number of attested examples for the six cases are Ellegård's as represented in Kroch's tables.

What is interesting about this data is that *do*-support was virtually obligatory in some cases, optional in others, and virtually non-existent in still others. This variability is found throughout the period of change. The subsequent development to ModE takes all of the optional cases to obligatory.

There is an apparent incompatibility between the variability of *do*-support at various stages of the language, and the theoretical account in terms of I^0 and V-to-I. Kroch (1989) suggests that at these intermediate points in the development of the language, there were two grammars, one with V-to-I for main verbs and the other without V-to-I. While such a formal solution does provide a way of accounting for apparent optionality, it is a rather crude tool, since in itself it cannot account for frequency data such as Figure 7.2.

It is possible to enrich the multiple grammar account with a device that determines the percentage of cases in which each grammar applies. Since the percentages are different for the different morphosyntactic contexts,

[29] There is actually some question whether 'contact' *do* has actually died out completely; see the discussion of example (23).

Date	Negative declaratives		Negative questions		Affirmative transitive adverbial & yes/no questions		Affirmative intransitive adverbial & yes/no questions		Affirmative wh-object questions		Affirmative declarative contact do	
	%	N	%	N	%	N	%	N	%	N	%	N
1650–1700	46.0	274	92.3	52	94.7	76	70.2	131	54.9	51	0.92	7426

FIGURE 7.2. Frequency of *do*-support in various environments at stage 1650–1700 (adapted from Kroch 1989)

this device must encode this grammatical information as coefficients of occurrence on individual rules. Since the percentages are different for different speakers (or different writers), the coefficients must vary from speaker to speaker.

Furthermore, in order to account for the development, we must assume a mechanism that adjusts the percentages over time, for each morphosyntactic context. Kroch has convincingly demonstrated that the rate of change in each case is precisely what we would expect if the various pieces of grammatical knowledge, e.g. *do*-support in questions, *do*-support in negative contexts, etc., spread through the population following a typical population dynamics pattern. Since each context is developing independently, there appear to be separate mechanisms that encode the frequency information for each context.[30]

[30] Moreover, since the rate of spread of each context of *do*-support is constant, Kroch concludes that the loss of V-to-I is spreading from one context to another. Thus the constant rate effect is often taken to be evidence for the existence of the rule V-to-I. However, it can equally well be taken to be evidence for the supplanting of V by V_{AUX} in the relevant set of constructions. This gradual spread of a property of constructions through the set of constructions as well as through the population is what a constructional theory would predict. This point is taken up in more detail in Chapter 8.

Recall the analogy drawn in Chapter 6 between constructions and viruses. Viewed from this perspective, the rate of spread of constructions has a direct counterpart in the spread of infection through a population. Jackson and Rogers (2007) show that the rate of diffusion of infection is a property of the network structure, including the particular nodes. They assume a view of the network and transmission of information that is essentially the same as that in Culicover and Nowak (2003) and Chapter 6. Given this, we would expect to find that similar innovations spread at the same rate through the same network, even those that are not formally or functionally related to one another.

This situation is of course precisely what we would expect if the knowledge in this area of the language developed not as a fully general grammatical rule, but as a more narrowly specified construction.[31] On such a scenario, a construction that is defined in terms of particular lexical items and particular grammatical contexts grows over time in terms of both, along the lines sketched out in Chapter 6. It extends to additional lexical items, and it spreads to other grammatical contexts. As it develops, it bleeds the existing rule, which continues to apply as the default. That is, English acquired *do*-support, but still has the construction that produces the appearance of V-to-I as the default.

It is also noteworthy that 'contact' *do* follows a very different path than do the other uses of *do*. It appears early, increases in use (to less than 10 percent of the possible environments) around 1550, and then becomes very rare by the period shown in Figure 7.2.

This use of *do* is also incompatible with the derivational V-to-I analysis, since there is no reason why *do* should be inserted if AH (or the equivalent) is already available in the language to derive V[TENSE], which clearly was the case. It seems, rather, that the 'contact' *do* is a distinct construction that exists side by side with the construction that licenses [TENSE] on the main verb, as discussed in §7.3.1 ((21)–(22)).

The plausibility of this account of the development of *do*-support is buttressed by the observation that it does not simply apply optionally in the language of individual writers, in the sense that it applies in less than 100 percent of the contexts in which it could apply. Rather, it appears to apply with respect to particular verbs, and not with others, as Ellegård's data makes very clear. For example, he writes (166) "Of Machyn's 370 'do'-instances, 216 involve the verb 'preach'; the simple verb 'preach' occurs only half a dozen times" and (167) "In *Polychronicom* there are 816 'do'-instances, 243 with 'slay' (and no finite 'slay' in the past tense), as well as 70 with 'succeed', 69 with 'write', 19 with 'eat', 8 with 'fight', 7 with 'hold', 7 with 'appear' and 4 with 'add'' and "we note that verbs or phrases singled out for preferential treatment are not always the same with different writers. We should add to the list *Cely Papers* /204/ 'do well understand', *Fitz James* 'appear', *Decaye of England* 'think', *King James Bible* 'eat'."

Ellegård shows that certain verbs, such as *know*, *say*, and *think*, "resisted the usage of *do*-support, at least in negatives, until as late as the nineteenth century" (Trudgill et al. 2002: 5; see also Nevalainen 2006: 576 and Nurmi 1999).

[31] Occasionally I use the term 'rule' to refer informally to constructions that are fully general. This is simply a terminological convenience.

Along the same lines, van Ostade (1987) looked at the occurrence in the early eighteenth century of *do*-support and non-*do*-support in comparable constructions, e.g. SAI and negation. The specific case are:

(42)
1. negative sentences without *do*
2. questions without *do*
3. SAI without *do*
4. *not*+non-finite (*I not like him*)
5. unemphatic *do*
6. *do*+adverbial+V
7. exclamatory *wh*-
8. *if*-less conditionals (with SAI)
9. SAI with initial adverbial
10. SAI with initial object

Her data show that even in the eighteenth century, the occurrence of *do*-support was far less uniform than it is in ModE. It is clear from van Ostade's data that *do*-support in the eighteenth century was spreading to the full range of potential syntactic environments, only some of which would turn out to preserve it in later centuries. The amount of spread is sensitive to the individual writer, to the style of writing, and to the main verb of the sentence. Sentences with main V-*not* (e.g., *I know not*) occur with different frequencies for different authors. Furthermore, such sentences occur with a small set of verbs (*know*, *doubt*, and a few others, varying by author), suggesting that by this point, main V-*not* had become a lexically constrained construction that was already on its way out.

The same verbs are involved in van Ostade's type 3 (SAI without *do*). But, crucially, the distribution for each author does not appear to correlate (van Ostade 1987: 167): "A high or low proportion of one construction does not necessarily correspond with a comparable proportion for the other one." This observation suggests further that main V-*not* and SAI without *do* are independent constructions in each author's grammar, and argues against the 'loss of V-to-I' interpretation of the changes.

Van Ostade also finds a number of uses of *do*-support that were on the verge of dying out. Again, this is what we would expect if there can be independent (but related) constructions in a grammar which sometimes merge into a more general construction, and sometimes splinter into distinct constructions with their own conditions. For instance, SAI in counterfactual conditionals, which is now restricted to *have* and (marginally) *be* in ModE, permitted *do* in the eighteenth century (van Ostade 1987: 44 and §3.4.8). However, Types 6 (*I do firmly believe*) and 7 (*What dreadful days do we live in*) are found in sufficient numbers to suggest that they are still active in the eighteenth century, and die out only later (p. 123).

These observations suggest further that *do*-support began as a lexically restricted construction. Its development as a fully general construction in English is a consequence of its spread through the population, and through the constructional lexicon, until it was generalized quite free of particular items and contexts. The absence of *do*-support in certain contexts, such as Type 6, may be explained in terms of a shift in the structural conditions of the construction. In the case of Type 7, however, a more natural explanation would be in terms of the contexts in which enabling conditions occur. That is, SAI no longer applies in *wh*-exclamatives, so there is no V_{AUX} in the construction, and hence no *do*-support. The next section takes up the shift from general V2 to conditioned SAI.

With the foregoing scenario in mind, let us consider more closely the development of *do*-support in constructional terms. The independent constructions are those tracked by Ellegård. In (43) I summarize just the syntactic conditions for each one and the corresponding conditions prior to the emergence of V_{AUX}. They are listed in the order in which they appear in Ellegård's data.

(43) a. affirmative transitive adverbial and yes/no questions
do[TENSE] NP V NP ~ V[TENSE] NP NP

b. negative questions
do[TENSE] NP *not* V ~ V[TENSE] NP *not*

c. affirmative intransitive adverbial and yes/no questions
do[TENSE] NP V ~ V[TENSE] NP

d. negative declaratives
do[TENSE] *not* V ~ V[TENSE] *not*

e. affirmative *wh*-object questions
NP[WH] *do*[TENSE] NP V ~ NP[WH] V[TENSE] NP

f. 'contact' *do*
do[TENSE] V ~ V^0[TENSE]

The counts for each construction arrived at by Ellegård are shown in Figure 7.3. Note that in the corpora studied by Ellegård, periphrastic *do* originated in negative yes-no questions before or around 1400, began to spread to affirmative yes-no questions, intransitive yes-no questions and negative declaratives by 1450, and spread to *wh*-questions by 1500. *Do* in affirmative declaratives, not shown in this graph, rose to about 10 percent of the possible cases by the middle of the sixteenth century, and then declined gradually towards 0 percent by 1750.

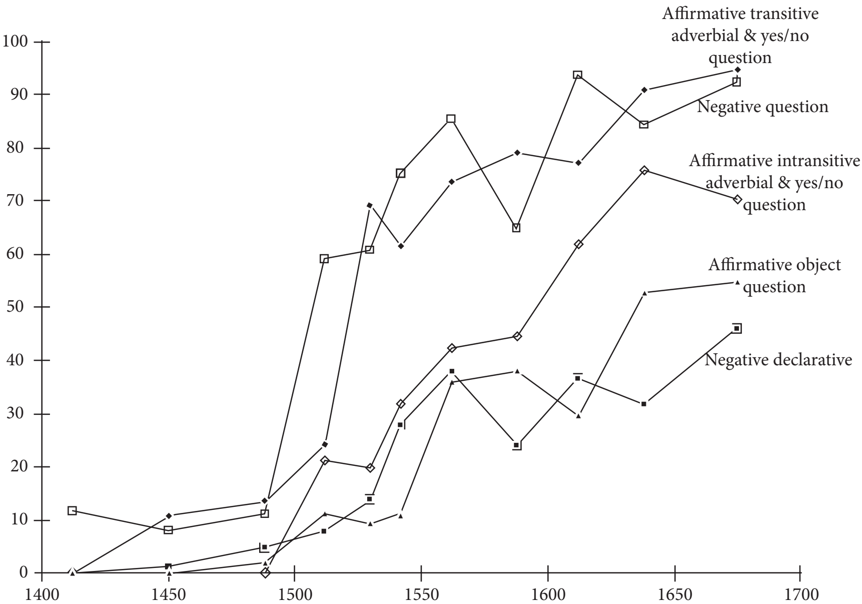


FIGURE 7.3. Occurrences of periphrastic *do* identified by Ellegård 1953 (from Kroch 1989: 22)

A plausible way to interpret the developments charted in Figure 7.3 is to understand *do* as an expletive V_{AUX} from the very beginning, and to take the emergence of *do*-support to be evidence of gradual encroachment of the category V_{AUX} over the correspondences previously expressed in terms of V[TENSE]. On this view, (43a) is a reflection of the replacement of (44a) by (44b), and similarly for the other cases.[32]

(44) a. V[TENSE] NP NP
b. V_{AUX}[TENSE] NP V NP

Given the emergence of similar constructions in other languages, e.g. Monnese, it is plausible that *do* is used to make it possible to restrict inversion to V_{AUX}. The virtue of this innovation, as noted above, is that it keeps the main verb near its complements.

Once established as an expletive for (43a), *do* is readily recruited for each stage of generalization of inversion to V_{AUX}, as well as to negation. On this view, the growth of *do*-support is the symptom of the victory of V_{AUX}[TENSE]

[32] Negative declaratives can be seen as loss of V2, since we get V[TENSE]-*not*-NP because there is a subject in initial position.

over V[TENSE] in one context after another. This course of events thus marks the transition in English from V2 to SAI, which I discuss in more detail in the next section.

7.4.3 Transition to SAI

The transition from V2 to SAI involves several critical changes. One is what Kiparsky calls 'positional licensing' of the subject, which I express in terms of the SUBJECT construction. Kiparsky (1997: 461) argues that there is a clear connection between the rise of a fixed position for the subject and the loss of overt case-marking.

> That there is a relationship between the loss of inflectional morphology and the development of rigid positional constraints is clear from comparative syntax. The most important point about this relationship is that it is not a vague correlation or tendency, as often assumed, but an exceptionless implication, which however holds in one direction only: lack of inflectional morphology implies fixed order of direct nominal arguments (abstracting away from Ā-movement of operators).... Every Germanic language which has lost case and agreement morphology, whether VO (English, Swedish, Danish, Norwegian) or OV (Dutch, West Flemish, Frisian, Afrikaans), has imposed a strict mutual ordering requirement on its nominal arguments, without changing the headedness of its VP. The order is always that subjects precede objects, and indirect objects (NPs, not PPs) precede direct objects.

A second change is the formation of a subclass of modal verbs that lack inflection for agreement in the present (and the past). This is a distinctive characteristic of ModE. The development of this subclass is well documented; see for example Denison (1993) and for a recent study, Bybee (2006: chapter 16). The changes that occurred in the transition to ModE were syntactic, semantic, and morphological:

- The modals lost their ability to take non-infinitival complements, and they lost their root meanings. For example, *cunnan* meant not only 'be able to', but 'know,' and *willan* meant 'wish, want.'
- The modals acquired epistemic meanings, having to do with possibility and necessity.
- The modals acquired contraction of *not* to *n't*. According to Denison (1993: 309), this is recorded "first from the fifteenth century, with assimilation of elision of part of the modal."
- The modals lost all person inflection.
- The infinitive lost its inflection.

Since, as main verbs, inflected modals functioned as the head of the S even in OE, we do not find any changes in their behavior with respect to word order in questions and negative sentences. But as they became reanalyzed as V_{AUX}, the door was opened for *do*-support as the dummy V_{AUX}.

In German, modal verbs do not govern VP-ellipsis; rather, they require a pronominal complement.

(45) Ich kann dieses Buch nicht verstehen, aber Peter *(es) kann.
I can this book not understand but Peter it can
'I can't understand this book, but Peter can (*it).'

The same use of pronouns is found in earlier forms of English (Denison 1993: 307). But Warner (1993: 112) gives many examples from OE of ellipsis with modal and auxiliary verbs. Here is just one.

(46) Wenst ðu đæt se godcunda anweald ne mihte afyrran đone anweald đam unrihtwisan kasere, . . . gif he wolde? Gise, la, gese; ic wat đæt he mihte, gif he wolde.
think you that the heavenly power NE could take-away the empire that righteous Caesar, . . . if he would? Yes, O yes, I know that he could, if he would!

On the view that ellipsis is possible only if the governing verb is an auxiliary, such data would suggest that the modals were already beginning to form a distinct subclass of verbs in OE; see Warner (1993: 113f.) for discussion. Warner also argues that OE already had pseudogapping, where the main verb is omitted but not the auxiliary or a portion of the VP. E.g.,

(47) se ðe wille godcundne wisdom secan ne mæg he
he who will heavenly wisdom seek-INF NE may he
hine wiđ ofermetta
it with arrogance
'He who will seek heavenly wisdom may not [seek it] with arrogance.' (Warner 1993: 114)

Denison (1993: 336) notes that each modal followed its own path from root verb to auxiliary, as shown by their co-occurrence with infinitival complements. "In the ModE period Warner notes that *will* and *can* are the last to lose the ability to take an object, an ability he correlates with the existence of non-finite forms . . . *must* and *shall*, on the other hand, seem to have attained unequivocal modalhood much sooner. *May* falls somewhere between the two pairs. And other items have shown varying degrees of modalhood in the different periods of English history."

Since the modal corresponds to the highest operator in CS, it is the leftmost verb, and hence the verb that appears to the left of the subject in V2 and to the left of negation. The respecialization of V2 to contexts with fronted scopal

	I: 1480–1530		II: 1580–1630		III: 1680–1730	
	no	%	no	%	no	%
XSV	4024	80.0	4338	81.5	4600	91.2
XVS	1009	20.0	987	18.5	444	8.8
total	5033	100.0	5325	100.0	5044	100.0

FIGURE 7.4. Frequency of the uninverted and inverted word orders (Baekken 2000: 394)

	I: 1480–1530		II: 1580–1630		III: 1680–1730		
	inv/tot	% inv	inv/tot	% inv	inv/tot	% inv	average
then	194/542	35.8	81/170	47.6	4/152	2.6	32.3
therefore	11/98	11.2	24/158	15.2	1/113	0.9	9.8
thus	15/71	21.1	18/49	36.7	7/60	11.7	22.2
yet	19/64	29.7	23/139	16.5	8/108	7.4	16.1

FIGURE 7.5. Frequency of inversion following initial *then*, *therefore*, *thus*, and *yet* (Baekken 2000: 400)

operators such as *wh*-phrases would have produced a situation in which a special subcategory of verbs, namely the modals and auxiliaries, had a different distribution than the main verbs. The existence of this special subcategory makes possible the formulation of constructions that specifically govern the distribution of this subcategory, e.g. SAI.

The third key change is the loss of full V2, which also displays the properties of constructional change. Baekken (2000: 394) gives the summary of the percentage of XVS and XSV word orders in Early ModE shown in Figure 7.4. As can be seen, there is a gradual loss of V2 from period I to period III. Baekken breaks down the data into various contexts in which V2 was found in OE and shows the percentage of cases in which V2 appears in these contexts in each of the three periods. These contexts are defined in terms of the type of the initial constituent X. For instance, Figure 7.5 shows the incidence of V2 with initial *then*, *therefore*, *thus*, and *yet*.

The importance of these numbers lies in the fact that V2 was not only not uniform during the transition, but it was also apparently sensitive to the particular lexical item in initial position. This variability reflects the constructional status of V2, in the sense that it was a syntactically complex correspondence, some of whose properties were idiosyncratic and others fully general.

Again, it is difficult to see how to capture such facts in terms of multiple grammars in the intended sense. But the changes lend themselves to a constructional account, if we assume that for each speaker each construction is defined over a somewhat different set of lexical items. In this case, for example, we may assume that there is a V2 construction, as in (48), where the specificity of the terms XP and V[TENSE] may vary from speaker to speaker and over time as the changes proceed.

(48) V2

$$\begin{array}{ll} \text{SYNTAX} & [_{S}\ \text{XP} \quad \text{V[TENSE]} \quad \ldots] \\ & \ \ \updownarrow \\ \text{CS} & \ \ \text{S}' \end{array}$$

In addition, there is a YES-NO-Q construction, in which V[TENSE] is initial. A comparison of the constructional ordering is given in (49). (I leave out COMPLEMENTIZER, since I am unable to determine its status in OE.)

(49) ModE: $\text{Topic} > \left\{ \begin{array}{l} \text{Subject} \\ \text{Yes-no-q} \\ \text{Wh-q} \end{array} \right\} > \text{Negation} > \text{Predicate} > \text{Object}$

OE: $\text{Topic} > \left\{ \begin{array}{l} \text{V2} \\ \text{Yes-no-q} \end{array} \right\} > \text{Negation} > \text{Predicate} > \text{Object}$

The constraints for OE are virtually the same as those for ModE, with two major differences. First, in OE, like German, V2 is the only option for positioning the tensed verb in a main clause.[33] This construction therefore requires that there be something in initial position in a declarative sentence. If there is no XP in initial position, the position of the tensed verb is licensed only by YES-NO-Q. In contrast, in English, the TOPIC construction links the topic to an IS representation, and the V2 effect is associated with interrogatives and other specialized constructions. Second, ModE has a construction SUBJECT that licenses the subject on the basis of position, while in OE, the subject is licensed by case, not position, as discussed above.

These differences suggest the following scenario. The independent emergence of SUBJECT, where the subject must precede the tensed verb, produced sentences in which there is no V2 (this is Kiparsky's 'positional licensing'). SUBJECT was restricted at first; sentences with pronominal subjects were more likely to require this ordering, and therefore they lacked V2, which began to apply in a more restricted set of conditions. When V2 lost ground in favor of

[33] I set aside here the fact that V2 is not a completely rigid requirement in German; see Müller (2005).

	I: 1480–1530		II: 1580–1630		III: 1680–1730	
	no.	%	no	%	no	%
XSV	78	72.2	29	16.9	11	6.1
XVS	23	22.8	143	83.1	169	93.9
total	101	100.0	172	100.0	180	100.0

FIGURE 7.6. Word order in structures with initial negative elements (Baekken 2000: 403)

	I: 1480–1530		II: 1580–1630		III: 1680–1730	
	inv/tot	% inv	inv/tot	% inv	inv/tot	% inv
ne	1/2	50.0	–	–	–	–
never	2/6	33.3	2/7	28.6	3/6	50.0
neither	10/10	100.0	91/91	100.0	31/33	93.9
nor	3/19	15.8	12/16	75.0	111/112	99.1

FIGURE 7.7. Frequency of inversion following initial *ne*, *never*, *neither*, *nor* (Baekken 2000: 405)

SUBJECT, there was no well-defined context from the learner's perspective that distinguished between V2 and non-V2. That is, the alternation appeared to be syntactically unconditioned, even though it was the consequence of positional licensing of subjects. Therefore the new constructions that emerged were indexed to specific lexical items in initial position, along the lines suggested by Baekken's data in Figure 7.5. Ultimately, learners associated a particular function with the 'residual' V2, giving rise to specialized constructions such as WH-QUESTION.[34]

The variability of V2 as it evolved into SAI is also demonstrated by the tables in Figures 7.6 and 7.7 from Baekken (2000) that track the development of negative inversion.

[34] There is in fact a long tradition in linguistics that takes the view that optionality evolves into respecialization. See for example, the discussion in Jäger (2006: 82ff.), and Kroch (1994). An important variant of this general idea is Wexler's Uniqueness Principle (Wexler 1981). This principle requires a one-to-one relationship between forms and meanings, and functions as a constraint governing the language learner. We may take this principle as part of the evaluation metric of UG (see Chapter 1), leading to the consequence that differentiated constructions take on distinct functions.

	I: 1480–1530		II: 1580–1630		III: 1680–1730	
	inv/tot	% inv	inv/tot	% inv	inv/tot	% inv
intrans	314/1111	28.3	227/992	22.9	137/1038	13.2
trans	232/1803	12.9	102/1468	6.9	13/1466	0.9
link	51/324	15.7	70/536	13.1	49/538	9.1

FIGURE 7.8. Inversion rates in structures with intransitive, transitive, and linking verbs (Baekken 2000: 413)

It is interesting to note that negative inversion became almost obligatory in this transition period as XSV became almost obligatory with non-negative X (Figure 7.6). But different initial negative elements contributed differentially to the total development of negative inversion, with *nor* becoming virtually obligatory and *never* rising to 50 percent (Figure 7.7).

Consider, finally, the data in Figure 7.8, also from Baekken (2000). This data shows that type of verb appears to have had an effect on the rate of inversion. Baekken suggests that what is going on here has to do with the relative weight of the postverbal material when there is or is not a direct object. She writes (p. 412):

From a pragmatic point of view, it is of considerable interest that inverted structures contain high rates of intransitive verbs . . . No doubt, this is connected with the principle of end weight: in such structures the post-verbal subject provides the required sentence-final weight; consequently, it may be assumed that the subject and the verb are inverted in such structures precisely because of pragmatic requirements such as end focus and end weight.

This data supports the view that during the transition there were two forces at play in the ordering of constituents, one having to do with grammatical function and the other to do with weight.[35] These are competing with one another, in the sense that given alternations such as (50) to express a declarative—

(50) a. XP NP V. . .
 b. XP V NP. . .

—it is not immediately obvious to a learner what characteristics of the correspondence determine whether the order should be (50a) or (50b). Is the ordering in (50a) due to the relative weight of the NP or to a constraint

[35] For a comprehensive discussion, see Warner (2007).

that says that the subject must be leftmost? Is (50b) to be chosen when the NP is 'heavy' and thus in focus?

If a rightmost position corresponds to focus, then it is more likely that the subject is focus in an intransitive than in a transitive, other things being equal, since in the intransitive the subject is the only thematic argument. Such cases are typically presentational—

(51) Into the room walked a man.

—and in ModE such inversion is completely ruled out when the verb is transitive (see Culicover and Levine 2001).

(52) *Into the room pushed the man the cart that had been discovered in the basement.

It is plausible, then, that the alternation in (50) was attributable to a range of more or less equally plausible but inconsistent conditions, including prosody, choice of lexical items, and case-marking. The result was the observed disorder in the early data followed by growing uniformity as a fixed set of conditions was ultimately settled on.

A related perspective is taken by Hinterhölzl (2004). He proposes that information structure may affect the order of constituents but not the underlying grammar. In the particular example that he discusses, he suggests that a shift from OV to VO may be analyzed not as a change in the grammar of the language, but in the ordering. Only when the shift has reached a certain threshold does the grammar change.

This effect can be achieved directly in terms of constructions. There are two constructions in competition—in the current case, V2, where the subject follows the tensed verb, and SUBJECT, where it precedes the tensed verb. As the latter rule grows in strength, because of the influence of information structure on constituent order, the former becomes weaker, to the point at which it is reinterpreted as YES-NO-QUESTION. The scenario is a transitional one, where an alternation is conditioned at first by processing factors and later becomes reinterpreted as a set of constructions, each of which serves a different semantic function.[36]

[36] A comparable transitional situation appears to exist in contemporary English. Consider the alternation between the genitive and *of* exemplified in (i).

(i) a. the committee's chairman
    b. the chairman of the committee

Kreyer (2003) shows through a corpus study that this alternation is conditioned by a number of factors. There is a strong effect of weight and syntactic complexity; heavy possessors are preferred in the *of*-construction, e.g.,

	1600–1635		1635–1670		1670–1705	
	do/total	% *do*	*do*/total	% *do*	*do*/total	% *do*
xsv	47/1380	3.4	39/879	4.4	16/904	18
xvs	30/121	24.8	12/23	52.2	8/30	26.7
total	77/1501	5.1	51/902	5.7	24/934	2.6

FIGURE 7.9. Proportion of periphrastic *do* in affirmative declarative clauses with transitive verbs: XSV and XVS (Baekken 2002: Table 3)

	1600–1635		1635–1670		1670–1705	
	do/total	% *do*	*do*/total	% *do*	*do*/total	% *do*
xsv	16/864	1.9	18/459	3.9	5/451	1.1
xvs	15/206	7.3	9/89	10.1	4/170	2.4
total	31/1070	2.9	27/548	4.9	9/621	1.4

FIGURE 7.10. Proportion of periphrastic *do* in affirmative declarative clauses with intransitive verbs: XSV and XVS (Baekken 2002: Table 4)

Finally, an interesting wrinkle in V2 > SAI involves affirmative declarative *do* (Baekken 2002). As already noted, this use of *do* peaked in the sixteenth century and declined through the seventeenth century. However, it does not decline to the same extent in inverted and uninverted sentences. It is retained more in inverted than in uninverted sentences. Moreover, it is retained more in transitives than in intransitives. I reproduce Baekken's Tables 3 and 4 for the period 1600–1705 in Figures 7.9 and Figure 7.10.

At this point in the development of the language, *do*-support is becoming a reflex of SAI, while affirmative declarative *do* in (X)SV is being supplanted by the tensed V. At the same time XSV (that is, the SUBJECT construction) in declaratives is winning out over V2. These data show that at this point, SAI had not yet become restricted to questions, but was able to extend to the affirmative declarative, in which case we see *do*-support.

(ii) a. The authority of the early Kings of Sussex.
b. ?The early Kings of Sussex's authority (Kreyer 2003: 194)

Beyond this there is a lexico-semantic factor. The greater the "degree of human involvement" the greater the likelihood that the possessor will be realized as a genitive. The two factors compete but the processing one is stronger: "if, for example, the choice of the genitive led to extreme difficulties of processing, an *of*-construction would be used, regardless of lexical or semantic factors that might indicate genitive" (194).

To summarize, then, at the point where V2 is variable, the main factors conditioning the alternation are whether or not the subject NP is pronominal and whether or not the verb is intransitive. These are purely formal factors that do not correspond to the interpretation, and are therefore likely to be overlooked by language learners, leading ultimately to replacement of the formal conditions by semantic conditions. Language learners seek to identify semantic correlates of formal variation, as they formulate constructions that account for the form–meaning correspondences. In this case, they identify V2 as inversion triggered by meaningful elements in clause-initial position, and topicalization with lack of inversion as indicative of a distinct discourse function.

7.5 Summary

I have argued in this chapter that an analysis in terms of constructions provides an attractive explanation of the shift from V2 to SAI and the development of *do*-support in English. The essential properties of the analysis are:

- *do* is an expletive V_{AUX}
- V_{AUX} emerged as a distinct syntactic category
- V_{AUX} emerged an explicit term in a set of constructions
- as SUBJECT began to supplant V2, the residue of V2 was reformulated as distinct constructions such as WH-QUESTION.

The emergence of a particular formative such as auxiliary *do* is not itself explained in this account. But once it exists in the language, it can be recruited as needed to supply V_{AUX} for any construction that requires it. The successive growth of *do*-support in various syntactic contexts documented by Ellegård provides evidence that *do*-support originated as the response to a set of constructions, each of which made a transition from the term V to the term V_{AUX}. The discussion of constructional complexity in Chapter 2 offers some insight into how constructions that begin as correspondences that are narrowly defined in terms of specific lexical items and syntactic configurations can grow to the point where they are fully general. This growth is fueled by the pressure to reduce complexity in a particular area of the grammar through generalization from specific lexical items to more general categories. At the same, the simulation model of Chapter 6 shows how a construction can spread through a population if the configuration of the social network is favorable, even if introduction of the new construction counts as added complexity in the course of its development.

8

Integrating constructions, complexity, and change

In this concluding chapter I look at how constructions, complexity, and the social network interact to produce variation, bringing together a number of ideas developed in the preceding chapters. To do this, I sketch out in some detail a preliminary account of word order variation in Continental West Germanic (CWG) varieties. I focus on the ordering among verbs in a verbal sequence, e.g. auxiliary (AUX) (e.g. *haben* 'have,' *sein* 'be'), modal (MOD) (e.g. *können* 'can'), and main verbs (V). The picture that I argue for has the following components.

(i) *Constructions*: Individual word order variants (e.g. AUX-V vs V-AUX) are distinct but related constructions, in the sense of Chapter 2. The related constructions express the same correspondence between form (syntactic structure) and meaning (conceptual structure). For example, AUX-V and V-AUX correspond to **TENSE**(F_V) where **TENSE** corresponds to AUX and F_V corresponds to V.

(ii) *Epidemiology*: The constructions are present in a population of speakers in much the same way as viruses are present in a population, as suggested in Chapter 6. They may reside in a speaker without overt symptoms, that is, they are accepted as 'possible' or 'acceptable,' but they are not produced, or they may achieve a level of strength within an individual to the point that they are manifested overtly and perhaps quite generally. Furthermore, they spread from speaker to speaker in much the same way as viruses do, i.e. through contact.

(iii) *Complexity*: The extent to which a particular variant is present in a population, and is able to spread further through the population, is a product of both non-linguistic and linguistic factors, as discussed in Chapter 6 (and see also Culicover and Nowak 2003; Culicover et al. 2003). The non-linguistic factors have to do with social and physical geography, which determine interactions among individuals. The linguistic factors have to do with the relative complexity of constructional alternatives in expressing the same meaning, along the lines of Chapters 2, 3, and 4.

(iv) *Variability*: From a formal perspective, all of the logically possible orderings of the daughters of a phrasal category are of equal complexity. A plausible minimal hypothesis, then, is that all of the logical orderings of the daughters of a phrase are in principle possible and will be attested in the linguistic evidence available to at least some learners, other things being equal. And even non-attested orders will be available to speakers as possibilities through innovation under the influence of complexity factors.

(v) *Autonomy*: A constructional change that is motivated by functional pressures (such as complexity) may become generalized in such a way that the functional motivation is no longer evident in the synchronic grammar.

(vi) *Change* occurs in a number of ways: (1) through the *diffusion* of constructions from speaker to speaker, (2) through the *generalization* of constructions from restricted to more general categories of elements (e.g. reordering of AUX-V to V-AUX generalizes to MOD-V > V-MOD and then V_{CAUS}-V > V-V_{CAUS}), (3) through the formation of *syntactic units* from high-frequency collocations (e.g. $[_{VP}\ [_{VP}\ldots$ V2] V1] > $[_{VP}\ldots$ [V2+V1]]). Such units participate in constructions, and their correspondence with CS is licensed through operations of domain union (e.g. Bouma and van Noord 1998). (4) through the *dissociation* of order from structure, which allows reordering under the influence of *formal* and *processing complexity*, allowing for V2+V1 > V1+V2 but also the opposite change.

8.1 CWG verb clusters

There is rich synchronic and diachronic data on CWG two- and three-verb clusters, which show considerable variability in ordering, as well as sensitivity to lexical subcategories. Typical two-verb clusters are V-$\begin{Bmatrix}\text{MOD}\\ \text{AUX}\end{Bmatrix}$ (2-1), which is the order in Standard German, and $\begin{Bmatrix}\text{MOD}\\ \text{AUX}\end{Bmatrix}$-V (1-2). To illustrate I use standard German forms, even though not all orders are possible in standard German.

(1) Maria glaubt, daß
Maria believes that

a. sie die Arie singen wird. [2-1]
she the aria sing will
'. . . she will sing the aria.'

b. sie die Arie wird singen. [1-2]

(2) Maria glaubt, daß
  a. sie das Buch gelesen hat. [2-1]
     she the book read has
  b. sie das Buch hat gelesen. [1-2]

Two-verb clusters may also consist of V-V, where V1 is a verb of perception or causation as in (3a,b), or a verb like *versuchen* 'try' that takes an infinitival complement with *zu* 'to' as in (3c).

(3) a. Maria glaubt daß sie Peter die Arie singen hört. [2-1]
     Maria believes that she Peter the aria sing hears
     'Maria believes that she hears Peter sing the aria.'
  b. Maria glaubt daß sie Peter die Arie singen lässt. [2-1]
     Mary believes that she Peter the aria singen lets/makes
     'Mary believes that she lets/makes Peter sing the aria.'
  c. Timo sagt daß Marie das Buch zu lesen versucht. [2-1]
     Timo says that Marie the book to read tries
     'Timo say that Marie tries to read the book.'

The most common three-verb clusters (made up of combinations of MOD, AUX, and main V) are 3-2-1 (4a) and 1-2-3 (4b), which reflect harmonic branching, but 1-3-2 (4c) and 3-1-2 (4d) are also widely attested; 2-3-1 (4e) is rare and 2-1-3 (4f) is very rare (and sometimes claimed to be non-existent).

(4) Maria glaubt, daß
  a. sie Peter die Arie singen hören wird. [3-2-1]
     she Peter the aria sing hear will
     '. . . she will hear Peter sing the aria.'
  b. sie Peter die Arie wird hören singen. [1-2-3]
  c. sie Peter die Arie wird singen hören. [1-3-2]
  d. sie Peter die Arie singen wird hören. [3-1-2]
  e. sie Peter die Arie hören singen wird. [2-3-1]
  f. sie Peter die Arie hören wird singen. [2-1-3]
  (Schmid and Vogel 2004)

Three-verb clusters may consist of MOD-V-V, as in (4), but also AUX-MOD-V (5a), AUX-V-V (5b), and MOD-AUX-V (5c). The AUXs *haben* 'have' and *sein* 'be' normally select past participle complements, but when V1 is AUX and V2 is MOD, in many varieties MOD is infinitival (as in (5a)), not past participial. This is the so-called Infinitivus Pro Participio (IPP) construction (Schmid 2005). In some dialects, IPP also occurs with AUX-V-V for certain V (Bader and Schmid 2009b).

(5) a. Maria sagt, daß sie das Buch hat lesen wollen.
Maria says that she the book has read-INF want-INF
[AUX-V-MOD, 1-3-2]
'Maria says that she wanted to read the book.'

b. Maria sagt, daß sie das Buch zu lesen versucht hat.
Maria says that she the book to read tried has
[V-V-AUX, 3-2-1]
'Maria says that she tried to read the book.'

c. Maria sagt, daß sie das Buch gelesen haben könnte.
Maria says that she the book read-PST.PRT have-INF could
[V-AUX-MOD, 3-2-1]
'Maria says that she could have read the book.'

The variability of CWG verb clusters in embedded clauses has been documented and reviewed by many scholars, including Zwart (1995, 1996), Wurmbrand (2004, 2005), Bader and Schmid (2009a,b,c), Bader et al. (2009), Schmid and Vogel (2004), and Seiler (2004). It appears that the order in Standard German is 3-2-1, and in Swiss German and Dutch it is typically 1-2-3, although there are other possibilities. The other options, especially 1-3-2 and 3-1-2 are widely found. The order 1-3-2 in IPP is typically obligatory in some 3-2-1 varieties, including standard German.

(6) a. ... daß Peter die Arie hat singen müssen
that Peter the aria has sing must
'... that Peter had to sing the aria.'

b. * ... daß Peter die Arie singen gemusst hat.
that Peter the arie sing must-PST.PRT has

c. * ... daß Peter die Arie singen müssen hat.

This pattern holds even though *gemusst hat* is a well-formed sequence that appears in other constructions, e.g.

(7) ... daß Peter aufs Klo gemusst hat.
that Peter to-the toilet must-PST.PRT has
'... that Peter had to go to the toilet.'

Moreover, CWG displays verb projection raising (VPR), where a verb phrase with its arguments appears in extraposed position (Haegeman and van Riemsdijk 1986).

(8) a. ... da Jan wilt een hus kopen [West Flemish]
that Jan wants a house to-buy
'... that Jan wants to buy a house.'

b. . . . dass er wett es Buech läse [Zürich German]
that he wants a book to read
'. . . that he wants to read a book.'

It would be interesting if each language variety permitted one ordering for all clusters regardless of the number and category of verbs, but in fact multiple orderings are possible in many varieties. If we found that only one ordering was permitted per variety, then we would explore the basis for a 'parametric' account, in which parameters that govern the ordering of AUX and the head of its complement, MOD and the head of its complement, and V and the head of its complement could take on one of two values. If the value of the ordering of AUX is [COMPLEMENT-FOLLOWS], for example, then the two-verb cluster ordering would be AUX-V, AUX-MOD, but if it is [COMPLEMENT-PRECEDES], then the two-verb cluster ordering would be V-AUX and MOD-AUX (the ordering in Standard German). The ordering 1-3-2 would be derived by combining 1-2 and 3-2, 3-1-2 would be impossible, and the ordering 2-3-1 would be a combination of 2-3 and 2-1. However, such a simple picture is in conflict with the facts—the order 2-3-1 is rare and 3-1-2 is not uncommon.

Moreover, the picture is complicated by the fact that many varieties appear to permit more than one order, both in two- and three-verb clusters. Sapp (2011: 108) summarizes the distribution of two-verb clusters in Figure 8.1 (headings renamed for consistency).

Dialect	MOD V	AUX V
Standard German	2–1	2–1
German & Austrian dialects (Wurmbrand)	2–1	2–1
S and W Austria	1–2 / (2–1)	1–2 / 2–1
N Austria	2–1	2–1
E Austria	2–1	1–2 / 2–1
Bavarian	2–1	2–1 / (1–2)
Swabian	2–1	2–1 / (1–2)
Alsatian	2–1 / (1–2)	2–1
Swiss	1–2/(2–1)	2–1 / (1–2)

FIGURE 8.1. Word order in two-verb clusters (Sapp 2011: 108)

Syntagm (group)	2–1	1–2	Total tokens
PTC + AUX (perfect, subjunctive, passive)	887 (92.4%)	73 (7.6%)	960 (72.4%)
INF + MOD	227 (74.9%)	76 (25.1%)	303 (22.9%)
INF + *tun* 'do'	27 (96.4%)	1 (3.6%)	28 (2.1%)
INF + LEX	13 (59.1%)	9 (40.9%)	22 (1.7%)
PTC + *kriegen* 'get'	8 (100%)	0 (0%)	8 (0.6%)
INF + *lassen* 'let/make'	4 (100%)	0 (0%)	4 (0.3%)
modal perfect	1 (100%)	0 (0%)	1 (0.1%)
Totals	1,167 (88%)	159 (12%)	1,326 (100%)

FIGURE 8.2. Two-verb clusters (Dubenion-Smith 2010: 112)

Crucially, these orders are "attested" in the various regions. Sapp highlights the fact that for each variety and for each subtype of cluster, there are different preferences for word orders. For example, he cites Patocka (1997) as observing that in the Austrian dialects, the passive with *werden*-V.PAST.PRT is uniformly 2-1, while *haben*-V.PAST.PRT permits both orders in some cases.

Along similar lines, Dubenion-Smith (2010: 112) found the judgments in Figure 8.2 in an analysis of spoken West Central German[1] from the *Datenbank Gesprochenes Deutsch* compiled between 1955 and 1970. Dubenion-Smith's more detailed figures, not reproduced here, show that for AUX-V, MOD-V, MOD-*tun*, and V-V both of the orders are attested, with a strong preference for 2-1. Dubenion-Smith takes the useful step of breaking down the data into specific subtypes of two-verb clusters, which allows the variability to show through clearly. While the totals show that both orders are possible, the breakdown shows that the 1-2 order is more common for MOD-V (as a percentage of the total cases) than it is for the other subtypes.

Similar variability is found in three-verb clusters. Consider the data in Figure 8.3 (from Schmid 2002) showing possible three-verb clusters in Zürich German, broken down by the type of verb that occupies the V2 slot. (The verb types are exemplified by *lassen* 'let/make' (Causative), *müssen* 'must' (Modal), *sehen* 'see' (Perception Verb), *helfen* 'help' (Benefactive), *bleiben* 'stay' (Durative), *beginnen* 'begin' (Inchoative), *versuchen* 'try' (Control Verb).)

According to Schmid, "Zürich German shows the largest variation of verbal order patterns of all languages . . . With the exception of order 312 (only possible with a special stress pattern), all logically possible patterns are

[1] The area bounded roughly by Karlsruhe to the South, Darmstadt to the East, Kassel to the North, and Aachen to the West.

	Perfect, V2: PastP	Perfect, V2: IPP	Future
Causative	*	321, 123, 132	321, 123, 132
Modal	*	?321, 123, 132	?321, 123, 132
PV	321, ?123, 213	?231, 123	321, 123, 132
Benefactive	321, 231, 123, 132, 213	231, 123, 213	321, 231, 123, 132
Durative	321	*	321, 132
Inchoative	231,213	*	231, 123, 213
CV	321, 123, 213	*	321, 123, 132, 213

FIGURE 8.3. Overview of verb order patterns in Zürich German (Schmid 2002)

Language	MOD–MOD–V FIN–INF–INF	AUX–MOD–V FIN–INF–INF	AUX–MOD–V FIN–IPP–INF	MOD–AUX–V FIN–INF–PART	AUX–AUX–V FIN–PART–PART
Afrikaans	1–2–3	1–2–3	2–3–1	1–3–2 3–1–2[❶]	N/A
Dutch	1–2–3	1–2–3	1–2–3	1–2–3 3–1–2 1–3–2 [3–2–1]	?3–1–2 ?1–3–2 [?others]
Frisian	3–2–1	3–2–1	3–2–1 [1–2–3/❷]	3–2–1	3–2–1
German (Standard)	3–2–1	3–2–1 1–3–2	1–3–2	3–2–1	3–2–1
German and Austrian dialects	3–2–1 1–3–2	3–2–1 1–3–2 3–1–2	1–3–2 3–1–2 3–2–1 [1–2–3]	3–2–1 1–3–2	3–2–1
Swiss dialects	1–2–3 3–2–1 1–3–2 3–1–2	N/A	1–2–3 1–3–2 3–1–2	1–2–2 3–2–1 3–1–2	3–2–1
West Flemish	1–2–3		1–2–3[❸] 2–3–1[❸]	1–3–2 3–1–2	3–2–1 1–3–2

Adjustments (Table 2)

[...] orders that are attested but very restricted in the language for which they are listed and could not be verified; these orders will not be considered as possible orders for these languages, but they are listed in the table to indicate that a further refinement and dialect separation is necessary.

❶: '3–1–2' is only possible when '2' is a passive auxiliary

❷: IPP is not obligatory; some speakers do not use IPP; '3–2–1' is possible for constructions with and without IPP (cf. speakers consulted by P. Ackema); '1–2–3' is mentioned in Hoekstra and Taanman (1995) and Ijbema (1997) for certain constructions (perception verbs and aspectual auxiliaries) in West Frisian.

❸ : '1–2–3' is obligatory when the auxiliary is in the past or has a negative marker attached
'2–3–1' is obligatory when the auxiliary is non-finite

FIGURE 8.4. Three-verb clusters (Wurmbrand 2005: Table 2)

confirmed by my informant."[2] But what her summary also shows is that different orders are attested with different verbs.

[2] However, according to another native speaker informant (Martin Saltzman, p.c.), not all of these orders are actually active in this dialect. It is entirely possible, although of course difficult

The data in Figure 8.4 from Wurmbrand (2005) show similar variability across the CWG varieties. Note that in Wurmbrand's summary, V2 is never initial in the cluster, while this possibility appears to exist in Schmid's data from Zürich German as confirmed by Martin Salzmann (p.c.), a native speaker. It appears, therefore, that it is in general impossible to say that a particular variety "has" a particular order in the two- or three-verb clusters, although there are certain varieties that only permit a single order.

8.2 Derivational accounts

The variability noted in the previous section is unfamiliar to a speaker of English (which lacks verb clusters entirely) or Standard German (which has clusters). In English the only possible orders are 1-2 and 1-2-3, while in standard German the only possible orders are 2-1, 3-2-1, and 1-3-2 in IPP. The regularity of the standard languages, as well as the standard methodology in syntactic theory, is consistent with derivational accounts of verb clusters.

On standard theoretical assumptions, there is a uniform underlying branching order, and orders that deviate from are derived by movement. Assume for the sake of illustration that this order is 2-1. In the two-verb case, the order 1-2 is derived by attaching V2 to the right of V1 (by verb-raising or VR) to form a cluster, as in (9) (Evers 1975; Haegeman and van Riemsdijk 1986).

(9) $[_{VP}\ [_{VP} \ldots \mathrm{V2}]\ \mathrm{V1}] \Rightarrow [_{VP}\ [_{VP} \ldots t_{V2}]\ [\mathrm{V1{+}V2}]]$

But since even the 2-1 order is a cluster, there must also be a VR operation in 2-1 dialects that adjoins V2 to the left of V1, as shown in (10).

(10) $[_{VP}\ [_{VP} \ldots \mathrm{V2}]\ \mathrm{V1}] \Rightarrow [_{VP}\ [_{VP} \ldots t_{V2}]\ [\mathrm{V2{+}V1}]]$

However, another logical possibility is that VR is optional. When it applies, we get the order 1-2 as in (9), and when it does not, we get the order 2-1, but no cluster. It is then logically possible that there are varieties with 1-2 clusters, and other varieties with 1-2 but no 1-2 clusters. There is as far as I know no empirical evidence to suggest that the latter type of variety exists.

The situation is further complicated by the fact that both orders can be derived from underlying 1-2. Suppose that the V-final ordering is derived by preposing the projection of V2 after the cluster is formed. The derivation of 2-1 is shown in (11), and the derivation of 1-2 is shown in (12).

to confirm, that the judgments of Schmid's consultant were influenced by exposure to standard German and to other dialects of Swiss German.

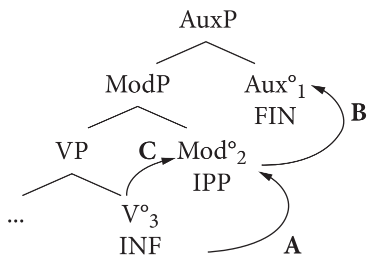
(11) $[_{VP}\ V1\ [_{VP}\ V2\ldots]] \Rightarrow [_{VP}\ V2{+}V1\ [_{VP}\ t_{V2}\ldots]] \Rightarrow [_{VP}\ t_{V2}\ldots][_{VP}\ V2{+}V1]$

(12) $[_{VP}\ V1\ [_{VP}\ V2\ldots]] \Rightarrow [_{VP}\ V1{+}V2\ [_{VP}\ t_{V2}\ldots]] \Rightarrow [_{VP}\ t_{V2}\ldots][_{VP}\ V1{+}V2]$

Such derivations illustrate the fact that in the absence of very strict constraints on underlying structures and possible movements, any order can be derived from any underlying structure (Rochemont and Culicover 1997).

The derivation of three-verb clusters proceeds along the same lines, using iteration. A typical example from Wurmbrand (2005) is given in (13) (Wurmbrand's (19)).

(13) a. *Head-final structure*



b. *Possible derivations*

no overt movement:	3–2–1
only A:	2–3–1
only B:	3–1–2
only C:	3–2–1 (vacuous)
A and B:	1–2–3
C and B:	1–3–2

Each operation is responsible for adjoining a lower V to a higher V, either to the left or to the right. By a selective application of these operations it is possible to derive all of the observed orders. It is even possible to derive 2-1-3, by simply adjoining 3 to the right of 1 in an underlying 3-2-1 structure. Wurmbrand notes that such a derivation could be ruled out by a constraint that blocks movement of one head over another, in this case, 3 over 2 to adjoin to the right of 1. But 2-1-3 could be derived from 1-2-3 by adjoining 2 to the left of 1, which is a legitimate movement under a VR analysis. Significantly, 2-1-3 occurs (apparently obligatorily) in Zürich German (Lötscher 1978: 9) and Pennsylvania German (Louden 1990: 475), when V2 is an inchoative and in the past participial form.

(14) a. wo s aagfange$_2$ hat$_1$ ragne$_3$ [Zürich German]
where it started has rain
'where it started to rain'

b. wu s kschtert$_2$ hot$_1$ regere$_3$ [Pennsylvania German]
where it started has rain

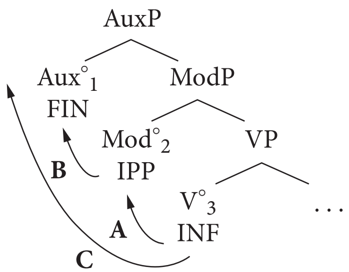
Lötscher also cites an example where V2 is perceptual.

(15) Wo mer aber de vatter vor de beiz gsee$_2$ händ$_1$ staa$_3$,
when we however the father before the pub seen have stand,
simer wider umgcheert.
we again turned-back
'But when we saw the father standing in front of the pub, we turned back again.'

Louden proposes that such cases may be derived by extraposition of the main verb to the right. In the constructional analysis outlined in below §8.4, 2-1-3 would be a special construction, where V2 is lexical (not modal).

Assuming an underlying right-branching structure, a different set of adjunctions will derive all of the attested orders, including 2-1-3 (Wurmbrand's (20)).

(16) a. *Head-initial structure*



b. *Possible derivations*

no overt movement :	1–2–3
only A :	1–3–2
only C :	3–1–2
A and B :	3–2–1
C and B :	2–3–1
B and C :	3–2–1
	2–3–1 ('tucking in')

Wurmbrand notes that 2-1-3 could be ruled out on this approach by a stipulation, e.g. the "middle" verb cannot move by itself, or if it does move, the lower verb is not licensed. It is difficult to take any of these options seriously as explanations, in part because they are ad hoc stipulations, and in part because there are restricted (but obligatory) attested instances of 2-1-3.

While VPR is another way of deriving various orders, the structures that result from VPR are not instances of verb clusters, because the complements of the lowest verb appear medially in the verbal sequence (see (8)). There is considerable evidence that verbal clusters are constituents—see Haider (2003) for a review of the linguistic evidence. For example, clusters cannot be broken up by other material, e.g. adverbs such as *wohl tatsächlich schlecht* in (17).

(17) ... daß die Theorie wohl tatsächlich schlecht formuliert (*)
that the theory possibly indeed badly formulated
worden (*) sein (*) mag
been be may
(Haider 2003: 94)

And 3-2 clusters can be left dislocated and topicalized as constituents:

(18) a. [Wiedererkennen können]$_i$ (das$_i$) müßte er sie schon
identify be-able (this) must he her well
b. [Vorsingen lassen müssen]$_i$ (das$_i$) wird man ihn schon
audit let must (this) shall one him well
'One will have to let him audit.'
c. [Übersehen haben]$_i$ (das$_i$) wird man sie sicher nicht
overlooked have (this) will one her surely not
d. [Übersehen worden]$_i$ (das$_i$) ist sie noch nie
overlooked been (this) has she never ever
e. [Zu reparieren versucht]$_i$ (das$_i$) hat man ihn nicht
to repair tried (this) has one it not
(Haider 2003: 95)

The left-dislocation cases with *das* are particularly telling, since these clusters would appear to be base-generated.

In fact, even when the linear order corresponds to a plausible canonical underlying structure, the verbs form a cluster. Consider the examples in (19) from Bayer (2004). Bayer shows that it is impossible to extrapose a constituent to a position between two verbs in the order 2-1, even though such a constituent is extraposable when the two verbs are not immediately adjacent to one another. Example (19b) is impossible because it contains the sequence **entscheidern dafür versucht* 'tried to decide on that.'

(19) a. Ich habe [mich dafür zu entscheidern] versucht.
I have REF there-for to decide tried
'I have tried to decide on it.'
b. *Ich habe [[mich ~~dafür~~ zu entscheidern] dafür] versucht.
c. Ich habe versucht [[mich ~~dafür~~ zu entscheidern] dafür].
d. Ich habe [[mich ~~dafür~~ zu entscheidern] dafür]$_i$
I have REF there-for to decide there-for
schon mehrmals erfolglos t_i versucht.
already more-than-once successlessly tried
'I have already more than once without success tried to decide on it.'

So while VPR undoubtedly exists in some varieties, this in itself cannot preclude the possibility of constructional verbal clusters per se.

Wurmbrand concludes that while derivational accounts are capable of accounting for all of the attested orders, there is no empirical basis for arguing that one particular order is underlying.[3] Moreover, it is unclear why such movements should occur in the first place.

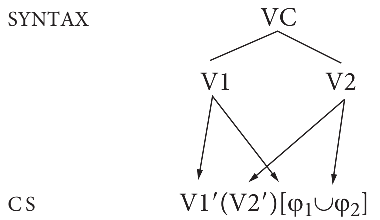
However, there is a very strong argument that favors a derivational approach. The hierarchical structure illustrated in (13) maps directly into a semantic interpretation in which each phrasal projection is an argument of the higher verbal head. If, for example, the heads in (13) are *have*, *must*, and *read*, the corresponding semantic interpretation is (20), under the assumption that *have* corresponds to past tense.

(20) **PAST**(**MUST**(**READ**(. . .)))

The semantic interpretation thus appears to be a compositional correspondence of the individual verbs arranged in the standard hierarchical structure.

But it is not necessary to assume this hierarchical structure in order to get the correct compositional interpretation of the verbs. Two or more verbs can be composed to form a CS representation in which the features of the verbs are added together. Bouma and van Noord (1998) show how to do this domain union in the HPSG formalism; their analysis can be translated into a constructional account. What is required is a construction that treats two verbs as a single constituent, which I label 'VC' here. φ_1 and φ_2 are the features of Vl and V2, respectively.

(21)



This construction licenses clusters such as *will lesen* 'wants to read.' Since *lesen* (V2 in (21)) selects a direct object, this feature will be a property of the composite V. Since embedded VP are V-final in CWG, a direct object will be licensed in the position before *will lesen* just as it is licensed in the position before *lesen* in a simple VP, or in the case of the cluster *lesen will*.[4]

(22) a. . . . daß er das Buch will lesen.
b. . . . daß er das Buch lesen will.

[3] See Rochemont and Culicover (1997) for arguments that in principle, surface orders underdetermine underlying order.

[4] The order in (22b) is of course derivable directly from an underlying hierarchical structure in which *das Buch lesen* is a VP complement of *will*. Such an analysis would require a left-branching structure, which is assumed to be possible in many derivational approaches to verb clusters.

I conclude without further discussion that derivation does not provide a plausible explanation for the range of possible verbal clusters or their distribution.

8.3 Optimality accounts

Having ruled out a derivational account, the question remains, How do we explain the CWG verbal clusters? One promising answer appeals to Optimality theory (OT). On an OT account, all orders are available and evaluated according to how seriously they violate universal constraints. The ordering (or orderings) that violates the least highly ranked constraint (or constraints) is the preferred one.

One proposal along these lines is due to Schmid and Vogel (2004). Some key assumptions in their OT account tie in with views in the mainstream literature about what is the most 'natural' relationship between linear order and syntactic structure.

- Rightward binary branching is preferred (following Kayne 1994).
- Asymmetric c-command (A higher than B) produces left-to-right ordering (A < B) (Kayne 1994).
- Linearization of constituents that are at the same level in the structure is ambiguous, since they mutually c-command.

These can be expressed to some extent as OT constraints, rather than absolute requirements/prohibitions.

Schmid and Vogel assume that the canonical structure is rightward-branching 1-2-3, where each verb takes the phrase headed by the next lowest verb as its complement. They assume that, given a basic hierarchical rightward-branching structure, the verbs get linearized in order to satisfy the constraints. So the structure corresponds to the meaning, in the sense that the higher verb scopes over its complement. But the structure is not necessarily preserved in the linearization (e.g. 2-1-3 is derived from underlying V1-[V2-[V3. . .]]).

The key constraint implements a preference for rightward branching of AUX, MOD, and V (the components of an 'extended projection'), following Kayne (1994).

(23) **MAP-left-right(V^0) (MAPlr(V^0))**

The heads of an extended projection of V are linearized in a left-to-right fashion, i.e., if head A asymmetrically c-commands head B at LF, then the PF correspondent of A precedes the one of B at PF.

If this constraint is ranked highest, then the ordering will be 1-2-3. Note that this constraint has the consequence that harmonic branching, where each head is on the same side of its complement (in this case, precedes), is optimal. Harmonic branching is also most highly valued in non-OT analyses, e.g. Hawkins (1982). Harmonic branching is an important component of any explanation for the possible three-verb clusters, given that the orders 1-2-3 and 3-2-1 are most common and, one might presume, least complex.

In an OT account everything has to be formulated in terms of constraints. So, for example, whether complements precede or follow their heads is the responsibility of two constraints, one for 'precede' and other for "follow." If "precede" is more highly ranked, then the ordering is complement–head, and if "follow" is more highly ranked, then the ordering is head–complement. Here are Schmid and Vogel's constraints for linearizing sequences of V's.

(24) **MAP(complement before head) (MAPch)**
If A and B are sister nodes at LF, and A is a head and B is a complement, then the correspondent of B precedes the one of A at PF.

MAP(head before complement) (MAPhc)
If A and B are sister nodes at LF, and A is a head and B is a complement, then the correspondent of A precedes the one of B at PF.

These constraints optimize 1-2-3 and 3-2-1. If MAPlr(V^0) is ranked highest, it further optimizes 1-2-3. But if MAPch is ranked highest we get the 3-2-1 order of Standard German.

(25)

	MAPlr(V^0)	MAPch	MAPhc
321: V MOD AUX	***		**
231: MOD V AUX	**	*	*
123: AUX MOD V		**	
132: AUX V MOD	*	*	*
312: V AUX MOD	**	*	**
213: MOD AUX V	*	**	*

(26)

	MAPch	MAPlr(V^0)	MAPhc
321: V MOD AUX		***	**
123: AUX MOD V	**		
etc.			

Ranking MAPhc low means that direct objects will precede their heads, in any case.

There is no way to get any of the other orders using just these three constraints. So it is necessary for Schmid and Vogel to introduce other constraints that have the function of positioning something other than the complement before head or head before complement—these constraints should place a particular head in initial or final position. For example, if the basic order is 1-2-3, but there is a highly ranked constraint that says 'a focused head is initial,' there will be an alternative 3-1-2-~~3~~ order. But there will also be a 2-1-~~2~~-3 order. And if the basic order is 3-2-1, such a constraint will derive 1-3-2 and 2-3-1. Alternatively, if there was a constraint that said that the focused element should be at the right, this would derive 2-1-3 and 3-1-2 from 3-2-1, and 2-3-1 and 1-3-2 from 1-2-3. So there can be lots of alternative orders, regardless of the basic order, as shown in (27):

(27)

Basic order	Focus left	Focus right
3-2-1	2-3-1, 1-3-2	2-1-3, 3-1-2
1-2-3	2-1-3, 3-1-2	2-3-1, 1-3-2

In other words, all of the orders will be derived. So there have to be highly ranked constraints whose function is to rule out the non-occurring orders in each variety.

While it is possible to formulate OT constraints and rankings that will derive just the observed orderings, and combinations of orderings, the number of constraints in Schmid and Vogel's analysis actually exceeds the number of possible linear orders. In conceptual terms, at least, the analysis would be simpler if each possible order was simply licensed as an individual construction, and dialects were described in terms of the orderings that they permit. Like the derivational account, the OT constraints and rankings do not appear to constitute an explanation of what is seen in the data.

This being said, the OT account does capture one crucial property of the data that does not follow naturally from the derivational account without further assumptions, and that is the fact that 1-2-3 and 3-2-1 are default orders and reflect harmonic branching. The challenge before us, then, is to capture this property of the OT account while avoiding the technical idiosyncrasies that arise when we require that all of the ordering facts be captured in terms of OT constraints.

8.4 A constructional account

I propose here a constructional account of verb clusters that falls within the general framework sketched out at the beginning of this chapter. Individual variants of a word order (e.g. AUX-V vs V-AUX) are distinct but related constructions, in the sense of Chapter 2. The related constructions express the same correspondence between form (syntactic structure) and meaning (conceptual structure).

Moreover, and crucially, verb clusters are phrasal constituents. The category of a verb cluster is VC. The structure of a typical subordinate clause is given in (28).

(28) Maria glaubt, [$_S$ daß sie [$_{VP}$ das Buch [$_{VC}$ gelesen hat]]]
Maria believe that she the book read.PAST.PRT has

The category VC is not a subcategory of V, it is *sui generis*. It is specially reserved for (and arises from) verb sequences in VP-final position, and the interpretation in terms of domain union (along the lines of (21)) is reserved for just this structure. This restriction accounts for the fact that verb clusters do not appear in the V2 position, which is reserved for finite verbs. Example (81b) is not even remotely possible in any variety of CWG, as far as I know.

(29) a. Maria glaubt, [$_S$ sie hat das Buch gelesen].
b. *Maria glaubt, [$_S$ sie [gelesen hat] das Buch].

Following the observation by Bayer (2004), I assume that verb cluster formation is a constructional innovation that is made possible by general properties of verbal sequences that arise from analytic syntactic structures, i.e. those in which a VP is the complement of a higher V. Following Schmid and Vogel (2004), the default orders will be those in which the branching direction is consistent. Thus, the clusters 1-2 and 1-2-3 reflect the underlying structures in (30),

(30) a. [$_{VP}$ XP [$_{VP}$ V1 [$_{VP}$ V2]]]
b. [$_{VP}$ XP [$_{VP}$ V1 [$_{VP}$ V2 [$_{VP}$ V3]]]]

while the clusters 2-1 and 3-2-1 reflect the underlying structures in (31).

(31) a. [$_{VP}$ [$_{VP}$ XP V2] V1]
b. [$_{VP}$ [$_{VP}$ [$_{VP}$ XP V3] V2] V1]

From this perspective it is easy to see why CWG has verb clusters, while English does not.[5] If we start out with strict V-final VP, as in (32)—

[5] There is no syntactic evidence that English has verb clusters, in the sense that only one order is possible for any pair of verbs. However, contractions such as *would've* may reflect clustering of some type.

(32) $[_{\mathrm{VP}} \ldots \mathrm{V}]$

—and embed it, we get strict V-final clusters, as in (33)—

(33) $[_{\mathrm{VP}}\ [_{\mathrm{VP}} \ldots \mathrm{V}]\ \mathrm{V}$

—which blocks V-X-V. This is because the specifier of VP is on the left, and cannot intervene between verbs. Thus, it is natural for a language to grammaticalize the VP-final sequence V-V as a cluster, perhaps for processing efficiency. Whenever there are two Vs in final position in the VP, they will be adjacent to one another. Cluster formation of this type is not possible in English because in English, Spec,VP is on the left edge of VP, which separates the V's. This is the position of negation, adverbs, and so on.[6]

Formulated as a construction, the cluster preempts the analytic structure, which explains Bayer's examples in (19). That is, the greater specificity of the cluster construction requires that it be applied before the more general hierarchical branching construction for head–complement.

It is appealing to view the innovation of verb clusters as a device for simplifying the interpretation of the hierarchical structure, by eliminating it in favor of the unitary mechanism associated with domain union. The function of the auxiliary verb is to mark the tense of the main verb. In the two-verb cluster with order 2-1, if the main verb denotes a relation and the tense denotes a time, then computation of the verb with its complements produces a CS representation, e.g. **EAT**(CAKE), the set of cake-eatings or the property of cake-eating, depending on one's semantic theory. Interpreting expression with respect to **PAST** takes all instances of cake-eating into the set of past-cake-eatings (or the property of past-cake-eating). On the other hand, interpreting first **PAST** with respect to **EAT** produces the relation of past-eating, which may then be specified by application to CAKE. While both are feasible, the second has one fewer computational step (although it does require holding the complement(s) a little longer in memory while the verb and tense are composed). I leave for future research the question of whether there is a complexity motivation for clustering along these or other lines.

Where it is allowed, the formation of clusters is in effect the emergence in a language of the non-phrasal category 'verbal cluster' (VC). A VC is a constituent, but it is not headed, and its category is idiosyncratic to the language. When this category is constructionally licensed, it opens up the possibility of alternative linear orders. These linear orders may come about as a consequence of a bias. So to take a simple example, if 2-1 is interpreted syntactically

[6] Pritty Patel-Grosz has informed me that Hindi, a V-final language, shows the kind of ordering in verb clusters seen in CWG.

as a cluster, along the lines sketched out above, and if the order 1-2 is a response to a bias, the possibility exists of the 1-2 construction emerging in the language and ultimately taking hold.

Once a category VC has been introduced into the language, the possibility exists for reordering of the constituents of this category.[7] The default orders will be 1-2-3 and 3-2-1 because of the canonical branching order, but other orders may come into being under the influence of other, non-grammatical factors. In this regard it is of interest to note the following insightful observation of Lötscher (1978), translated by Schmid and Vogel (2004).[8]

> The additional complication, that one single rule type is hardly sufficient to account for word order, must be taken into account as well. Rather, there are at least three interacting but primarily independent kinds of rules: first, grammatical rules [. . .] that determine an order more or less arbitrarily [. . .]; performance rules [. . .]; and last, functional rules [. . .] that allow for certain functional relations in a sentence in the sense of the topic-comment distinction. (Lötscher 1978: 10–11)

Now the question arises, What are the performance or functional factors that explain variation in word order in the clusters? I suggest in the next section that there are two biases, one that motivates positioning the main verb adjacent to its complements, and one that motivates ordering the auxiliaries and modals according to their scope in CS. Both are at play in all language varieties, but one or the other may take precedence due to social factors that are operating in the social network.

Before turning to the biases, however, it is important to highlight a number of other facts about verb clusters that lend support to the view that verb clusters are the province of constructions in the sense developed in this book.

[7] A natural question is whether a reordering explanation is possible in an analysis where the verbs in a cluster are heads of individual projections. This is in fact the derivational approach. As Wurmbrand (2005) notes, the derivational approach is not explanatory, since it relies heavily on arbitrary assumptions about structure, movements, adjunction and triggering in order to get the observed orders (2005: 294): "Many interesting accounts have been suggested addressing the question of how verb clusters are derived. However, what still appears to be an open question is the question of why the elements of a verb cluster are inverted in certain languages and constructions."

[8] "Dabei muß die zusätzliche Komplikation berücksichtigt werden, daß für die Erklärung der Wortstellung wohl kaum ein einziger Regeltyp vorausgesetzt werden kann. Vielmehr lassen sich mindestens drei interagierende, aber primär voneinander unabhängige Arten von Regeln ansetzen: Erstens grammatisch bedingte Regeln—Regeln wie die der Adjektivstellung bei Nomen, der Verbzweitstellung usw., die mehr oder weniger willkürlich eine Abfolge bestimmen, die aber auch anders aussehen könnten; performanzbedingte Regeln—Regeln, welche es ermöglichen, performanzmäßig allzu komplexe Strukturen in leichter zu verarbeitende Strukturen umzubilden, wie etwa die verschiedenen Extrapositionsregeln in vielen Sprachen; endlich funktional bedingte Regeln—Regeln, deren Zweck die Ermöglichung des Ausdrucks von bestimmten funktionalen Satzverhältnissen im Sinne der Thema-Rhema-Unterscheidung ist."

As noted above in the citations of Sapp, Dubenion-Smith, and Wurmbrand's summaries, the occurrence of particular orders in verb clusters in many varieties is sensitive to the lexical categories of the participating verbs. The following implicational hierarchy appears to hold for two-verb clusters.

(34)	AUX-V.PAST.PRT | V.PAST.PRT-V	<	[e.g. gelesen haben]
	MOD-V.INF | V.INF-MOD	<	[e.g. lesen müssen]
	V-V.INF | V.INF-V		[e.g. lesen lassen]

The direction of generalization is precisely what we expect to see in constructions that begin with a restricted element. In this case, we surmise that the earliest clusters were based on finite AUX, perhaps for reasons discussed above. AUX is non-lexical, so the next natural generalization is to all V [NON-LEXICAL], which comprises AUX and MOD; the last stage of generalization is to V, but the generalization may proceed by adding elements to the set that appear to fit the pattern until the full category is achieved. So the move from V[NON-LEXICAL] might proceed to add a single lexical verb, e.g. *lassen* 'make, let,' and then move to other verbs that take verbal complements such as *kriegen* 'get,' *bleiben* 'remain,' *gehen* 'go,' *sehen* 'see,' and so on. Generalization to a new kind of verb increases coverage, in the sense of Chapter 2, while extension of the set of verbs reduces strain.

8.5 Two biases

I turn now to the biases mentioned in the preceding section. To get an insight into the biases that affect word order in verb clusters, consider the following quotation from Haider (2003: 119–20):

> In a right-branching structure, the parser can unambiguously identify the top-most node of the projection after encountering the first element of the projection, V1. [The left-branching structure], however, is not parser-friendly. The parser would have to guess how many brackets there might be, because their number—or in other words, the depth of embedding of the left-most element—depends on the number of verbs to come. General top-down information on the possible structure of a VP will not help guessing, because the number of auxiliaries is not context dependent. (Haider 2003: 8)

Haider argues that because of "parser unfriendliness" there must be a constraint that requires branching nodes to follow the heads, thereby ruling out the left-branching structure. Instead, there must be a cluster in V-final structures. (See also Dubenion-Smith 2010 for a similar idea.)

A slightly different way to put this is that there is nothing wrong with 3-2-1 order per se, but 1-2-3 order is computationally less complex with respect to a particular aspect of processing and interpretation. The tense or modal operator expressed by the first verb must take scope over the VP that is its argument. In general, scope-taking elements tend to precede what is in their scope. It appears that something along the lines of the following is correct (see for example Kroch 1974).

(35) **Scope-Order principle**
The preferred scope ordering of operators corresponds to the left-to-right ordering of the phrases in the surface structure of the sentence.

With this in mind, consider what steps are required in processing a 2-1 order, exemplified in (36).

(36) ... daß sie das Buch lesen will.
that she the book to-read wants
'... that she wants to read the book.'

For simplicity, let us suppose that the marginal costs associated with each step are equal; let this measure be **c**, which we will call the 'scope bias.' At the point at which the word *lesen* 'read' is processed, the partial representation is (37).

(37) **READ**(PRO, BOOK)

A potential problem is that *sie* is singular and the singular form of *lesen* is *liest*. However, *sie* has the same form as *Sie* 'they, you (polite),' and the infinitival form and the 3.PL inflected form are the same. So at the point of *lesen*, there is no way to predict that there will be another verb, as Haider points out. Either we hold *lesen* in memory as an ambiguous form, or we decide that it agrees with *sie*.

In either case we have to do more processing when we get to *will* 'want,' which is 1|3.SG. If we refused to make a commitment, then we had to hold the form in memory, at a cost of **c**. If we decided that *lesen* was 3.PL, then we have to change the 3.PL|2.POLITE interpretation of *sie* to 3.SG, which also costs **c**. In either case, we end up with (38).

(38) **WANT**(PRO[3.FEMALE]$^{\alpha}$, **READ**(α,BOOK))

Suppose now that the order is 1-2. In this case there is no marginal cost. Processing ... *daß sie das Buch* is the same in both cases. Then we get to *will*, which tells us that the subject is 3.SG.

(39) **WANT**(PRO[3.FEMALE]$^{\alpha}$, **F**(α,BOOK))

And when we get to *lesen*, all we have to do is fill in the value of **F** as **READ** and link the object thematic role to BOOK.

Processing the order 2-1 is thus apparently more complex than processing the order 1-2, by the amount **c**. So, we might ask, why aren't all of the dialects 1-2(-3) (not to mention all languages)? The answer, as suggested earlier, is that there is also a bias that favors 2-1 and 3-2-1. As discussed at length by Hawkins (1994, 2004), languages tend to favor linearizations in which the subcategorized arguments of a head and closely related adjuncts are as close to the head as possible. Hawkins's reasoning is that the computational cost, measured in terms of memory load, of constructing the CS representation is greater when the heads are more distant from one another. A similar measure of complexity is argued for by Gibson (1998, 2000) and Grodner and Gibson (2005).

Consider again the processing cost of 1-2 versus 2-1. Let us assume that holding an argument or an adjunct in memory until its head is encountered incurs a cost of **c**′. Let us call this the 'dependency bias.' In the 2-1 variant of our example (88), there is no cost, since the direct object *das Buch* is adjacent to the verb *lesen.* But in the 1-2 variant, the verb *will* intervenes:

(40) . . . daß sie das Buch | will lesen.

It is possible to hypothesize at the point that *das Buch* is processed that it is an argument, probably a direct object, of some verb, and so corresponds to **F**(α,BOOK). But just as holding **READ**(α,BOOK) until *will* is processed incurs a cost, so does holding **F**(α,BOOK) until *lesen* is processed. Hence in this latter respect, processing 1-2 is more costly than processing 2-1 by **c**′. And when we deal with more complex VPs, the processing costs are correspondingly greater for both orders.

Order 1-2 is favored by the scope bias, and 2-1 is favored by the dependency bias. As suggested in Chapter 7, competing biases do not necessarily cancel one another. They are both active in the population, and in principle one or the other can acquire an advantage through an accidental property of the topology of the network. That is, a critical mass of speakers can shift to one order for non-linguistic reasons, making it the only or the preferred order.

Furthermore, both orders may be active as long as the constructional alternatives are allowed by the theory, even if one of them is not actually in use by speakers. Suppose, for example, that we have a uniform 1-2 population, and suppose that there is a subgroup of 1-2 that is isolated from contact with the larger population. The order 2-1 is available to this subgroup (as it is to the larger population) because of the principle that all orders of the daughters are possible. It is possible that enough members of this subgroup would spontaneously succumb to the appeal of 2-1 (under the dependency bias) to pass it on to the next generation. And then given enough contact between this subgroup and the larger population, it is possible that the innovation could spread beyond the subgroup.

Finally, recall the relatively rare 2-1-3 order, exemplified in (14) and (15), repeated here.

(14) a. wo s aagfange$_2$ hat$_1$ ragne$_3$ [Zürich German]
where it started has rain
'where it started to rain'

b. wu s kschtert$_2$ hot$_1$ regere$_3$ [Pennsylvania German]
where it started has rain

(15) Wo mer aber de vatter vor de beiz gsee$_2$ händ$_1$ staa$_3$, simer wider umgcheert.
when we however the father before the pub seen have stand, we again turned-back
'But when we saw the father standing in front of the pub, we turned back again.'

Martin Salzmann (p.c.) observes, in connection with the similar example (41)—

(41) dass i en gsee ha schaffe
that I him seen have.1SG work.INF
'that I saw him work'

—that 2-1-3 is not at all surprising in such a case, given the dependency bias. The subject of the most deeply embedded V3 is also the subject of the perception verb V2, and in (15) and (41) this verb moves closer to its dependent.

In the inchoative case, one could argue that 'it' is the argument of 'start.' That is, it is a control verb. Examples such as (42) show that the subject of 'rain' is not a dummy NP but a quasi-argument, and thus capable of control.

(42) It rained [without PRO snowing].

Given this, it is plausible that the verb 'start' moves closer to its argument in the perception cases (14) and (15) as a consequence of the dependency bias.

8.6 Clusters in the network

Consider now the fact that language varieties may contain several orderings in two- and three-verb clusters. As we saw in earlier chapters, it is entirely possible for a language to maintain a range of alternative constructions, even though some are more complex in certain respects than others. The various constructions live in the 'body linguistic,' like viruses (Chapter 6). They spread through contact, both within an individual lexicon and across the population.

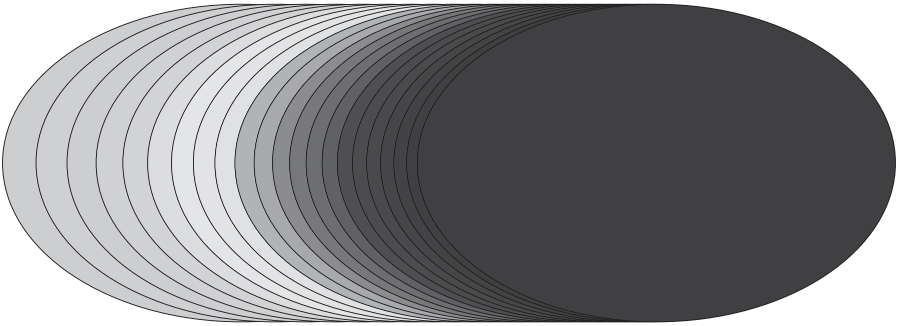


FIGURE 8.5. Distribution of competing constructions over region

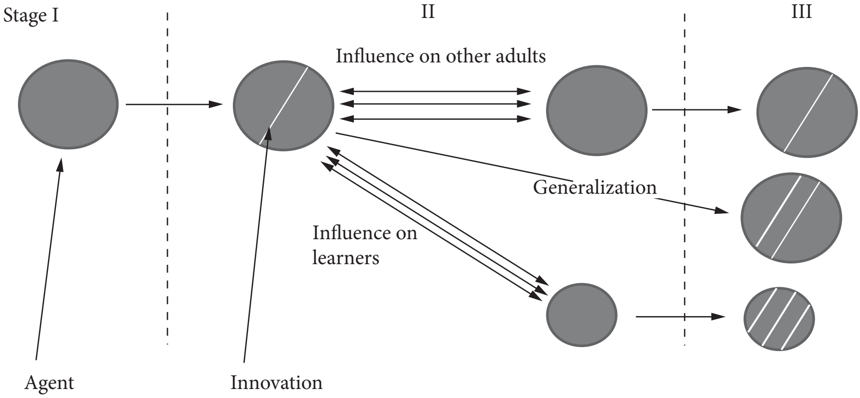


FIGURE 8.6. Spread of innovation to adults and children across generations

The effect is a picture along the lines of Figure 8.5, where one construction is prevalent in one region, a second construction is prevalent in another region, and the two are both found with differing strengths in intermediate regions.

Figure 8.6 illustrates how constructions spread. In stage I we have a population with a particular construction in the grammar. In stage II there is a constructional innovation in this population. It is conveyed to other adult members of the population, and to children (symbolized by the small circle). In stage III more members of the population have acquire the construction, some of the members of the population that originally had the construction have generalized it, and children have generalized it even further.

On this view of constructional propagation we can see that it is possible for a single speaker to "have" several constructions with different verb cluster linearizations. There can be a construction or constructions that are used in normal discourse (analogous to symptomatic viruses), a construction or constructions that are used for special situations such as contrast or repair, and constructions

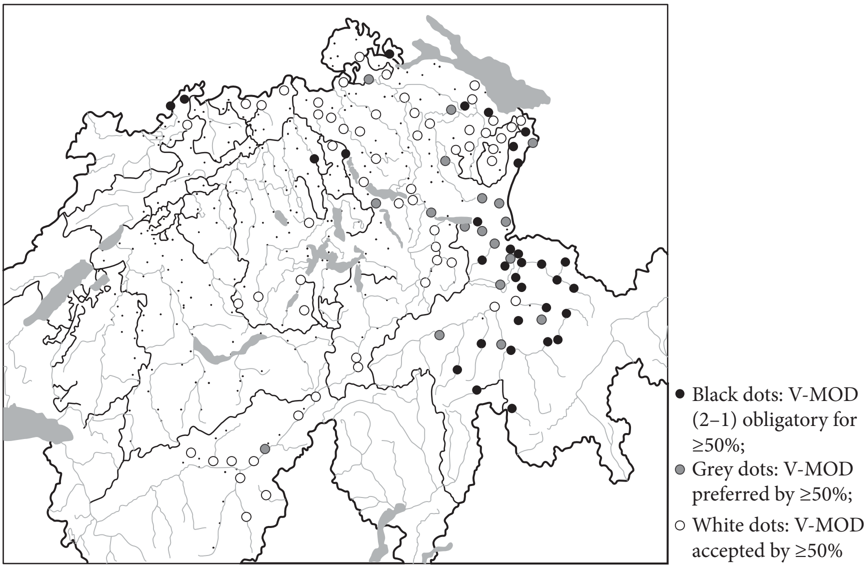


FIGURE 8.7. Swiss two-verb clusters (Seiler 2004). From the research project on Swiss German dialect syntax at the University of Zurich (principal investigator: Elvira Glaser; Swiss National Science Foundation 11-68244)

that may be judged acceptable or possible, but are not used in discourse (analogous to asymptomatic viruses). This last group of constructions are those that the speaker has heard sufficiently often to recognize and accept but not sufficiently often that they are a part of speakers' active grammar.

The variability that we see in language varieties arises from the fact that different speakers are exposed to different constructions with different frequencies, and from the fact that there is generalization within an individual speaker of the construction that is defined in terms of a particular verb or verb class. This variability reflects the fact that constructions spread and generalize through the individual lexicon, as well as through the population.

Seiler (2004) provides an elegant summary of cluster variation in Swiss German dialects that nicely illustrates these points. The spread of verb clusters in Swiss German the general pattern given in Figure 8.5. Consider Figure 8.7. This figure shows that in general the order 2-1 is obligatory in the eastern part of the area, it is preferred over the order 1-2 as we move west, and it is accepted (in the sense discussed above) as we move still further west. A similar pattern is shown by the distribution of three-verb clusters, particularly 3-1-2 versus 1-2-3, as shown in Figure 8.8.

One interpretation of this pattern is that 3-1-2 is an innovation in the 1-2-3 region. It occurs in those areas where 2-1 is obligatory or strongly preferred.

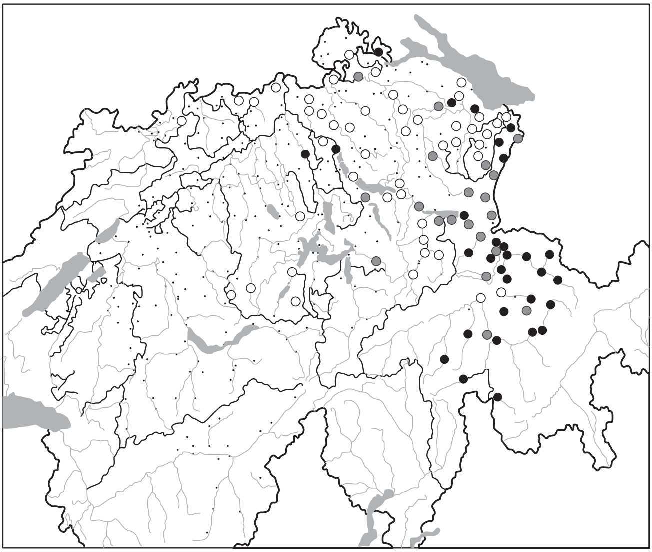


FIGURE 8.8. Swiss three-verb clusters (Seiler 2004). From the research project on Swiss German dialect syntax at the University of Zurich (principal investigator: Elvira Glaser; Swiss National Science Foundation 11-68244)

West ←——→ East

	I	II	III	IV	V
AUX,V	1–2	2–1	2–1	2–1	2–1
MOD,V	1–2	1–2	2–1	2–1	2–1
V_1,V_2	1–2	1–2	1–2	2–1	2–1
AUX,MOD,V	1–2–3	1–2–3	1–2–3	1–2–3	3–1–2

FIGURE 8.9. Distribution of V clusters (Seiler 2004)

We can interpret both the 2-1 and 3-1-2 orders in terms of the dependency bias. In fact, the pattern can be seen as spreading not only to the population from east to west, but through the lexicon, as Figure 8.9 from Seiler (2004) shows.

The pattern that we see is just that of Figure 8.5, where one pattern (VC1) is moving west, and the other (VC2) is moving east, as shown in Figure 8.10. Using the computational simulation described in Chapter 6 we can model this spread of properties; see Figure 8.11.

As Seiler says, "First, I have shown that the ordering of elements in western dialects is strictly ascending (1-2-3), but the more we move eastwards the more the tendency for ascending ordering weekends. Second, the ordering of

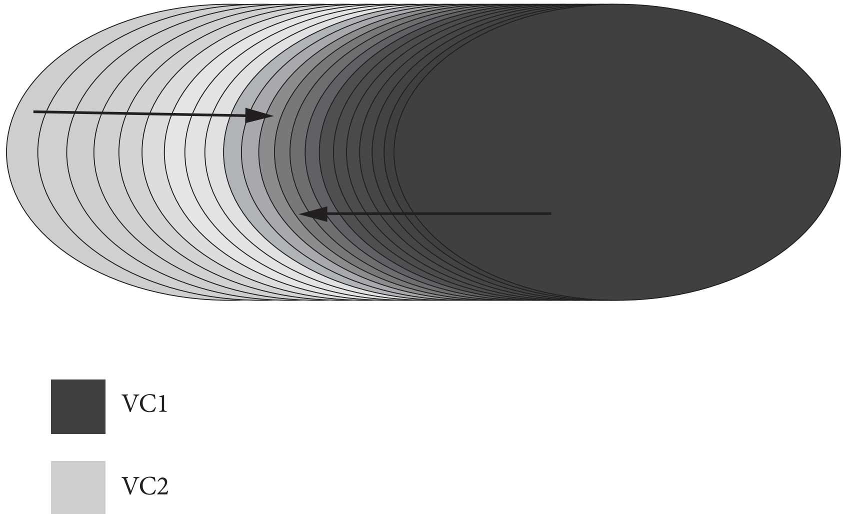


FIGURE 8.10. Distribution of verb cluster patterns over region

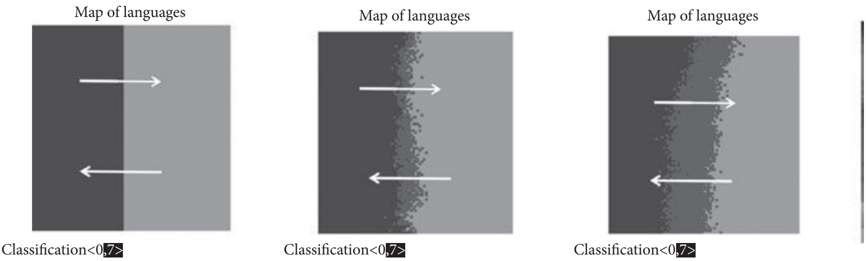


FIGURE 8.11. Spread of properties in two directions

elements is sensitive to the category of the head. Auxiliaries tend most to be set at the right edge of the cluster. **This tendency is much weaker with modal verbs and almost absent with lexical verbs as heads of a cluster.**" (Emphasis mine—PWC)

One way to interpret this pattern is that in this dialect area, the 1-2(-3) pattern of verb clusters was originally very widespread. Then the 2-1 order began as an innovation stimulated by the dependency bias, where 1 is AUX, and subsequently extended to MOD and finally V. On this view, the order 3-1-2 is a final generalization of 2-1, extending the change motivated by the dependency bias.

A pattern similar to that found in Swiss German is seen in Heerlen Dutch (Cornips 2009). In this variety, 2-1 is preferred for V1=AUX and 1-2 is strongly preferred for V1=MOD. One interpretation of this state of affairs is that there is a change going on from 1-2 to 2-1 which is applying first to V1=AUX. This interpretation is consistent with that of Cornips (2009: 212) who argues for basic 1-2 on the basis of the correlations between order preferences of individual speakers.

In West Flemish, 2-1 is required for V1=AUX and 1-2 is required for V1=MOD (Haegeman 1994). Again, this pattern suggests a shift from 1-2 to 2-1 beginning (and ending) with AUX-V.

This pattern of change reflects essentially the same pattern seen with *do* support in early modern English, as documented by Ellegård and analyzed by Kroch and discussed in Chapter 7: the change begins with a specific element and then generalizes to a broader range of similar cases.

Let us conclude with some observations about the three-verb clusters. The following quote from Wurmbrand (2005: 199) shows how specifically constructional the various orders are:

> West Flemish allows all five orders attested in verb clusters, however, the distribution of these orders is severely restricted by the type of construction: the '1-2-3' order is only possible in the double modal construction and certain IPP-constructions; the '1-3-2' order is only possible in participle constructions; the '3-1-2' order is only possible in the modal-auxiliary-participle construction; the '3-2-1' order is only possible in double auxiliary constructions; and finally, the '2-3-1' order is restricted to IPP constructions.

Observations such as these, and those summarized in Figure 8.3 regarding Swiss German, suggest that there are complex factors at play that must be taken into account, factors such as those alluded to by Lötscher (1978). But it is possible to make some reasonable conjectures about which three-verb clusters should occur in a variety, given the basic two-verb pattern and the biases.

If a variety permits 1-2 or 2-1 for a particular pair of verbs, then by iteration, it should also permit 1-(2-3) and (3-2)-1 respectively for longer combinations of these verbs. Hence we would not expect to find 3-2-1 in the absence of 2-1, or 1-2-3 in the absence of 1-2. If a variety mixes 1-2 and 2-1, as West Flemish and Afrikaans do for V1=AUX and MOD, we expect 2-3-1 where V1=AUX, V2=MOD, since V follows MOD and AUX is maximally final in the sequence. Wurmbrand (2005) identifies West Flemish as 2-3-1 in the IPP construction; e.g., (43) from Haegeman 1994.

(43) ... da Valère nie nor us will$_2$-en kom$_3$-en eet$_1$
that Valery not to house want-INF come-INF has
'... that Valery did not want to come home.'

2-3-1 clusters with a range of verbs that are well documented in Afrikaans (Biberauer n.d.), e.g.,

(44) a. . . . dat hy die medisyne kon drink het [modal]
that he the medicine could.INF drink.INF have
'. . . that he could drink the medicine'

b. . . . dat hy hom die medisyne maak/laat drink
that he him the medicine make.INF/let.INF drink.INF
het [causative]
have
'. . . that he made/let him drink the medicine'

c. . . . dat hy haar hoor roep het [perception]
that he her hear.INF call.INF have
'. . . that he heard her call'

d. . . . dat ek haar die bokse help dra het [benefactive]
that I her the boxes help.INF carry.INF have
'. . . that I helped her carry the boxes'

e. . . . dat die mense bly staan het [durative]
that the people remain.INF stand.INF have
'. . . that the people remained standing'

f. . . . dat dit ophou reën het [inchoative]
that it stop.INF rain.INF have
'. . . that it has stopped raining'

g. . . . dat hy probeer voorgee het [control]
that he try.INF pretend.INF have
'. . . that he tried to pretend'

h. . . . dat hy die boek gaan lees het [motion]
that he the book go.INF buy.INF have
'. . . that he went to buy the book'

i. . . . dat hy die boek loop (en) koop het [linking]
that he the book walk.INF and buy.INF have
'. . . that he went and bought the book'

j. . . . dat hy die boek sit en lees het [linking]
that he the book sit and read have
'. . . that he was sitting and reading the book'

According to Biberauer, this 2-3-1 pattern alternates with 1-2-3, and is possible only when the finite verb is a form of *hebben* 'have,' that is, it is IPP. Thus it appears that in Afrikaans the 2-3-1 pattern has generalized into a distinct construction, one that is restricted to a particular finite verb but is very free with regards to V2.

I have already suggested that 3-1-2 may be motivated by the dependency bias applying in a 1-2-3 variety. The coexistence of MOD-MOD-V and

V-MOD-AUX in West Flemish may be because the scope bias for two modals is stronger than the dependency bias, which emerges when there is only one modal.

Wurmbrand (2005: 11) notes that in Dutch 1-2-3 may alternate with 3-1-2 and 1-3-2, which can be seen as a consequence of the dependency bias moving the main verb towards the beginning of the cluster; a similar pattern is found in Afrikaans. But we do not find 3-1-2 in Standard German and Frisian, where 3-2-1 is the dominant order in three-verb clusters.

Since both biases are operative, we expect to find alternative orders even for the same types of sequences. At the same time, we cannot predict that a particular pattern will hold since we do not know precisely what the social factors are that are responsible for the growth in certain orders. It is always possible that a particular order can fully generalize, e.g. as in standard German, which is always (3)-2-1 (except for IPP).

Another contributor to complexity in the CWG clusters is the phenomenon of IPP, noted earlier. In standard German, IPP shows a 1-3-2 pattern, even though the canonical order is 3-2-1 (6). But in other varieties, IPP does not require 1-3-2. In Viennese Austrian vernacular IPP is 3-2-1—

(45) daß er sie nicht fragen können hätte
that he her not ask.INF can.INF have.SUBJ
'that he could not have asked her'
(Haider 2003: 124)

—while in Dutch, IPP is 1-2-3—

(46) dat Jan het boek heeft kunnen lezen
that Jan the book has can.INF read.INF
'that Jan was able to read the book'
(Wurmbrand 2005: 195)

For additional complications of IPP, see among others De Vos (2002) and Zwart (2007).

Restrictions on three-verb clusters also appear to be sensitive to focus and other discourse- and weight-related factors; see Lötscher (1978), Sapp (2005, 2006, 2011), Dubenion-Smith (2010), and Bader and Schmid (2009b,c). It is entirely plausible that general focus-ordering preferences in Germanic that order the constituents of a phrase apply to VC phrases. For instance, scrambling in the Mittelfeld in German places the focused constituent rightmost in VP (immediately before V in embedded clauses). A definite NP must scramble to the left in order to avoid being interpreted as focus (Winkler 2005: 213–14), but an indefinite can appear in the internal position because it is new information. It is conceivable that the rightmost position in the verbal sequence could be similarly used to express focus of one of the verbs in a cluster, as Schmid and Vogel (2004) suggest. Sapp (2006) found that new information and contrastive focus was more likely to appear with 1-2 clusters than with 2-1 clusters in his Early New High German corpus. In three-verb clusters where there is focus in the VP the order 3-2-1 is dispreferred over the alternatives. Dubenion-Smith (2010) found a similar, albeit weak, effect in his corpus. However, the precise mechanisms that underlie such word order alternations are poorly understood. Moreover, the interactions between various competing factors are complex and it is difficult if not impossible to identify a single factor as responsible for a given word order (cf. Wasow 2002).

With all this being said, it does appear that a picture can be drawn that makes plausible sense of word order variation in CWG verb clusters. The picture makes crucial use of constructions as an explanatory device, one that restricts the set of possible reorderings to the daughters of a phrase. The explanation draws as well on constructional biases (dependency and scope) that give preference to some orderings over others. These biases operate within a social network to create subgroups of speakers who favor one or another alternative constructional variant. It is important to interpret the variation data somewhat non-categorically, so that some options may be accepted by speakers and form part of their grammars in some sense (due to contact or innovation), but are not responsible for their productions.

It is also clear that the account that we have arrived at is by no means predictive, nor could it be. There are simply too many unknowns regarding the internal states of actual speakers of a language, the topology of the social networks in which they interact, the frequencies with which particular constructions are used, the strength of the biases for various alternatives, and the presence of processing and other factors that might affect how the language develops over time. In this respect the study of language variation, that is, what is possible, what is impossible, and why what we see is the way it is, faces much the same issue as other complex systems such as economics and biological evolution. There are too few pieces of data and too many possible outcomes to be able to prove with any certainty that there is a causal connection between them. But I hope to have at least shed some light on some of the ways in which the complexity of the linguistic representation and complexity of processing the form–meaning correspondence may play a role in explaining why a language takes the form that it does.

8.7 Summary

This book has investigated the idea that the form of the grammar of a natural language is determined in part by a resistance to complexity. I have argued that a useful measure of formal complexity can be stated in terms of constructions. Languages change in the direction of reducing constructional complexity, by generalizing the classes to which constructions apply, and eliminating idiosyncrasies. At the same time, change may result in new areas of complexity, as constructions become respecialized.

Another type of complexity, processing complexity, plays a role in determining the acceptability of syntactic configurations, and thus plays a role in how we construct grammars and our grammatical theory. The key point is that in the absence of high frequency of occurrence, a complex configuration will be judged unacceptable by native speakers. Such cases have typically been characterized as 'ungrammatical' in linguistic theory, leading to attempts to formulate the constraints on complex configurations in grammatical terms. I have argued that these cases should not be ruled out as ungrammatical, but rather should be explained in terms of the embodiment of processing complexity.

A central theme of the complexity scenario is that the representation of language exists not only in the mind, but in the social network. Other things being equal, grammatical idiosyncrasies and grammatical but highly complex configurations are vulnerable to elimination under the pressure to increase simplicity and generality. But other thing are rarely if ever equal. In particular, an idiosyncrasy or complex configuration can resist elimination if it receives sufficient support from the social network, either in the form of high frequency of occurrence, or insulation from competition with simpler and more general alternatives due to social and geographical factors.

I concluded in this chapter by situating the main ideas of this book within a broader perspective, and suggesting possible future lines of development. On the model offered in Chapter 2 and Culicover and Jackendoff (2005), the shape that a language (and a grammar) takes is determined by these primary factors:

- universal grammar/evaluation metric;
- 'concrete' minimalism in learning (exemplified by ***CAMiLLe*** in §6.3);
- the social network.

Each of these contributes to grammar construction in the learner, and to the persistence, diffusion, or loss of particular linguistic properties in a population over time.

Universal grammar consists of the structures and principles that constrain grammars; I include in this category universal conceptual structure (Jackendoff 1990, 2002) and the laws governing the correspondences between CS and strings of words (Culicover and Jackendoff 2005). The grammars of individual languages consist of sets of constructions. The most general constructions are ultimately realizations of UG structures; the others express more or less idiosyncratic correspondences between syntactic forms and CS representations. The most idiosyncratic constructions are in all likelihood remnants of earlier more general constructions that have been supplanted through change. The more general constructions typically carve out a more or less systematic chunk of CS representation and specify a syntactic configuration that may be used to express it.

One of the obvious concerns raised by this approach to constructions, especially in light of the central role of syntactic uniformity found in MGG, is that it appears to leave open the possibility that language can vary in arbitrary and unbounded ways. I address this concern in Chapter 2. In constructionalist approaches to grammar this issue is managed by taking into account the extent to which a construction shares certain properties with other constructions, including those that are characterized in terms of very general rules (e.g. the structure of the VP) (see Kay and Fillmore 1999: 30 for one example).

More generally, following Culicover and Jackendoff (2005: chapter 1), English has a hierarchy of VP constructions, which range from the very specific and idiomatic (e.g. *kick the bucket*) to the very general ($[_{VP}$ V ...]). Above the language-specific construction hierarchy is Universal Grammar (UG) that stipulates endocentricity as the default state of affairs for phrases. On this view, UG is not actually part of the grammar of any particular language. Through the evaluation metric, it is a guide to the construction of a grammar. It sets defaults from which general rules of grammar may depart, at a cost, just as the rules of any given language set defaults from which particular constructions of the language may depart.

Along related lines, Culicover (1999) (see also Hawkins 1994) argues that constraints, such as Ross's (1967) constraints, are measures of markedness that can be violated by particular constructions, but at a cost. They are not universal prohibitions against certain configurations, but reflections of the relative complexity of these configurations. I propose (Chapter 5), that such constraints should not be in the grammar or grammatical theory per se. They are "universal" to the extent that they reflect universals of processing complexity, but not grammatical universals.

The overall picture that emerges is one in which each level of the hierarchy, from UG on down, establishes a standard against which the complexity of deviations are measured. The extent of deviation from the standard may then be understood as contributing to the level of complexity in the grammar. This notion of complexity offers the potential for explaining language change in terms of the pressure to simplify (i.e. to achieve "economy"), while leaving open the possibility of deviation from the ideal at a cost.[9] The cost, in turn, translates into frequency of occurrence (among languages), and perhaps learnability and processing complexity, as discussed in Chapter 5.

A theory of constructions, then, must incorporate an evaluation metric, in order to explain what is natural and what is not. But this is true of any theory of grammar. In the current framework, the evaluation metric distinguishes grammars in terms of the deviation of general rules from UG and deviations of constructions from general rules (Chapters 3 and 4). This is in fact the classical sense of the evaluation metric introduced by Chomsky (1965), rather than in the framework of economy of derivation, which has been the focus of much recent work.

Turning to the second point, 'concrete minimalism' refers to the conservative strategy of a learner to construct only those hypotheses about the form–meaning correspondences that are warranted by the evidence and universal grammar. On the view that universal grammar is relatively impoverished (see Culicover 1999 for arguments), minimal generalization leads to a situation in which learners are always beginning from constructions. They move to rules only when the weight of the evidence warrants generalization beyond the evidence. Precisely what justifies and constrains generalization is an open question, and a fundamental one, but the evidence from language acquisition appears to be consistent with this general view.

Finally, language exists in a social network. The social and physical topology of the network determines to some extent the stability of clusters of linguistic properties quite independent of their substantive properties. This point is demonstrated in Chapter 6, and holds more generally for properties of agents that are transmitted to other agents in a social network (see e.g. Latané and Nowak 1997). One interesting and important consequence of the fact that language is situated in a social network is that the co-occurrence of linguistic properties may be a reflection of their distribution in the network, and not of their inherent content. On the other hand, the possibility of bias, due to formal or processing complexity as well as to social factors, also plays a role in the distribution of linguistic properties and combinations of linguistic properties.

[9] This idea goes back to early work on markedness in generative grammar, e.g. Chomsky (1964).

References

Abeillé, Anne and Robert D. Borsley (2008), "Comparative correlatives and parameters," *Lingua*, 118: 1139–57.

Abels, Klaus (2003), "Successive cyclicity, anti-locality, and adposition stranding." Unpublished doctoral dissertation. Storrs, Conn.: University of Connecticut.

Abney, Steven (1987), "The noun phrase in its sentential aspect." Unpublished doctoral dissertation. Cambridge, Mass.: MIT.

Ackema, Peter, Ad Neeleman, and Fred Weerman (1993), "Deriving functional projections," in Amy Schafer (ed.), *Proceedings of the Twenty-Third Meeting of the Northeast Linguistics Society* (Amherst, Mass.: GLSA), 17–31.

Alexopolou, Theodora and Frank Keller (2003), "Linguistic complexity, locality and resumption," in Gina Garding and Mimu Tsujimura (eds), *Proceedings of the Twenty-Second West Coast Conference on Formal Linguistics* (Somerville, Mass.: Cascadilla Press), 15–28.

Allen, Cynthia L. (1980a), *Topics in Diachronic English Syntax* (New York: Garland).

—— (1980b), "Movement and deletion in Old English," *Linguistic Inquiry*, 11: 261–323.

Allwood, Jens (1982), "The Complex NP Constraint in Swedish," in Elisabet Engdahl and Eva Ejerhed (eds), *Readings on Unbounded Dependencies in Scandinavian Languages* (Stockholm: Almqvist and Wiksell), 15–32.

Ambridge, Ben and Adele E. Goldberg (2008), "The island status of clausal complements: Evidence in favor of an information structure explanation," *Cognitive Linguistics*, 19: 357–89.

Ariel, Mira (1999), "Cognitive universals and linguistic conventions: The case of resumptive pronouns," *Studies in Language*, 23: 217–69.

—— (2004), "Accessibility marking: Discourse functions, discourse profiles, and processing cues," in Ted J. M. Sanders and Morton Ann Gernsbacher (eds), *Accessibility in Text and Discourse Processing: A Special Issue of Discourse Processes, A Multidisciplinary Journal*, 37/2: 91–116.

Auwera, Johann van der (1999), "Periphrastic 'Do': Typological Prolegomena," in G. A. J. Tops, D. Devriendt, and G. Geukens (eds), *Thinking English Grammar: To Honour Xavier Dekeyser, Professor Emeritu* (Leuven: Peeters), 457–70.

Bach, Kent and Robert M. Harnish (1979), *Linguistic Communication and Speech Acts* (Cambridge, Mass.: MIT Press).

Bader, Markus and Tanja Schmid (2009a), "Minimality in verb-cluster formation," *Lingua*, 119: 1458–81.

—— —— (2009b), "Verb clusters in colloquial German," *Journal of Comparative Germanic Linguistics*, 12: 175–228.

—— —— (2009c), "CAT meets GO: 2-verb clusters in German," in Jeroen van Craenenbroeck (ed.), *Alternatives to Cartography* (Berlin: Mouton de Gruyter), 203–44.

Bader, Markus, Tanja Schmidt, and Janna Häussler (2009), "Optionality in verb-cluster formation," in Susanne Winkler and Sam Featherston (eds), *The Fruits of Empirical Linguistics. Volume 2: Product* (Berlin: Walter de Gruyter), 37–58.

Baekken, Bjoerg (2000), "Inversion in Early Modern English," *English Studies*, 5: 393–421.

—— (2002), "'Yet this fillie doth many times assault the brauest minds': Affirmative declarative *do* in 17th-century English," *Nordic Journal of English Studies*, 1: 317–37.

Baker, Mark, Kyle Johnson, and Ian Roberts (1989), "Passive arguments raised," *Linguistic Inquiry*, 20: 219–51.

Battistella, Edward L. (1996), *The Logic of Markedness* (New York: Oxford University Press).

Bayer, Joseph (1995), *Directionality and Logical Form: On the Scope of Focusing Particles and Wh-in-Situ* (Dordrecht: Kluwer Academic Publishers).

—— (2002), "Decomposing the left periphery: Dialectal and cross-linguistic evidence," in Horst Lohnstein and Susanne Trissler (eds), *Proceedings of the Israeli Association of Theoretical Linguistics (Iatl)*, 18 (Berlin: Mouton de Gruyter), 59–96.

—— (2004), "Was beschränkt die Extraktion? Subject–Objekt vs. Topik–Focus," in Franz Josef D'Avis (ed.), *Deutsche Syntax: Empirie Und Theorie* (Göteborg: Acta Universitatis Gothoburgensis), 233–57.

Beck, Sigrid (1997), "On the semantics of comparative conditionals," *Linguistics and Philosophy*, 20: 229–32.

Beecher, Henry (2006), "Pragmatic licensing of sluiced prepositional phrases." Unpublished manuscript. San Diego, CA: University of California, San Diego.

Benincà, Paola and Cecilia Poletto (2004), "A Case of *do*-Support in Romance," *Natural Language and Linguistic Theory*, 22: 51–94.

Bennis, Hans (1986), *Gaps and Dummies* (Dordrecht: Foris Publications).

Bever, Thomas G. and David J. Townsend (2001), *Sentence Comprehension: The Integration of Habits and Rules* (Cambridge, Mass.: MIT Press).

Biberauer, Theresa (n.d.) "Disharmonic word order, quirky morphology and the Afrikaans verb cluster." Unpublished manuscript. Available online at <http://www.hum.au.dk/engelsk/engsv/nyvad-abstracts/biberauer-ho.pdf>.

Bies, Ann E. (1996), "Syntax and discourse factors in Early New High German: Evidence for verb-final word order." Unpublished Masters thesis. Philadelphia. University of Pennsylvania.

Birner, Betty J. (1996), *The Discourse Function of Inversion in English* (New York: Garland Publishing).

—— and Gregory Ward (1998), *Information Status and Non-canonical Word Order in English* (Amsterdam and Philadelphia: John Benjamins).

Black, James R. and Virginia Motapanyane (eds) (1996), *Micro-Parametric Syntax and Dialect Variation* (Amsterdam: John Benjamins).

Bloom, Paul (2000), *How Children Learn the Meanings of Words* (Cambridge, Mass.: MIT Press).

Bobaljik, Jonathon David (1995), "Morphosyntax: The syntax of verbal inflection." Unpublished doctoral dissertation. Cambridge, Mass.: MIT.

—— and Phil Branigan (2007), "Eccentric agreement and multiple case checking," in Alana Johns, Diane Massam, and Juvenal Ndayiragije (eds), *Ergativity* (Dordrecht: Springer), 47–77.

Boland, Julie E. (2005), "Cognitive mechanisms and syntactic theory," in Anne Cutler (ed.), *Twenty-First Century Psycholinguistics: Four Cornerstones* (Mahwah, NJ: Lawrence Erlbaum Associates), 23–42.

Booij, Geert (2008), "Constructional idioms as products of linguistic change: The *aan + Infinitive* construction in Dutch," in Alexander Bergs and Diewald Diewald (eds), *Constructions and Language Change* (Berlin: Mouton de Gruyter), 79–104.

Borer, Hagit (1984), *Parametric Syntax* (Dordrecht: Foris Publications).

Borsley, Robert D. and Frederick J. Newmeyer (2009), "On subject-auxiliary inversion and the notion 'purely formal generalization'," *Cognitive Linguistics*, 20: 135–43.

Bouma, Gosse and Gertjan van Noord (1998), "Word order constraints on Germanic verb clusters," in Erhard Hinrichs, Tsuneko Nakazawa, and Andreas Kathol (eds), *Complex Predicates in Nonderivational Syntax* (New York: Academic Press), 43–71.

Brighton, Henry (2005), "Composition, linguistic evolution and induction by minimum description length," in Markus Werning and Edouard Machery (eds), *The Compositionality of Concepts and Meanings: Applications to Linguistics, Psychology and Neuroscience* (Frankfurt: Ontos Verlag), 13–39.

—— Kenneth Smith, and Simon Kirby (2005), "Language as an evolutionary system," *Physics of Life Reviews*, 2: 177–226.

Brown, Roger and Camille Hanlon (1970), "Derivational complexity and the order of acquisition in child speech," in John R. Hayes (ed.), *Cognition and the Development of Language* (New York: Wiley), 11–53.

Burn, John (1766), *A Practical Grammar of the English Language* (Glasgow: Archibald McLean).

Bybee, Joan (2006), *Frequency of Use and the Organization of Language* (Oxford: Oxford University Press).

Chater, Nick and Paul Vitányi (2007), "'Ideal learning' of natural language: Positive results about learning from positive evidence," *Journal of Mathematical Psychology*, 51: 135–63.

Chen, Rong (2003), *English Inversion: A Ground-Before-Figure Construction* (Berlin and New York: Mouton de Gruyter).

Chomsky, Noam (1957), *Syntactic Structures* (The Hague: Mouton).

—— (1964), *Current Issues in Linguistic Theory* (The Hague: Mouton).

—— (1965), *Aspects of the Theory of Syntax* (Cambridge, Mass.: MIT Press).

—— (1972), "Remarks on nominalization," in Richard Jacobs and Peter Rosenbaum (eds), *Readings in English Transformational Grammar* (Waltham, Mass.: Ginn and Co), 184–221.

Chomsky, Noam (1973), "Conditions on transformations," in Steven Anderson and Paul Kiparsky (eds), *Festschrift for Morris Halle* (New York: Holt, Rinehart and Winston), 232–86.

—— (1977), "On wh movement," in Peter W. Culicover, Thomas Wasow, and Adrian Akmajian (eds), *Formal Syntax* (New York: Academic Press), 71–132.

—— (1981a), "Markedness and core grammar," in Adriana Belletti, L. Brandi, and Luigi Rizzi (eds), *Theory of Markedness in Generative Grammar* (Pisa: Scuola Normale Superiore), 123–46.

—— (1981b), *Lectures on Government and Binding* (Dordrecht: Foris Publications).

—— (1986a), *Barriers* (Cambridge, Mass.: MIT Press).

—— (1986b), *Knowledge of Language: Its Nature, Origin, and Use* (New York: Praeger).

—— (1993), "A minimalist program for linguistic theory," in Kenneth Hale and Samuel Jay Keyser (eds), *The View From Building Twenty* (Cambridge, Mass.: MIT Press), 1–52.

—— (1995), *The Minimalist Program* (Cambridge, Mass.: MIT Press).

—— (2005), "Three factors in language design," *Linguistic Inquiry*, 36: 1–22.

—— and Morris Halle (1968), *The Sound Pattern of English* (New York: Harper and Row).

—— and Howard Lasnik (1977), "Filters and control," *Linguistic Inquiry*, 8: 425–504.

———— (1993), "Principles and Parameters theory," in Joachim Jacobs, Arnim von Stechow, and Theo Vennemann (eds), *Syntax: An International Handbook of Contemporary Research* (Berlin: Walter de Gruyter), 506–69.

Chung, Sandra, William Ladusaw, and James McClosky (1995), "Sluicing and logical form," *Natural Language Semantics*, 3: 239–82.

Cinque, Guglielmo (1990), *Types of A-Bar Dependencies* (Cambridge, Mass.: MIT Press).

—— (1999), *Adverbs and Functional Heads: A Cross-Linguistic Perspective* (Oxford: Oxford University Press).

Clifton, Charles and Lyn Frazier (1989), "Comprehending sentences with long-distance dependencies," in Michael K. Tanenhaus and Gregory N. Carlson (eds), *Linguistic Structure in Language Processing* (Dordrecht: Kluwer Academic Publishers), 272–317.

Cornips, Leonie (2009), "Empirical syntax: Idiolectal variability in two-and three-verb clusters in regional Standard Dutch and Dutch dialects," in Andreas Dufter, Juerg Fleischer, and Guido Seiler (eds), *Describing and Modeling Variation in Grammar* (Berlin and New York: Mouton de Gruyter), 203–24.

Cover, Thomas M., Peter Gacs, and Robert M. Gray (1983), "Kolmogorov's contributions to information theory and algorithmic complexity," *The Annals of Probability*, 17: 840–65.

Craig, Collette (1977), *The Structure of Jacaltec* (Austin: University of Texas Press).

Culicover, Peter W. (1971), "Syntactic and semantic investigations." Unpublished doctoral dissertation. Cambridge, Mass.: MIT.

—— (1973), "On the coherence of syntactic descriptions," *Journal of Linguistics*, 9: 35–51. Reprinted in Peter W. Culicover (forthcoming), *Explaining Syntax: Representations, Structures and Computation* (Oxford: Oxford University Press).

Culicover, Peter W. (1991), "Topicalization, inversion, and complementizers in English," in Denis Delfitto, Martin Everaert, Arnold Evers, and Frits Stuurman (eds), *Going Romance, and Beyond: Fifth Symposium on Comparative Grammar*, OTS Working Papers (Utrecht: University of Utrecht), 1–43. Reprinted in Peter W. Culicover (forthcoming), *Explaining Syntax: Representations, Structures and Computation* (Oxford: Oxford University Press).

—— (1993a), "Evidence against ECP accounts of the *that-t* effect," *Linguistic Inquiry*, 24: 557–61. Reprinted in Peter W. Culicover (forthcoming), *Explaining Syntax: Representations, Structures and Computation* (Oxford: Oxford University Press).

—— (1993b), "Degrees of freedom," in Eve V. Clark (ed.), *Proceedings of the Child Language Research Forum* (Stanford: Department of Psychology), 30–7.

—— (1996), "Adaptive learning and concrete minimalism", in C. Koster and F. Wijnen (eds), *Proceedings of the Groningen Assembly on Language Acquisition (GALA 1995)*, Center for Language and Cognition, Groningen, 165–74.

—— (1999), *Syntactic Nuts* (Oxford: Oxford University Press).

—— (2001), "Parasitic gaps: A history," in Peter W. Culicover and Paul Postal (eds), *Parasitic Gaps* (Cambridge, Mass.: MIT Press), 3–68.

—— (2008a), "Squinting at Dali's Lincoln—How to think about language," *Proceedings of the Forty-First Annual Meeting of the Chicago Linguistics Society* (Chicago: Chicago Linguistics Society), 109–28.

—— (2008b), "The rise and fall of constructions and the history of English *do*-support," *Journal of Germanic Linguistics*, 20: 1–52.

—— (2009), "Simpler Syntax," in Bernd Heine and Heiko Narrog (eds), *The Oxford Handbook of Linguistic Analysis* (Oxford: Oxford University Press).

—— (2010), "Parataxis and Simpler Syntax," in Marie-José Béguelin, Mathieu Avanzi, and Gilles Corminboeuf (eds), *Actes Du 1er Colloque International De Macrosyntaxe* (Pieterlen, Switzerland: Peter Lang), 121–38.

—— (2011), "A reconsideration of English relative constructions," *Constructions*, 2. Available online at <http://elanguage.net/journals/index.php/constructions/index>.

—— (forthcoming), *Explaining Syntax: Representations, Structures and Computation* (Oxford: Oxford University Press).

—— and Johannes Dellert (2008), "Going postal: The architecture of English change-of-state verbs," unpublished manuscript, the Otio State University.

—— and Ray Jackendoff (1997), "Syntactic coordination despite semantic subordination," *Linguistic Inquiry*, 28: 195–217.

—— —— (1999), "The view from the periphery: The English comparative correlative," *Linguistic Inquiry*, 30: 543–71.

—— —— (2005), *Simpler Syntax* (Oxford: Oxford University Press).

—— —— (2012), "A domain-general cognitive relation and how language expresses it," *Language*, 88: 305–40.

—— and Robert D. Levine (2001), "Stylistic inversion and the *that-t* effect in English: A reconsideration," *Natural Language and Linguistic Theory*, 19: 283–310.

Culicover, Peter W. and Louise McNally (eds) (1998), *The Limits of Syntax* (New York: Academic Press).

—— and Andrzej Nowak (2002), "Learnability, markedness, and the complexity of constructions," in Pierre Pica and Johan Rooryk (eds), *Language Variation Yearbook*, vol. 2 (Amsterdam: John Benjamins), 5–30.

—— —— (2003), *Dynamical Grammar* (Oxford: Oxford University Press).

—— —— and Wojtech Borkowski (2003), "Linguistic theory, explanation and the dynamics of language," in John Moore and Maria Polinsky (eds), *The Nature of Explanation in Linguistic Theory* (Stanford, CA: CSLI Publications), 83–104.

—— —— and Katherine Woznicki (2008), "Adventures with ***CAMiLLe***: Investigating the architecture of the language faculty through computational simulation," *OSU Working Papers in Linguistics*, 58: 1–29. Reprinted in Peter W. Culicover (forthcoming), *Explaining Syntax: Representations, Structures and Computation* (Oxford: Oxford University Press).

—— and Michael S. Rochemont (1990), "Extraposition and the Complement Principle," *Linguistic Inquiry*, 21: 23–48.

—— and Wilkins, Wendy (1986). "Control, PRO and the Projection Principle," *Language*, 62: 12–153.

—— and Susanne Winkler (2008), "English focus inversion constructions," *Journal of Linguistics*, 44: 625–58. Reprinted in Peter W. Culicover (forthcoming), *Explaining Syntax: Representations, Structures and Computation* (Oxford: Oxford University Press).

—— —— (2010), "Freezing: A conspiracy." Unpublished manuscript. Columbus, OH and Tübingen: Ohio State University and University of Tübingen.

Dąbrowska, Ewa (2008), "Questions with long-distance dependencies: A usage-based perspective," *Cognitive Linguistics*, 19: 391–425.

Dahl, Östen (2004), *The Growth and Maintenance of Linguistic Complexity* (Amsterdam: John Benjamins).

—— (2009), "Testing the assumption of complexity invariance: The case of Elfdalian and Swedish," in Geoffrey Sampson, David Gil, and Peter Trudgill (eds), *Linguistic Complexity: A Comprehensive Definition and Survey* (Oxford: Oxford University Press), 50–63.

Davies, Mark (2011), "Google Books (American English) Corpus (155 billion words, 1810–2009)." Available online at <http://googlebooks.byu.edu/>.

Dayal, Veneeta (2004), "Number marking and (in)definiteness in kind terms," *Linguistics and Philosophy*, 27: 393–450.

Dehé, Nicole (2005), "The optimal placement of *up* and *ab*: A comparison," *Journal of Comparative Germanic Linguistics*, 8: 185–224.

—— and Yordanka Kavalova (2007), "Parentheticals: An introduction," in Nicole Dehé and Yordanka Kavalova (eds), *Parentheticals* (Amsterdam: John Benjamins), 1–22.

Demberg, Vera and Frank Keller (2008), "Data from eye-tracking corpora as evidence for theories of syntactic processing complexity," *Cognition*, 109: 193–210.

Den Besten, Hans and Jerold A. Edmondson (1983), "The verbal complex in Continental West Germanic," in Werner Abraham (ed.), *On the Formal Syntax of the Westgermania* (Amsterdam: John Benjamins), 155–216.

Den Dikken, Marcel (2005), "Comparative correlatives comparatively," *Linguistic Inquiry*, 36: 497–533.

Denison, David (1993), *English Historical Syntax* (London and New York: Longmans).

De Vos, Mark (2003), "Past participles in Afrikaans dialects and Dutch," in Gina Garding and Mimu Tsujimura (eds), *Proceedings of the 22nd West Coast Conference on Formal Linguistics* (Somerville, Mass.: Cascadilla Press), 519–32.

Dowty, David (1991), "Thematic proto-roles and argument selection," *Language*, 67: 547–619.

Dryer, Matthew (2011), "Chapter 81: Order of subject, object and verb," in Matthew Dryer and Martin Haspelmath (eds), *The World Atlas of Language Structures Online* (Munich: Max Planck Digital Library).

Dubenion-Smith, Shannon A. (2010), "Verbal complex phenomena in West Central German: Empirical domain and multi-causal account," *Journal of Germanic Linguistics*, 22: 99–191.

Ellegård, Alvar (1953), *The Auxiliary Do: The Establishment and Regulation of Its Use in English* (Stockholm: Almqvist and Wiksell).

Elsig, Martin and Shana Poplack (2009), "Transplanted dialects and language change: Question formation in Québec," in Andreas Dufter, Jorg Fleischer, and Guido Seiler (eds), *Describing and Modeling Variation in Grammar* (Berlin and New York: Mouton de Gruyter), 255–69.

Emonds, Joseph E. (1970), *Root and Structure Preserving Transformations* (Bloomington, IN: Indiana University Linguistics Club).

—— (1972), "Evidence that indirect object movement is a structure-preserving rule," *Foundations of Language*, 8: 546–61.

—— (1976), *A Transformational Approach to English Syntax* (New York: Academic Press).

Enfield, Nicholas J. (2008), "Transmission biases in linguistic epidemiology," *Journal of Language Contact*, 2: 299–310.

Engdahl, Elisabet (1997), "Relative clause extractions in context," *Working Papers in Scandinavian Syntax*, 60: 51–79.

Erteschik-Shir, Nomi (1997), *The Dynamics of Focus Structure* (Cambridge: Cambridge University Press).

Everett, Daniel L. (2005), "Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language," *Current Anthropology*, 46: 621–46.

Evers, Arnold (1975), "V-raising in Dutch and German." Unpublished doctoral dissertation. Utrecht: University of Utrecht.

Fanselow, Gisbert and Stefan Frisch (2006), "Effects of processing difficult on judgements of acceptability," in Gisbert Fanselow, Caroline Féry, Ralf Vogel, and Matthias Schlesewsky (eds), *Gradience in Grammar* (Oxford: Oxford University Press), 291–316.

—— Reinhold Kliegl, and Matthias Schlesewsky (2005), "Syntactic variation in German wh-questions: Empirical investigations of weak crossover violations and long wh-movement," *Linguistic Variation Yearbook*, 5: 37–63.

Featherston, Sam (2001), *Empty Categories in Sentence Processing* (Amsterdam: John Benjamins).
—— (2005a), "Universals and grammaticality: Wh-constraints in German and English," *Linguistics*, 43: 667–711.
—— (2005b), "The decathlon model of empirical syntax," in Stephan Kepser and Marga Reis (eds), *Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives* (Berlin and New York: Mouton de Gruyter), 187–208.
Fillmore, Charles J. (1988), "The mechanisms of Construction Grammar," *Proceedings of the 14th Annual Meeting of the Berkeley Linguistics Society*, 14: 35–55.
—— Paul Kay, and Mary Catherine O'Connor (1988), "Regularity and idiomaticity in grammatical constructions: The case of *let alone*," *Language*, 64: 501–39.
Fischer, Olga, Ans van Kemenade, Willem Kopman, and Wim van der Wurrff (2000), *The Syntax of Early English* (Cambridge: Cambridge University Press).
Fodor, Janet D. (1978), "Parsing strategies and constraints on transformations," *Linguistic Inquiry*, 9: 427–73.
—— (1998), "Unambiguous triggers," *Linguistic Inquiry*, 29: 1–36.
—— and Atsu Inoue (1994), "The diagnosis and cure of garden paths," *Journal of Psycholinguistic Research*, 23: 407–34.
—— —— (2000), "Garden path reanalysis: Attach (anyway) and revision as last resort," in Marica De Vincenzi and Vincenzo Lombardo (eds), *Cross-Linguistic Perspectives on Sentence Processing* (Dordrecht: Kluwer), 21–61.
Fodor, Jerry A. (1983), *The Modularity of Mind* (Cambridge, Mass: MIT Press).
—— Thomas Bever, and Merrill Garrett (1974), *The Psychology of Language: An Introduction to Psycholinguistics and Generative Grammar* (New York: McGraw-Hill).
Fox, Danny and Howard Lasnik (2003), "Successive cyclic movement and island repair: The difference between sluicing and VP ellipsis," *Linguistic Inquiry*, 34: 143–54.
Frank, Robert (2004), "Restricting grammatical complexity," *Cognitive Science*, 28: 669–97.
—— and G. Satta (1998), "Optimality theory and the generative complexity of constraint violability," *Computational Linguistics*, 24: 307–15.
Frazier, Lynn (2008), "Processing ellipsis: A processing solution to the undergeneration problem," in Charles B. Chang and Hannah J. Haynie (eds), *Proceedings of the 26th West Coast Conference on Formal Linguistics* (Somerville, Mass.: Cascadilla Press), 21–32.
Freidin, Robert (2004), "Syntactic Structures Redux," *Syntax*, 7: 101–27.
Gergel, Remus (2008a), "Comparative inversion: A diachronic study," *The Journal of Comparative Germanic Linguistics*, 11: 191–211.
—— (2008b), "Comparatives and inversion in English: A (necessarily) diachronic account," *University of Pennsylvania Working Papers in Linguistics*, 14/1: 177–90.
Gibson, Edward (1998), "Linguistic complexity: Locality of syntactic dependencies," *Cognition*, 68: 1–76.

Gibson, Edward (2000), "The dependency locality theory: A distance-based theory of linguistic complexity," in Yasushi Miyashita, Alec Marantz, and Wayne O'Neil (eds), *Image, Language, Brain* (Cambridge, Mass.: MIT Press), 95–126.

—— and Tessa Warren (2004), "Reading time evidence for intermediate linguistic structure in long-distance dependencies," *Syntax*, 7: 55–78.

Gil, David (2008), "How complex are isolating languages?" in Matti Miestamo, Kaius Sinnemäki, and Fred Karlsson (eds.), *Language Complexity: Typology, Contact, Change* (Amsterdam: John Benjamins), 109–31.

Givón, Talmy (2009), *The Genesis of Syntactic Complexity: Diachrony, Ontogeny, Neuro-Cognition, Evolution* (Amsterdam: John Benjamins).

—— and Masayoshi Shibatani (2009), *Syntactic Complexity: Diachrony, Acquisition, Neuro-Cognition, Evolution* (Amsterdam and Philadelphia: John Benjamins).

Goldberg, Adele E. (1995), *Constructions: A Construction Grammar Approach to Argument Structure.* Chicago: University of Chicago Press.

—— (2006), *Constructions at Work: Constructionist Approaches in Context* (Oxford: Oxford University Press).

—— (2009), "Constructions work," *Cognitive Linguistics*, 20: 201–24.

—— and Alex Del Giudice (2005), "Subject-auxiliary inversion," *The Linguistic Review*, 22: 411–28.

Grodner, Daniel J. and Edward A. F. Gibson (2005), "Consequences of the serial nature of linguistic input for sentential complexity," *Cognitive Science*, 29: 261–91.

Grünwald, Peter D. (1996), "A minimum description length approach to grammar inference," in Stefan Wermter, Ellen Riloff, and Gabriele Scheler (eds), *Symbolic, Connectionist, and Statistical Approaches to Learning for Natural Language Processing*, Springer Lecture Notes in Artificial Intelligence 1040 (Berlin: Springer-Verlag), 203–16.

Haegeman, Liliane (1994), "Verb raising as verb projection raising: Some empirical problems," *Linguistic Inquiry*, 25: 509–22.

—— and Henk van Riemsdijk (1986), "Verb projection raising, scope and the typology of rules affecting verbs," *Linguistic Inquiry*, 17: 417–66.

Haider, Hubert (2003), "V-clustering and clause union," in Pieter A. M. Seuren and Gerard Kempen (eds), *Verb Constructions in German and Dutch* (Amsterdam: John Benjamins), 91–126.

—— and Inger Rosengren (2003), "Scrambling: Nontriggered chain formation in OV languages," *Journal of Germanic Linguistics*, 15: 203–67.

Hale, John T. (2003), "The information conveyed by words in sentences," *Journal of Psycholinguistic Research*, 32: 101–23.

—— (2011), "What a rational parser would do," *Cognitive Science*, 35: 399–443.

Hamburger, Henry (1980), "A deletion ahead of its time," *Cognition*, 8: 389–416.

Harris, Martin and Nigel Vincent (1980), "On zero relatives," *Linguistic Inquiry*, 11: 805–7.

Haspelmath, Martin (1997), *Indefinite Pronouns* (Oxford: Oxford University Press).

Hawkins, John A. (1982), "Cross-category harmony, X-bar and the predictions of markedness," *Journal of Linguistics*, 18: 1–35.

—— (1994), *A Performance Theory of Order and Constituency* (Cambridge: Cambridge University Press).

—— (1999), "Processing complexity and filler gap dependencies across grammars," *Language*, 75: 244-85.

—— (2001), "Why are categories adjacent?" *Journal of Linguistics*, 37: 1–34.

—— (2004), *Complexity and Efficiency in Grammars* (Oxford: Oxford University Press).

Hinterhölzl, Roland (2004), "Language change versus grammar change what diachronic data reveal about the distinction," in Eric Fuss and Carola Trips (eds), *Diachronic Clues to Synchronic Grammar* (Amsterdam: John Benjamins), 131–60.

—— and Svetlana Petrova (2005), "Rhetorical relations and verb placement in early Germanic languages: Evidence from the Old High German Tatian translation (9th century)," in Manfred Stede, Christian Chiarcos, Michael Grabsik, and Luuk Lagerwert (eds), *Salience in Discourse: Multidisciplinary Approaches to Discourse* (Münster: Stichting/Nodus), 71–9.

Hirtle, Walter (1997), "Do auxiliary—A meaningful support and operator," *Lingua*, 100: 111–49.

Hoffmann, Thomas (2011), *Preposition Placement in English* (Cambridge: Cambridge University Press).

Hofmeister, Philip (2007), "Representational complexity and memory retrieval in language comprehension." Unpublished doctoral dissertation. Stanford, CA: Stanford University.

—— Inbal Arnon, T. Florian Jaeger, Ivan A. Sag, and Neal Snider (submitted), "The source ambiguity problem: Distinguishing the effects of grammar and processing on acceptability judgments," *Language and Cognitive Processes*.

—— Laura Staum Casasanto, and Ivan A. Sag (forthcoming), "Islands in the grammar? Standards of evidence," in Jon Sprouse and Norbert Hornstein (eds), *Experimental Syntax and Island Effects* (Cambridge: Cambridge University Press).

—— and Ivan A. Sag (2010), "Cognitive constraints and island effects," *Language*, 86: 366–415.

Huddleston, Rodney A. and Geoffry K. Pullum (eds) (2002), *The Cambridge Grammar of the English Language* (Cambridge: Cambridge University Press).

Hunter, Patricia J. and Gary D. Prideaux (1983), "Empirical constraints on the verb-particle construction in English," *Journal of the Atlantic Provinces Linguistic Association*, 5: 3–15.

Jackendoff, Ray (1975), "Morphological and semantic regularities in the lexicon," *Language*, 51: 639–71.

—— (1977), *X-Bar Syntax: A Study of Phrase Structure* (Cambridge, Mass.: MIT Press).

—— (1983), *Semantics and Cognition* (Cambridge, Mass.: MIT Press).

—— (1990), *Semantic Structures* (Cambridge, Mass.: MIT Press).

Jackendoff, Ray (1992), "Babe Ruth homered his way into the hearts of America," in Timothy Stowell and Eric Wehrli (eds), *Syntax and the Lexicon 26* (New York: Academic Press), 155–78.

—— (1997), *The Architecture of the Language Faculty* (Cambridge, Mass.: MIT Press).

—— (2002a), "English particle constructions, the lexicon, and the autonomy of syntax," in Ray Jackendoff, Nicole Dehé, Silke Urban, and Andrew McIntyre (eds), *Verb-Particle Explorations* (Berlin: Walter de Gruyter), 67–94.

—— (2002b), *Foundations of Language* (Oxford: Oxford University Press).

—— (2007), *Language, Consciousness, Culture* (Cambridge, Mass.: MIT Press).

—— and Peter W. Culicover (1972), "A reconsideration of Dative Movement," *Foundations of Language*, 73: 397–412. Reprinted in Peter W. Culicover (forthcoming), *Explaining Syntax: Representations, Structures and Computation* (Oxford: Oxford University Press).

Jackson, Matthew O. and Brian W. Rogers (2007), "Relating network structure to diffusion properties through stochastic dominance," *The BE Journal of Theoretical Economics*, 7: 1–13.

Jacobson, Pauline (1994), "Binding connectivity in copular sentences," in Mary Harvey and Lynn Santelman (eds), *Proceedings of the Fourth Conference on Semantics and Linguistic Theory*, Cornell Working Papers in Linguistics (Cornell: CLC Publications) 161–78.

Jacobsson, Bengt (1951), *Inversion in English: With Special Reference to the Early Modern English Period* (Uppsala: Almqvist & Wiksell).

Jaeger, T. Florian (2010), "Redundancy and reduction: Speakers manage syntactic information density," *Cognitive Psychology*, 61: 23–62.

—— and Harry Tily (2010), "On language 'utility': Processing complexity and communicative efficiency," *Wiley Interdisciplinary Reviews: Cognitive Science*, 2: 323–35.

Jaeggli, Oswaldo (1986), "Passive," *Linguistic Inquiry*, 17: 587–622.

Jäger, Andreas (2006), *Typology of Periphrastic 'Do' Constructions* (Bochum: Brockmeyer).

Jäger, Gerhard (2002), "Some notes on the formal properties of bidirectional optimality theory," *Journal of Logic, Language, and Information*, 11: 427–51.

Jakobson, R. (1968), *Child Language Aphasia and Phonological Universals* (Berlin: Walter de Gruyter).

Johnson, David E. and Shalom Lappin (1997), "A critique of the Minimalist Program," *Linguistics and Philosophy*, 20: 273–333.

—— (1999), *Local Constraints vs. Economy* (Stanford, CA: CSLI).

Johnson, Kyle (1991), "Object positions," *Natural Language and Linguistic Theory*, 9: 577–636.

Juola, Patrick (2008), "Assessing linguistic complexity," in Matti Miestamo, Kaius Sinnemäki, and Fred Karlsson (eds), *Language Complexity: Typology, Contact, Change* (Amsterdam: John Benjamins), 89–108.

Kathol, Andreas (1997), "On the non-existence of true parasitic gaps in Standard German," in Peter W. Culicover and Paul M. Postal (eds), *Parasitic Gaps* (Cambridge, Mass.: MIT Press), 315–38.

Kay, Paul (2002), "An informal sketch of a formal architecture for Construction Grammar," *Grammars*, 5: 1–19.

—— and Charles J. Fillmore (1999), "Grammatical constructions and linguistic generalizations: The what's X doing Y? construction," *Language*, 75: 1–33.

Kayne, Richard S. (1983), "Connectedness," *Linguistic Inquiry*, 24: 223–49.

—— (1984), *Connectedness and Binary Branching* (Dordrecht: Foris Publications).

—— (1994), *The Antisymmetry of Syntax* (Cambridge, Mass.: MIT Press).

Kean, Mary Louise (1980), *The Theory of Markedness in Generative Grammar* (Bloomington, IN: Indiana University Linguistics Club).

Keller, Frank (2000), "Gradience in grammar: Experimental and computational aspects of degrees of grammaticality." Unpublished doctoral dissertation. Edinburgh: University of Edinburgh.

Kemenade, Ans van (1987), *Syntactic Case and Morphological Case in the History of English*. Dordrecht: Foris Publications.

—— (1997), "V2 and embedded topicalization in Old and Middle English," in Ans van Kemenade and Nigel Vincent (eds), *Parameters of Morphosyntactic Change* (Cambridge: Cambridge University Press), 326–52.

King, Ruth and Yves Roberge (1990), "Preposition stranding in Prince Edward Island French," *Probus*, 2: 351–69.

Kiparsky, Paul (1973), "'Elsewhere' in phonology," in Stephen R. Anderson and Paul Kiparsky (eds), *A Festschrift for Morris Halle* (New York: Holt, Rinehart & Winston), 93–106.

—— (1996), "The shift to head-initial VP in Germanic," in Höskuldur Þráinsson, Samuel David Epstein, and Steve Peter (eds), *Studies in Comparative Germanic Syntax II* (Dordrecht: Kluwer Academic Publishers), 140–79.

—— (1997), "The rise of positional licensing," in Ans van Kemenade and Nigel Vincent (eds), *Parameters of Morphosyntactic Change* (Cambridge: Cambridge University Press), 460–94.

Kluender, Robert (1992), "Deriving island constraints from principles of predication," in Helen Goodluck and Michael Rochemont (eds), *Island Constraints: Theory, Acquisition and Processing* (Dordrecht: Kluwer Academic Publishers), 223–58.

—— (1998), "On the distinction between strong and weak islands: A processing perspective," in Peter W. Culicover and Louise McNally (eds), *The Limits of Syntax* (New York: Academic Press), 241–79.

—— (2004), "Are subject islands subject to a processing account?" in Benjamin Schmeiser, Vineeta Chand, Ann Kelleher, and Angelo Rodriguez (eds), *Proceedings of Wccfl 23* (Somerville, MA: Cascadilla Press), 101–25.

Kobele, Gregory M. (2007), "Parsing elliptical structure." Unpublished manuscript. Berlin: Institut für deutsche Sprache und Linguistik, Humboldt Universität zu Berlin. Available online at: <http://home.uchicago.edu/~gkobele/files/2007-Kobele07ParsingEllipsis.pdf>.

Koopman, Hilda (1984), *The Syntax of Verbs* (Dordrecht: Foris Publications).

Kreyer, Rolf (2003), "Genitive and *of*-construction in modern written English: Processability and human involvement," *International Journal of Corpus Linguistics*, 8: 169–207.

Krifka, Manfred (1999), "Additive particles under stress," in A. Lawson and D. Strolovitch (eds), *Proceedings of the Eighth Conference on Semantics and Linguistic Theory*, Cornell Working Papers in Linguistics (Cornell: CLC Publications), 111–28.

—— (2007), "Basic notions of information structure," in Caroline Féry, Gisbert Fanselow, and Manfred Krifka (eds), *Interdisciplinary Studies on Information Structure* (Potsdam: Universitätsverlag Potsdam), 13–56.

Kroch, Anthony S. (1974), *The Semantics of Scope in English.* Unpublished doctoral dissertation (Cambridge, Mass.: MIT).

—— (1989), "Reflexes of grammar in patterns of language change," *Language Variation and Change*, 1: 199–244.

—— (1994), "Morphosyntactic variation," *Proceedings of the Thirtieth Meeting of the Chicago Linguistic Society* (Chicago: Chicago Linguistic Society), 180–201.

—— (2003), "Syntactic change," in Mark Baltin and Chris Collins (eds), *Handbook of Contemporary Syntactic Theory* (Oxford: Blackwell), 699–729.

—— and Ann Taylor (2000), "Verb-object order in Early Middle English," in Susan Pintzuk, George Tsoulas, and Anthony Warner (eds), *Diachronic Syntax: Models and Mechanism* (Oxford: Oxford University Press) 132–63.

Kuno, Susumu (1973). "Constraints on internal clauses and sentential subjects," *Linguistic Inquiry*, 4: 363–85.

Kurtzman, Howard S. and Loren E. Crawford (1991), "Processing parasitic gaps," *Proceedings of the Twenty-First Annual Meeting of the North Eastern Linguistics Society*, 217–31.

Kusters, Wouter (2003), *Linguistic Complexity: The Influence of Social Change on Verbal Inflection* (Utrecht: LOT).

Laenzinger, Christopher (2005), "French adjective ordering: Perspectives on DP-internal movement types," *Lingua*, 115: 645–89.

Lakoff, George (1965), *Irregularity in Syntax* (New York: Holt, Rinehart & Winston).

Lappin, Shalom (1996), "The interpretation of ellipsis," in Shalom Lappin (ed.), *Handbook of Contemporary Semantic Theory* (Oxford: Blackwell Publishers), 145–75.

Lasnik, Howard (1989), "On certain substitutes for negative data," in Robert J. Matthews and William Demopoulos (eds), *Learnability and Linguistic Theory* (Dordrecht: Kluwer Academic Publishers), 89–105.

—— (1999a), "Pseudogapping puzzles," in Shalom Lappin and Elabbas Benmamoun (eds), *Fragments: Studies in Ellipses and Gapping* (Oxford: Oxford University Press), 141–74.

—— (1999b), "Verbal morphology: *Syntactic Structures* meets the Minimalist Program," in Howard Lasnik (ed.), *Minimalist Analysis* (Oxford: Blackwell Publishers), 97–119.

—— (2001), "When can you save a structure by destroying it?" in M. Kim and U. Strauss (eds), *Proceedings of the North Eastern Linguistic Society 31* (Amherst, Mass.: GLSA), 301–20.

Latané, Bibb and Andrzej Nowak (1997), "Self-organizing social systems: Necessary and sufficient conditions for the emergence of consolidation, clustering, and continuing diversity," in George Barnett and Franklin Boster (eds), *Progress in Communication Sciences* (Norwood, NJ: Ablex), 43–74.

Levine, Robert D., Thomas Hukari, and Michael Calcagno (2001), "Parasitic gaps in English: Some overlooked cases and their theoretical implications," in Peter W. Culicover and Paul Postal (eds), *Parasitic Gaps* (Cambridge, Mass.: MIT Press), 181–222.

—— and Ivan A. Sag (2003), "Some empirical issues in the grammar of extraction," in Stefan Mueller (ed.), *Proceedings of the HPSG03 Conference, Michigan State University, East Lansing* (Stanford, CA: CSLI Publications).

Levy, Roger (2005), "Probabilistic models of word order and syntactic discontinuity." Unpublished doctoral dissertation. Stanford University.

—— Evelina Fedorenko, Michael Breen, and Edward Gibson (2012), "The processing of extraposed structures in English," *Cognition*, 122: 12–36.

—— and Frank Keller (2011), "Expectation and locality effects in German verb-final structures." Unpublished manuscript. UC San Diego and Edinburgh University.

Lewis, Richard L. (1998), "Leaping off the garden path: Reanalysis and limited repair parsing," in Janet D. Fodor and Fernanda Ferreira (eds), *Reanalysis in Sentence Processing* (Boston: Kluwer Academic Publishers), 247–85.

—— (2000), "Specifying architectures for language processing: Process, control, and memory in parsing and interpretation," in Matthew Crocker, Martin Pickering, and Charles Clifton (eds), *Architectures and Mechanisms for Language Processing* (Cambridge: Cambridge University Press), 56–89.

—— Shravan Vasishth, and Julie A. Van Dyke (2006), "Computational principles of working memory in sentence comprehension," *Trends in Cognitive Sciences*, 10: 447–54.

Lidz, Jeffrey and Sandra Waxman (2004), "Reaffirming the poverty of the stimulus argument: A reply to the replies," *Cognition*, 93: 157–65.

Lötscher, Andreas (1978), "Zur Verbstellung im Zürichdeutschen und in anderen Varianten des Deutschen," *Zeitschrift für Dialektologie und Linguistik*, 45: 1–29.

Louden, Mark L. (1990), "Verb raising and the position of the finite verb in Pennsylvania German," *Linguistic Inquiry*, 21: 470–77.

MacWhinney, Brian (1995), *The CHILDES Project : Tools for Analyzing Talk* (Hillsdale, NJ: Lawrence Erlbaum).

Mailhammer, Robert (2006), 'On the origin of the Germanic strong verb system', *Sprachwissenschaft*, 31: 1-52.

—— (2007a), *The Germanic Strong Verbs: Foundations and Development of a New System* (Berlin: Mouton de Gruyter).

—— (2007b), "Islands of resilience: The history of the German strong verbs from a systemic point of view," *Morphology*, 17: 77–108.

McCawley, James D. (1970), "English as a VSO language," *Language*, 46: 286–99.

McWhorter, John H. (2001), "The world's simplest grammars are creole grammars," *Linguistic Typology*, 5: 125–66.

McWhorter, John H. (2004), "What happened to English?," in Werner Abraham (ed.), *Focus on Germanic Typology* (Berlin: Akademie Verlag), 19–60.

—— (2005), "Review of 'Issues in the study of pidgin and creole languages'," *Journal of Pidgin and Creole Languages*, 20: 211–18.

—— (2007), *Language Interrupted: Signs of Non-Native Acquisition in Standard Language Grammars* (New York: Oxford University Press).

Merchant, Jason (2001), *The Syntax of Silence* (Oxford: Oxford University Press).

—— (2003), "Subject-auxiliary inversion in comparatives and PF output constraints," in Kerstin Schwabe and Susanne Winker (eds), *The Interfaces: Deriving and Interpreting Omitted Structures* (Amsterdam: John Benjamins), 55–77.

—— (2008), "Variable island repair under ellipsis," in Kyle Johnson (ed.), *Topics in Ellipsis* (Cambridge: Cambridge University Press), 132–53.

Michaelis, Laura A. (2006), "Construction grammar," in Keith Brown (ed.), *The Encyclopedia of Language and Linguistics*, vol. 3, second edition (Oxford: Elsevier), 73–84.

Miestamo, Matti (2006), "On the feasibility of complexity metrics," in Krista Kerge and Maria-Maren Sepper (eds), *Finest Linguistics: Proceedings of the Annual Finnish and Estonian Conference of Linguistics. Tallinn, May 6–7, 2004* (Tallinn: Tallinn University Press), 11–26.

—— (2008), "Grammatical complexity in cross-linguistic perspective," in Matti Miestamo, Kaius Sinnemäki, and Fred Karlsson (eds), *Language Complexity: Typology, Contact, Change* (Amsterdam: John Benjamins), 23–41.

—— (2009), "Implicational hierarchies and grammatical complexity," in Geoffrey Sampson, David Gil, and Peter Trudgill (eds), *Language Complexity as an Evolving Variable* (Oxford: Oxford University Press), 80–97.

—— Kaius Sinnemäki, and Fred Karlsson (eds) (2006), *Language Complexity: Typology, Contact, Change* (Amsterdam: John Benjamins).

Miller, George A. and Noam Chomsky (1963), "Finitary models of language users," in R. D. Luce, R. R. Bush, and E. Galanter (eds), *Handbook of Mathematical Psychology*, vol. 2 (New York: Wiley), 419–91.

Müller, Stefan (2005), "Zur Analyse der scheinbar mehrfachen Vorfeldbesetzung," *Linguistische Berichte*, 203: 29–62.

—— (2006), "Phrasal or lexical constructions?" *Language*, 82: 850–83.

Munn, Alan (2001), "Explaining parasitic gap restrictions," in Peter W. Culicover and Paul Postal (eds), *Parasitic Gaps* (Cambridge, Mass.: MIT Press), 369–92.

Nevalainen, Terttu (1997), "Recycling inversion: The case of initial adverbs and negators in early modern English," *Studia Anglica Posnaniensia* 31: 203–14.

—— (2006), *An Introduction to Early Modern English.* Oxford: Oxford University Press.

Newmeyer, Frederick J. (2000), *Language Form and Language Function* (Cambridge, Mass.: MIT Press).

—— (2005), *Possible and Probable Languages: A Generative Perspective on Linguistic Typology* (Oxford: Oxford University Press).

—— (2007), " 'More complicated and hence, rarer': A look at grammatical complexity and cross-linguistic rarity," in Simin Karimi, Vida Samiian, and Wendy K. Wilkins (eds), *Phrasal and Clausal Architecture: Syntactic Derivation and Interpretation. In Honor of Joseph E. Emonds* (Amsterdam: John Benjamins), 221–42.

Nübling, Damaris, Antje Dammel, Janet Duke, and Renata Szczepaniak (2006), *Historische Sprachwissenschaft Des Deutschen: Eine Einführung in Die Prinzipien Des Sprachwandels* (Tübingen: Gunter Narr Verlag).

Nunberg, Geoffrey, Ivan A. Sag, and Thomas Wasow (1994), "Idioms," *Language*, 70: 491–538.

Nurmi, Arja (1999), *A Social History of Periphrastic Do* (Helsinki: Société Néophilologique de Helsinki).

Ostade, Ingrid Tieken-Boon van (1987), *The Auxiliary Do in Eighteenth-Century English: A Sociohistorical-Linguistic Approach* (Dordrecht: Foris Publications).

Ott, Daniel (2009), "Multiple NP split: A distributed deletion analysis," *Groninger Arbeiten zur germanistischen Linguistik*. Available online at: <http://gagl.eldoc.ub.rug.nl/FILES/root/2009-48/2009-48-03/2009-48-03.pdf>.

Partee, Barbara (1992), "Syntactic categories and semantic type," in Michael Rosner and Roderick Johnson (eds), *Computational Linguistics and Formal Semantics* (Cambridge: Cambridge University Press), 97–125.

Patocka, Franz (1997), *Satzgliedstellung in Den Bairischen Dialekten Österreichs* (Frankfurt am Main: Peter Lang).

Perlmutter, David M. and Postal, Paul M. (1983). "Toward a universal characterization of passivization", in David M. Perlmutter (ed), *Studies in Relational Grammar* (Chicago: University of Chicago Press), 1–29.

Pesetsky, David (1989), "The Earliness Principle." Unpublished manuscript. Cambridge, Mass.: MIT.

—— and Esther Torrego (2000), "T-to-C movement: Causes and consequences," in Michael Kenstowicz (ed.), *Ken Hale, a Life in Language* (Cambridge, Mass.: MIT Press), 355–426.

Petrova, Svetlana (2006), "A discourse-based approach to verb placement in early West-Germanic," in Shinichiro Ishihara, Michaela Schnmitz, and Anne Schwarz (eds), *Interdisciplinary Studies on Information Structure*, 5 (Potsdam: Universitätsverlag), 153–85.

Phillips, Colin (2006), "The real-time status of island phenomena," *Language*, 82: 795–823.

—— (2011), "Some arguments and non-arguments for reductionist accounts of syntactic phenomena," *Language and Cognitive Processes*, 26: 1–32.

Pickering, Martin and Guy Barry (1991), "Sentence processing without empty categories," *Language and Cognitive Processes*, 6: 229–59.

Pintzuk, Susan (1991), *Variation and Change in Old English Word Order*. Unpublished doctoral dissertation. Philadelphia: University of Pennsylvania.

—— and Anthony Kroch (1989), "The rightward movement of complements and adjuncts in the Old English of Beowulf," *Language Variation and Change*, 1: 115–43.

—— and Ann Taylor (2006), "Loss of OV order," in Ans van Kamenade and Bettelou Los (eds), *The Handbook of the History of English* (Malden, Mass.: Blackwell Publishing), 249–78.

Pollard, Carl and Ivan A. Sag (1994), *Head-Driven Phrase Structure Grammar* (Chicago: University of Chicago Press).

Pollock, Jean-Yves (1989), "Verb movement, Universal Grammar and the structure of IP," *Linguistic Inquiry*, 20: 365–424.

Postal, Paul M. (1969), "Anaphoric islands," in Robert I. Binnick, Alice Davison, Georgia Green, and Jerry L. Morgan (eds), *Papers From the Fifth Regional Meeting of the Chicago Linguistic Society* (Chicago: University of Chicago), 205–35.

—— (1993), "Parasitic gaps and the across-the-board phenomenon," *Linguistic Inquiry*, 24: 735–54.

—— (1998), *Three Investigations of Extraction* (Cambridge, Mass.: MIT Press).

Progovac, Ljiljana (2009), "Layering of grammar: Vestiges of protosyntax in modern-day languages," in Geoffrey Sampson, David Gil, and Peter Trudgill (eds), *Language Complexity as an Evolving Variable* (Oxford: Oxford University Press), 203–12.

Pullum, Geoffrey K. and Barbara C. Scholz (2002), "Empirical assessment of stimulus poverty arguments," *The Linguistic Review*, 18: 9–50.

Pustejovsky, James (1995), *The Generative Lexicon* (Cambridge, Mass.: MIT Press).

—— and Boris Boguraev (1993), "Lexical knowledge representation and natural language processing," *Artificial Intelligence*, 63: 193–223.

Regier, Terry and Susanne Gahl (2004), "Learning the unlearnable: The role of missing evidence," *Cognition*, 93: 147–55.

Reich, Ingo (2007), "From phases to 'Across the board movement'," in Malcolm Elliot, James Kirby, Osamu Sawada, Eleni Staraki, and Suwon Yoon (eds), *Proceedings of the 43rd Annual Meeting of the Chicago Linguistics Society* (Chicago: Chicago Linguistic Society), 217–32.

Reinhart, Tanya and Eric Reuland (1991), "Anaphors and logophors: An argument structure perspective," in Jan Koster and Eric Reuland (eds), *Long Distance Anaphora* (Cambridge: Cambridge University Press), 283–321.

—— (1993), "Reflexivity," *Linguistic Inquiry*, 24: 657–720.

Riemsdijk, Henk van (1978), *A Case Study in Syntactic Markedness* (Dordrecht: Foris Publications).

Rizzi, Luigi (1996), "Residual verb second and the wh-criterion," in Adriana Belletti and Luigi Rizzi (eds), *Parameters and Functional Heads* (Oxford: Oxford University Press), 63–90.

—— (1997), "The fine structure of the left periphery," in Liliane Haegeman (ed.), *Handbook of Generative Syntax* (Dordrecht: Kluwer Academic Publishers), 281–338.

Roberts, Craige (forthcoming), "Topics," in Claudia Maienborn, Klaus von Heusinger, and Paul Portner (eds), *Semantics: An International Handbook of Natural Language Meaning* (Berlin: Mouton de Gruyter).

Rochemont, Michael S. (1978), "A theory of stylistic rules in English." Unpublished doctoral dissertation. Amherst, Mass.: University of Massachusetts.

—— (1986), *Focus in Generative Grammar* (Amsterdam: John Benjamins).

—— and Peter W. Culicover (1990), *English Focus Constructions and the Theory of Grammar* (Cambridge: Cambridge University Press).

Rochemont, Michael S. and Peter W. Culicover (1997), "Deriving dependent right-adjuncts in English," in Dorothee Beermann, David LeBlanc, and Henk van Riemsdijk (eds), *Rightward Movement* (Amsterdam: John Benjamins), 279–300.

Rooth, Mats E. (1992), "A theory of focus interpretation," *Natural Language Semantics*, 1: 75–116.

Rosenbaum, Peter (1967), *The Grammar of English Predicate Complement Constructions* (Cambridge, Mass.: MIT Press).

Ross, John R. (1967), "Constraints on variables in syntax." Unpublished doctoral dissertation. Cambridge, Mass.: MIT.

—— (1969), "Guess who," in Robert I. Binnick, Alice Davison, Georgia M. Green, and Jerry L. Morgan (eds), *Proceedings of the Fifth Annual Meeting of CLS* (Chicago: Chicago Linguistics Society), 252–86.

Rothstein, Susan D. (1983), "The syntactic forms of predication." Unpublished doctoral dissertation. Cambridge, Mass.: MIT.

Sabel, Joachim (2000), "Partial wh-movement and the typology of wh-questions," in Uli Lutz, Gereon Müller, and Arnim von Stechow (eds), *Wh-Scope Marking* (Amsterdam: John Benjamins), 409–46.

—— (2006), "Impossible infinitival interrogatives and relatives," in Patrick Brandt and Eric Fuss (eds), *Form, Function, and Grammar: A Festschrift Presented to Guenther Grewendorf on Occasion of His 60th Birthday* (Berlin: Akademie Verlag), 243–54.

Sag, Ivan A. (1983), "On parasitic gaps," *Linguistics and Philosophy*, 6: 35–45.

—— (1997), "English relative clause constructions," *Journal of Linguistics*, 33: 431–84.

—— (forthcoming), "Sign-based construction grammar: An informal synopsis," in Hans C. Boas and Ivan A. Sag (eds), *Sign-Based Construction Grammar* (Stanford, CA: CSLI), 39–170.

—— Philip Hofmeister, Inbal Arnon, Neal Snider, and Florian Jaeger (2008), "Processing accounts for superiority effects." Unpublished manuscript. Stanford, CA: Stanford University.

—— and Carl Pollard (1991), "An integrated theory of complement control," *Language*, 67: 63–113.

Sampson, Geoffrey, David Gil, and Peter Trudgill (eds) (2009), *Language Complexity as an Evolving Variable* (Oxford: Oxford University Press).

Sapp, Christopher D. (2005), "Factors Favoring Aux-V Order in 15th-Century German Subordinate Clauses," in Dorian Roehrs, Ock-Hwan Kim, and Yoshihisa Kitagawa (eds), *Syntax and Beyond*, Indiana University Working Papers in Linguistics, 5 (Bloomington, IN: Indiana University Linguistics Club), 149–69.

—— (2006), "Verb order in subordinate clauses from Early New High German to Modern German." Unpublished doctoral dissertation. Bloomington, IN: Indiana University.

—— (2007), "Focus and verb order in Early New High German: Historical and contemporary evidence," in Sam Featherston and Wolfgang Sternefeld (eds), *Roots: Linguistics in Search of its Evidential Base* (Berlin: Walter de Gruyter), 299–318.

Sapp, Christopher D. (2011), *Verb Order in Subordinate Clauses from Early New High German to Modern German* (Amsterdam: John Benjamins).

Schallert, Oliver (2007), "The word order in the right periphery: OV-, VO-, and 'mixed' OV/VO-orders in Old High German," Workshop on Information Structure and Word Order Variation in Older Germanic, Berlin. Available online at: <http://www.uetzgenfatz.priv.at/Dokumente/Talks%20Webseite/Vortrag%20IS-Workshop%20Berlin.pdf>.

Schmid, Tanja (2005), *Infinitival Syntax. Infinitivus Pro Participio as a Repair Strategy* (Amsterdam and Philadelphia: John Benjamins).

—— and Ralf Vogel (2004), "Dialectal variation in German 3-verb clusters: A surface-oriented optimality theoretic account," *Journal of Comparative Germanic Linguistics*, 7: 235–74.

Schwarzschild, Roger (1999), "GIVENness, AVOIDF and other constraints on the placement of accent," *Natural Language Semantics*, 7: 141–77.

Seiler, Guido (2004), "On three types of dialect variation and their implications for linguistic theory: Evidence from verb clusters in Swiss German dialects," in Bernd Kortmann (ed.), *Dialectology Meets Typology: Dialect Grammar From a Cross-Linguistic Perspective* (Berlin and New York: Mouton de Gruyter), 367–99.

Shosted, R. K. (2006), "Correlating complexity: A typological approach," *Linguistic Typology*, 10: 1–40.

Sims, Andrea D. (2009), "Why paradigmatic gaps are, and aren't, the result of competing morphological patterns," in Malcolm Elliot, James Kirby, Osamu Sawada, and Suwon Yoon (eds), *Proceedings From the Annual Meeting of the Chicago Linguistic Society*, vol. 43 (Chicago: Chicago Linguistics Society), 267–81.

Stabler, Edward P (2004), "Varieties of crossing dependencies: Structure dependence and mild context sensitivity," *Cognitive Science*, 28: 699–720.

Steedman, Mark (2000), *The Syntactic Process* (Cambridge, Mass.: MIT Press).

Sternefeld, Wolfgang (2001), "Semantic vs. syntactic reconstruction," in Hans Kamp, A. Rossdeutscher, and Christian Rohrer (eds), *Linguistic Form and Its Computation* (Stanford: CSLI Publications), 145–82.

—— (2006), *Syntax* (Tübingen: Stauffenburg Verlag).

Tagliamonte, Sali A., Jennfier Smith, and Helen Lawrence (2005), "No taming the vernacular! Insights from the relatives in northern Britain," *Language Variation and Change*, 17: 75–112.

Taylor, Ann (2005), "Prosodic evidence for incipient VO order in Old English," *English Language and Linguistics*, 9: 139–56.

Tenenbaum, Joshua B., Thomas L. Griffiths, and Charles Kemp (2006). "Theory-based Bayesian models of inductive learning and reasoning," *Trends in Cognitive Sciences*, 10: 309–18.

Tesar, Bruce and Paul Smolensky (1998), "Learnability in optimality theory," *Linguistic Inquiry*, 29: 229–68.

Tomasello, Michael J. (2000), "Do young children have adult syntax," *Cognition*, 74: 209–53.

—— (2003), *Constructing a Language* (Cambridge, Mass.: Harvard University Press).

Trubetskoy, Nicholai S. (1939), *Grundzüge Der Phonologie* (Prague: Travaux du Cercle Linguistique de Prague).

Truckenbrodt, Hubert (2006), "On the semantic motivation of syntactic verb movement to C in German," *Theoretical Linguistics*, 32: 257–306.

Trudgill, Peter (1989), "Contact and isolation in linguistic change," in Leiv Earl Breivik and Ernst Hakon Jahr (eds), *Language Change: Contributions to the Study of Its Causes* (Berlin: Mouton de Gruyter), 227–37.

—— (1992), "Dialect typology and social structure," in Ernst Hakon Jahr (ed.), *Language Contact: Theoretical and Empirical Studies* (Berlin: Mouton de Gruyter), 195–211.

—— (1995), "Grammaticalization and social structure: Non-standard conjunction formation in East Anglian English," in Frank Palmer (ed.), *Grammar and meaning: Essays in honor of Sir John Lyons* (Cambridge: Cambridge University Press), 136–47.

—— (2004a), "Linguistic and social typology," in J. K. Chambers, Peter Trudgill, and Natalie Schilling-Estes (eds), *The Handbook of Language Variation and Change* (Oxford: Blackwell Publishing), 707–28.

—— (2004b), "On the complexity of simplification," *Linguistic Typology*, 8: 384–8.

—— (2009), "Sociolinguistic Typology and Complexification," in Geoffrey Sampson, David Gil, and Peter Trudgill (eds), *Language Complexity as an Evolving Variable* (Oxford: Oxford University Press), 98–109.

—— (2011), *Sociolinguistic Typology: Social Determinants of Linguistic Complexity* (Oxford: Oxford University Press).

—— Terttu Nevalainen, and Ilse Wischer (2002), "Dynamic *have* in North American and British Isles English," *English Language and Linguistics*, 6: 1–15.

Vasishth, Shravan (2011), "Integration and prediction in head-final structures," in Hiroko Yamashita, Yuki Hirose, and Jerry Packard (eds), *Processing and Producing Head-Final Structures* (Dordrecht: Springer Verlag), 349–67.

—— and Richard L. Lewis (2006), "Argument-head distance and processing complexity: Explaining both locality and antilocality effects," *Language*, 82: 767–94.

Vat, Jan (1978), "On footnote 2: Evidence for the pronominal status of *þær* in Old English relatives," *Linguistic Inquiry*, 9/4: 695–716.

Visser, Fredericus Theodorus (1973), *An Historical Syntax of the English Language* (Leiden: E. J. Brill).

Ward, Gregory, Betty J. Birner, and Rodney Huddleston (2002), "Information packaging," in Rodney Huddleston and Geoffrey K. Pullum (eds), *The Cambridge Grammar of the English Language* (Cambridge: Cambridge University Press), 1363–447.

Warner, Anthony R. (1993), *English Auxiliaries* (Cambridge: Cambridge University Press).

—— (2007), "Parameters of variation between verb-subject and subject-verb order in late Middle English," *English Language and Linguistics*, 11: 81–111.

Wasow, Thomas (1997), "Remarks on grammatical weight," *Language Variation and Change*, 9: 81–106.

—— (2002), *Postverbal Behavior* (Stanford: CSLI Publications).

Wexler, Kenneth (1978), "A review of John R Anderson's *Language, Memory and Thought*," *Cognition*, 6, 327–51.

—— (1981), "Some issues in the theory of learnability," in C. L. Baker and J. McCarthy (eds), *The Logical Problem of Language Acquisition* (Cambridge, Mass.: MIT Press), 30–52.

—— and Peter W. Culicover (1980), *Formal Principles of Language Acquisition* (Cambridge, Mass.: MIT Press).

Wichmann, Søren and Jan Wohlgemuth (2008), "Loan verbs in a typological perspective," in Thomas Stolz, Dik Bakke, and Rosa Palomo (eds), *Aspects of Language Contact: New Theoretical, Methodological and Empirical Findings with Special Focus on Romancisation Processes* (Berlin and New York: Mouton de Gruyter), 89–121.

Williams, Edwin (1980), "Predication," *Linguistic Inquiry*, 1: 203–38.

Winkler, Susanne (2005), *Ellipsis and Focus in Generative Grammar* (Berlin and New York: Mouton de Gruyter).

Wulff, Stephanie (2008), *Rethinking Idiomaticity: A Usage-Based Approach* (London and NewYork: Continuum).

—— (2009), "Converging evidence from corpus and experimental data to capture idiomaticity," *Corpus Linguistics and Linguistic Theory*, 5: 131–59.

Wurmbrand, Susi (2004), "West Germanic verb clusters: The empirical domain," in Katalin E. Kiss and Henk van Riemsdijk (eds), *Verb Clusters: A Study of Hungarian, German, and Dutch* (Amsterdam: John Benjamins), 43–85.

—— (2005), "Verb clusters, verb raising, and restructuring," in Martin Everaert and Henk van Riemsdijk (eds), *The Syntax Companion* (Oxford: Blackwell), 227–341.

Zanuttini, Raffaela (1991), "Syntactic properties of sentential negation: A comparative study of Romance languages." Unpublished doctoral dissertation. Philadelphia, PA: University of Pennsylvania.

Zwart, C. Jan-Wouter (1995), "A note on verb clusters in the Stellingwerf dialect," in Marcel den Dikken and Kees Hengeveld (eds), *Linguistics in the Netherlands 1995* (Amsterdam and Philadelphia: John Benjamins), 215–26.

—— (1996), "Verb clusters in Continental West Germanic dialects," in James R. Black and Virginia Motapanyane (eds), *Microparametric Syntax and Dialect Variation* (Amsterdam and Philadelphia: John Benjamins), 229–58.

—— (2007), "Some notes on the origin and distribution of the IPP-effect," *Groninger Arbeiten zur Germanistischen Linguistik*, 45: 77–99.

Index

A-over-A condition 8, 15
Abeillé, A. 115
Abels, K. 184
Abney, S. 216
absolute complexity 12, 22
Ackema, P. 220
across-the-board (ATB) extraction 154–7, 162
Affix Hopping (AH) 212–19, 234, 239
Afrikaans 243, 258, 278–280
Alexopoulou, T. 146
Allen, C. 183
Allwood, J. 145, 186
Alsatian 256
Ambridge, B. 146
Ariel, M. 146
attainability 15
Austrian 256–8, 280
Auwera, J. van der 210, 236n

Bach, K. 127
Bader, M. 254–5, 280
Baekken, B. 245–8, 250
Baker, M. 27
Barry, G. 146
bare argument ellipsis (BAE) 112–13, 132–5
Battistella, E. 14n, 201
Bavarian 256
Bayer, J. 79n, 184, 262, 267–8
Beck, S. 118
Beecher, H. 113n
Benincà, P. 234–5
Bennis, H. 178
Bever, T. 23
bias 16, 201–6, 268–81, 284
Biberauer, T. 278–9
Bies, A. 101
Birner, B. 88
Black, J. 108
Bloom, P. 188
Bobaljik, J. 212, 216n, 219
Boguraev, B. 49
Boland, J. 36
Booij, G. 136
Borer, H. 88, 216
Borsley, R. 89, 115
Bouma, G. 253–4, 263
Brighton, H. 15
Brown, R. 17
Burn, J. 223n
Bybee, J. 143, 243

CaMiLLe 189–94, 282
case-marking 11, 29, 112, 209, 211, 243, 249
category, syntactic 55–6, 70, 88, 92, 141, 190, 231, 251, 256, 267–70
chain 17, 19, 64, 69, 71, 75–7, 111, 139, 142, 146–59, 161, 163–7, 169–71, 181–7, 207
Chater, N. 140n
Chen, R. 88
Chomsky, N. 3, 6–9, 12, 14–18, 62–4, 66, 73, 78–9, 87, 90, 108, 140, 141, 145–6, 154, 169, 182, 184–6, 194–5, 212, 215, 217, 284
Chung, S. 132
Cinque, G. 19, 220
Clifton, C. 146
collocation 205–6, 253
comparative inversion (CI) 87, 89–105
complex NP constraint (CNPC) 142, 144–5, 147, 161, 166, 173, 186
conceptual structure (CS) 28–33, 36, 70, 73, 136, 141, 151, 252, 262
concessives 121–6
Connectivity construction 75

constituent order 10, 11, 33, 53, 90, 98, 107, 220, 227, 230, 249
construction *see* individual constructions
constructional lexicon 37, 41, 107, 122, 241
contact 177, 179–80, 194–6, 201, 204, 207–8, 211, 252, 272–3, 281
core-periphery distinction 7–9, 16, 22, 26, 66, 108, 136
Cornips, L. 277–8
Cover, T. 15
Craige, C. 184
Crawford, L. 147n, 157

Dąbrowska, E. 146
Dahl, Ö. 12
Dali, S. 4–8
Danish 71, 109, 184, 243
dative 31–2
Dayal, V. 123
De Vos, M. 280
Dehé, N. 45n, 124
Del Giudice, A. 40n, 89, 90n
Demberg, V. 24, 25
Den Besten, H. 216
Den Dikken, M. 115, 116n
Denison, D. 243, 244
dependency bias 272–3, 276, 280
discharge 27, 87, 216–19
do-periphrasis 210–11, 234–6, 243–4, 242, 250
do-support 128–31, 209–19, 223, 225–6, 231–44, 250–1
Dowty, D. 43n, 223
Dryer, M. 11n
Dubenion-Smith, S. 257, 270, 280–1
Dutch 184, 220, 243, 255, 258, 277, 280

economy 10, 12, 16–18, 283
Edmondson, J. 216
Ellegård, A. 237, 239–42, 251, 278
ellipsis 32, 35, 36, 91, 92, 98, 100, 112, 113, 132, 133, 135, 166–72, 210, 214, 222, 232, 244; *see also* bare argument ellipsis, VP ellipsis
Ellipsis construction 222

Elsig, M. 27n
Emonds, J. 44, 47
Emph construction 228, 230
Enfield, N. 195
Engdahl, E. 186
epidemiology 19–196, 238, 252, 273–275
Erteschik-Shir, N. 24
Everett, D. 14
Evers, A. 259
extended projection principle (EPP) 90, 99–100, 106
extraposition 79, 151, 220, 255, 261, 262

Fanselow, G. 173
feasibility 15
Featherston, S. 143, 147, 173
Feldman, A. 193
Fillmore, C. 26, 29, 283
Fischer, O. 212
focus 30, 34, 35, 61, 63, 70, 75, 86, 88–91, 98–108, 132–5, 178, 225–7, 248–9, 266, 280–1
focus inversion 63, 86, 90, 100–7
Fodor, J. A. 17
Fodor, J. D. 143, 146, 158
formal complexity 3, 12–13, 16, 20, 22, 28, 109, 177, 186, 202, 210, 282
Fox, D. 169, 171
Francis, E. 159, 163
Frank, R. 25n
Frazier, L. 146
French 55, 184, 212, 216
Freidin, R. 212
Frisch, S. 173
Frisian 243, 258, 280
functionalism 12

Gahl, S. 140n
gapping 232
Gbadi 184
Gergel, R. 101
German 27, 31 5, 40, 71, 100, 109, 155, 156, 186, 204–12, 216, 220, 234, 236, 244, 246, 253, 255–9, 265, 270, 278, 280
Gibson, E. 20, 24, 146, 147, 272

Gil, D. 14
Givón, T. 19
Goldberg, A. 31, 40n, 89, 90n, 146
Grodner, D. 24, 272
Grünwald, P. 15

Haegeman, L. 255, 259, 278
Haider, H. 156, 261, 262, 270, 271, 280
Hale, J. 24
Halle, M. 14, 62
Hamburger, H. 192
Hanlon, C. 17
harmonic branching 254, 265–6
Harnish, R. 127
Harris, M. 79
Haspelmath, M. 123
Hawkins, J. 20–3, 43n, 142–3, 181–2, 184, 210, 211, 236, 265, 272, 283
heavy NP shift (HNPS) 43, 94–6
Hindi 268n
Hinterhölzl, R. 101, 249
Hirtle, W. 211
Hoffmann, T. 62n, 63n, 73n
Hofmeister, P. 77n, 142, 144, 147, 165
Huddleston, R. 210
Hunter, P. 22n

idioms 9, 27, 30, 32, 37–9, 43–6, 52, 109, 123, 136, 183, 184, 190, 194
idiosyncrasy 3, 7, 10, 16, 26, 37–8, 47, 52, 63, 67, 83, 86, 107, 109, 118, 123, 177, 179, 180–2, 190, 282
imperative 126–31, 136, 192–3
information structure 19, 30, 37, 43, 101, 228, 249
inheritance 37, 62, 63, 66, 68–70, 83, 114, 126, 130, 131
Inoue, A. 146
inversion *see* focus inversion, negative inversion, subject aux inversion (SAI)
IPP 254–5, 258–61, 278–80
island constraints 8, 18, 25, 55, 73, 77, 139, 140, 142, 143, 144 , 147, 153–7, 161, 162, 166, 168–9, 186, 260, 264, 283, 285; *see also* complex NP constraint, superiority condition, *that-t* effect, *wh*-island condition
Italian 55, 234

Jaceltec 184
Jackendoff, R. 28–9, 32–4, 36, 37n, 39, 42n, 43–4, 53, 62, 64n, 72, 75, 81n, 83, 98, 112, 11516, 120, 132, 135, 164, 167, 183, 215n, 282–3
Jackson, M. 239n
Jacobson, P. 75
Jacobsson, B. 211n
Jaeger, F. 24, 25, 80, 144, 146, 236
Jaeggli, O. 27
Jäger, A. 210, 211, 235–6, 247n
Jäger, G. 25n
Jakobson, R. 201
Johnson, D. 17
Johnson, K. 88n
Juola, P. 15

Kathol, A. 156
Kay, P. 29, 283
Kayne, R. 22, 154–5, 216, 229, 264
Kean, M.-L. 201
Keller, F. 24, 25, 143, 146
Kemenade, A. van 180, 212, 220
King, R. 184
Kiparsky, P. 99, 211, 212, 220, 243, 246
Klima, E. 128
Kluender, R. 24, 25, 142, 144, 146, 147n, 148, 154, 170, 186, 206
Kobele, G. 146
Kolmogorov complexity 15
Koopman, H. 184
Kreyer, R. 249n
Krifka, M. 102n, 103
Kroch, A. 101, 146, 212, 220–1, 233–4, 237, 238, 242, 247, 271, 278
Kuno, S. 39
Kurtzman, H. 147n, 157
Kusters, W. 20

Laenzinger, C. 55
Lakoff, G. 47
lambda reduction 73, 75–6, 78–9, 81–2, 114
language acquisition 10–11, 15, 17, 18, 25, 139, 140, 141, 177, 187–8, 207, 233, 284
language change 10–11, 55, 60, 178, 186–7, 210, 221, 283
language evolution 15
Lappin, S. 17, 168
Lasnik, H. 79, 84, 98, 108, 127, 128, 140n, 166, 168–9, 171–2, 212
Latané, B. 284
learner 7, 10, 13, 14, 15 and, 21, 28, 67, 140, 177–8, 186–90, 206, 220, 247, 248, 251, 253, 274, 282, 284
left dislocation 262
Levine, R. 62n, 88, 155, 159, 161, 164
Levy, R. 24, 24, 144, 146
Lewis, R. 20, 24, 25, 146, 150
lexicon 29, 33, 37–41, 54, 107, 114, 122, 195, 235, 241, 273, 270, 276
Lidz, J. 140n
Lötscher, A. 260, 269, 278
Louden, M. 260–1

Mailhammer, R. 205n
markedness 7–9, 10, 12, 14–16, 18, 22, 63, 185, 201, 283, 284
McCawley, J. 33
McNally, L. 4
McWhorter, J. 12, 13, 14, 19, 178, 211n
memory 10, 11, 13, 17, 20–3, 25, 150, 152, 171, 186, 268, 271–2
Merchant, J. 90–1, 109, 166–7
Merge 26, 218
Meurers, D. 62n
Michaelis, L. 29
Miestamo, M. 12, 14, 23
Miller, G. 3, 146
Minimalist Program (MP) 6, 18, 19, 30, 217
modal, as category 243–5
Monnese 234–5
Motapanyane, V. 108
Müller, S. 29n, 246n
Munn, A. 156, 157

N MODIFICATION construction 73–4, 79–84
NEGATION construction 130, 222, 229, 231–2, 246
negative inversion 40, 89, 247–8
Nevalainen, T. 211n
Newmeyer, F. 11n, 108n, 182
Noord, G. van 253, 254, 263
Norwegian 71, 184, 243
Nowak, A. 22n, 143, 144, 188, 238n, 252, 284
Nübling, D. 205
Nunberg, G. 52n

OBJECT construction 30–1, 222, 229, 231, 246
Old English (OE) 180–2, 184–6, 211–12, 220, 226, 233 235, 243–6
optimality theory (OT) 25, 264, 292
Ostade, I. van 240
Ott, D. 45n
OV to VO order 211, 248

parasitic gaps 94–6, 153–66
Partee, B. 84n
Patel-Grosz, P. 268n
Patocka, F. 257
Pennsylvania German 260–1
Perlmutter, D. 34
Pesetsky, D. 87n, 215
Petrova, S. 101
Phillips, C. 143, 158, 165
Pickering, M. 146
pied-piping 74, 77–8, 111, 113, 182, 184
Pintzuk, S. 101, 211, 220
Poletto, C. 234–5
Pollard, C. 81n
Pollock, J.-Y. 212, 216
Poplack, S. 27n
Postal, P. 19, 34, 156

PREDICATE construction 218, 221, 228, 231, 246
p(reposition)-stranding 178, 180–6
Prideaux, G. 22n
processing complexity 3–4, 12, 16–17, 20, 23–5, 142–4, 147, 153, 169–70, 173, 177–8, 186, 202, 208, 210, 282–4
Progovac, L. 19
pseudogapping 94, 98–100, 214, 232, 244
Pullum, G. 141n, 210
Pustejovsky, J. 49

raising to subject 47–8
Regier, T. 140n
Reich, I. 146
Reinhart, T. 141
relative clauses *see* relatives, RELATIVE construction
relative complexity 12, 17, 22, 55
RELATIVE construction 72, 73, 84, 185
relatives 61–74, 76, 79–80, 82–6, 107, 156, 178–81, 183–5
  finite 76ff
  infinitival 63–8, 70, 82–3, 107, 179–83
  reduced 61, 65, 71, 76
  zero 64–9, 71, 74, 79–80, 82, 86, 151, 180–1, 184–5, 208
Reuland, E. 141
Riemsdijk, H. van 184–5, 255, 259
Rizzi, L. 211n, 216, 227
Roberge, Y. 184
Roberts, C. 228n
Rochemont, M. 37n, 88, 93, 220, 260, 263n
Rogers, B. 239n
Rooth, M. 99
Rosenbaum, P. 47
Rosengren, I. 156
Ross, J. 73n, 77n, 139, 142, 166n
Rothstein, S. 90n
Russian 186, 206–7

Sabel, J. 71, 82, 206
Sag, I. 27n, 29, 61, 62, 64, 66–9, 71, 72n, 74, 81n, 85, 142, 147, 155–6
Salzmann, M. 258n, 259, 273
Sampson, G. 14, 19n, 279
Sapp, C. 101, 256, 257, 270, 280–1
Satta, S. 25n
Schallert, O. 101
Schmid, T. 254–5, 257–8, 264–7, 269, 280–1
Scholz, B. 141n
Schuler, W. 62n
Schwarzschild, R. 99
scope bias 271–272, 279
scrambling 209, 220, 280
Seiler, G. 255, 275–6
Shibatani, M. 19
Shosted, R. 20
Simpler Syntax 7, 9, 27, 29, 34, 39, 41, 107, 111–13, 141, 151, 167, 169, 173, 186, 246, 249–51
Sims, A. 16
sluice-stranding 10, 109, 113–14, 136
sluicing 109–13, 136, 166–9
Smolensky, P. 25n
social network 10 , 177, 184, 196, 201–4, 207, 210, 226, 251–2, 269, 281–4
Spanish 55, 126
Stabler, E. 146
Sternefeld, W. 36n, 62n, 75, 78, 220
strain 53–6, 180–1, 270
subjacency 185
subject aux inversion (SAI) 17, 39, 40, 41, 61, 87, 88, 89, 90, 91, 93, 95, 96, 104, 131, 209–11, 214–17, 222, 229–34, 240–3, 245–51
SUBJECT construction 31, 213, 39, 40, 100, 130, 221, 224, 227, 229, 250
subjunctive 127, 211, 257
superiority condition 17–18
surprisal 24, 25, 144, 146, 173
Swabian 256
Swedish 71, 109, 145, 154–6, 161, 184, 186, 206, 243
Swiss German 255, 256, 258–9, 275–8
syntactic nuts 9, 132, 135, 209

Syntactic Nuts 6, 8, 9, 26, 28, 109, 113, 114, 121, 122, 132, 207
Syntactic Structures 6, 213

Tagliamonte, S. 79, 151n
Taylor, A. and 211, 212, 220, 233
Tenenbaum, J. 140n
Tesar, B. 25n
that-t effect 163–4
thematic roles 11, 21, 22, 32, 36, 79, 80, 136, 150–1, 192, 194, 223, 271
Tomasello, M. 187, 190
TOPIC construction 228–9
topicalization 34–5, 73, 79, 135, 150, 209, 214, 228, 232, 234, 251, 262
Torrego, E. 87n
Townshend, D. 23
Trubetskoy, N. 16, 201
Truckenbrodt, H. 27n
Trudgill, P. 14, 179, 239

unacceptability 3–4, 144–7, 150, 153, 170–3
Universal Grammar (UG) 7, 28, 42–43, 115, 123, 139, 157, 187, 283, 284
universals 7, 15, 16, 26–9, 76, 85, 87, 126, 130, 136, 139–42, 188, 191, 264, 282–4

V-to-I 212–13, 221, 233, 235, 237–240
V2 27, 89, 209, 211, 212, 216, 220, 233–4, 241–7, 249–51
V2 construction 246
variable binding 72–3, 111
Vasishth, S. 24, 25
Vat, J. 178
Vata 184
verb clusters 213, 253–81
verb-particle construction 23, 44–7, 49–50, 52
verb projection raising (VPR) 255, 261–2
Vincent, N. 79
virus *see* epidemiology
Visser, F. 223n
Vitányi, P. 140n
Vogel, R. 254, 255, 264–7, 269, 281
VP constructions 41–7
VP ellipsis (VPE) 91–2, 100, 102, 132, 135–136, 138, 169–72, 214, 244

Waxsman, S. 140n
Ward, G. 88
Warren, T. 24
Wasow, T. 12, 24, 43n, 211, 281
weight 43, 68, 105, 248–9, 280
West Flemish 243, 255, 258–9
Wexler, K. 55n, 140n, 141, 247n
wh-island condition 17
wh-movement 9, 108, 153, 206–7
WH-QUESTION construction 110, 112, 185, 230–1, 247, 252
wh-questions 17, 40, 51, 71, 94, 96–7, 100, 110, 149, 185, 187, 194, 230–1, 234, 241, 247, 251
Wichmann, S. 236
Wilkins, W. 81n
Williams, E. 90n
Winkler, S. 4, 86n, 88, 89n, 90, 94, 95n, 98–100, 139n, 280
Wohlgemuth, J. 236
Wulff, S. 52n
Wurmbrand, S. 255, 258–62, 269n, 270, 278, 280

X-bar theory 216, 218

YES-NO QUESTION construction 40, 230, 246, 248–9

Zanuttini, R. 234
Zürich German 256–60, 273
Zwart, C. 255, 280