Syntactic Theory
A Formal Introduction

Ivan A. Sag & Thomas Wasow

ISBN: 1-57586-160-7

Contents

Preface xi

1 Introduction 1
1.1 Two Conceptions of Grammar 1
1.2 An Extended Example 4
1.3 Remarks on the History of the Study of Grammar 9
1.4 Why Study Syntax? 12
1.5 Conclusion 21
1.6 Further Reading 22

2 Some Simple Theories of Grammar 23
2.1 Introduction 23
2.2 Two Simplistic Syntactic Theories 24
2.3 Context-free Phrase Structure Grammar 29
2.4 Applying Context-Free Grammar 34
2.5 Trees Revisited 38
2.6 Worksection on Phrase Structure Grammar 40
2.7 Heads 40
2.8 Subcategorization 41
2.9 Transitivity and Agreement 43
2.10 Conclusion 45
2.11 Further Reading 46

3 Analyzing Features of Grammatical Categories 47
3.1 Introduction 47
3.2 Feature Structures 48
3.3 The Linguistic Application of Feature Structures 54
3.4 The Head Feature Principle 60
3.5 Trees and Phrase Structure 63
3.6 Conclusion 73
3.7 Further Reading 74

4 Complex Feature Values 75
4.1 Introduction 75
4.2 Complements 76
4.3 The COMPS Feature 77
4.4 Specifiers 83
4.5 The Valence Principle 86
4.6 Subject-Verb Agreement Revisited 88
4.7 Determiner-Noun Agreement 91
4.8 Worksection on Case Marking 96
4.9 Summary 99
4.10 Further Reading 100

5 Semantics 101
5.1 Introduction 101
5.2 Semantics and Pragmatics 102
5.3 Linguistic Meaning 104
5.4 How Semantics Fits In 110
5.5 Modification 114
5.6 The Semantic Principles 115
5.7 Coordination Revisited 118
5.8 Summary 124
5.9 Further Reading 125

6 How the Grammar Works 127
6.1 A Factorization of Grammatical Information 127
6.2 A Detailed Example 131
6.3 Appendix: Well-Formed Structures 144

7 Binding Theory 147
7.1 Introduction 147
7.2 Binding Theory of Chapter 1 Revisited 148
7.3 A Feature-Based Formulation of Binding Theory 149
7.4 Imperatives 161
7.5 The Argument Realization Principle Revisited 167
7.6 Further Reading 170

8 The Structure of the Lexicon 171
8.1 Introduction 171
8.2 Lexical Types 172
8.3 Lexical Rules 184
8.4 Derivational Rules 194
8.5 Conclusion 198
8.6 Further Reading 199

9 Realistic Grammar 201
9.1 Introduction 201
9.2 The Grammar So Far 202
9.3 Constraint-Based Lexicalism 218
9.4 Incremental Processing 219
9.5 A Performance-Plausible Competence Grammar 222
9.6 Universal Grammar: A Mental Organ? 227
9.7 Conclusion 231
9.8 Further Reading 231

10 The Passive Construction 233
10.1 Introduction 233
10.2 Basic Data 233
10.3 The Verb *Be* in Passive Sentences 239
10.4 An Example 240
10.5 Further Reading 249

11 Nominal Types: Dummies and Idioms 251
11.1 Introduction 251
11.2 *Be* Revisited 251
11.3 The Existential *There* 253
11.4 Extraposition 257
11.5 Idioms 265
11.6 Conclusion 269
11.7 Further Reading 269

12 Infinitival Complements 271
12.1 Introduction 271
12.2 The Infinitival *To* 271
12.3 The Verb *Continue* 273
12.4 The Verb *Try* 279
12.5 Subject Raising and Subject Control 284
12.6 Object Raising and Object Control 286
12.7 Further Reading 293

13 Auxiliary Verbs 295
13.1 Introduction 295
13.2 Basic Facts about Auxiliaries 296
13.3 The NICE Properties 302
13.4 Auxiliary *Do* 303
13.5 Analyzing the NICE Properties 305
13.6 Conclusion 316
13.7 Further Reading 316

14 Variation in the English Auxiliary System 317
14.1 Introduction 317
14.2 Auxiliary Behavior in the Main Verb *Have* 318
14.3 African American Vernacular English 320
14.4 Conclusion 334
14.5 Further Reading 334

15 Long-Distance Dependencies 335
15.1 Introduction 335
15.2 Some Data 335
15.3 Formulating the Problem 337
15.4 The Feature GAP 338
15.5 Subject Gaps 348
15.6 The Coordinate Structure Constraint 351
15.7 Conclusion 354
15.8 Further Reading 354

16 From Rule to Sign 355
16.1 Taking Stock 355
16.2 Words and Signs 356
16.3 The Organization of the Lexicon 359
16.4 Phrasal Signs 369
16.5 Conclusion 383

Appendix A: Summary of the Grammar 385

Appendix B: Generative Grammar 411

Glossary 429

References 447

Subject Index 459

Name Index 479

1

Introduction

1.1 Two Conceptions of Grammar

The reader may wonder, why would a college offer courses on grammar – a topic that is usually thought of as part of junior high school curriculum (or even GRAMMAR school curriculum)? Well, the topic of this book is not the same thing that most people probably think of as grammar.

What is taught as grammar in primary and secondary school is what linguists call 'prescriptive grammar'. It consists of admonitions not to use certain forms or constructions that are common in everyday speech. A prescriptive grammar might contain rules like:

> Be sure to never split an infinitive.
>
> Prepositions are bad to end sentences with.

As modern linguists we think that prescriptive grammar is for the most part a pointless activity. We view human language as a phenomenon amenable to scientific investigation, rather than something to be regulated by the decrees of authorities. Your seventh grade math teacher might have discussed the time the Indiana legislature almost passed a bill establishing the value of π as 3, and everybody in class no doubt laughed at such foolishness. Linguists regard prescriptive grammar as silly in much the same way: Natural phenomena simply cannot be legislated.

Of course, we do not deny the existence of powerful social and economic reasons for learning the grammatical norms of educated people.[1] How these norms get established and their influence on the evolution of languages are fascinating questions in sociolinguistics and historical linguistics, but they are beyond the scope of this book. Similarly, we will

[1] By the same token, there may well be good economic reasons for standardizing a decimal approximation to π (though 3 is almost certainly far too crude an approximation for most purposes).

not address issues of educational policy, except to say that in dismissing traditional (prescriptive) grammar instruction, we are not denying that attention to linguistic structure in the classroom can turn students into more effective speakers and writers. Indeed, we would welcome more enlightened grammar instruction in the schools. See Nunberg (1983) for an insightful discussion of these issues.

So, if modern grammarians don't worry about split infinitives and the like, then what do they study? It turns out that human languages are amazingly complex systems, whose inner workings can be investigated in large part simply by consulting the intuitions of native speakers. Here are some examples from English.

Example 1 The adjectives *unlikely* and *improbable* are virtually synonymous: we talk about unlikely or improbable events or heroes, and we can paraphrase *It is improbable that Lee will be elected* by saying *It is unlikely that Lee will be elected.* This last sentence is synonymous with *Lee is unlikely to be elected.* So why does it sound so strange to say **Lee is improbable to be elected*? (In keeping with standard linguistic practice, we will use an asterisk to mark an expression that is not well formed – that is, that doesn't 'sound good' to our ears.)

Example 2 The sentences *They saw Pat with Chris* and *They saw Pat and Chris* are near paraphrases. But if you didn't catch the second name, it would be far more natural to ask *Who did they see Pat with?* than it would be to ask **Who did they see Pat and?* Why do these two nearly identical sentences differ with respect to how we can question their parts? Notice, by the way, that the question that sounds well formed (or 'grammatical' in the linguist's sense) is the one that violates a standard prescriptive rule. The other sentence is so blatantly deviant that prescriptivists would never think to comment on the impossibility of such sentences. Prescriptive rules typically arise because human language use is innovative, leading languages to change. If people never use a particular construction – like the bad example above – there's no point in bothering to make up a prescriptive rule to tell people not to use it.

Example 3 The two sentences *Something disgusting has slept in this bed* and *Something disgusting has happened in this bed* appear on the surface to be grammatically completely parallel. So why is it that the first has a passive counterpart: *This bed has been slept in by something disgusting*, whereas the second doesn't: **This bed has been happened in by something disgusting*?

These are the sorts of questions contemporary grammarians try to

answer. The first two will eventually be addressed in this text, but the third will not.[2] The point of introducing them here is to illustrate a fundamental fact that underlies all modern work in theoretical syntax:

> Every normal speaker of any natural language has acquired an immensely rich and systematic body of unconscious knowledge, which can be investigated by consulting speakers' intuitive judgments.

In other words, knowing a language involves mastering an intricate system full of surprising regularities and idiosyncrasies. Languages are phenomena of considerable complexity, which can be studied scientifically. That is, we can formulate general hypotheses about linguistic structure and test them against the facts of particular languages.

The study of grammar on this conception is a field in which hypothesis-testing is particularly easy: the linguist can simply ask native speakers whether the predictions regarding well formedness of crucial sentences are correct.[3] The term 'syntax' is often used instead of 'grammar' in technical work in linguistics. While the two terms are sometimes interchangeable, 'grammar' may also be used more broadly to cover all aspects of language structure; 'syntax', in contrast, refers only to the ways in which words combine into phrases, and phrases into sentences – the form or structure of well formed expressions. Although the boundaries are not sharp, 'syntax' contrasts with 'semantics' (the study of linguistic meaning), 'morphology' (the study of word structure), and 'phonology' (the study of the sound patterns of language) in ways that 'grammar' does not.

⚠ *This symbol before a problem indicates that it should not be skipped. The problem involves something that will either be elaborated upon or else simply incorporated into subsequent chapters.*

---

[2]For extensive discussion of the third question, see Postal (1986).

[3]This methodology is not without its pitfalls. Judgments of acceptability show considerable variation across speakers. Moreover, they can be heavily influenced by context, both linguistic and nonlinguistic. Since linguists rarely make any effort to control for such effects, not all of the data employed in the syntax literature should be accepted without question. On the other hand, many judgments are so unequivocal that they can clearly be relied on. In more delicate cases, many linguists have begun to supplement judgments with data from actual usage, by examining grammatical patterns found in written and spoken corpora. The use of multiple sources and types of evidence is always a good idea in empirical investigations. See Schütze (1996) for a detailed discussion of methodological issues surrounding the use of judgment data in syntactic research.

⚠ **Problem 1: Judging Examples**

Indicate whether each of the following examples is acceptable or unacceptable. If it is unacceptable, give an intuitive explanation of what is wrong with it, i.e. whether it:

a. fails to conform to the rules of English grammar,
b. is grammatically well-formed, but bizarre in meaning (if so, explain why), or
c. contains a feature of grammar that occurs only in a particular variety of English, for example, slang, or a regional dialect; if so, identify the feature. Is it stigmatized in comparison with 'standard' English?

If you are uncertain about any judgments, feel free to consult with others. Nonnative speakers of English, in particular, are encouraged to compare their judgments with others.

(i) Kim and Sandy is looking for a new bicycle.
(ii) Have you the time?
(iii) I've never put the book.
(iv) The boat floated down the river sank.
(v) It ain't nobody goin to miss nobody.
(vi) Terry really likes they.
(vii) Chris must liking syntax.
(viii) Aren't I invited to the party?
(ix) They wondered what each other would do.
(x) There is eager to be fifty students in this class.
(xi) They persuaded me to defend themselves.
(xii) Strings have been pulled many times to get people into Harvard.
(xiii) This is the kind of problem that my doctor is easy to talk to about.
(xiv) A long list of everyone's indiscretions were published in the newspaper.
(xv) Which chemical did you mix the hydrogen peroxide and?
(xvi) There seem to be a good feeling developing among the students.

1.2 An Extended Example

To get a feel for the sort of research syntacticians conduct, consider the following question:

In which linguistic environments do English speakers normally use reflexive pronouns (i.e. forms like *herself* or *ourselves*), and where does it sound better to use a nonreflexive pronoun (e.g. *her*, *she*, *us*, or *we*)?

To see how to approach an answer to this question, consider, first, some basic examples:

(1) a. *We like us.
b. We like ourselves.
c. She likes her. [where, she $\neq$ her]
d. She likes herself.
e. Nobody likes us.
f. *Nobody likes ourselves.
g. *Ourselves like us.
h. *Ourselves like ourselves.

These examples suggest a generalization along the following lines:

Hypothesis I: *A reflexive pronoun can appear in a sentence only if that sentence also contains a preceding expression that has the same reference (i.e. a preceding* COREFERENTIAL *expression); a nonreflexive pronoun cannot appear in a sentence that contains such an expression.*

The following examples are different from the previous ones in various ways, so they provide a first test of our hypothesis:

(2) a. She voted for her. [she $\neq$ her]
b. She voted for herself.
c. We voted for her.
d. *We voted for herself.
e. *We gave us presents.
f. We gave ourselves presents.
g. *We gave presents to us.
h. We gave presents to ourselves.
i. *We gave us to the cause.
j. We gave ourselves to the cause.
k. *Nobody told us about us.
l. Nobody told us about ourselves.
m. *Nobody told ourselves about us.
n. *Nobody told ourselves about ourselves.

These examples are all predicted by Hypothesis I, lending it some initial plausibility. But here are some counterexamples:

(3) a. We think that nobody likes us.
b. *We think that nobody likes ourselves.

According to our hypothesis, our judgments in (3a,b) should be reversed. Intuitively, the difference between these examples and the earlier ones is that the sentences in (3) contain subordinate clauses, whereas (2) and (1) contain only simple sentences.

Problem 2: Applying Hypothesis I
But it isn't actually the mere presence of the subordinate clauses in (3) that makes the difference. To see why, consider the following, which contain subordinate clauses but are covered by Hypothesis I.

(i) We think that she voted for her. [she $\neq$ her]
(ii) We think that she voted for herself.
(iii) *We think that herself voted for her.
(iv) *We think that herself voted for herself.

a. Explain how Hypothesis I accounts for the data in (i)-(iv).
b. What is it about the subordinate clauses in (3) that makes them different from those in (i)-(iv) with respect to Hypothesis I?

Given our investigation so far, then, we might revise Hypothesis I to the following:

Hypothesis II: *A reflexive pronoun can appear in a clause only if that clause also contains a preceding, coreferential expression; a nonreflexive pronoun cannot appear in any clause that contains such an expression.*

For sentences with only one clause (such as (1)-(2)), Hypothesis II makes the same predictions as Hypothesis I. But it correctly permits (3a) because *we* and *us* are in different clauses, and it rules out (3b) because *we* and *ourselves* are in different clauses.

However, Hypothesis II as stated won't work either:

(4) a. Our friends like us.
b. *Our friends like ourselves.
c. Those pictures of us offended us.
d. *Those pictures of us offended ourselves.
e. We found a letter to us in the trash.
f. *We found a letter to ourselves in the trash.

What's going on here? The acceptable examples of reflexive pronouns have been cases (i) where the reflexive pronoun is functioning as an object of a verb (or the object of a preposition that goes with the verb) and (ii) where the antecedent – that is, the expression it is coreferential

with – is the subject or a preceding object of the same verb. If we think of a verb as denoting some sort of action or state, then the subject and objects (or prepositional objects) normally denote the participants in that action or state. These are often referred to as the ARGUMENTS of the verb. In the examples in (4), unlike many of the earlier examples, the reflexive pronouns and their antecedents are not arguments of the same verb (or, in other words, they are not COARGUMENTS). For example in (4b), *our* is just part of the subject of the verb *like*, and hence not itself an argument of the verb; rather, it is *our friends* that denotes participants in the liking relation. Similarly, in (4e) the arguments of *found* are *we* and *a letter to us*; *us* is only part of an argument of *found*.

So to account for these differences, we can consider the following:

Hypothesis III: *A reflexive pronoun must be an argument of a verb that has another preceding argument with the same reference. A nonreflexive pronoun cannot appear as an argument of a verb that has a preceding coreferential argument.*

Each of the examples in (4) contains two coreferential expressions (*we, us, our*, or *ourselves*), but none of them contains two coreferential expressions that are arguments of the same verb. Hypothesis III correctly rules out just those sentences in (4) in which the second of the two coreferential expressions is the reflexive pronoun *ourselves*.

Now consider the following cases:

(5) a. Vote for us!
b. *Vote for ourselves!
c. *Vote for you!
d. Vote for yourself!

In (5d), for the first time, we find a well formed reflexive with no antecedent. If we don't want to append an *ad hoc* codicil to Hypothesis III,[4] we will need to posit a hidden subject (namely, *you*) in imperative sentences.

Similar arguments can be made with respect to the following sentences.

(6) a. We appealed to him_1 to vote for him_2. [$\text{him}_1 \neq \text{him}_2$]
b. We appealed to him to vote for himself.
c. We appealed to him to vote for us.

(7) a. We appeared to him to vote for him.

[4]For example, an extra clause that says: 'unless the sentence is imperative, in which case a second person reflexive is well formed and a second person nonreflexive pronoun is not.' This would rule out the offending case but not in any illuminating way that would generalize to other cases.

b. *We appeared to him to vote for himself.
c. We appeared to him to vote for ourselves.

In (6), the pronouns indicate that *him* is functioning as the subject of *vote*, but it looks like it is the object of the preposition *to*, not an argument of *vote*. Likewise, in (7), the pronouns suggest that *we* should be analyzed as an argument of *vote*, but its position suggests that it is an argument of *appeared*. So, on the face of it, such examples are problematical for Hypothesis III, unless we posit arguments that are in some sense missing. We will return to the analysis of such cases in later chapters.

⚠ **Problem 3: Reciprocals**

English has a 'reciprocal' expression *each other* (think of it as a single word for present purposes), which behaves in some ways like a reflexive pronoun. For example, a direct object *each other* must refer to the subject, and a subject *each other* cannot refer to the direct object:

(i) They like each other.
(ii) *Each other like(s) them.

A. Construct examples parallel to those in (1)–(3), replacing the reflexives with reciprocals. Is the basic behavior of reciprocals similar to that of reflexives?
B. Construct examples parallel to those in (5)–(7), replacing the reflexives with reciprocals. Is the behavior of reciprocals similar to that of reflexives in imperative sentences and in sentences containing *appeal* and *appear*?
C. Are there any constraints that the reciprocal imposes on its antecedent that reflexives don't impose? [*Hint: what change to (1d) and (6b) did you have to make in order to construct the corresponding well formed reciprocal sentence?*]
D. Consider the following contrast:

They lost each other's books.
*They lost themselves' books.

Discuss how such examples bear on the applicability of Hypothesis III to reciprocals. [*Hint: before you answer the question, think about what the verbal arguments are in the above sentences.*]

You can see that things get quite complex quite fast, requiring abstract notions like 'coreference', being 'arguments of the same verb', and allowing arguments to be missing from the sentence but 'understood',

for purposes of the rules for pronoun type. And we've only scratched the surface of this problem. For example, all the versions of the rules we have come up with so far predict that nonreflexive forms of a pronoun should appear only in positions where their reflexive counterparts are impossible. But this is not quite true, as the following examples illustrate:

(8) a. We wrapped the blankets around us.
b. We wrapped the blankets around ourselves.
c. We admired the pictures of us in the album.
d. We admired the pictures of ourselves in the album.

It should be evident by now that formulating precise rules characterizing where English speakers use reflexive pronouns and where they use nonreflexive pronouns will be a difficult task. We will return to this task in Chapter 7. Our reason for discussing it here was to emphasize the following points:

- Normal use of language involves the mastery of a complex system, which is not directly accessible to consciousness.
- Speakers' tacit knowledge of language can be studied by formulating hypotheses and testing their predictions against intuitive judgments of well formedness.
- The theoretical machinery required for a viable grammatical analysis could be quite abstract.

1.3 Remarks on the History of the Study of Grammar

The conception of grammar we've just presented is quite a recent development. Until about 1800, almost all linguistics was primarily prescriptive. Traditional grammar (going back hundreds, even thousands of years, to ancient India and ancient Greece) was developed largely in response to the inevitable changing of language, which is always (even today) seen by most people as its deterioration. Prescriptive grammars have always been attempts to codify the 'correct' way of talking. Hence, they have concentrated on relatively peripheral aspects of language structure. On the other hand, they have also provided many useful concepts for the sort of grammar we'll be doing. For example, our notion of parts of speech, as well as the most familiar examples (such as *noun* and *verb*) come from the ancient Greeks.

A critical turning point in the history of linguistics took place at the end of the eighteenth century. It was discovered at that time that there was a historical connection among most of the languages of Europe, as well as Sanskrit and other languages of India (plus some lan-

guages in between).[5] This led to a tremendous flowering of the field of historical linguistics, centered on reconstructing the family tree of the Indo-European languages by comparing the modern languages with each other and with older texts. Most of this effort concerned the correspondences between individual words and the sounds within those words. But syntactic comparison and reconstruction was also initiated during this period.

In the early twentieth century, many linguists, following the lead of the Swiss scholar Ferdinand de Saussure, turned their attention from the historical (or 'diachronic'[6]) study to the 'synchronic'[7] analysis of languages – that is, to the characterization of languages at a given point in time. The attention to synchronic studies encouraged the investigation of languages that had no writing systems, which are much harder to study diachronically since there is no record of their earlier forms.

In the United States, these developments led linguists to pay far more attention to the indigenous languages of the Americas. Beginning with the work of the anthropological linguist Franz Boas, American linguistics for the first half of the twentieth century was very much concerned with the immense diversity of languages. The Indo-European languages, which were the focus of most nineteenth-century linguistic research, constitute only a tiny fraction of the approximately five thousand known languages. In broadening this perspective, American linguists put great stress on developing ways to describe languages that would not forcibly impose the structure of a familiar language (such as Latin or English) on something very different; most, though by no means all, of this work emphasized the differences among languages. Some linguists, notably Edward Sapir and Benjamin Lee Whorf, talked about how language could provide insights into how people think. They tended to emphasize alleged differences among the thought patterns of speakers of different languages. For our purposes, their most important claim is that the structure of language can provide insight into human cognitive processes. This idea has wide currency today, and, as we shall see below, it constitutes one of the most interesting motivations for studying syntax.

In the period around World War II, a number of things happened to set the stage for a revolutionary change in the study of syntax. One was that great advances in mathematical logic provided formal tools that seemed well suited for application to studying natural languages.

[5]The discovery is often attributed to Sir William Jones who announced such a relationship in a 1786 address, but others had noted affinities among these languages before him.

[6]From the Greek: *dia* 'across' plus *chronos* 'time'

[7]*syn* 'same, together' plus *chronos*.

A related development was the invention of the computer. Though early computers were unbelievably slow and expensive by today's standards, some people immediately saw their potential for natural language applications, such as machine translation or voice typewriters.

A third relevant development around mid-century was the decline of behaviorism in the social sciences. Like many other disciplines, linguistics in America at that time was dominated by behaviorist thinking. That is, it was considered unscientific to posit mental entities or states to account for human behaviors; everything was supposed to be described in terms of correlations between stimuli and responses. Abstract models of what might be going on inside people's minds were taboo. Around 1950, some psychologists began to question these methodological restrictions, and arguing they made it impossible to explain certain kinds of facts. This set the stage for a serious rethinking of the goals and methods of linguistic research.

In the early 1950s, a young man named Noam Chomsky entered the field of linguistics. In the late '50s, he published three things that revolutionized the study of syntax. One was a set of mathematical results, establishing the foundations of what is now called 'formal language theory'. These results have been seminal in theoretical computer science, and they are crucial underpinnings for computational work on natural language. The second was a book called *Syntactic Structures* that presented a new formalism for grammatical description and analyzed a substantial fragment of English in terms of that formalism. The third was a review of B. F. Skinner's (1957) book *Verbal Behavior*. Skinner was one of the most influential psychologists of the time, an extreme behaviorist. Chomsky's scathing and devastating review marks, in many people's minds, the end of behaviorism's dominance in American social science.

Since about 1960, Chomsky has been the dominant figure in linguistics. As it happens, the 1960s were a period of unprecedented growth in American academia. Most linguistics departments in the United States were established in the period between 1960 and 1980. This helped solidify Chomsky's dominant position.

One of the central tenets of the Chomskyan approach to syntax, known as 'generative grammar', has already been introduced: hypotheses about linguistic structure should be made precise enough to be testable. A second somewhat more controversial one is that the object of study should be the unconscious knowledge underlying ordinary language use. A third fundamental claim of Chomsky's concerns the biological basis of human linguistic abilities. We will return to this claim in the next section.

Within these general guidelines there is room for many different theories of grammar. Since the 1950s, generative grammarians have explored a wide variety of choices of formalism and theoretical vocabulary. We present a brief summary of these in Appendix B, to help situate the approach presented here within a broader intellectual landscape.

1.4 Why Study Syntax?

Students in syntax courses often ask about the point of such classes: why should one study syntax?

Of course, one has to distinguish this question from a closely related one: why DO people study syntax? The answer to that question is perhaps simpler: exploring the structure of language is an intellectually challenging and, for many people, intrinsically fascinating activity. It is like working on a gigantic puzzle – one so large that it could occupy many lifetimes. Thus, as in any scientific discipline, many researchers are simply captivated by the complex mysteries presented by the data themselves – in this case a seemingly endless, diverse array of languages past, present and future.

This reason is, of course, similar to the reason scholars in any scientific field pursue their research: natural curiosity and fascination with some domain of study. Basic research is not typically driven by the possibility of applications. Although looking for results that will be useful in the short term might be the best strategy for someone seeking personal fortune, it wouldn't be the best strategy for a society looking for long-term benefit from the scientific research it supports. Basic scientific investigation has proven over the centuries to have long-term payoffs, even when the applications were not evident at the time the research was carried out. For example, work in logic and the foundations of mathematics in the first decades of the twentieth century laid the theoretical foundations for the development of the digital computer, but the scholars who did this work were not concerned with its possible applications. Likewise, we don't believe there is any need for linguistic research to be justified on the basis of its foreseeable uses. Nonetheless, we will mention three interrelated reasons that one might have for studying the syntax of human languages.

1.4.1 A Window on the Structure of the Mind

One intellectually important rationale for the study of syntax has been offered by Chomsky. In essence, it is that language – and particularly,

its grammatical organization – can provide an especially clear window on the structure of the human mind.[8]

Chomsky claims that the most remarkable fact about human language is the discrepancy between its apparent complexity and the ease with which children acquire it. The structure of any natural language is far more complicated than those of artificial languages or of even the most sophisticated mathematical systems. Yet learning computer languages or mathematics requires intensive instruction (and many students still never master them), whereas every normal child learns at least one natural language merely through exposure. This amazing fact cries out for explanation.[9]

Chomsky's proposed explanation is that most of the complexity of languages does not have to be learned, because much of our knowledge of it is innate: we are born knowing about it. That is, our brains are 'hardwired' to learn languages of certain types.

More generally, Chomsky has argued that the human mind is highly modular. That is, we have special-purpose 'mental organs' that are designed to do particular sorts of tasks in particular sorts of ways. The language organ (which, in Chomsky's view, has several largely autonomous submodules) is of particular interest because language is such a pervasive and unique part of human nature. All people use language, and (he claims) no other species is capable of learning anything much like human language. Hence, in studying the structure of human languages, we are investigating a central aspect of human nature.

This idea has drawn enormous attention not only from linguists but also from people outside linguistics, especially psychologists and philosophers. Scholars in these fields have been highly divided about Chomsky's innateness claims. Many cognitive psychologists see Chomsky's work as a model for how other mental faculties should be studied, while others argue that the mind (or brain) should be regarded as a general-purpose thinking device, without specialized modules. In philosophy, Chomsky provoked much comment by claiming that his work constitutes a modern version of Descartes' doctrine of innate ideas.

Chomsky's innateness thesis and the interdisciplinary dialogue it stimulated were major factors in the birth of the new interdisciplinary field of cognitive science in the 1970s. (An even more important factor was the rapid evolution of computers, with the concomitant growth

[8]See Katz and Postal (1991) for arguments against the dominant Chomskyan conception of linguistics as essentially concerned with psychological facts.

[9]Chomsky was certainly not the first person to remark on the extraordinary facility with which children learn language, but, by giving it a central place in his work, he has focused considerable attention on it.

of artificial intelligence and the idea that the computer could be used as a model of the mind.) Chomsky and his followers have been major contributors to cognitive science in the subsequent decades.

One theoretical consequence of Chomsky's innateness claim is that all languages must share most of their structure. This is because all children learn the languages spoken around them, irrespective of where their ancestors came from. Hence, the innate knowledge that Chomsky claims makes language acquisition possible must be common to all human beings. If this knowledge also determines most aspects of grammatical structure, as Chomsky says it does, then all languages must be essentially alike. This is a very strong universal claim.

In fact, Chomsky tends to use the term 'Universal Grammar' to mean the innate endowment that permits language acquisition. A great deal of the syntactic research since the late 1960s has been concerned with identifying linguistic universals, especially those that could plausibly be claimed to reflect innate mental structures operative in language acquisition. As we proceed to develop the grammar in this text, we will ask which aspects of our grammar are peculiar to English and which might plausibly be considered universal.

If Chomsky is right about the innateness of the language faculty, it has a number of practical consequences, especially in fields like language instruction and therapy for language disorders. For example, since there is evidence that people's innate ability to learn languages is far more powerful very early in life (specifically, before puberty) than later, it seems most sensible that elementary education should have a heavy emphasis on language, and that foreign language instruction should not be left until secondary school, as it is in most American schools today.

1.4.2 A Window on the Mind's Activity

If you stop and think about it, it's really quite amazing that people succeed in communicating by using language. Language seems to have a number of design properties that get in the way of efficient and accurate communication of the kind that routinely takes place.

First, it is massively ambiguous. Individual words, for example, often have not just one but a number of meanings, as illustrated by the English examples in (9).

(9) a. Leslie used a *pen*. ('a writing implement')
b. We put the pigs in a *pen*. ('a fenced enclosure')
c. They should *pen* the letter quickly. ('to write')
d. The judge sent them to the *pen* for a decade. ('a penitentiary')

(10) a. The cheetah will *run* down the hill. ('to move fast')
b. The president will *run*. ('to be a political candidate')
c. The car won't *run*. ('to function properly')
d. This trail should *run* over the hill. ('to lead')
e. This dye will *run*. ('to dissolve and spread')
f. This room will *run* $200 or more. ('to cost')
g. She can *run* an accelerator. ('to operate')
h. They will *run* the risk. ('to incur')
i. These stockings will *run*. ('to tear')
j. There is a *run* in that stocking. ('a tear')
k. We need another *run* to win. ('a score in baseball')
l. Fats won with a *run* of 20. ('a sequence of successful shots in a game of pool')

To make matters worse, many sentences are ambiguous not because they contain ambiguous words, but rather because the words they contain can be related to one another in more than one way, as illustrated in (11).

(11) a. Lee saw the student with a telescope.
b. I forgot how good beer tastes.

(11a) can be interpreted as providing information about which student Lee saw (the one with a telescope) or about what instrument Lee used (the telescope) to see the student. Similarly, (11b) can convey either that the speaker forgot how GOOD beer (as opposed to bad or mediocre beer) tastes, or else that the speaker forgot that beer (in general) tastes good. These differences are often discussed in terms of which element a word like *with* or *good* is modifying (the verb or the noun).

Lexical and modificational ambiguity interact to produce a bewildering array of (often comical) ambiguities, like these:

(12) a. Visiting relatives can be boring.
b. If only Superman would stop flying planes!
c. That's a new car dealership.
d. I know you like the back of my hand.
e. An earthquake in Romania moved buildings as far away as Moscow and Rome.
f. The German shepherd turned on its master.
g. I saw that gas can explode.
h. Max is on the phone now.
i. The only thing capable of consuming this food has four legs and flies.
j. I saw her duck.

Problem 4: Ambiguity
Give a brief description of each ambiguity illustrated in (12). We will return to many of these examples – or closely related ones – later in the book.

This is not the end of the worrisome design properties of human language. Many words are used to refer to different things on different occasions of utterance. Pronouns like *them*, *(s)he*, *this*, and *that* pick out different referents almost every time they are used. Even seemingly determinate pronouns like *we* don't pin down exactly which set of people the speaker is referring to (compare *We have two kids/a city council/a lieutenant governor/50 states/oxygen-based life here*). Moreover, although certain proper names like *Sally Ride* or *Sandra Day O'Connor* might reliably pick out the same person almost every time they are used, most conversations are full of uses of names like *Chris*, *Pat*, *Leslie*, *Sandy*, *Bo*, etc. that vary wildly in their reference, depending on who's talking to whom and what they're talking about.

Add to this the observation that some expressions seem to make reference to 'covert elements' that don't exactly correspond to any one word. So expressions like *in charge* and *afterwards* make reference to missing elements of some kind – bits of the meaning that have to be supplied from context. Otherwise, discourses like the following wouldn't make sense, or would at best be incomplete.

(13) a. I'm creating a committee. Kim – you're in charge. [in charge of what? – the committee]
b. Lights go out at ten. There will be no talking afterwards. [after what? – after ten]

The way something is said can also have a significant effect on the meaning expressed. A rising intonation, for example, on a one word utterance like *Coffee?* would very naturally convey 'Do you want some coffee?' Alternatively, it might be used to convey that 'coffee' is being offered as a tentative answer to some question (say, *What was Columbia's former number-one cash crop?*). Or even, in the right context, the same utterance might be used in seeking confirmation that a given liquid was in fact coffee. Intonational meaning can be vivified in striking ways.

Finally, note that communication using language leaves a great deal unsaid. If I say to you *Can you give me a hand here?* I'm not just requesting information about your abilities, I'm asking you to help me out. This is the unmistakable communicative intent, but it wasn't lit-

erally said. Other examples of such inference are similar, but perhaps more subtle. A famous example[10] is the letter of recommendation saying that the candidate in question has outstanding penmanship (and saying nothing more than that!).

Summing all this up, what we have just seen is that the messages conveyed by utterances of sentences are multiply ambiguous, vague, and uncertain. Yet somehow, in spite of this, those of us who know the language are able to use it to transmit messages to one another with considerable accuracy – far more accuracy than the language itself would seem to permit. Those readers who have any experience with computer programming or with mathematical logic will appreciate this dilemma instantly. The very idea of designing a programming language or a logical language whose predicates are ambiguous or whose variables are left without assigned values is unthinkable. No computer can process linguistic expressions unless it 'knows' precisely what the expressions mean and what to do with them.

The fact of the matter is that human language-users are able to do something that modern science doesn't understand well enough to replicate via computer. Somehow, people are able to use nonlinguistic information in such a way that they are never even aware of most of the unwanted interpretations of words, phrases, and sentences. Consider again the various senses of the word *pen*. The 'writing implement' sense is more common – that is, more frequent in the language you've been exposed to (unless you're a farmer or a prisoner) – and so there is an inherent bias toward that sense. You can think of this in terms of 'weighting' or 'degrees of activation' of word senses. In a context where farm animals are being discussed, though, the weights shift – the senses more closely associated with the subject matter of the discourse become stronger in this case. As people direct their attention to and through a given dialogue, these sense preferences can fluctuate considerably. The human sense selection capability is incredibly robust, yet we have only minimal understanding of the cognitive mechanisms that are at work. How exactly does context facilitate our ability to locate the correct sense?

In other cases, it's hard to explain disambiguation so easily in terms of affinity to the domain of discourse. Consider the following contrast:

(14) a. They found the book on the table.
b. They found the book on the atom.

[10]This example is one of many due to the late H. Paul Grice, the philosopher whose work forms the starting point for much work in linguistics on problems of PRAGMATICS, how people 'read between the lines' in natural conversation; see Grice (1989).

The preposition *on* modifies the verb in (14a) and the noun in (14b), yet it seems that nothing short of rather complex reasoning about the relative size of objects would enable someone to choose which meaning (i.e. which modification) made sense. And we do this kind of thing very quickly, as you can see from (15).

(15) After finding the book on the atom, Sandy went into class, confident that there would be no further obstacles to getting that term paper done.

When you read this sentence, there's no strong feeling that you were 'garden pathed', that is, derailed by an incorrect interpretation midsentence. The decision about how to construe *on the atom* is made well before the words *class* or *confident* are even encountered.

When we process language, we integrate encyclopedic knowledge, plausibility information, frequency biases, discourse information, and perhaps more. Although we don't yet know exactly how we do it, it's clear that we do it very quickly and reasonably accurately. Trying to model this integration is probably the most important research task facing the study of language in the coming millennium.

Syntax plays a crucial role in all this. It imposes constraints on how sentences can or cannot be construed. So the discourse context may provide a bias for the 'fenced enclosure' sense of *pen*, but it is the syntactic context that determines whether *pen* occurs as a noun or a verb. Syntax is also of particular importance to the development of language-processing models, because it is a domain of knowledge that can be characterized perhaps more precisely than some of the other kinds of knowledge that are involved.

When we understand how language processing works, we probably will also understand quite a bit more about how cognitive processes work in general. This in turn will no doubt enable us to develop better ways of teaching language. We should also be better able to help people who have communicative impairments (and more general cognitive disorders). The study of human language-processing is an important sub-area of the study of human cognition, and it is one that can benefit immensely from precise characterization of linguistic knowledge of the sort that syntacticians seek to provide.

1.4.3 Natural Language Technologies

Grammar has more utilitarian applications, as well. One of the most promising areas for applying syntactic research is in the development of useful and robust natural language technologies. What do we mean by 'natural language technologies'? Roughly, what we have in mind is any

sort of computer application that involves natural languages like English, Japanese, or Swahili in essential ways. These include devices that translate from one language into another (or perhaps more realistically, that provide translation assistance to someone with less than perfect command of a language), that understand spoken language (to varying degrees), that automatically retrieve information from large bodies of text stored on-line, or that help the disabled to communicate.

There is one application that obviously must incorporate a great deal of grammatical information, namely, grammar checkers for word processing. Most modern word processing systems include a grammar checking facility, along with a spell-checker. These tend to focus on the concerns of prescriptive grammar, which may be appropriate for the sorts of documents they are generally used on, but which often leads to spurious 'corrections'. Moreover, they typically depend on superficial pattern-matching for finding likely grammatical errors, rather than employing in-depth grammatical analysis. In short, grammar checkers can benefit from incorporating the results of research in syntax.

Other computer applications in which grammatical knowledge is clearly essential include those in which well formed natural language output must be generated. For example, reliable software for translating one language into another must incorporate some representation of the grammar of the target language. If it did not, it would either produce ill-formed output, or it would be limited to some fixed repertoire of sentence templates.

Even where usable natural language technologies can be developed that are not informed by grammatical research, it is often the case that they can be made more robust by including a principled syntactic component. For example, Stanford University's Center for the Study of Language and Information is developing software to reduce the number of keystrokes needed to input text. This has many potential uses, including facilitating the use of computers by individuals with motor disabilities or temporary impairments such as carpal tunnel syndrome. It is clear that knowledge of the grammar of English can help in predicting what words are likely to come next at an arbitrary point in a sentence. Software that makes such predictions and offers the user a set of choices for the next word or the remainder of an entire sentence – each of which can be inserted with a single keystroke – can be of great value in a wide variety of situations. Word prediction can likewise facilitate the disambiguation of noisy signals in continuous speech recognition and handwriting recognition.

But it's not obvious that all types of natural language technologies need to be sensitive to grammatical information. Say, for example,

we were trying to design a system to extract information from an on-line database by typing in English questions (rather than requiring use of a special database query language, as is the case with most existing database systems). Some computer scientists have argued that full grammatical analysis of the queries is not necessary. Instead, they claim, all that is needed is a program that can extract the essential semantic information out of the queries. Many grammatical details don't seem necessary in order to understand the queries, so it has been argued that they can be ignored for the purpose of this application. Even here, however, a strong case can be made for the value of including a syntactic component in the software.

To see why, imagine that we are using a database in a law office, containing information about the firm's past and present cases, including records of witnesses' testimony. Without designing the query system to pay careful attention to certain details of English grammar, there are questions we might want to ask of this database that could be misanalyzed and hence answered incorrectly. For example, consider our old friend, the rule for reflexive and nonreflexive pronouns. Since formal database query languages don't make any such distinction, one might think it wouldn't be necessary for an English interface to do so either. But suppose we asked one of the following questions:

(16) a. Which witnesses work with defendants who supervise them?
b. Which witnesses work with defendants who supervise themselves?

Obviously, these two questions will have different answers, so an English language 'front end' that didn't incorporate some rules for distinguishing reflexive and nonreflexive pronouns would sometimes give wrong answers.

In fact, it isn't enough to tell reflexive from nonreflexive pronouns: a database system would need to be able to tell different reflexive pronouns apart. The next two sentences, for example, are identical except for the plurality of the reflexive pronouns:

(17) a. List all witnesses for the defendant who represented himself.
b. List all witnesses for the defendant who represented themselves.

Again, the appropriate answers would be different. So a system that didn't pay attention to whether pronouns are singular or plural couldn't be trusted to answer correctly.

Even features of English grammar that seem useless – things that appear to be entirely redundant – are needed for the analysis of some sentences that might well be used in a human-computer interaction.

Consider, for example, English subject-verb agreement (a topic we will return to in some detail in Chapters 2–4). Since subjects are marked as singular or plural – *the dog* vs. *the dogs* – marking verbs for the same thing – *barks* vs. *bark* – seems to add nothing. We would have little trouble understanding someone who always left subject agreement off of verbs. In fact, English doesn't even mark past-tense verbs (other than forms of *be*) for subject agreement. But we don't miss agreement in the past tense, because it is semantically redundant. One might conjecture, therefore, that an English database querying system might be able simply to ignore agreement.

However, once again, examples can be constructed in which the agreement marking on the verb is the only indicator of a crucial semantic distinction. This is the case with the following pair:

(18) a. List associates of each witness who speaks Spanish.
b. List associates of each witness who speak Spanish.

In the first sentence, it is the witnesses in question who are the Spanish-speakers; in the second, it is their associates. These will, in general, not lead to the same answer.

Such examples could be multiplied, but these should be enough to make the point: Building truly robust natural language technologies – that is, software that will allow you to interact with your computer in YOUR language, rather than in ITS language – requires careful and detailed analysis of grammatical structure and how it influences meaning. Shortcuts that rely on semantic heuristics, guesses, or simple pattern-matching will inevitably make mistakes.

Of course, this is not to deny the value of clever engineering and statistical approximation. Indeed, the rapid emergence of natural language technology that is taking place in the world today owes at least as much to this as it does to the insights of linguistic research. Our point is rather that in the long run, especially when the tasks to be performed take on more linguistic subtlety and the accuracy of the performance becomes more critical, the need for more subtle linguistic analysis will likewise become more acute.

In short, although most linguists may be motivated primarily by simple intellectual curiosity, the study of grammar has some fairly obvious uses, even in the relatively short term.

1.5 Conclusion

In this chapter, we have drawn an important distinction between prescriptive and descriptive grammar. In addition, we provided an illustration of the kind of syntactic puzzles we will focus on later in the

text.[11] Finally, we provided an overview of some of the reasons people have found the study of syntax inherently interesting or useful. In the next chapter, we look at some simple formal models that might be proposed for the grammars of natural languages and discuss some of their shortcomings.

1.6 Further Reading

An entertaining and knowledgeable exposition of modern linguistics and its implications is provided by Pinker (1994). A somewhat more scholarly survey with a slightly different focus is presented by Jackendoff (1994). For discussion of prescriptive grammar, see Nunberg (1983) and Chapter 12 of Pinker's book (an edited version of which was published in *The New Republic*, January 31, 1994). For an overview of linguistic science in the nineteenth century, see Pedersen (1959). A succinct survey of the history of linguistics is provided by Robins (1967).

Among Chomsky's many writings on the implications of language acquisition for the study of the mind, we would especially recommend Chomsky (1959) and Chomsky (1972); a more recent, but much more difficult work is Chomsky (1986a). There have been few recent attempts at surveying work in (human or machine) sentence processing. J. A. Fodor, Bever, and Garrett (1974) is a comprehensive review of early psycholinguistic work within the Chomskyan paradigm, but it is now quite dated. Garrett (1990) and J. D. Fodor (1995) are more recent, but much more limited in scope.

[11] Our discussion of reflexive and nonreflexive pronouns borrows heavily from the presentation in Perlmutter and Soames (1979: chapters 2 and 3).

8

The Structure of the Lexicon

8.1 Introduction

Several chapters ago, we began with context-free grammar and motivated numerous modifications to develop a formalism better tailored to the description of natural language phenomena. In the course of doing this, we have put more and more of the descriptive burden of our theory into the lexicon. Lexical entries have evolved from simple pairings of phonological forms with grammatical categories into elaborate information structures, in which phonological forms are now paired with more articulated feature structure descriptions. This has permitted us to reduce our inventory of grammar rules to a few very general schemas, and to account for a range of syntactic phenomena.

Since our theory relies heavily on rich lexical representations, we need to consider what kind of internal organization the lexicon should have. In particular, we do not want to claim that all information contained in lexical entries is simply listed. A great deal of what we are now putting into lexical entries is predictable, so stipulating all of it would miss linguistically significant generalizations. For example, we handle subject-verb agreement by having the lexical entries for certain verbs (basically, those with the suffix *-s*) specify that their SPR values have to be third person and singular. Aside from that specification, these entries are essentially identical to those for other forms of what is intuitively the same verb: their part of speech is *verb*; they have the same COMPS values; and their semantics is virtually the same. This is no accident, nor is the fact that the same suffix is used to mark almost all third-person singular present tense verb forms.

Notice, by the way, that capturing such generalizations is motivated not only by general considerations of parsimony, but also by psychological considerations. On encountering a novel English verb (say, a recent coinage such as *email* or an obscure word like *cark*), any com-

petent speaker will add the suffix *-s* when using it in the present tense with a third person singular subject. In short, speakers know that there are systematic (or, as linguists say, 'productive') relationships among different forms of the same word, and our grammar should reflect this systematicity. The focus of the present chapter is to develop mechanisms for expressing regularities within the lexicon.

8.2 Lexical Types

One mechanism for reducing the redundancy of the lexicon has already been introduced: a hierarchy of types. Our initial motivation for introducing types was to define feature appropriateness, to avoid having to specify values for features that are irrelevant to particular classes (such as COUNT for prepositions). Later, we also used types for stating constraints on feature structures. In particular, in Chapter 4, we motivated the 'Nominal SPR Agreement' (NSA) constraint, which required unification of any noun's AGR specification with the AGR specification of the element on its SPR list – that is, nouns have to agree with their specifiers.

We also introduced the idea that some types are subtypes of others, in the following sense:

(1) If T_2 is a subtype of T_1, then
   a. every feature specified as appropriate for T_1 is also appropriate for T_2, and
   b. every constraint associated with T_1 affects all instances and subtypes of T_2.

Formulated in this way, the inheritance of constraints in our type hierarchy is MONOTONIC: constraints on supertypes affect all instances of subtypes, without exception. An intuitive alternative to this conception is to allow for DEFAULT INHERITANCE of constraints, according to which contradictory information associated with a subtype takes precedence over (or OVERRIDES) constraints that would otherwise be inherited from a supertype. Default inheritance allows a type system to express the idea that language embodies general constraints with exceptions – subclasses with idiosyncratic constraints that violate the general ones.

This intuitive idea is simple to express: we need only allow constraints associated with lexical types to be marked as defaults. Suppose a constraint associated with a lexical type T_i is marked as a default. Then this constraint holds of any lexical entry of type T_i for which it is not explicitly overridden. It could be overridden in one of two ways. First, a subtype of T_i might have a constraint associated with it that contradicts the default. That is, there could be a type T_j that is a sub-

type of T_i and has a constraint associated with it that is incompatible with the default. Then the constraint on T_j takes precedence. Second, the information stipulated in a particular lexical entry might contradict the default. That is, a particular instance of a maximal[1] type T_m (T_m a subtype of T_i) could contradict the default. In this case, too, the information associated with the lexical entry takes precedence over the default. But the default constraint is true of all instances of subtypes of T_i in which it is not overridden (as are all nondefault constraints associated with those types).

Natural languages exhibit a great many regularities that have exceptions that are naturally modeled in terms of type hierarchies of this sort. Returning to the example of agreement between nouns and specifiers, recall that NSA was originally motivated by considering what are standardly referred to as COMMON NOUNS, specifically, the noun *dog*. Names, which are often called PROPER NOUNS, are for the most part irrelevant to NSA because they do not take specifiers. This is illustrated in (2).

(2) a. Cameron skates.
b. $\left\{\begin{array}{l}\text{*A}\\ \text{*The}\end{array}\right\}$ Cameron skates.

Moreover, proper nouns are normally third-person singular.

(3) *Cameron skate.

These generalizations will be captured in our type system by introducing a type for proper nouns with a constraint (stated more formally below) specifying that the value of AGR must be of type *3sing* and that the ARG-ST (and hence both SPR and COMPS lists) must be empty. But there are exceptions to this constraint. In particular, there are several proper nouns in English naming mountain ranges that appear only in the plural and only with a determiner.

(4) a. The Andes are magnificent.
b. *The Ande is magnificent.
c. Hannibal crossed the Alps.
d. *Hannibal crossed Alps.

This is a typical situation: many broad and productive generalizations in languages have exceptions, either idiosyncratic lexical items or classes of idiosyncratic expressions. For this reason, we shall adopt the method of default constraint inheritance in our type hierarchy. This will allow us both to restrict the number of types that are required and

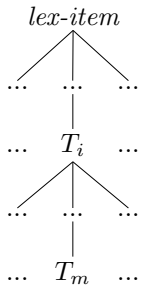
[1]Recall that a 'maximal' type is a type that has no subtypes.

to keep our constraints simple, without precluding the possibility that some instances or subtypes might be exceptions to the constraints.

Problem 1: Mountains and NSA
Does NSA apply to the lexical entries for proper nouns like *Alps, Andes*, and *Himalayas*? Explain your answer and provide data to justify it.

The lexicon itself can be treated in terms of a type hierarchy. By organizing the lexicon in this way, together with the use of default inheritance, as described above, we can minimize the stipulations associated with particular lexical entries and express the shared properties of different word classes, at the same time that we allow for idiosyncrasy of the kind we have been discussing. The overall conception of the lexicon is as shown in (5), where *lex-item* (short for *lexical-item*) encompasses all types of lexical items.

(5)



Each lexical item will belong to some maximal type T_m and T_m will have a family of supertypes T_i, that are intermediate between *lex-item* and T_m. The various intermediate types then correspond to intermediate levels of classification, where type constraints can express linguistic generalizations. Each type in the lexical hierarchy has constraints associated with it – some inviolable, and others that are default in nature. Since this is a default inheritance hierarchy, we can provide a natural account of the fact that lexical items have many properties in common but may differ from one another in terms of particular constraints that override the general constraints governing their supertypes. The idea is that each lexical item inherits constraints associated with the maximal type T_m that it belongs to and those associated with supertypes of T_m. A lexical item inherits all inviolable constraints and all compatible default constraints. Once a lexical hierarchy (with associated constraints) is put into place, the lexical entries that we write become highly streamlined (perhaps indicating no more than which maximal type a given item

is assigned to). All further grammatically relevant constraints (i.e. the rest of the constraints that need to be included in the lexical entry) are inherited automatically, according to the method just described.

The rest of this section is devoted to the details of the particular lexical hierarchy we are assuming to be part of our grammar. As indicated, we assume that *lex-item* is the most general type of feature structure. We posit two immediate subtypes of *lex-item*, which we call *word* and *lexeme*. These two types correspond to what, in everyday English, are two different uses of the term 'word'. In some contexts, people informally distinguish, for example, *runs* and *ran* as two different words: they are pronounced differently, have (subtly) different meanings, and have slightly different co-occurrence restrictions. But in other contexts, the same people would have no hesitation in referring to *runs* and *ran* as two forms of the word *run*. Clearly, these are two very different conceptions of a word: the first refers to a certain pairing of sounds and meanings, whereas the latter refers to a family of such pairings. In a formal theory of grammar, these two concepts must not be conflated. Our type *word* corresponds to the first usage (in which *runs* and *ran* are distinct words). The lexical entries that give rise to word structures must all be of type *word*.

But we also want to capture what people have in mind when they use 'word' in the second sense. That is, we want to be able to express the relationship between *runs* and *ran* (and *run* and *running*). We do this by means of the type *lexeme*. A *lexeme* can be thought of as an abstract proto-word, which, by means to be discussed in this chapter, gives rise to genuine words (that is, instances of the type *word*).

The most basic lexical entries then are pairings of a form and a lexeme, which give rise to a set of word entries.[2] Lexemes thus serve as the atoms from which all linguistic descriptions are built. In the next section (and much of the remainder of this book) we will discuss lexical rules, the mechanisms we employ to derive words from lexemes.

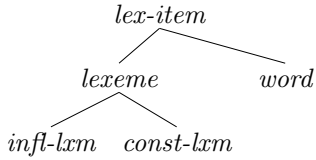
Many of the constraints we present here are stated solely in terms of ARG-ST specifications. These constraints affect lexemes directly, but they also indirectly affect words derived from those lexemes, as most of the lexical rules we will propose have the effect of preserving ARG-ST specifications. Because of the ARP, developed in the last chapter, the

[2]More precisely, basic lexical entries are a pair consisting of a form and a lexeme description; these are used to derive lexical entries that pair a form with a word description. For ease of exposition, we will sometimes blur the distinctions among 'lexeme', 'lexeme description', and 'lexical entry pairing a form with a lexeme description'. A similar blur will sometimes occur among 'word', 'word description', and 'lexical entry pairing a form and a word description'.

SPR and COMPS values in the relevant word structure will be appropriately constrained as well.

We have assumed that *lexeme* and *word* are the two immediate subtypes of the type *lex-item*. Among lexemes, we draw a further distinction between those that give rise to a set of inflected forms and those that do not inflect. That is, we posit *inflecting-lexeme* (*infl-lxm*) and *constant-lexeme* (*const-lxm*) as two subtypes of *lexeme*.[3] This gives rise to the type hierarchy sketched in (6).

(6)



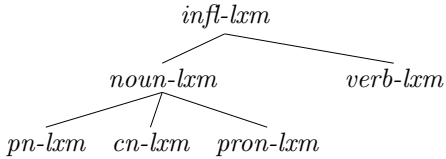
The feature ARG-ST is introduced on the type *lex-item*, so every lexeme and word has an argument structure. Lexemes also have the MODE value none by default:

(7) *lexeme*: $\begin{bmatrix} \text{SEM} & \begin{bmatrix} \text{MODE} & / \text{ none} \end{bmatrix} \end{bmatrix}$

Here we use the symbol '/' to indicate that a certain constraint is a default and hence can be overridden by some conflicting specification.[4]

We further classify inflecting lexemes in terms of the subtypes *noun-lexeme* (*noun-lxm*) and *verb-lexeme* (*verb-lxm*), with the former being further subclassified in terms of the types *proper-noun-lexeme* (*pn-lxm*), *common-noun-lexeme* (*cn-lxm*), and *pronoun-lexeme* (*pron-lxm*), as shown in (8).

(8)



We can see how constraints on lexical types work by considering as an example the constraints associated with some of the types in (8).

First, noun lexemes are governed by the constraint in (9).

[3]We assume that even noninflected words are derived from lexemes. Not much hangs on the choice to formalize our lexicon in this way, rather than to enter such words directly into the lexicon with no corresponding lexemes.

[4]This notion of defaults (as well as the '/' notation) is taken from Lascarides et al. (1996). See also Lascarides and Copestake (1999).

(9)

$$\textit{noun-lxm}: \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{noun} & \\ \text{AGR} & \begin{bmatrix} \text{PER} & / \text{ 3rd} \end{bmatrix} \\ \text{ANA} & / - \end{bmatrix} \end{bmatrix} \\ \text{ARG-ST} & / \langle\ \rangle \\ \text{SEM} & \begin{bmatrix} \text{MODE} & / \text{ ref} \end{bmatrix} \end{bmatrix}$$

This type constraint thus makes it a general but DEFEASIBLE (that is, overrideable) property of noun lexemes that, among other things, they are third person and not anaphors. These properties are then inherited by all instances of the subtypes of *noun-lxm* except when it is overridden, as described above. Default properties are also inherited by words derived by some lexical rule (unless the lexical rule in question explicitly changes the default specification).

(9) ensures that all nouns, whether they are proper nouns or common nouns, have a HEAD value of the type *noun*, and hence bear specifications for CASE (a feature appropriate for *noun* objects, but not, say, for *verb* objects). As just noted, the '−' default specification for ANA is inherited by instances of *noun-lxm*, but this value will be overridden in the entries for reflexive pronouns and reciprocals. The [PER / 3rd] specification will be overridden by certain pronoun lexemes. (9) also states that nouns by default have an empty ARG-ST list, that is, that they take no specifiers or complements. Common nouns will also override this constraint, a matter we turn to directly.[5] Finally, (9) also specifies the semantic mode of nouns as 'ref'. This SEM specification is a default, for reasons that will become clear in Chapter 11.

The subtypes of *noun-lxm* are constrained as follows:[6]

(10) a. $\textit{pron-lxm}: \begin{bmatrix} \ldots \end{bmatrix}$

b. $\textit{pn-lxm}: \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{AGR} & \begin{bmatrix} \text{NUM} & / \text{ sg} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$

[5]This particular constraint is overridden by subordinate types not discussed here (e.g. types we might use for nouns like *picture*, that select for PPs headed by *of*) or by nouns with idiosyncratic complement selection properties of one sort or another (e.g. *difference*, which selects an optional PP argument headed by *between* or *among*).

[6]The notation '[ ]' is used to indicate the obligatory presence of a synsem-struc, whose features are left unspecified.

c.

$$\textit{cn-lxm}: \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{AGR} & \boxed{1} \end{bmatrix} \\ \text{SPR} & \langle\, [\ ]\, \rangle \end{bmatrix} \\ \text{ARG-ST} & / \left\langle \begin{matrix} \text{DetP} \\ \begin{bmatrix} \text{AGR} & \boxed{1} \end{bmatrix} \end{matrix} \right\rangle \end{bmatrix}$$

There are no constraints particular to the type *pron-lxm* (used for personal pronouns and reflexives). However, the type *pn-lxm* (used for proper names) is specified as singular by default. By default inheritance of constraints from *noun-lxm*, instances of both these types are third person, and have an empty ARG-ST list. (10c) requires that common nouns by default select agreeing DetPs as their specifier, thus incorporating NSA and overriding the default empty ARG-ST specification.[7] Further subtypes of *cn-lxm* might be introduced to distinguish count and mass nouns,[8] constraining them to select for an appropriate ([COUNT +] or [COUNT −]) specifier. Note that (10b) need not identify the member of the SPR list with the sole member of the ARG-ST list, as this unification (in any lexical SDs that lexemes of this type give rise to) is already a consequence of the ARP.[9]

Given the type hierarchy and constraints just outlined, the rather complex set of specifications that we want to associate with a particular lexeme can be largely determined simply by associating the lexeme with the appropriate type in the lexical hierarchy. For example, from the fact that the lexical specification for *book* says it is assigned to the type

[7]Note that if lists are represented in terms of the features FIRST and REST (as they often are in computational work), then the proper formulation of the ARG-ST default constraint in (10) would be as in (i):

(i) $$\begin{bmatrix} \text{ARG-ST} & \begin{bmatrix} \text{FIRST} & \begin{matrix} \text{DetP} \\ \begin{bmatrix} \text{AGR} & \boxed{1} \end{bmatrix} \end{matrix} \\ \text{REST} & / \langle\ \rangle \end{bmatrix} \end{bmatrix}$$

[8]The distinction between count and mass nouns was discussed in Chapter 4.

[9]The alert reader may have noticed that (10b)'s claim that specifiers are obligatory for common nouns appears to be inconsistent with the existence of plural and mass NPs that lack determiners. The analysis of such NPs is the topic of Problem 3 below.

cn-lxm, it follows that that the lexeme *book* is associated with all the constraints shown in (11).

(11)
$$\begin{bmatrix} \textit{cn-lxm} & \\ \text{SYN} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{noun} & \\ \text{ANA} & - \\ \text{AGR} & \boxed{1}\begin{bmatrix}\text{PER} & \text{3rd}\end{bmatrix} \end{bmatrix} \\ \text{SPR} & \langle\ [\ ]\ \rangle \end{bmatrix} \\ \text{ARG-ST} & \left\langle \begin{matrix} \text{DetP} \\ \begin{bmatrix}\text{AGR} & \boxed{1}\end{bmatrix} \end{matrix} \right\rangle \\ \text{SEM} & \begin{bmatrix}\text{MODE} & \text{ref}\end{bmatrix} \end{bmatrix}$$

All that remains to be stipulated about this lexeme is its phonological form and the particular predication in its semantic restriction – its form and its idiosyncratic meaning – and perhaps also that it selects for a [COUNT +] determiner. The rest of its grammatical properties follow from the fact that it is assigned to a particular type in our system of grammatical descriptions. This is precisely what lexical stipulation should be reduced to whenever possible.

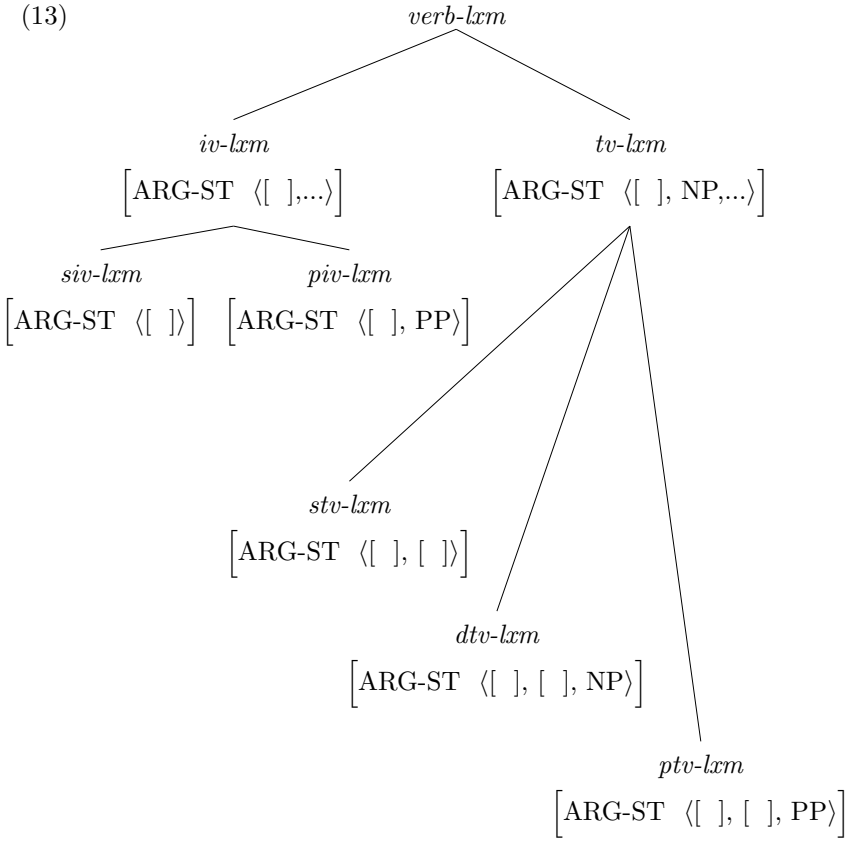
We require that any instance of the type *verb-lxm* must have: a HEAD value of type *verb*, a SPR list that contains one element, and a MODE value of 'prop'. In addition, the argument structure of a lexeme of this type begins by default with an NP.[10] This information is specified in the following constraint on the type *verb-lxm*:

(12)
$$\textit{verb-lxm}: \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{HEAD} & \textit{verb} \\ \text{SPR} & \langle\ [\ ]\ \rangle \end{bmatrix} \\ \text{ARG-ST} & /\ \langle\ \text{NP},\ldots\ \rangle \\ \text{SEM} & \begin{bmatrix}\text{MODE} & \text{prop}\end{bmatrix} \end{bmatrix}$$

The various subtypes of *verb-lxm*, distinguished by their ARG-ST specifications, are organized as shown in (13).

[10]In Chapter 11, we discuss verbs that take non-NP subjects, and hence override this default, which also lends itself to the formulation in fn. 6.

(13)



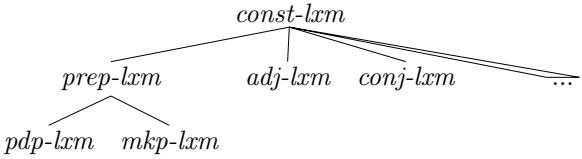
Here we have introduced the types *intransitive-verb-lexeme* (*iv-lxm*) and *transitive-verb-lexeme* (*tv-lxm*) at the first level of subhierarchy. The intransitive verb lexemes are then further classified in terms of the subtypes *strict-intransitive-verb-lexeme* (*siv-lxm*) and *prepositional-intransitive-verb-lexeme* (*piv-lxm*). The former type of verb takes no complements at all (e.g. *sleep*); the latter takes a PP complement (e.g. *rely*). Similarly, the transitive verb lexemes are subclassified into *strict-transitive-verb-lexeme* (*stv-lxm*, e.g. *devour*), *ditransitive-verb-lexeme* (*dtv-lxm*, e.g. *hand*), and *prepositional-transitive-verb-lexeme* (*ptv-lxm*, e.g. *put*). As before, these types and their associated constraints allow us to eliminate lexical stipulation in favor of type-based inference. Thus from the simple statement that *give* belongs to the type *dtv-lxm*, we can infer that it has all the properties shown in (14):

(14)
$$\begin{bmatrix} \textit{dtv-lxm} & \\ \text{SYN} & \begin{bmatrix} \text{HEAD} & \textit{verb} \\ \text{SPR} & \langle\ [\ ]\ \rangle \end{bmatrix} \\ \text{ARG-ST} & \langle\ \text{NP, NP, NP}\ \rangle \\ \text{SEM} & \begin{bmatrix} \text{MODE} & \text{prop} \end{bmatrix} \end{bmatrix}$$

Because all instances of *verb-lxm* are specified as [SPR $\langle\ [\ ]\ \rangle$], words formed from these lexemes will also be so specified. This in turn interacts with the ARP to ensure that in verbal word structures, the first member of the ARG-ST list must also be on the SPR list. Without this [SPR $\langle\ [\ ]\ \rangle$] constraint, there would be another way to satisfy the ARP, namely with an empty SPR list and all the arguments on the COMPS list. As noted in the final section of the previous chapter, this consequence must be avoided (to ensure, for example, that the the verb's first argument is realized as a subject, i.e. before the VP), as it is in the present analysis.

Let us turn now to noninflecting lexemes, that is, the various subtypes of the type *const-lxm* illustrated in (15).

(15)



These correspond to the various kinds of lexeme that undergo no inflectional rules in English.[11] Here the types *preposition-lexeme*, *adjective-lexeme*, and *conjunction-lexeme* have their obvious abbreviations. Once again, we specify general constraints on elements of each subtype of *const-lxm*. For example, the general properties of prepositions are expressed in terms of the following type constraint:

(16) $$\textit{prep-lxm}\colon \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{HEAD} & \textit{prep} \end{bmatrix} \end{bmatrix}$$

In Chapter 7 we distinguished two kinds of prepositions – those that function as predicates and those that serve as argument markers. This distinction corresponds to the two types *predicational-preposition-lexeme* (*pdp-lxm*) and *marking-preposition-lexeme* (*mkp-lxm*) in (15). Recall that in our earlier discussion we distinguished these prepositions in terms

[11]The type *adj-lxm* arguably should be classified as a subtype of *infl-lxm*, rather than as a subtype of *const-lxm*, in light of the fact that many adjectival lexemes give rise to comparative and superlative forms, e.g. *tall*, *taller*, *tallest*. We will not pursue this matter here.

of the feature P-OBJ, which is present only in argument-marking prepositions and is unified with the sole member of the ARG-ST list. We may now incorporate this distinction into our lexicon by attaching the following type constraints to the two subtypes of *prep-lxm*:

(17) a.

$$\textit{pdp-lxm}: \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{MOD} & \left\{\text{NOM, VP, none}\right\} \\ \text{P-OBJ} & \text{none} \end{bmatrix} \\ \text{SPR} & \langle\, [\ ]\, \rangle \end{bmatrix} \\ \text{ARG-ST} & \langle\, \text{NP}\ ,\ \text{NP}\, \rangle \end{bmatrix}$$

b.

$$\textit{mkp-lxm}: \begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{MOD} & \text{none} \\ \text{P-OBJ} & \boxed{1} \end{bmatrix} \\ \text{SPR} & \langle\ \rangle \end{bmatrix} \\ \text{ARG-ST} & \langle\, \boxed{1}\, \rangle \end{bmatrix}$$

The constraint in (17a) allows predicational prepositions to function as modifiers of nominal phrases or VPs, but it also allows for phrases that modify nothing, as is the case when such a PP occurs as the complement of a verb, as in some of the cases discussed in Chapter 7:[12]

(18) I wrapped the blanket [around me].

Note that the first argument of a predicational preposition must be identified with the SPR element, in accordance with the ARP. What plays these roles in (18) is the NP *the blanket*, which is also an argument of the verb *wrapped*. This is the first time we have seen one constituent serving as an argument of more than one predicate at the same time. This is a common phenomenon, however, as we will see in subsequent chapters, especially in Chapter 12.

The argument-marking prepositions, because of the constraint in (17b), project a nonmodifying [SPR $\langle\ \rangle$] PP whose P-OBJ value is identified with the category of the preposition's NP object:

(19) He talks [to himself].

As described in Chapter 7, the P-OBJ value of such a PP is of the same rank as the synsem-struc of the whole PP. Hence it can outrank or be outranked by elements of the ARG-ST list of the higher verb (e.g. *talks*). This allows the objects of argument-marking prepositions to enter into binding relations with other NPs. Finally, recall that some

[12]Note in addition that nothing in our analysis blocks the projection of subject-saturated PPs like [*My blanket* [*around me*]]. In English, these occur only in restricted circumstances, e.g. as 'absolutive' clauses:

(i) [My blanket around me], I was ready to curl up for the night.

prepositions, for example, *around*, behave either as predicational or as argument-marking. Hence the following example is also well formed.

(20) I wrapped the blanket [around myself].

This pattern of optional reflexivization is now neatly accounted for by allowing *around* to live a double life (via two separate lexical entries) as either a predicational or argument-marking preposition.

Some readers may have noticed that our type hierarchy posits two distinct types corresponding roughly to each of the traditional parts of speech. In addition to *noun, verb, adjective, preposition*, etc. – the subtypes of *pos* introduced in Chapter 3 – we now have *noun-lxm, verb-lxm, adj-lxm, prep-lxm*, etc., which are subtypes of *lexeme*. A careful examination of the way we use these two sets of types reveals that they serve rather different functions in our grammar. The subtypes of *pos* specify which features are appropriate for particular categories of words and phrases. These distinctions manifest themselves in the head features, so that the *pos* subtypes show up in our feature structures as specifications of which type of HEAD value we have. The subtypes of *lexeme*, on the other hand, introduce constraints on what combinations of feature values are possible, for example, the NSA or the constraint that verbs take NP subjects (SPRs). These typically involve argument structure (and/or valence features) as well as HEAD features. Consequently, the *pos* subtypes (*noun, verb*, etc.) frequently appear inside of the constraints associated with the *lexeme* subtypes (*noun-lxm, verb-lxm*, etc.).

The type hierarchy simplifies our descriptions in two ways: it saves us from having to assign values to features where they would do no work, for example, PER (person) in prepositions; and it allows us to stipulate common combinations of feature values only once, using (default) inheritance to account for their distribution. The hierarchy contains two sets of types corresponding roughly to the traditional parts of speech then, because the hierarchy serves these two separate functions.

Problem 2: Parts of Speech and Types
What would happen if we tried to eliminate the *pos* subtypes like *noun* and *verb*? To answer this, you will need to consider where the features currently associated with the *pos* subtypes would have to be declared, and what consequences this would have for our feature structures. Be explicit.

Up to this point, we have made no mention of CASE specifications in our lexical type hierarchy. Thus nothing yet guarantees that NPs in English must be accusative except when they are the subject of a

finite verb form. One might think this is a constraint on lexemes, but in fact the various lexical rules we will formulate, for example, the Passive Lexical Rule, never have the effect of transferring accusative case to their first argument (the subject), as we will see in Chapter 10. For this reason, we will treat the assignment of accusative case as a fact about words, not about lexemes. The easiest way to do this is to add the following constraint to our definition of lexical satisfaction.

(21) Case Constraint

Any NP in a noninitial position of a word's ARG-ST list is [CASE acc].

This principle allows us to keep our constraints on verbal lexemes just as we formulated them above, with no mention of case. Thus it is unnecessary to specify lexically the accusative case for most objects, providing a significant improvement on the analysis of English case suggested in the worksection in Chapter 4.[13] Notice, however, that (21) is a one-way implication: it says that certain NPs are accusative, but it says nothing about which NPs are not accusative. The nominative case, characteristic of subjects, will need to be specified in some other way (a point to which we return later in this chapter).

Finally, it must be stressed that the Case Constraint is specific to English. Many other languages exhibit far more complex case systems; see, for example, the problems on Icelandic and Wambaya in Chapter 4.

8.3 Lexical Rules

The lexical rule is a mechanism for further reducing redundancy and stipulation in the lexicon by using information in one lexical entry as the basis for generating another lexical entry. The idea is that the lexicon contains two types of entries: basic ones (lexemes) and those that are 'based on' or 'derived from' these. The lexical rules are used for deriving predictably related lexical entries, such as inflected forms of verbs and nouns.

It is traditional to think of words (or at least certain kinds of words) as being built up from smaller units through the addition of affixes. We have followed this tradition by using our notion of type to distinguish *lexeme* from *word*. For most nouns and verbs, we will assume that there is only one basic lexical entry, which contains a description of type *lexeme*. We then derive all the nominal and verbal *word* entries by applying lexical rules to these *lexeme* entries.

Lexical rules have the general form '$X \Rightarrow Y$'. The intended inter-

[13]But argument-marking prepositions, since they have only one NP argument, are not covered by the Case Constraint; they must specify lexemically that that argument is [CASE acc].

pretation of this is that for any lexical entry that is consistent with the description in X (which we will sometimes refer to as the 'input' to the rule), there is another lexical entry (the 'output' of the rule) that includes the description in Y.[14] The input and output entries are identical, except in those ways that the rule specifies. This breaks down into several special cases:

- If a feature F_1 is given conflicting values in X and Y, the input and output entries must have different values for F_1.
- If X specifies a value v_2 for a feature F_2, but Y says nothing about F_2, then the rule applies only to entries whose value for F_2 is v_2; in this case, both input and output entries have the same value for F_2 (namely v_2).
- If Y specifies that feature F_3 has value v_3 but X says nothing about F_3, the situation is slightly more complex. In this case, the output's value for F_3 is the unification of v_3 with the input's value for F_3. Thus,
    - if the input's value for F_3 is compatible with v_3 (including where the input does not even mention F_3), the output's value for F_3 consists of both the input's value and v_3 (more precisely, their unification); but
    - if the input's value for F_3 is incompatible with v_3, then there is no output (that is, the rule cannot apply).

A lexical rule applies to a lexical entry (of type *lexeme*) and produces as output a new lexical entry whose (morphological) form, syntactic category and meaning are systematically related to the input. Lexical rules often add morphological endings in English, but not always.

Our first example of a lexical rule in fact does nothing to the phonological form of the base; but it adds a restriction requiring that the specifier's AGR value is [NUM sg].

(22) Singular Noun Lexical Rule

$$\left\langle \boxed{1}, \begin{bmatrix} \textit{noun-lxm} \end{bmatrix} \right\rangle \Rightarrow$$

$$\left\langle \boxed{1}, \begin{bmatrix} \textit{word} & \\ \text{SYN} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{AGR} & \begin{bmatrix} \text{NUM} & \text{sg} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle$$

This rule produces lexical entries containing singular forms of both count and mass nouns without mentioning anything about their form

[14] More precisely, if the left-hand side of a lexical rule specifies a value v for a feature F, then the value of F in any input to the rule must unify with v. This includes instances in which the input's value for F is underspecified or disjunctive.

or meaning.[15] Notice that the rule output only specifies the AGR value of the noun itself. But by NSA (which we incorporated into the constraints on the type *cn-lxm*, as discussed in the previous section), this value is identified with the AGR value of the noun's specifier. Thus an output of the rule in (22), such as the one illustrated in (23), must combine with a singular determiner.

(23) Example output of the Singular Noun Lexical Rule

$$\left\langle \text{book}\ ,\ \begin{bmatrix} \textit{word} & \\ \text{SYN} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{noun} & \\ \text{ANA} & - \\ \text{AGR} & \boxed{1}\begin{bmatrix} \text{PER} & \text{3rd} \\ \text{NUM} & \text{sg} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{ARG-ST} & \left\langle \overset{\text{DetP}}{\begin{bmatrix} \text{COUNT} & + \\ \text{AGR} & \boxed{1} \end{bmatrix}} \right\rangle \\ \text{SEM} & \begin{bmatrix} \text{MODE} & \text{ref} \\ \text{INDEX} & i \\ \text{RESTR} & \left\langle \begin{bmatrix} \text{RELN} & \text{book} \\ \text{SIT} & s \\ \text{INSTANCE} & i \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \right\rangle$$

Now consider the closely related rule that maps nominal lexemes into lexical entries for their corresponding plural forms:

(24) Plural Noun Lexical Rule

$$\left\langle \boxed{1}\ ,\ \begin{bmatrix} \textit{noun-lxm} & \\ \text{ARG-ST} & \langle\ [\text{COUNT}\ +]\ \rangle \end{bmatrix} \right\rangle \ \Rightarrow$$

$$\left\langle \text{F}_{NPL}(\boxed{1})\ ,\ \begin{bmatrix} \textit{word} & \\ \text{SYN} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{AGR} & \begin{bmatrix} \text{NUM} & \text{pl} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle$$

[15]The careful reader will have noted that the Singular Noun Lexical Rule as formulated will also apply to proper nouns and pronouns. It may at first seem a bit odd that pronouns undergo the rule, since not all pronouns are singular. But the lexeme entries for pronouns specify their number, so that inherently plural pronouns like *them* cannot undergo rule (22). And if we were to limit (22) to common nouns, we would have no mechanism allowing pronoun lexemes and proper noun lexemes to give rise to words.

Here, F_{NPL} is a morphological function that applies to nominal bases in English, giving their plural forms. In most cases, it suffixes an *-s*, though its full specification would also include a listing of irregular plurals. Thus, $F_{NPL}(cat) = cats$, but $F_{NPL}(child) = children$. The rule says that for every nominal lexeme that takes a [COUNT +] specifier, there is a corresponding lexical entry for a plural noun whose form is dictated by the function F_{NPL}. The requirement that the input take a [COUNT +] specifier keeps the rule from applying to mass nouns like *furniture*, so that there is no word **furnitures*.[16]

A complete formulation of both these lexical rules would require the introduction of a fundamental difference between the semantics of singular and plural nouns. Working out the semantics of singular and plural nouns – which would have to include a semantic analysis of the count/mass distinction – is beyond the scope of this book.

⚠ **Problem 3: Plural and Mass NPs Without Specifiers**

There is a problem with our treatment of common nouns. The type *cn-lxm* requires common nouns to have nonempty SPR lists, and this requirement is preserved in the Plural Noun Lexical Rule above. But specifiers are optional for plural nouns and mass nouns.

A. Give at least three examples showing the optionality of specifiers for plural and mass nouns.

Two obvious approaches to this problem are the following:

(i) allow empty SPR lists in the lexical entries for plural and mass nouns (as we did in the treatment of mass nouns sketched in Chapter 4, section 7); or

(ii) introduce a new grammar rule to account for NPs with plural or mass heads and no specifiers.

Alternative (i) would involve modifying the Plural Noun Lexical Rule, as well as introducing a new subtype of *cn-lxm* for mass nouns. The rule in alternative (ii) is analogous to the Imperative Rule given in Chapter 7, in that it would have only one constituent on the right hand side, and its function would be to license a constituent without a specifier, although its daughter has a nonempty SPR list.

B. The ARP does not manage to sufficiently constrain what alternative (i) would allow. Explain why. [*Hint: Assume that alterna-*

[16]Note that plural pronouns are listed as belonging to *pron-lxm*, and hence can also undergo the Plural Noun Lexical Rule (assuming we put parentheses around the [COUNT +] element on the input's ARG-ST list). This is as desired, since we need to be able to build words from such lexemes. The lack of an *-s* inflection on these words can be accounted for by defining $F_{NPL}(us)$ as *us*, $F_{NPL}(they)$ as *they*, etc.

tive (i) involves making the determiner an optional member of the noun's ARG-ST. Consider all the ways that an ARG-ST element could be realized on the SPR or COMPS list. You may also want to construct examples of common nouns with complements but no specifiers.]

C. Formulate the rule required for alternative (ii). Be as precise as you can.
[*Hint: The trickiest part is stating the disjunction of plural or mass nouns. The place to do this is in the SPR list of the feature structure description on the right hand side of the rule.*]

We posit additional lexical rules to generate entries for the various inflected forms of verbs. Since the forms differ in their distributions, it is useful to have a feature whose value distinguishes among them. We have already seen this feature, which we call FORM, in the analysis of imperatives presented in the previous chapter. Later, we will use this same feature (with different values) to distinguish among prepositions and to mark certain fixed expressions. For now, we will identify just the values of FORM that correspond to verbal inflectional categories for which we need to introduce lexical rules. These are:[17]

(25)

inf	The bare uninflected form, as in *Andy would* **eat** *rice*, *Andy tried to* **eat** *rice*, or **Eat** *rice!*
fin	'Finite', i.e. present or past tense, as in *Andy* **eats** *rice* or *Andy* **ate** *rice*
prp	'Present participle', suffixed with *-ing*, usually following some form of *be*, as in *Andy is* **eating** *rice*
psp	'Past participle' (or 'perfect participle'), the form that follows *have*, as in *Andy has* **eaten** *rice*

[17]Other works that have made use of a feature FORM to distinguish verbal inflectional categories have not always posited precisely these values. In particular, what we call 'inf' has often been called 'base' (with [FORM inf] being reserved for the *to* of infinitives). In addition, a value 'ger' (for 'gerund') has sometimes been proposed for a kind of word not covered here. Like present participles, gerunds are suffixed with *-ing*, but unlike present participles, gerunds head phrases that have the distribution of NPs. The occurrences of *singing* in (i)–(iii) are present participles; those in (iv)–(vi) are gerunds.

(i) The birds are singing.
(ii) Anyone singing in class will be punished.
(iii) Ashley began singing Christmas carols in October.
(iv) Ashley's singing Christmas carols in October annoyed Jordan.
(v) We denied singing during class.
(vi) Don't even think about singing!

The analysis of gerunds is beyond the scope of this text. Hence, we will not consider the question of whether there should be a FORM value for gerunds.

pass ‘Passive’, as in *Rice was* **eaten** *by Andy*
(to be discussed in Chapter 10)

Since FORM is a head feature, VP nodes carry the FORM value of their lexical heads. This permits us to specify in the ARG-ST value of certain elements that they take VP complements whose lexical heads have a particular inflection. For example, the *to* that occurs before verbs (which we will analyze in more detail in Chapter 12) takes [FORM inf] VP complements, and the auxiliary verb *have* takes [FORM psp] VP complements. Such co-occurrence restrictions will be discussed in detail in subsequent chapters. We mention them here only to indicate why FORM is needed, and why it is classified as a head feature.

We are now in a position to state several lexical rules for verbal inflections.

(26) 3rd-Singular Verb Lexical Rule

$$\left\langle \boxed{3}, \begin{bmatrix} \textit{verb-lxm} & \\ \text{SEM} & \begin{bmatrix}\text{RESTR} & \boxed{a}\end{bmatrix} \end{bmatrix} \right\rangle \Rightarrow$$

$$\left\langle \mathrm{F}_{3SG}(\boxed{3}), \begin{bmatrix} \textit{word} & \\ \text{SYN} & \begin{bmatrix}\text{HEAD} & \begin{bmatrix}\text{FORM} & \text{fin}\end{bmatrix}\end{bmatrix} \\ \text{ARG-ST} & \left\langle \begin{bmatrix}\text{CASE} & \text{nom} \\ \text{AGR} & \textit{3sing}\end{bmatrix}, \ldots \right\rangle \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & \boxed{2} \\ \text{RESTR} & \boxed{a} \oplus \left\langle \begin{bmatrix}\text{RELN} & \text{t-overlap} \\ \text{ARG1} & \boxed{2} \\ \text{ARG2} & \text{now}\end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \right\rangle$$

As with the Plural Noun Lexical Rule, we have glossed over the morphological component of the 3rd-Singular Verb Lexical Rule by simply giving it a name, F_{3SG}. The semantic effect of this rule is clear, however: it preserves the semantics of the input, except for the value of the feature RESTRICTION, which is augmented to include a further predication requiring that the INDEX value – the situation described by the verb – be in the temporal-overlap (t-overlap) relation with ‘now’, an atomic value that will always designate the time of an utterance. Thus, according to this rule, using a 3rd singular form of a verb lexeme imposes the requirement that the situation introduced by the verb be located in some temporal interval that overlaps the time of the utterance. This is perhaps only a first attempt at the semantics of present tense verb forms, but our goal in presenting it is to illustrate how a

precise semantic analysis, however refined, can be integrated into lexical rules of the sort outlined here. What the rule in (26) says, then, is that for any verbal lexeme, there is a corresponding finite verb that takes a nominative, third-person singular subject. Further, the morphology and semantics of the latter verb are systematically related to those of the base verb.

Problem 4: *-s*
In most cases, F_{3SG} has the same effect as F_{NPL}, namely, that of suffixing *-s*. In fact, both suffixes have multiple pronunciations, and the conditions under which they are pronounced like *s*, like *z*, or like *iz* are identical. (They depend on phonological properties of the preceding sound.) Nevertheless, these two morphological functions are not identical. Why?
[*Hints: 1. Remember that a function is single-valued, i.e. it specifies only one output for each input. 2. Consider elements that can be used as both nouns and verbs.*]

We turn next to the rule that generates the entries for finite verbs with subjects other than third-person singular NPs. Because the type distinction we have drawn between the AGR values *3sing* and *non-3sing* already distinguishes third singular NPs from all others, this rule is almost identical to the last one, as shown in (27).

(27) Non-3rd-Singular Verb Lexical Rule

$$\left\langle \boxed{3}, \begin{bmatrix} \textit{verb-lxm} & \\ \text{SEM} & \begin{bmatrix} \text{RESTR} & \boxed{a} \end{bmatrix} \end{bmatrix} \right\rangle \Rightarrow$$

$$\left\langle \boxed{3}, \begin{bmatrix} \textit{word} & \\ \text{SYN} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{FORM} & \text{fin} \end{bmatrix} \end{bmatrix} \\ \text{ARG-ST} & \left\langle \begin{bmatrix} \text{CASE} & \text{nom} \\ \text{AGR} & \textit{non-3sing} \end{bmatrix}, \ldots \right\rangle \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & \boxed{2} \\ \text{RESTR} & \boxed{a} \oplus \left\langle \begin{bmatrix} \text{RELN} & \text{t-overlap} \\ \text{ARG1} & \boxed{2} \\ \text{ARG2} & \text{now} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \right\rangle$$

No change in form is introduced, the nominative subject of the output must have a *non-3sing* AGR value (see Chapter 4 for further discussion), and the situation picked out by the verb must overlap the utterance time. A typical output, illustrated in (28), can never combine with a third-person singular subject and will project a sentence whose semantics includes the restrictions illustrated in addition to those that arise from the particular dependents of the verb or from modifiers.

(28)

$$\left\langle \text{give}, \begin{bmatrix} \textit{word} & \\ \text{SYN} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{verb} & \\ \text{FORM} & \text{fin} \end{bmatrix} \end{bmatrix} \\ \text{ARG-ST} & \left\langle \begin{matrix} \text{NP}_i \\ \begin{bmatrix} \text{AGR} & \textit{non-3sing} \\ \text{CASE} & \text{nom} \end{bmatrix} \end{matrix}, \text{NP}_j, \text{NP}_k \right\rangle \\ \text{SEM} & \begin{bmatrix} \text{MODE} & \text{prop} \\ \text{INDEX} & s \\ \text{RESTR} & \left\langle \begin{bmatrix} \text{RELN} & \text{give} \\ \text{SIT} & s \\ \text{GIVER} & i \\ \text{GIVEN} & j \\ \text{GIFT} & k \end{bmatrix}, \begin{bmatrix} \text{RELN} & \text{t-overlap} \\ \text{SIT} & u \\ \text{ARG1} & s \\ \text{ARG2} & \text{now} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \right\rangle$$

The two rules just discussed create lexical entries that license the present tense forms of verbs. The next rule creates entries for the past tense forms. English makes no distinction between singular and plural

in past tense forms (aside from *was* vs. *were*);[18] hence only one rule is needed.

(29) Past-Tense Verb Lexical Rule

$$\left\langle \boxed{2}\,,\ \begin{bmatrix} \textit{verb-lxm} & \\ \text{SEM} & \begin{bmatrix}\text{RESTR} & \boxed{a}\end{bmatrix} \end{bmatrix} \right\rangle \Rightarrow$$

$$\left\langle \mathrm{F}_{PAST}(\boxed{2})\,,\ \begin{bmatrix} \textit{word} & \\ \text{SYN} & \begin{bmatrix}\text{HEAD} & \begin{bmatrix}\text{FORM} & \text{fin}\end{bmatrix}\end{bmatrix} \\ \text{ARG-ST} & \left\langle \begin{bmatrix}\text{CASE} & \text{nom}\end{bmatrix}, \ldots \right\rangle \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & \boxed{3} \\ \text{RESTR} & \boxed{a} \oplus \left\langle \begin{bmatrix} \text{RELN} & \text{t-precede} \\ \text{ARG1} & \boxed{3} \\ \text{ARG2} & \text{now} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \right\rangle$$

(29) posits a function F_{PAST} to account for the morphological relation between verbal lexemes and their past tense forms; in most cases, this consists of suffixing *-ed*, though there are many exceptions (such as *sleep*/*slept*, *eat*/*ate*, and *put*/*put*). Like the lexical rules for present tense

[18]Of course, something must be said about this exception and about the first-person singular form *am*. The fact that *be* makes finer distinctions among its verb forms than other verbs does not justify making these distinctions throughout the rest of the verbal system in English. Rather, it is more parsimonious to make *be* an exception to some of these lexical rules, and to stipulate the individual forms in the lexicon or to posit highly specialized lexical rules for the forms of *be*. (The latter course may be desirable because, as we shall see in Chapters 10–14, there appear to be several different *be* lexemes in English). We will not go into the question of what kind of formal machinery to use to specify that certain lexical entries are exceptions to certain lexical rules, though some such mechanism is surely needed irrespective of *be*.

The inflectional paradigm of *be* looks quite confusing at first, with one form (*am*) that goes only with first-person subjects and others (*are, were*) that go only with subjects that are second person or plural. The situation looks a bit less arbitrary if we make use of the hierarchy of subtypes of *non-3sing* suggested in footnote 10 of Chapter 4. That hierarchy makes available a type *1sing* that is the AGR value we need for the specifier of *am*. It also provides a type *non-1sing* encompassing just second-person and plural AGR values (that is, it excludes just the first-person singular and third-person singular values). This is precisely the AGR value we need for the specifier of *are* and *were*. To accommodate *was*, we would need either two lexical entries (one requiring a *1sing* specifier and another requiring a *3sing* specifier) or a lexical entry with a disjunctive SPR value.

verbs, (29) requires its subject to be nominative (to rule out examples like **Him died*); but unlike the present tense rules, it puts no number or person restrictions on the subject, since English past tense verbs exhibit no agreement with their subjects. The semantic effect of the rule is parallel to that of the two present tense rules, except that the relation on the restriction added is 't-precede', rather than 't-overlap'. That is, the situation denoted by the index of the verb temporally precedes the time of utterance if the verb is in the past tense. Again, this is only an approximation of the semantics of the English past tense, but it is good enough for our purposes.

Finally, we need a trivial lexical rule for noninflecting lexemes.

(30) Constant Lexeme Lexical Rule

$$\langle \boxed{1}, [\textit{const-lxm}] \rangle \quad \Rightarrow \quad \langle \boxed{1}, [\textit{word}] \rangle$$

This rule does nothing except promote such lexemes to wordhood so that the resulting entries can give rise to word SDs.

⚠ **Problem 5: More Lexical Rules**

Write the lexical rules that generate the infinitival (base), present participle, and past participle entries for verbs in English. You should be able to specify the syntactic effects of these rules quite precisely and (for the most part) straightforwardly. For the semantics, you will need to invent some relation names (that is, some new values for the feature RELN).

⚠ **Problem 6: Coordination and Verbal Form**

FORM values must be preserved in coordinate structures. That is, if a VP[FORM $\boxed{1}$] is a coordinate structure, then each conjunct must also be VP[FORM $\boxed{1}$]. Demonstrate that this is true by constructing examples with coordinate VPs of various kinds, for example, the VP head of a finite sentence or the VP complement of verbs like *will*, *have*, or (progressive) *be*.

Problem 7: Coordination and Tense

Notice that we have not introduced distinct FORM values to distinguish between past and present tense verbs – we treat both as [FORM fin]. Using the result of the preceding problem, explain why this decision is right or wrong. Be sure to consider examples where finite VPs that differ in tense are coordinated.

8.4 Derivational Rules

Each of the lexical rules in the previous section maps entries of type *lexeme* into lexical entries for inflected words. Such rules are usually called INFLECTIONAL rules. In contrast to these, languages have another kind of lexical rule (called a DERIVATIONAL rule) that relates lexemes to lexemes. This is pretty clearly what we want for certain elements built up by prefixes and suffixes from basic lexemes. So noun lexemes like *driver* or *eater* might be derived by the following lexical rule.

(31) Agent Nominalization Lexical Rule

$$\left\langle \boxed{2}\,,\ \begin{bmatrix} \textit{verb-lxm} & \\ \text{ARG-ST} & \langle\ \text{NP}_i\ (,\ \boxed{1}\text{NP})\ \rangle \\ \text{SEM} & \begin{bmatrix}\text{INDEX} & s\end{bmatrix} \end{bmatrix} \right\rangle \ \Rightarrow$$

$$\left\langle \text{F}_{-er}(\boxed{2})\,,\ \begin{bmatrix} \textit{cn-lxm} & \\ \text{ARG-ST} & \left\langle \text{DetP}\ (,\ \begin{bmatrix}\text{P-OBJ} & \boxed{1} \\ \text{FORM} & \text{of}\end{bmatrix}^{\text{PP}})\right\rangle \\ \text{SEM} & \begin{bmatrix}\text{INDEX} & i\end{bmatrix} \end{bmatrix} \right\rangle$$

Here the function F_{-er} adds the appropriate suffix to the form of the rule output. The input is a verbal lexeme whose subject's index i is identified with the index of the nominal output. Note that the change in type from *verb-lxm* to *cn-lxm* has many side effects in terms of values of head features and in terms of the MODE value within the semantics. The RESTR value remains unchanged, however, as the information present in the input is compatible with the type constraints associated with the output type.

The ARG-ST values in (31) deserve some comment. The input must be either a strictly transitive or a strictly intransitive verb. Thus we correctly rule out agent nominals of such verbs as *rely* or *put*. The output, like other common nouns, takes a determiner. In addition, if the input is a transitive verb, the agent nominal may take a PP complement whose object is identified with the object of the input verb. This is for agent nominals such as *the discoverer of oxygen* and *a builder of bridges*.[19]

[19]Notice that in formulating this rule, we have used the FORM value 'of' to indicate that the preposition heading this PP must be *of*. We return to the matter of FORM values for prepositions in Chapter 10.

Consider, for example, the lexical entry for the verbal lexeme *drive*, the semantics of which is a proposition whose RESTR value contains a driving predication, with the role of driver assigned to the referent of the verb's subject. Applying the Agent Nominalization Lexical Rule to this lexeme yields an entry for the nominal lexeme *driver*, whose index is restricted to be the driver in a driving predication (since the RESTR value is unchanged). This lexical entry can now undergo both our nominal lexical rules, and so we derive entries for the singular form *driver* and its plural analog *drivers*.

There are further semantic constraints that must be placed on our derivational rule, however. For example, the subject in the input verb has to be sufficiently agentive – that is, it must play an active (usually volitional) role in the situation. That's why nominalizations like *knower* or *resembler* sound funny. But the formulation in (31) is a reasonable first pass at the problem, and it gives you an idea of how phenomena like this can be analyzed within our framework.

There are many other cross-categorial relations that work this way in English. Noun lexemes, both common and proper, can be converted into verbal lexemes:

(32) a. Sandy *porched* the newspaper without difficulty.
b. The senator *houdinied* his way out of the accusations.
c. They have been *computering* me to death all morning.

This kind of derivation without morphological change, an instance of what is often called ZERO DERIVATION, could be handled by one or more derivational rules.

Finally, note that lexical rules are a traditional way of approaching the problem of valence alternations, that is, the fact that many verbs allow systematically related valence patterns. Among the most famous of these is the dative alternation illustrated in (33) – (34).

(33) a. Birch gave Brooke a book.
b. Birch gave a book to Brooke.

(34) a. Birch handed Brooke a book.
b. Birch handed a book to Brooke.

Rather than list entries for two distinct verbal lexemes for *give*, *hand*, and a family of related elements, it makes much more sense to list only one (with one of the two valence patterns fixed) and to derive the other by a derivational rule. Note however, that there are certain other verbs or particular idiomatic uses that appear in only one of the two valence patterns.

(35) a. Kris donated a book to the library.

b. *Kris donated the library a book.

(36) a. Dale gave Brooke a hard time.
b. ??Dale gave a hard time to Brooke.

These underline once again the need for a theory of exceptions to lexical rules and lexical irregularity.

Other famous examples of valence alternation are illustrated in (37)–(41).

(37) a. The police sprayed the protesters with water.
b. The police sprayed water on the protesters. ('*spray/load*' alternations)

(38) a. The students drove cars.
b. These cars drive easily. ('middle' uses)

(39) a. Pat sneezed.
b. Pat sneezed the napkin off the table. ('caused motion' uses)

(40) a. The horse kicked me.
b. The horse kicked me black and blue. ('resultative' uses)

(41) a. Pat yelled.
b. Pat yelled her way into the meeting. (the 'X's way' construction)

All these patterns of valence alternation are governed by both semantic and syntactic constraints of the kind that could be described by finely tuned lexical rules.

⚠ Problem 8: Arguments in Japanese

As noted in Chapter 7, Japanese word order differs from English in a number of ways, including the fact that it is a 'Subject-Object-Verb' (SOV) language. Here are a few relevant examples ('nom', 'acc', and 'dat' stand for nominative, accusative, and dative case, respectively).[20]

(i) *hitorino otoko-ga sono hon-o yonda*
one man-nom that book-acc read-past
'One man read that book'
[cf. *Yonda hitorino otoko-ga sono hon-o
*Hitorino otoko-ga yonda sono hon-o
*Otoko-ga hitorino sono hon-o yonda
*Hitorino otoko-ga hon-o sono yonda

[20]The examples marked with '*' here are unacceptable with the indicated meanings. Some of these might be well formed with some other meaning of no direct relevance; others might be well formed with special intonation that we will ignore for present purposes.

*Hitorino otoko-ni/-o sono hon-o yonda
*Hitorino otoko-ga sono hon-ga/-ni yonda.]

(ii) *Hanako-ga hon-o yonda*
Hanako-nom book-acc read-past
'Hanako read the book(s)'
[cf. *Yonda Hanako-ga hon-o
*Hanako-ga yonda hon-o
*Hanako-ni/-o hon-o yonda
*Hanako-ga hon-ni/-ga yonda.]

(iii) *sensei-ga Taroo-ni sono hon-o ageta*
sensei-nom Taroo-dat that book-acc gave-past
'The teacher(s) gave that book to Taroo'
[cf. *Ageta sensei-ga Taroo-ni sono hon-o
*Sensei-ga ageta Taroo-ni sono hon-o
*Sensei-ga Taroo-ni ageta sono hon-o
*Sensei-o/-ni Taroo-ni sono hon-o ageta
*Sensei-ga Taroo-ga/-o sono hon-o ageta
*Sensei-ga Taroo-ni sono hon-ga/-ni ageta.]

(iv) *Sono hon-ga akai-desu*
that book-nom red-copula
'That book is red'

As the contrasting ungrammatical examples show, the verb must appear in final position in Japanese. In addition, we see that verbs select for NPs of a particular case, much as in English. In the following tasks, assume that the nouns and verbs of Japanese are inflected words, derived by lexical rule from the appropriate lexemes.

A. Given the data illustrated here, how could the Head-Specifier and Head-Complement rules be revised to deal with Japanese? Explain the effect of the difference(s) you have posited.
B. Give the lexical entry for each of the verbs illustrated in (i)–(iv) (treating *akai-desu* as a simple intransitive verb). The data given only permit you to specify some features; leave others unspecified. Make sure your entries interact with the rules you formulated in Task 1 to account for the above data.
C. Assume that nouns like *Taroo*, *hon*, etc. are entered in the Japanese lexicon as nominal lexemes. Give lexical entries for these lexemes (again limiting the features specified to those for which you have data).
D. Formulate three lexical rules for deriving the inflected forms (the words ending in *-ga*, *-o*, and *-ni*) from the nominal lexemes.

E. Explain the role of the Head Feature Principle in accounting for the case-selecting properties of verbs in (i)–(iv).

As you know from Problem 3 of Chapter 7, both subjects and objects in Japanese can in general be omitted. When such an element is omitted, the sentence usually has a meaning that we would express in English by using a pronoun corresponding to the unexpressed element:

(v) *Taroo-ga yonda*
Taroo-nom read-past
'Taroo read it/them'

(vi) *hon-o yonda*
book-acc read-past
'He/she/they read the book(s)'

(vii) *yonda*
read-past
'He/she/it/they read it/them'

(viii) *Taroo-ga tatai-ta.*
Taroo-NOM hit-PST
'Taroo$_i$ hit him$_j$/her$_j$/them$_j$/*himself$_i$.'

F. Sketch an account of this phenomenon that uses lexical rules. Be explicit about the verbs' values for SPR, COMPS, and ARG-ST. You should assume that all complements can be omitted, not just those illustrated in (v) – (viii); i.e. your lexical rules should apply with maximum generality. Following the suggestion in Chapter 7, Problem 2, you should assume that the ARP is a default principle in Japanese (unlike English).

8.5 Conclusion

An important insight, going back at least to Saussure, is that all languages involve arbitrary (that is, unpredictable) information. Most clearly, the association between the forms (sounds) and meanings of words is purely conventional, in the vast majority of cases. A grammar of a language must list these associations somewhere. The original conception of the lexicon in modern linguistics was simply as the repository of such arbitrary information.

This conception did not last long, however. Beginning in the early years of transformational grammar, linguists began enriching their conception of the lexicon to include information that was not idiosyncratic

to individual words. This trend continued in a great deal of research carried out within a variety of grammatical frameworks.

In this text, we have to some extent recapitulated this history. We began with context-free grammar in which the lexicon contained only idiosyncratic information, and we gradually enriched our lexical representations, including more and more information – much of it systematic and predictable – about the grammatical and semantic properties of words. Indeed, most of the information needed to determine the well-formedness of sentences is now encoded into our lexical entries.

With the increased expressiveness and concomitant complexity of lexical entries came a need to express succinctly certain generalizations about words. In this chapter, we have examined two formal mechanisms for capturing such generalizations. Structuring the lexicon in terms of an inheritance hierarchy of types has made it possible to factor out information common to many lexical entries, thereby greatly reducing lexical redundancy. By allowing inherited information to be overridden, we can encode default values for features, while still allowing for lexical idiosyncrasy. The second mechanism, the lexical rule, is an older idea, going back to work in transformational grammar of the 1970s. We will make extensive use of lexical rules in subsequent chapters. In fact, many of the phenomena that provided the motivation for transformations in the 1950s and 1960s can be reanalyzed in our theory using lexical rules. These include the passive construction – the topic of Chapter 10 – and many of the properties of the English auxiliary verb system, which we treat in Chapter 13.

8.6 Further Reading

An important early paper on lexical rules is Jackendoff (1975). The idea of combining lexical rules with an inheritance hierarchy was first put forward by Flickinger et al. (1985). See also Pollard and Sag (1987, chap. 8). Briscoe et al. (1993) is a collection of more recent papers about lexical hierarchies, default inheritance, and related issues.

9

Realistic Grammar

9.1 Introduction

In the preceding eight chapters, we have laid out the theory that we will apply to more complex data in the remainder of this book. Perhaps surprisingly, only one more grammatical principle and one more grammar rule are required; the other extensions we will present in subsequent chapters (aside from a minor reformulation of the ARP in chapter 15) all concern the lexicon. Before we proceed, however, it is useful to consolidate the components of our treatment of English grammar and to reflect on the strategy we have adopted for solving syntactic problems – to reflect on the motivation for the design of grammar.

As we noted briefly in Chapter 2, syntacticians rely heavily on considerations of parsimony: the desirability of 'capturing generalizations' is given great weight in choosing between analyses. This concern with providing elegant descriptions is not unique to this field, though it probably figures more prominently in linguistic argumentation than elsewhere. It is natural to ask, however, whether a grammar whose design has been shaped in large measure by concern for parsimony corresponds straightforwardly to the way linguistic knowledge is represented in the minds of language users. We argue in this chapter that the available psycholinguistic evidence fits rather well with the conception of grammar that we have been developing in this book.

First, however, we turn to a summary of our grammar to date. The next section of this chapter gives a formal presentation of everything we have covered so far, including types, lexical entries, grammar rules, the well-formedness definitions (incorporating various principles), and lexical rules.

Section 9.2.1 presents the type hierarchy in terms of a table that provides details of the features and constraints associated with each type, as well as an indication of each type's immediate supertype (IST) in the

hierarchy. Constraints on the possible values of features are introduced by enclosing the list of possible values in curly brackets, following the feature name. Almost all of the types and constraints listed in Section 9.2.1 have been introduced in earlier chapters. We have added almost nothing that is new. Two new types are *gram-cat* (for 'grammatical category') and *sem-struc* (for 'semantic structure). The first of these is the type of the value of SYN – that is, feature structures of type *gram-cat* have the features HEAD, SPR, and COMPS. The second is the type of the value of SEM, which consists of the features MODE, INDEX, and RESTR. Both of these have been implicit in the discussion of the preceding chapters. The type *verb* lists no features or constraints associated with it, but we have put in the placeholder '[...]', because this will change in subsequent chapters.

Section 9.2.2 gives some sample lexical entries. It is worth noting that most of what we have to stipulate in our entries is semantic. By virtue of having a richly structured lexicon, we are able to limit the amount of syntactic information that has to be listed in individual entries, thereby greatly reducing redundant stipulation.

Section 9.2.3 lists our familiar grammar rules from Chapter 6, together with the Imperative Rule introduced in Chapter 7.

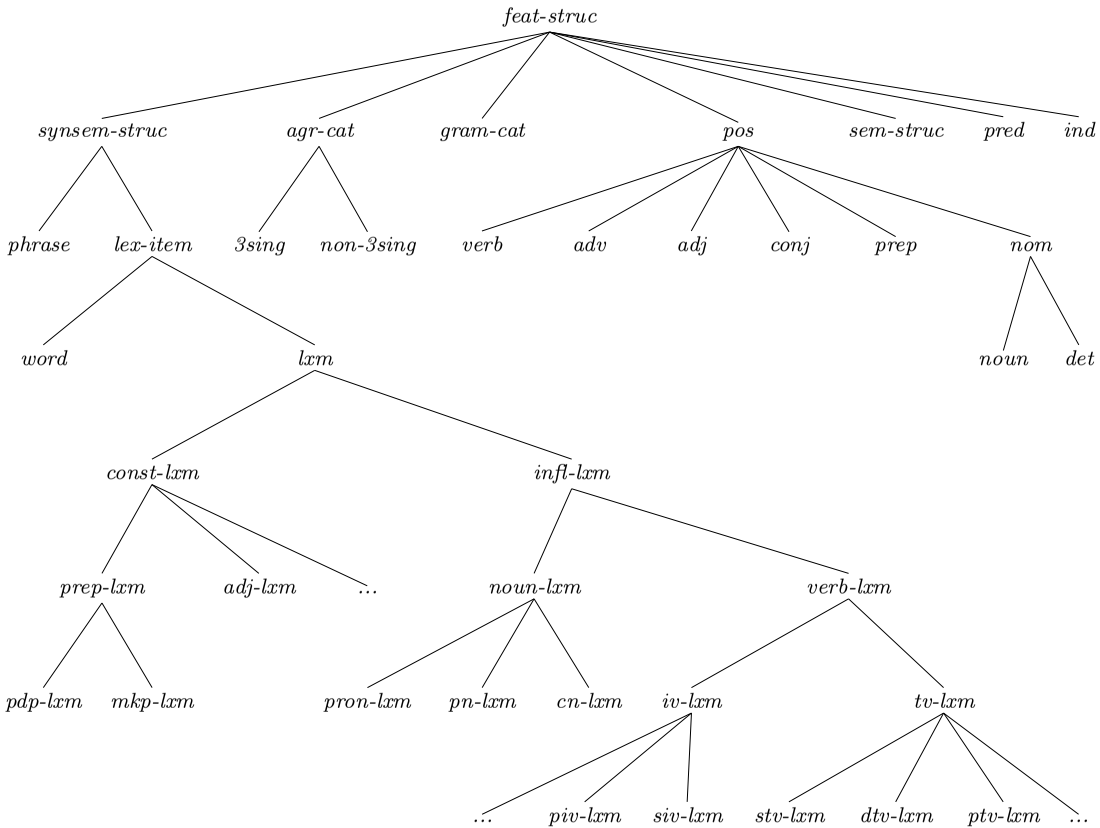
Section 9.2.4 gives the formal definitions of well-formed tree structure and lexical and phrasal satisfaction, incorporating all of the general principles of grammar we have adopted so far. This version is slightly modified from the one given in Chapter 6, in that the Argument Realization Principle has been incorporated into the first part (that is, the definition of lexical satisfaction). In addition, our binding theory and the Anaphoric Agreement Principle have been built in. This version of the definition also provides something that was only promised in Chapter 6: an initial symbol for our grammar. It says, in effect, that a well-formed stand-alone expression in English must be a finite sentence.

Section 9.2.5 lists the lexical rules that were presented in Chapter 8.

9.2 The Grammar So Far

The following pages contain a summary of the type hierarchy developed in the preceding chapters.[1]

[1]We use the notation 'list(τ)' to indicate a (possibly empty) list, all of whose members are of type τ.



9.2.1 Feature Declarations and Type Constraints

SOME GENERAL TYPES		
TYPE	FEATURES/CONSTRAINTS	IST
feat-struc		
synsem-struc	$\begin{bmatrix} \text{SYN} & \textit{gram-cat} \\ \text{SEM} & \textit{sem-struc} \end{bmatrix}$	*feat-struc*
phrase		*synsem-struc*
lex-item	$[\text{ARG-ST} \quad \text{list}(\textit{synsem-struc})]$	*synsem-struc*
word		*lex-item*
lexeme	$\begin{bmatrix} \text{SEM} & \begin{bmatrix} \text{MODE} & / \text{ none} \end{bmatrix} \end{bmatrix}$	*lex-item*
infl-lxm		*lexeme*
const-lxm		*lexeme*
noun-lxm	$\begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{noun} & \\ \text{AGR} & \begin{bmatrix} \text{PER} & / \text{ 3rd} \end{bmatrix} \\ \text{ANA} & / - \end{bmatrix} \end{bmatrix} \\ \text{ARG-ST} & / \langle \ \rangle \\ \text{SEM} & [\text{MODE} \quad / \text{ ref}] \end{bmatrix}$	*infl-lxm*
pron-lxm	$\begin{bmatrix} \ldots \end{bmatrix}$	*noun-lxm*
pn-lxm	$\begin{bmatrix} \text{SYN} & [\text{HEAD} \quad [\text{AGR} \quad [\text{NUM} \quad / \text{ sg}]]] \end{bmatrix}$	*noun-lxm*
cn-lxm	$\begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{HEAD} & [\text{AGR} \quad \boxed{1}] \\ \text{SPR} & \langle [\ ] \rangle \end{bmatrix} \\ \text{ARG-ST} & / \left\langle \begin{matrix} \text{DetP} \\ \begin{bmatrix} \text{AGR} & \boxed{1} \end{bmatrix} \end{matrix} \right\rangle \end{bmatrix}$	*noun-lxm*
verb-lxm	$\begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{HEAD} & \textit{verb} \\ \text{SPR} & \langle [\ ] \rangle \end{bmatrix} \\ \text{ARG-ST} & / \langle \text{ NP , ... } \rangle \\ \text{SEM} & [\text{MODE} \quad \text{prop}] \end{bmatrix}$	*infl-lxm*

MORE LEXEME TYPES		
TYPE	FEATURES/CONSTRAINTS	IST
iv-lxm		*verb-lxm*
siv-lxm	$\begin{bmatrix}\text{ARG-ST} & \langle\ [\ ]\ \rangle\end{bmatrix}$	*iv-lxm*
piv-lxm	$\begin{bmatrix}\text{ARG-ST} & \langle\ [\ ]\ ,\ \text{PP}\ \rangle\end{bmatrix}$	*iv-lxm*
tv-lxm	$\begin{bmatrix}\text{ARG-ST} & \langle\ [\ ]\ ,\ \text{NP}\ ,\ \ldots\ \rangle\end{bmatrix}$	*verb-lxm*
stv-lxm	$\begin{bmatrix}\text{ARG-ST} & \langle\ [\ ]\ ,\ [\ ]\ \rangle\end{bmatrix}$	*tv-lxm*
dtv-lxm	$\begin{bmatrix}\text{ARG-ST} & \langle\ [\ ]\ ,\ [\ ]\ ,\ \text{NP}\ \rangle\end{bmatrix}$	*tv-lxm*
ptv-lxm	$\begin{bmatrix}\text{ARG-ST} & \langle\ [\ ]\ ,\ [\ ]\ ,\ \text{PP}\ \rangle\end{bmatrix}$	*tv-lxm*
prep-lxm	$\begin{bmatrix}\text{SYN} & \begin{bmatrix}\text{HEAD} & \textit{prep}\end{bmatrix}\end{bmatrix}$	*const-lxm*
adj-lxm, *conj-lxm*, *det-lxm*,...	$\begin{bmatrix}\ldots\end{bmatrix}$	*const-lxm*
pdp-lxm	$\begin{bmatrix}\text{SYN} & \begin{bmatrix}\text{HEAD} & \begin{bmatrix}\text{MOD} & \begin{Bmatrix}\text{NOM,}\\ \text{VP,}\\ \text{none}\end{Bmatrix}\\ \text{P-OBJ} & \text{none}\end{bmatrix}\\ \text{SPR} & \langle[\ ]\rangle\end{bmatrix}\\ \text{ARG-ST} & \langle\text{NP, NP}\rangle\end{bmatrix}$	*prep-lxm*
mkp-lxm	$\begin{bmatrix}\text{SYN} & \begin{bmatrix}\text{HEAD} & \begin{bmatrix}\text{MOD} & \text{none}\\ \text{P-OBJ} & \boxed{1}\end{bmatrix}\\ \text{SPR} & \langle\ \rangle\end{bmatrix}\\ \text{ARG-ST} & \langle\ \boxed{1}\ \rangle\end{bmatrix}$	*prep-lxm*

OTHER GRAMMATICAL TYPES		
TYPE	FEATURES/CONSTRAINTS	IST
gram-cat	$\begin{bmatrix} \text{HEAD} & \textit{pos} \\ \text{COMPS} & \text{list}(\textit{synsem-struc}) \\ \text{SPR} & \text{list}(\textit{synsem-struc}) \end{bmatrix}$	*feat-struc*
pos	$\begin{bmatrix} \text{FORM} & \left\{\text{fin, to,...}\right\} \end{bmatrix}$	*feat-struc*
verb	$\begin{bmatrix} \text{...} \end{bmatrix}$	*pos*
prep	$\begin{bmatrix} \text{P-OBJ} & \left\{\textit{synsem-struc}, \text{none}\right\} \\ \text{MOD} & \left\{\begin{matrix} \text{NOM,} \\ \text{VP,} \\ \text{none,} \\ \text{...} \end{matrix}\right\} \end{bmatrix}$	*pos*
nominal	$\begin{bmatrix} \text{AGR} & \textit{agr-cat} \end{bmatrix}$	*pos*
noun	$\begin{bmatrix} \text{ANA} & \left\{+, -\right\} \\ \text{CASE} & \left\{\text{nom, acc}\right\} \end{bmatrix}$	*nominal*
det	$\begin{bmatrix} \text{COUNT} & \left\{+, -\right\} \end{bmatrix}$	*nominal*
adv	$\begin{bmatrix} \text{MOD} & \left\{\text{VP, none,...}\right\} \end{bmatrix}$	*pos*
adj	$\begin{bmatrix} \text{MOD} & \left\{\text{NOM, none}\right\} \end{bmatrix}$	*pos*
conj		*pos*

OTHER GRAMMATICAL TYPES		
TYPE	FEATURES/CONSTRAINTS	IST
agr-cat	$\begin{bmatrix} \text{PER} & \{\text{1st, 2nd, 3rd}\} \\ \text{NUM} & \{\text{sg, pl}\} \end{bmatrix}$	*feat-struc*
3sing	$\begin{bmatrix} \text{PER} & \text{3rd} \\ \text{NUM} & \text{sg} \\ \text{GEND} & \{\text{masc, fem, neut}\} \end{bmatrix}$	*agr-cat*
non-3sing		*agr-cat*
sem-struc	$\begin{bmatrix} \text{MODE} & \{\text{prop, ques, dir, ref, none}\} \\ \text{INDEX} & \textit{index} \\ \text{RESTR} & \text{list}(\textit{predication}) \end{bmatrix}$	*feat-struc*
predication	$\begin{bmatrix} \text{RELN} & \{\text{love,walk,...}\} \\ \text{...} & \end{bmatrix}$	*feat-struc*
index		*feat-struc*

9.2.2 The Basic Lexicon

Here are some sample lexical entries that are part of the basic lexicon. We have included only information that is not inherited from higher types. For example, the first entry does not contain a specification for the feature ANA, because it is of type *pron-lxm*, a subtype of *noun-lxm*, which has the specification [ANA / –]. As before, the notation '...' indicate things we haven't dealt with but which a complete grammar would have to.

9.2.2.1 Nouns

(1)

$$\left\langle \text{she}\ ,\ \begin{bmatrix} \textit{pron-lxm} & \\ \text{SYN} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{CASE} & \text{nom} \\ \text{AGR} & \begin{bmatrix} \text{NUM} & \text{sg} \\ \text{GEND} & \text{fem} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & i \\ \text{RESTR} & \left\langle \begin{bmatrix} \text{RELN} & \text{female} \\ \text{SIT} & s \\ \text{INSTANCE} & i \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \right\rangle$$

(2)

$$\left\langle \text{him} , \begin{bmatrix} \textit{pron-lxm} & \\ \text{SYN} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{CASE} & \text{acc} \\ \text{AGR} & \begin{bmatrix} \text{NUM} & \text{sg} \\ \text{GEND} & \text{masc} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & i \\ \text{RESTR} & \left\langle \begin{bmatrix} \text{RELN} & \text{male} \\ \text{SIT} & s \\ \text{INSTANCE} & i \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \right\rangle$$

(3)

$$\left\langle \text{themselves} , \begin{bmatrix} \textit{pron-lxm} & \\ \text{SYN} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{CASE} & \text{acc} \\ \text{ANA} & + \\ \text{AGR} & \begin{bmatrix} \text{NUM} & \text{pl} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & i \\ \text{RESTR} & \left\langle \begin{bmatrix} \text{RELN} & \text{group} \\ \text{SIT} & s \\ \text{INSTANCE} & i \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \right\rangle$$

(4)

$$\left\langle \text{Kim} , \begin{bmatrix} \textit{pn-lxm} & \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & i \\ \text{RESTR} & \left\langle \begin{bmatrix} \text{RELN} & \text{name} \\ \text{SIT} & s \\ \text{NAME} & \text{Kim} \\ \text{NAMED} & i \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \right\rangle$$

(5)

$$\left\langle \text{book} , \begin{bmatrix} \textit{cn-lxm} & \\ \text{ARG-ST} & \langle\, [\text{COUNT} +] \,\rangle \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & i \\ \text{RESTR} & \left\langle \begin{bmatrix} \text{RELN} & \text{book} \\ \text{SIT} & s \\ \text{INSTANCE} & i \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \right\rangle$$

9.2.2.2 Verbs

(6)

$$\left\langle \text{die} , \begin{bmatrix} \textit{siv-lxm} & \\ \text{ARG-ST} & \langle [\ \ ]_i \rangle \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & s \\ \text{RESTR} & \left\langle \begin{bmatrix} \text{RELN} & \text{die} \\ \text{SIT} & s \\ \text{CORPSE} & i \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \right\rangle$$

(7)

$$\left\langle \text{love} , \begin{bmatrix} \textit{stv-lxm} & \\ \text{ARG-ST} & \langle [\ \ ]_i , [\ \ ]_j \rangle \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & s \\ \text{RESTR} & \left\langle \begin{bmatrix} \text{RELN} & \text{love} \\ \text{SIT} & s \\ \text{LOVER} & i \\ \text{LOVED} & j \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \right\rangle$$

(8)

$$\left\langle \text{give} , \begin{bmatrix} \textit{dtv-lxm} & \\ \text{ARG-ST} & \langle [\ \ ]_i , [\ \ ]_j , [\ \ ]_k \rangle \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & s \\ \text{RESTR} & \left\langle \begin{bmatrix} \text{RELN} & \text{give} \\ \text{SIT} & s \\ \text{GIVER} & i \\ \text{GIVEN} & j \\ \text{GIFT} & k \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \right\rangle$$

(9)

$$\left\langle \text{give} , \begin{bmatrix} \textit{ptv-lxm} & \\ \text{ARG-ST} & \langle [\ \ ]_i , [\ \ ]_k , [\text{P-OBJ}\ \ \text{NP}_j] \rangle \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & s \\ \text{RESTR} & \left\langle \begin{bmatrix} \text{RELN} & \text{give} \\ \text{SIT} & s \\ \text{GIVER} & i \\ \text{GIVEN} & j \\ \text{GIFT} & k \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \right\rangle$$

9.2.2.3 Miscellaneous

(10)
$$\left\langle \text{the} , \begin{bmatrix} \textit{det-lxm} \\ \text{ARG-ST} & \langle\ \rangle \\ \text{SEM} & [...] \end{bmatrix} \right\rangle$$

(11)
$$\left\langle \text{few} , \begin{bmatrix} \textit{det-lxm} \\ \text{SYN} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{COUNT} & + \\ \text{AGR} & \begin{bmatrix} \textit{non-3sing} \\ \text{NUM} & \text{pl} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{ARG-ST} & \langle\ \rangle \\ \text{SEM} & [...] \end{bmatrix} \right\rangle$$

(12)
$$\left\langle \text{'s} , \begin{bmatrix} \textit{det-lxm} \\ \text{ARG-ST} & \langle\ \text{NP}\ \rangle \\ \text{SEM} & [...] \end{bmatrix} \right\rangle$$

(13) a.
$$\left\langle \text{to} , \begin{bmatrix} \textit{mkp-lxm} \\ ... \end{bmatrix} \right\rangle$$

b.
$$\left\langle \text{in} , \begin{bmatrix} \textit{pdp-lxm} \\ \text{ARG-ST} & \langle \text{NP}_i, \text{NP}_j \rangle \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & s \\ \text{RESTR} & \left\langle \begin{bmatrix} \text{RELN} & \text{in} \\ \text{SIT} & s \\ \text{CONTAINER} & j \\ \text{CONTAINED} & l \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \right\rangle$$

(14)
$$\left\langle \text{and} , \begin{bmatrix} \textit{conj-lxm} \\ \text{ARG-ST} & \langle\ \rangle \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & s \\ \text{RESTR} & \left\langle \begin{bmatrix} \text{RELN} & \text{and} \\ \text{SIT} & s \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \right\rangle$$

9.2.3 The Grammar Rules

- Head-Complement Rule

$$\begin{bmatrix} \textit{phrase} & \\ \text{COMPS} & \langle\ \rangle \end{bmatrix} \rightarrow \text{H}\begin{bmatrix} \textit{word} & \\ \text{COMPS} & \left\langle \boxed{1},\ldots,\boxed{n} \right\rangle \end{bmatrix} \boxed{1} \ldots \boxed{n}$$

- Head-Specifier Rule

$$\begin{bmatrix} \textit{phrase} & \\ \text{SPR} & \langle\ \rangle \end{bmatrix} \rightarrow \boxed{1} \quad \text{H}\begin{bmatrix} \textit{phrase} & \\ \text{SPR} & \langle\ \boxed{1}\ \rangle \end{bmatrix}$$

- Head-Modifier Rule

$$[\textit{phrase}] \rightarrow \text{H}\boxed{1}[\textit{phrase}] \quad \begin{bmatrix} \textit{phrase} & \\ \text{MOD} & \boxed{1} \end{bmatrix}$$

- Coordination Rule

$$\begin{bmatrix} \text{SYN} & \boxed{0} \\ \text{IND} & s_0 \end{bmatrix} \rightarrow$$

$$\begin{bmatrix} \text{SYN} & \boxed{0} \\ \text{IND} & s_1 \end{bmatrix} \ldots \begin{bmatrix} \text{SYN} & \boxed{0} \\ \text{IND} & s_{n-1} \end{bmatrix} \begin{bmatrix} \text{HEAD} & \textit{conj} \\ \text{IND} & s_0 \\ \text{RESTR} & \left\langle \left[\text{ARGS}\ \langle s_1 \ldots s_n \rangle \right] \right\rangle \end{bmatrix} \begin{bmatrix} \text{SYN} & \boxed{0} \\ \text{IND} & s_n \end{bmatrix}$$

- Imperative Rule

$$\begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \begin{bmatrix} \textit{verb} & \\ \text{FORM} & \text{fin} \end{bmatrix} \\ \text{SPR} & \langle\ \rangle \\ \text{COMPS} & \langle\ \rangle \\ \text{SEM} & \begin{bmatrix} \text{MODE} & \text{dir} \\ \text{INDEX} & s \end{bmatrix} \end{bmatrix} \rightarrow \begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \begin{bmatrix} \textit{verb} & \\ \text{FORM} & \text{inf} \end{bmatrix} \\ \text{SPR} & \left\langle \text{NP}\left[\text{PER} \quad \text{2nd}\right] \right\rangle \\ \text{SEM} & \left[\text{INDEX} \quad s\right] \end{bmatrix}$$

9.2.4 Well-Formedness Definitions

In the following definitions, we take for granted the notions of resolved feature structure and tree structure.[2] We also presuppose the notion of feature structure satisfaction.[3]

(15) Well-Formed Tree Structure

Φ is a well-formed tree structure just in case (1) every local subtree in Φ satisfies some lexical entry η or some grammar rule ρ and (2) Φ satisfies the Anaphoric Agreement Principle.

(16) Anaphoric Agreement Principle
Coindexed elements agree (i.e. share a single AGR value).

(17) Lexical Satisfaction
A word structure:

$$\begin{array}{c} F \\ | \\ \omega \end{array}$$

satisfies a lexical entry $\langle \omega, \delta \rangle$ just in case:

1. F is of type *word* and F satisfies δ,
2. Argument Realization Principle
   F satisfies the following feature structure description:

$$\begin{bmatrix} \text{SYN} & \begin{bmatrix} \text{SPR} & \boxed{a} \\ \text{COMPS} & \boxed{b} \end{bmatrix} \\ \text{ARG-ST} & \boxed{a} \oplus \boxed{b} \end{bmatrix}$$

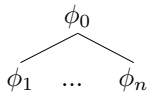
   and
3. Case Constraint
   Any NP in a noninitial position of F's ARG-ST list is [CASE acc].

[2]As noted in Chapter 6, a resolved feature structure ϕ is a directed graph of a particular kind, where: (i) ϕ is assigned to a maximally specific type τ, (ii) ϕ is specified for every feature F that is appropriate for τ and (iii) ϕ's value for each such feature F is maximally specific. A tree structure is a configuration of nodes that satisfies a number of conditions, e.g. (i) that each node have only one mother, (ii) each nonterminal node is a resolved feature structure, and (iii) each terminal node is a lexical form.

[3]Intuitively, a resolved feature structure ϕ satisfies a feature structure description δ when all the information specified in δ is true of ϕ.

(18) Phrasal Satisfaction

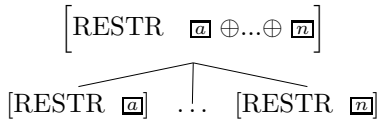
A local subtree $\Phi =$



satisfies a grammar rule $\rho = \delta_0 \rightarrow \delta_1 \ldots \delta_n$ just in case:[4]

1. the sequence $\langle \phi_0, \phi_1, ...\phi_n \rangle$ satisfies the description $\langle \delta_0, \delta_1, \ldots \delta_n \rangle$,
2. Φ satisfies the Semantic Compositionality Principle and the Binding Theory,
3. if ρ is a headed rule, then Φ satisfies the Head Feature Principle and the Semantic Inheritance Principle,
4. if ρ is a headed rule other than the Head-Complement Rule, then Φ satisfies Part A of the Valence Principle, and
5. if ρ is a headed rule other than the Head-Specifier Rule, then Φ satisfies Part B of the Valence Principle.

(19) Semantic Compositionality Principle



(20) The Binding Theory

Principle A: An [ANA +] synsem-struc must be outranked by a coindexed synsem-struc.
Principle B: An [ANA −] synsem-struc must not be outranked by a coindexed synsem-struc.

[4]Note that clause (1) here must speak of a sequence of resolved feature structures satisfying a sequence description. This is because of identities that must hold across members of the sequence, e.g. those required by the rules themselves or by the principles given below.

(21) Head Feature Principle

$$\begin{array}{c} [\text{HEAD } \boxed{1}] \\ \ldots \quad \delta_h \quad \ldots \\ \begin{bmatrix} \text{HEAD} & \boxed{1} \end{bmatrix} \end{array}$$

(22) Semantic Inheritance Principle

$$\begin{array}{c} \begin{bmatrix} \text{MODE} & \boxed{4} \\ \text{INDEX} & \boxed{5} \end{bmatrix} \\ \ldots \quad \delta_h \quad \ldots \\ \begin{bmatrix} \text{MODE} & \boxed{4} \\ \text{INDEX} & \boxed{5} \end{bmatrix} \end{array}$$

(23) Valence Principle

A

$$\begin{array}{c} \begin{bmatrix} \text{COMPS} & \boxed{b} \end{bmatrix} \\ \ldots \quad \delta_h \quad \ldots \\ \begin{bmatrix} \text{COMPS} & \boxed{b} \end{bmatrix} \end{array}$$

B

$$\begin{array}{c} \begin{bmatrix} \text{SPR} & \boxed{c} \end{bmatrix} \\ \ldots \quad \delta_h \quad \ldots \\ \begin{bmatrix} \text{SPR} & \boxed{c} \end{bmatrix} \end{array}$$

(24) Initial Symbol Definition

A CFG must include a statement of which category (or categories) can start a tree derivation. The analogue of this 'initial symbol' definition in our grammar is a constraint guaranteeing that for a phrase to stand alone, its synsem-struc must satisfy the following description:

$$\begin{bmatrix} \textit{phrase} & \\ \text{SYN} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{verb} & \\ \text{FORM} & \text{fin} \end{bmatrix} \\ \text{SPR} & \langle\ \rangle \\ \text{COMPS} & \langle\ \rangle \end{bmatrix} \end{bmatrix}$$

9.2.5 Lexical Rules

(25) Singular Noun Lexical Rule

$$\left\langle \boxed{1}\ ,\ \begin{bmatrix} \textit{noun-lxm} \end{bmatrix} \right\rangle \ \Rightarrow$$

$$\left\langle \boxed{1}\ ,\ \begin{bmatrix} \textit{word} & \\ \text{SYN} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{AGR} & \begin{bmatrix} \text{NUM} & \text{sg} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle$$

(26) Plural Noun Lexical Rule

$$\left\langle \boxed{1}\ ,\ \begin{bmatrix} \textit{noun-lxm} & \\ \text{ARG-ST} & \langle\ [\text{COUNT}\ +]\ \rangle \end{bmatrix} \right\rangle \ \Rightarrow$$

$$\left\langle \mathrm{F}_{NPL}(\boxed{1})\ ,\ \begin{bmatrix} \textit{word} & \\ \text{SYN} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{AGR} & \begin{bmatrix} \text{NUM} & \text{pl} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle$$

(27) 3rd-Singular Verb Lexical Rule

$$\left\langle \boxed{3}, \begin{bmatrix} \textit{verb-lxm} & \\ \text{SEM} & \begin{bmatrix} \text{RESTR} & \boxed{a} \end{bmatrix} \end{bmatrix} \right\rangle \Rightarrow$$

$$\left\langle \mathrm{F}_{3SG}(\boxed{3}), \begin{bmatrix} \textit{word} & \\ \text{SYN} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{FORM} & \text{fin} \end{bmatrix} \end{bmatrix} \\ \text{ARG-ST} & \left\langle \begin{bmatrix} \text{CASE} & \text{nom} \\ \text{AGR} & \textit{3sing} \end{bmatrix}, \ldots \right\rangle \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & \boxed{2} \\ \text{RESTR} & \boxed{a} \oplus \left\langle \begin{bmatrix} \text{RELN} & \text{t-overlap} \\ \text{ARG1} & \boxed{2} \\ \text{ARG2} & \text{now} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \right\rangle$$

(28) Non-3rd-Singular Verb Lexical Rule

$$\left\langle \boxed{3}, \begin{bmatrix} \textit{verb-lxm} & \\ \text{SEM} & \begin{bmatrix} \text{RESTR} & \boxed{a} \end{bmatrix} \end{bmatrix} \right\rangle \Rightarrow$$

$$\left\langle \boxed{3}, \begin{bmatrix} \textit{word} & \\ \text{SYN} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{FORM} & \text{fin} \end{bmatrix} \end{bmatrix} \\ \text{ARG-ST} & \left\langle \begin{bmatrix} \text{CASE} & \text{nom} \\ \text{AGR} & \textit{non-3sing} \end{bmatrix}, \ldots \right\rangle \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & \boxed{2} \\ \text{RESTR} & \boxed{a} \oplus \left\langle \begin{bmatrix} \text{RELN} & \text{t-overlap} \\ \text{ARG1} & \boxed{2} \\ \text{ARG2} & \text{now} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \right\rangle$$

(29) Past-Tense Verb Lexical Rule

$$\left\langle \boxed{2}, \begin{bmatrix} \textit{verb-lxm} & \\ \text{SEM} & \begin{bmatrix}\text{RESTR} & \boxed{a}\end{bmatrix} \end{bmatrix} \right\rangle \Rightarrow$$

$$\left\langle \mathrm{F}_{PAST}(\boxed{2}), \begin{bmatrix} \textit{word} & \\ \text{SYN} & \begin{bmatrix}\text{HEAD} & \begin{bmatrix}\text{FORM} & \text{fin}\end{bmatrix}\end{bmatrix} \\ \text{ARG-ST} & \left\langle \begin{bmatrix}\text{CASE} & \text{nom}\end{bmatrix}, \ldots \right\rangle \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & \boxed{3} \\ \text{RESTR} & \boxed{a} \oplus \left\langle \begin{bmatrix} \text{RELN} & \text{t-precede} \\ \text{ARG1} & \boxed{3} \\ \text{ARG2} & \text{now} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \right\rangle$$

(30) Present Participle Lexical Rule

$$\left\langle \boxed{3}, \begin{bmatrix} \textit{verb-lxm} & \\ \text{SEM} & \begin{bmatrix}\text{RESTR} & \boxed{a}\end{bmatrix} \end{bmatrix} \right\rangle \Rightarrow$$

$$\left\langle \mathrm{F}_{PRP}(\boxed{3}), \begin{bmatrix} \textit{word} & \\ \text{SYN} & \begin{bmatrix}\text{HEAD} & \begin{bmatrix}\text{FORM} & \text{prp}\end{bmatrix}\end{bmatrix} \\ \text{SEM} & \begin{bmatrix}\text{RESTR} & \boxed{a} \oplus \ldots \end{bmatrix} \end{bmatrix} \right\rangle$$

(31) Past Participle Lexical Rule

$$\left\langle \boxed{2}, \begin{bmatrix} \textit{verb-lxm} & \\ \text{SEM} & \begin{bmatrix}\text{RESTR} & \boxed{a}\end{bmatrix} \end{bmatrix} \right\rangle \Rightarrow$$

$$\left\langle \mathrm{F}_{PSP}(\boxed{2}), \begin{bmatrix} \textit{word} & \\ \text{SYN} & \begin{bmatrix}\text{HEAD} & \begin{bmatrix}\text{FORM} & \text{psp}\end{bmatrix}\end{bmatrix} \\ \text{SEM} & \begin{bmatrix}\text{RESTR} & \boxed{a} \oplus \ldots \end{bmatrix} \end{bmatrix} \right\rangle$$

(32) Infinitival Lexical Rule

$$\left\langle \boxed{2}\,,\ \begin{bmatrix} \textit{verb-lxm} & \\ \text{SEM} & \begin{bmatrix}\text{RESTR} & \boxed{a}\end{bmatrix} \end{bmatrix} \right\rangle \ \Rightarrow$$

$$\left\langle \boxed{2}\,,\ \begin{bmatrix} \textit{word} & \\ \text{SYN} & \begin{bmatrix}\text{HEAD} & \begin{bmatrix}\text{FORM} & \text{inf}\end{bmatrix}\end{bmatrix} \\ \text{SEM} & \begin{bmatrix}\text{RESTR} & \boxed{a} \oplus \ldots\end{bmatrix} \end{bmatrix} \right\rangle$$

(33) Constant Lexeme Lexical Rule

$$\langle\, \boxed{1}\,,\ [\textit{const-lxm}]\,\rangle \quad \Rightarrow \quad \langle\, \boxed{1}\,,\ [\textit{word}]\,\rangle$$

(34) Agent Nominalization Lexical Rule

$$\left\langle \boxed{2}\,,\ \begin{bmatrix} \textit{verb-lxm} & \\ \text{ARG-ST} & \langle\, \text{NP}_i\ (,\ \boxed{1}\text{NP})\ \rangle \\ \text{SEM} & \begin{bmatrix}\text{INDEX} & s\end{bmatrix} \end{bmatrix} \right\rangle \ \Rightarrow$$

$$\left\langle \text{F}_{-er}(\boxed{2})\,,\ \begin{bmatrix} \textit{cn-lxm} & \\ \text{ARG-ST} & \left\langle \text{DetP}\ (,\ \begin{bmatrix}\text{P-OBJ} & \boxed{1} \\ \text{FORM} & \text{of}\end{bmatrix}^{\text{PP}})\right\rangle \\ \text{SEM} & \begin{bmatrix}\text{INDEX} & i\end{bmatrix} \end{bmatrix} \right\rangle$$

9.3 Constraint-Based Lexicalism

We turn now to some reflections on the relationship between the sort of grammatical descriptions in this text and what is known about the mental processes underlying human language comprehension and production. We believe it is possible to draw at least preliminary conclusions about how the flexibility and incrementality of human language processing – language comprehension in particular – should inform the design of grammars that aspire to be models of human linguistic knowledge. Adopting the familiar terminology of Chomsky (1965), we will suggest that competence grammars, if they are ever to be directly embedded within realistic models of how language is used (what Chomsky would call linguistic 'performance'), should embody certain design properties that make them consistent with these basic facts about processing.

Let us start with three basic observations about the grammar we have been developing:

1. It is SURFACE ORIENTED. Our grammar (like standard context-free grammars) provides a reasonably simple structure that is directly associated with the string of words that constitute each sentence. The ancillary structure that has to be computed to ascertain whether a given sentence is grammatical expresses information that is straightforwardly derivable from properties of the words in the string. No additional abstract structures are posited.
2. It is CONSTRAINT-BASED. There are no operations that destructively modify any representations. The principles of the theory, the grammar rules, and the lexical entries that result once the grammar is 'closed' under lexical rules are all just constraints that interact so as to define a set of phrase structures – those that simultaneously satisfy the relevant constraints of our grammar.
3. It is STRONGLY LEXICALIST. We have localized most grammatical and semantic information within lexical entries. These lexical entries furthermore correspond directly to the words present in the sentence, which can be viewed as the key elements that drive the construction of the syntactic and semantic structure of the sentence.

Each of these three design properties, we believe, plays an important role when we turn to embedding our theory of linguistic competence within a theory of performance – specifically a theory of how language is processed. Any theory that has these design properties exemplifies a viewpoint that we will refer to as CONSTRAINT-BASED LEXICALISM (CBL).

9.4 Incremental Processing

We don't have to venture into a psycholinguistic laboratory to convince ourselves that language processing is highly incremental. We saw this already in Chapter 1, when we considered examples like (35).

(35) After finding the book on the atom, Sandy went into class, confident that there would be no further obstacles to getting that term paper done.

When we hear such a sentence, we process it as it comes – more or less word by word – building structure and partial interpretation incrementally, using what nonlinguistic information we can to make the right decisions at certain points, for example, when we encounter the PP *on the atom* and have to decide whether it modifies VP or NOM. We make

this decision 'on-line' it seems, using a plausibility assessment of the meaning that would result from each structure.

Psycholinguists have shown us that sentence processing sometimes goes astray. Garden path examples like (36a,b) are as remarkable today as they were when they were first brought to the attention of language researchers.[5]

(36) a. The horse raced past the barn fell.
b. The boat floated down the river sank.

On first encountering such sentences, almost all English speakers judge them to be totally ungrammatical. However, after seeing them juxtaposed to fully well-formed examples like (37), speakers recognize that sentences like (36) are grammatical, though very hard to process.

(37) a. The horse that was raced past the barn fell.
b. The horse taken to the hospital died.
c. The boat that was floated down the river sank.
d. The boat seen down the river sank.

Experimental researchers thought at first that these garden paths showed that certain purely linguistic processing strategies (like trying to build an S out of the NP *the horse* and a VP beginning with *raced past*) were automatic - virtually impossible to turn off. But modern psycholinguistics has a very different story to tell.

First, note that in the right context, one can eliminate the garden path effect even with the sentences in (36). The right context can even make the NOM-modifying interpretation of *raced past the barn* the most natural one:[6]

(38) The horse that they raced around the track held up fine. The horse that was raced down the road faltered a bit. And the horse raced past the barn fell.

The context here highlights the need to identify one horse among many, which in turn favors the meaning of the NOM-modifying structure of (36a).

Moreover, if we keep the same potential for ambiguity, but change the words, we can eliminate the garden path effect even without an elaborate preceding context. Consider examples like (39a,b).

(39) a. The evidence examined by the judge turned out to be unreliable.
b. The thief arrested by the police turned out to be our nephew.

[5]By Bever (1970).
[6]This kind of effect is discussed by Crain and Steedman (1985).

As shown in a number of studies,[7] examples like these present no more processing difficulty than their unambiguous counterparts in (40).

(40) a. The evidence that was examined by the judge turned out to be unreliable.

b. The thief who was arrested by the police turned out to be our nephew.

That is, the examples in (39), even in the absence of a prior biasing context, do not cause garden path effects.

The explanation for this difference lies in the relevant nonlinguistic information. Evidence (or, say, a particular piece of evidence) is a pretty implausible examiner and the sentence built out of a subject NP *the evidence* and a VP headed by *examined* would require precisely that interpretation. (Similar claims hold for (39b), given that a thief is unlikely to be the one doing the arresting.) That is, it is a fact about the world that examiners are animate, and since evidence is inanimate, that hypothesis about the interpretation of the sentence is implausible. The fact that the decision to reject that interpretation (and hence the associated sentential structure) is made so quickly as to be imperceptible (i.e. so as to produce no noticeable garden path effect) is evidence that language comprehension is working in a highly integrative and incremental fashion. Linguistic and nonlinguistic constraints on the interpretation are interleaved in real time.

Language understanding appears to be a process of constraint satisfaction. Competing interpretations exist in parallel, but are active to varying degrees. A particular alternative interpretation is active to the extent that evidence is available to support it as the correct interpretation of the utterance being processed. Note, by the way, that frequency has a very important role to play here. The reason the *horse raced past the barn* example is such a strong garden path is that *raced* occurs much more frequently as a finite verb form than as the passive participle of the transitive use of *race*, which is precisely what the NOM-modifying reading requires. Ambiguity resolution is a continuous process, where inherent degrees of activation, for example, those correlating with gross frequency, fluctuate as further evidence for particular interpretations become available. Such evidence may in principle stem from any aspect of the sentence input or the local or discourse context. A garden-path sentence is one that has an interpretation strongly supported by initial evidence that later turns out to be incorrect.

[7]See, for example, Trueswell, Tanenhaus, and Garnsey (1992), Pearlmutter and MacDonald (l992), and Tabossi et al. (1994).

9.5 A Performance-Plausible Competence Grammar

9.5.1 Surface-Orientation

We know that sentence processing proceeds in a more or less word-by-word (or even syllable-by-syllable) fashion. In comprehending utterances, hearers use their knowledge of language to build partial hypotheses about the intended meaning. These hypotheses become more or less active, depending on how plausible they are, that is, depending on how well their meaning squares with the hearers' understanding of what's going on in the discourse.

Sometimes the process even gets short-circuited. We have all had the experience of completing someone else's utterance (a phenomenon that is, incidentally, far more common than one might imagine, as shown, e.g. by Wilkes-Gibbs (1986)) or of having to wait for someone to finish an utterance whose completion had already been made obvious by context. One striking example of this is English 'echo questions', as illustrated in the following kind of dialogue:[8]

(41)

[Speaker A:] Senora Maria Consuelo Bustamante y Bacigalupo is coming to dinner tomorrow night.

[Speaker B:]

ght?

o did you say is coming to dinner tomorrow ni

h

w

* * * * * * * * * * * * * * * * * * *

In a dialogue like this, it is quite likely that Speaker A may comprehend the intent of Speaker B's utterance well before it is complete, somewhere in the region indicated by the asterisks. Presumably, this is possible precisely because Speaker A can recognize that the remainder of B's utterance is a repetition of A's own utterance and can graft that bit of content onto the partial analysis A has performed through word-by-word processing of B's utterance. What examples like this show is that a partial linguistic analysis (e.g. the partial linguistic analysis of *who did you*, *who did you say* or *who did you say is*) is constructed incrementally, assigned a (partial) interpretation, and integrated with information from the context to produce an interpretation of a complete utterance even before the utterance is complete. Amazing, if you think about it!

So if a grammar is to be realistic, that is, if it is to be directly embedded in a model of this kind of incremental and integrative language

[8]The relative height of the type here is meant to indicate variation in fundamental frequency, i.e. pitch.

processing, then it needs to characterize linguistic knowledge in a way that allows for the efficient incremental computation of partial analyses. Moreover, the partial grammatical analyses have to be keyed in to partial linguistic meanings, because these are what interacts with other factors in processing.

The kind of grammar we are developing seems quite compatible with these performance-driven design criteria. The lexical information that comes with each word provides information about the structure of the sentence directly, that is, about the phrases that the words are part of and about the neighboring phrases that they combine with syntactically. In addition, the words of our grammar provide partial information about the meaning of those phrases, and hence, since all phrases are built up directly from the component words and phrases in a context-free manner, there is useful partial semantic information that can be constructed incrementally, using our surface-oriented grammar.

Of course this is no rigorous proof that there is a precise performance model based on a grammar like ours, but the context-free-like architecture of the theory and the hybrid syntactic-semantic nature of the lexical data structures are very suggestive. Incremental computation of partial semantic structures, the key to modeling integrative sentence processing, seems to fit in well with our grammar.

9.5.2 Constraint-Based Grammar

In addition to the incremental and integrative nature of human language processing, we can also observe that there is no fixed order in which particular kinds of information are considered. For example, it is not the case that syntactic information (e.g. agreement information that might rule out a particular parse) is always consulted before semantic information (e.g. semantic incompatibility that would favor or disfavor some potential interpretation of an utterance). In fact, it is possible to make an even stronger claim. In examples like (42), early accessing of morphological information allows the number of sheep under discussion to be determined incrementally, and well before the nonlinguistic knowledge necessary to select the 'fenced enclosure' sense of *pen*, rather than its 'writing implement' sense.

(42) The sheep that was sleeping in the pen stood up.

In (43), on the other hand, the relevant information about the world – that sheep might fit inside a fenced enclosure, but not inside a writing implement – seems to be accessed well before the relevant morphological information constraining the number of sheep.[9]

[9]This pair of examples is due to Martin Kay.

(43) The sheep in the pen had been sleeping and were about to wake up.

So the information accessed in on-line language processing is typically made available in an order determined by the input stream, not by the constructs of any grammatical theory. In comprehending these examples, for example, a hearer accesses morphological information earlier in (42) and later in (43) precisely because the order of access is tied fairly directly to the order of the words being processed. A theory positing a fixed order of access – for example, one that said all strictly linguistic processing must be completed before nonlinguistic knowledge could be brought to bear on utterance interpretation – would not be able to account for the contrast between (42) and (43).

Finally, we know that for the most part linguistic information functions fairly uniformly in many diverse kinds of processing activity, including comprehension, production, translation, playing language games, and the like. By 'fairly uniformly' we mean that the set of sentences reliably producible[10] by a given speaker-hearer is similar – in fact bears a natural relation (presumably proper inclusion) – to, the set of sentences that that speaker-hearer can comprehend. This might well have been otherwise. That there is so close and predictable a relation between the production activity and the comprehension activity of any given speaker of a natural language militates strongly against any theory on which the production grammar is independent from the comprehension grammar, for instance. This simple observation suggests rather that the differences between, say, comprehension and production should be explained by a theory that posits distinct processing regimes making use of a single linguistic description. And that description should therefore be a process-neutral grammar of the language, which can serve each kind of process that plays a role in on-line linguistic activity.[11]

Observations of this sort, namely, that linguistic descriptions are used in an order-independent fashion, lead naturally to the conclusion that the constructs of linguistic theory should have an order-independent character. That is, the architecture of grammar should be such that linguistic constraints can be independently deployed in different situa-

[10]That is, sentences short enough to utter in a real language-use situation. We also intend to rule out production errors.

[11]The fact that comprehension extends beyond systematic production can be explained in terms of differences of process - not differences of grammar. Speakers that stray far from the grammar of their language run a serious risk of not being understood; yet hearers that allow grammatical principles to relax when necessary will understand more than those that don't. There is thus a deep functional motivation for the two kinds of processing to differ as they appear to.

tions. Finally, given that linguistic knowledge is process-independent, there should be no bias within a linguistic theory – whether overt or hidden, intentional or inadvertent - toward one kind of processing, rather than another.

Grammars whose constructs are truly process-neutral, then hold the most promise for the development of processing models. And the best known way to ensure process-neutrality is to formulate a grammar as a declarative system of constraints. Such systems of constraints fit well into models of processing precisely because all the information they provide is in principle on an equal footing. What these observations add up to is a view of grammar as a set of constraints, each expressing partial information about linguistic structures, rather than a system employing destructive operations of any kind. Moreover, we have also seen that these constraints should exhibit certain further properties, such as order-independence, if performance-compatibility is to be achieved. The grammar we've been developing has just these design properties – all the constructs of the grammar (lexical entries, grammar rules, even our general principles) are nothing more than constraints that produce equivalent results no matter what order they are applied in.

9.5.3 Strong Lexicalism

Our theory partitions grammatical information into a number of components whose interaction determines the well-formedness of particular examples. By far the richest locus of such information, however, is the lexicon. Our grammar rules are simple in their formulation and general in their application, as are such aspects of our formal theory as the Head Feature Principle and the Valence Principle. Most of the details we need in order to analyze individual sentences are codified in the lexical entries (though much of it need not be stipulated, thanks to lexical rules and inheritance through the type hierarchy).

However, other divisions of grammatical labor are conceivable. Indeed, a number of theories with highly articulated rule systems and relatively impoverished lexicons have been developed in considerable detail (e.g. early transformational grammar and Generalized Phrase Structure Grammar, both of which are described briefly in Appendix B). We have argued for strong lexicalism on the basis of linguistic adequacy (along with general considerations of elegance and parsimony). It turns out, moreover, that the psycholinguistic evidence on language processing points in the same direction. Investigations of syntactic ambiguity resolution and garden path effects have shown that both phenomena are sensitive to a variety of types of information. That is, the difficulty listeners exhibit in resolving such ambiguities and overcoming garden paths

is influenced by factors other than syntactic structure. These include semantic compatibility and pragmatic plausibility, type and valence of the words involved, and the frequencies with which individual words occur in particular constructions (see Tanenhaus and Trueswell (1995) for a survey of relevant results).

For example, a sentence beginning with the sequence NP_1 - V - NP_2 can be continued in a number of ways. NP_2 could be the object of the verb, or it could be the subject of a complement sentence. This is illustrated in (44a), which can be continued as in (44b) or (44c).

(44) a. Lou forgot the umbrella ...
b. Lou forgot the umbrella was broken.
c. Lou forgot the umbrella in the closet.

Hence a listener or reader encountering (44a) must either postpone the decision about whether to attach the NP *the umbrella* to the VP, or decide prematurely and then have to reanalyze it later. Either way, this places a burden on the parser in at least some cases. Various experimental paradigms have been used to verify the existence of this parsing difficulty, including measuring reading times and tracking the eye movements of readers.

However, not all verbs that could appear in place of *forgot* in (44a) can appear in both of the contexts in (44b) and (44c). This is illustrated in (45).

(45) a. Lou hoped the umbrella was broken.
b. *Lou hoped the umbrella in the closet.
c. *Lou put the umbrella was broken.
d. Lou put the umbrella in the closet.

The increased parsing load in (44a) is reduced greatly when the valence of the verb allows for no ambiguity, as in (45). This has been demonstrated using the methods used to establish the complexity of the ambiguity in the first place (see Trueswell et al. (1993)). This provides strong evidence that people use lexical valence information incrementally as they process sentences.

Similarly, listeners use semantic and pragmatic information about the verb and the following NP to choose between possible attachment sites for the NP. For example, though *learn* may take either an NP object or a sentential complement, illustrated in (46), when the immediately following NP is not the sort of thing one can learn, people do not exhibit the level of complexity effects in parsing that show up in (44).

(46) a. Dana learned the umbrella was broken.
b. Dana learned a new theorem in class.

The same sort of effect of lexical meaning on parsing shows up with PP attachment ambiguities, like those in (47).

(47) a. The artist drew the child with a pencil.
b. Lynn likes the hat on the shelf.

In (47a), the pencil could be either the artist's instrument or something in the child's possession; in (47b), *on the shelf* could identify either Lynn's preferred location for the hat, or which hat it is that Lynn likes. The structural ambiguity of such sentences causes parsing complexity, but this is substantially mitigated when the semantics or pragmatics of the verb and/or noun strongly favors one interpretation, as in (48).

(48) a. The artist drew the child with a bicycle.
b. Lynn bought the hat on the shelf.

Furthermore, the information that we have been led to posit in our lexical entries has independently been found to play a role in language processing. After reviewing a number of studies on the factors that influence syntactic ambiguity resolution, MacDonald et al. (1994) discuss what information they believe needs to be lexically specified to account for the psycholinguistic results. Their list includes:

- valence;
- 'coarse-grained semantic information' (i.e. the sort of information about who did what to whom that is given in our SEM feature); and
- 'grammatically relevant features' such as 'tense..., finiteness..., voice (active or passive), number..., person..., and gender...'.

They also mention grammatical category, which we represent in our lexical entries by means of types (specifically, the subtypes of *pos*). In short, the elements MacDonald et al. list correspond remarkably well to the information that we list in our lexical entries.

9.6 Universal Grammar: A Mental Organ?

In the preceding sections we have argued that the design features of our grammatical theory comport well with existing evidence about how people process language. There is yet another psycholinguistic consideration that has played a central role in much work in generative grammar, namely, learnability. In this section, we briefly address the question of evaluating our theory by this criterion.

As noted in Chapter 1, Chomsky has argued that the most remarkable fact about human language – and the one he thinks linguists should be primarily concerned with explaining – is that virtually all children become fluent speakers of a language, with little apparent effort or in-

struction. The puzzle, as Chomsky sees it, is how people can come to know so much about language so quickly and easily. His solution in a nutshell is that people's knowledge of language is for the most part innate, not learned. This entails that much linguistic structure – namely, those aspects that are innate – must be common to all languages. Consequently, a central goal of much work in modern syntactic theory has been to develop a conception of universal grammar rich enough to permit the descriptions of particular languages to be as simple as possible.

Chomsky's strong claims about the role of innate knowledge in language acquisition are by no means uncontroversial among developmental psycholinguists. In particular, many scholars disagree with his position that the human language faculty is highly task-specific – that is, that people are born with a 'mental organ' for language which is distinct in its organization and functioning from other cognitive abilities (see, e.g. Bates and MacWhinney (1989) and Tomasello (1992) for arguments against Chomsky's position).

There can be little doubt that biology is crucial to the human capacity for language; if it were not, family pets would acquire the same linguistic competence as the children they are raised with. It is far less clear, however, that the human capacity for language is as specialized as Chomsky says. A range of views on this issue are possible. At one end of the spectrum is the idea that the language faculty is a fully autonomous module, unrelated to general cognitive capacity. At the other end is the idea that there are no specifically linguistic abilities – that our capacity to learn language arises essentially as a side-effect of our general intelligence or of other abilities. Chomsky's view is close to the former; Tomasello (1992) argues for something close to the latter. Other scholars have defended views somewhere in between.

The participants in this debate often seem to be talking past one another. Opponents of task-specificity tend to take a simplistic view of linguistic structure, emphasizing basic communicative functions while ignoring the intricacies of syntax that are the bread and butter of generative grammar. On the other hand, proponents of task-specificity have a tendency to leap from the complexity of their analyses to the conclusion that the knowledge involved must be innate and unique to language.

We find much of the argumentation on both sides of this controversy unconvincing, and hence we take no position in this book. Nevertheless, the theory presented here can contribute to its resolution. By making syntactic and semantic analyses explicit, we make it possible to formulate more precisely what is at issue in the debate over task-specificity. Moreover, providing formal representations of our data structures and their interactions permits us to see more clearly where there could be

analogues in other cognitive domains. Our position is that the grammatical constructs we have been developing in this text are well suited to a theory of universal grammar, whether or not that theory turns out to be highly task-specific, and that the explicitness of our proposals can be helpful in resolving the task-specificity question.

To justify this claim, we will consider various components of our theory, namely: the phrase structure rules, the features and their values, the type hierarchy with its feature declarations and constraints, the definition of phrasal satisfaction (incorporating the Head Feature Principle, the Valence Principle, and the two semantic principles), the binding theory, and the lexical rules. We will find that most of these have elements that are very likely universal, and that our formulations do not prejudge the issue of task-specificity.

- Our grammar rules (with the exception of the Imperative Rule) are sufficiently general that, aside from their linear ordering of the constituents, they are natural candidates for universality. It would not be hard to factor out the ordering, so that versions of these rules could be posited as part of universal grammar. The sort of hierarchical structure induced by the rules, which we represent with trees, is arguably not unique to language: it also seems appropriate, for example, to aspects of mathematical reasoning. On the other hand, the concepts of 'head', 'complement', 'specifier', and 'modifier', which are crucial to our formulation of the rules, appear to be specialized to language. If it should turn out, however, that they can be shown to be instances of some more generally applicable cognitive relations, this would in no way undermine our analysis.
- Most of the features we have posited have obvious cross-linguistic application. It seems at least plausible that a more fully worked out version of the theory presented here could include an inventory of features from which the feature structures of all languages must be constructed. In later chapters, we will identify the values of some features with particular English words, a practice inconsistent with saying that the set of possible feature values is part of universal grammar. It might be possible, however, to restrict feature values to come from either the set of morphological forms of the language or a universally specifiable set. Some features (e.g. PER, GEND, COUNT) clearly reflect properties of the world or of human thought, whereas others (e.g. ANA, P-OBJ, FORM) seem specifically linguistic. Our treatment is neutral on the question of whether grammatical features will ultimately be reducible to more general aspects of cognition.

- The types we have proposed could arguably be drawn as well from a fixed universal inventory. The feature declarations associated with the types are likewise probably quite similar across languages. The constraints introduced by some types (such as NSA), on the other hand, appear to be more specific to the particular language. Our types are arranged in a default inheritance hierarchy, a kind of structure that very likely plays an important role in how people organize many kinds of information. Indeed, the use of such hierarchies in linguistics was inspired by earlier work in artificial intelligence, which suggested this sort of structure for taxonomies of concepts. The particular types we have posited appear task-specifically linguistic, though again, this is not built into our theory.
- Our definition of phrasal satisfaction involves both universal and English-specific elements. As noted earlier, the Argument Realization Principle may well differ across languages. And clearly, the Case Constraint as we have formulated it applies only to English. On the other hand, the Head Feature Principle and the two semantic principles are intended to apply to all languages. Some parts of the Phrasal Satisfaction definition make reference to specifically linguistic constructs (such as grammar rules, heads, and particular features), but the idea of unifying information from diverse sources into a single structure has nonlinguistic applications as well.
- All languages evidently have some binding principles, and they are quite similar. Characteristically, there is one type of element that must be bound within a local domain and another type that cannot be locally bound. But there is cross-language variation in just what counts as 'local' and in what can serve as the antecedents for particular elements. Our particular binding theory is thus not part of universal grammar. Ideally, a grammatical theory would delineate the range of possible binding principles, of which the ones presented in Chapter 7 would be instances. While these principles appear to be quite language-specific, it is conceivable that they might be explained in terms of more general cognitive principles governing identity of reference.
- The lexical rules presented in the previous chapter are clearly parochial to English. A general theory of what kinds of lexical rules are possible would be desirable, but is beyond the scope of this book. The contents of these rules are quite specific to language, but their general form is one that one might expect to find in many domains: if a database contains an object of form X, then it also contains one of form Y.

To sum up this superficial survey of the components of our theory: it contains many elements (the grammar rules, the definition of Well-Formed Tree Structure, the features and types) that are plausible candidates for playing a role in a theory of universal grammar. Moreover, some elements (the binding principles, some lexical rules) probably have close analogues in many other languages. Although our central purpose in this book is to present a precise framework for the development of descriptively adequate grammars for human languages, rather than to account for the puzzle of language learnability through the development of a highly restrictive theory of universal grammar, the framework we have presented here is nevertheless quite compatible with the latter goal.

Further, our grammatical theory suggests a number of parallels between the kinds of information structures needed to account for linguistic competence and those employed in other cognitive domains. However, we need not commit ourselves on the question of task-specificity; rather, we offer the hope that increasingly precise linguistic descriptions like those that are possible within the framework developed here will help to clarify the nature of this controversy and its resolution.

9.7 Conclusion

Chomsky's famous distinction between knowledge of language ('competence') and use of language ('performance') has allowed syntacticians to concentrate on relatively tractable problems, by abstracting away from many features of the way people actually speak. But most generative grammarians agree that an optimal theory of competence will play a role in explaining many features of linguistic performance. To the extent that a theory of grammar attains this ideal, we call it 'realistic'.

We have argued in this chapter that the theory we are developing in this book does well by this criterion. Our theory, by virtue of being surface-oriented, constraint-based, and strongly lexicalist, has properties that fit well with what we know about how people process utterances and extract meaning from them. Our understanding of the mechanisms that underlie linguistic performance is incomplete at present, and many of the points discussed in this chapter remain controversial. Nevertheless, a preliminary examination of what is known about processing provides grounds for optimism about our approach to syntactic theory. Considerations of learnability also support such a favorable assessment.

9.8 Further Reading

Many of the issues raised in this chapter are discussed at a relatively elementary level in the essays in Gleitman and Liberman (1995).

Appendix B
Generative Grammar: an historical overview

As noted in Chapter 2, the theory of grammar developed in this text is most closely related to the framework known as 'Head-driven Phrase Structure Grammar', or HPSG. HPSG is one of a number of frameworks for syntactic analysis that have been developed within the Chomskyan paradigm, broadly conceived. The intellectual tradition it represents is eclectic in its orientation, synthesizing ideas from several approaches to the study of language. To clarify these connections, we provide here a brief history of generative grammar, including a survey of its major branches,[1] with notes on their relevant contribution or influence to the ideas presented within this text.

The basic concept of generative grammar is simply a system of rules that defines in a formally precise way (i.e. 'generates') a set of sequences (strings over some vocabulary of words or 'formatives') that represent the well-formed sentences of a given language. Thus both of the systems considered in Chapter 2 (the regular expression (finite-state) and context-free phrase structure grammars) are generative grammars, as is the grammar summarized in Appendix A.

Generative syntax began in the 1950s when Noam Chomsky and others he influenced developed and formalized a theory of grammar based on the notion of 'transformation'.[2] The architecture of a transformational generative grammar defines sentence well-formedness indirectly:

[1]The dates given in parentheses are roughly the periods during which a substantial number of researchers have been or are still active in developing the theory.

[2]A somewhat different though related concept of transformation can be found in the work of Zellig Harris, beginning in the 1950s. A number of relevant papers are collected in Harris 1970.

first, base (or 'underlying' or 'kernel') structures are generated via a system of phrase structure rules;[3] and then transformational rules apply successively to map these phrase structures into other phrase structures. The sentences of the language, then, are just those that can be derived by applying transformational rules to the base structures according to a particular regime, e.g. a regime of ordered transformations, at least some of which are obligatory. A transformational derivation thus involves a sequence of phrase structures (or structural descriptions), the first of which is a base structure and the last of which is a phrase structure (usually called a 'surface structure') whose word string corresponds to a sentence of the language:

(106) $S_1 \rightsquigarrow S_i \rightsquigarrow \ldots \rightsquigarrow S_i \rightsquigarrow S_n$

BASE STRUCTURE — SURFACE STRUCTURE

Transformational generative grammar (which has dominated the mainstream of syntactic theory from the 1960s through the 1990s) has changed significantly over the years, in ways we sketch below. Yet, despite considerable evolution within this framework, the notion of transformational derivation has been present in one guise or another in virtually every formulation of transformational grammar.[4] Similarly, other commonalities remain in the practice of transformational grammarians, such as the treatment of sublexical entities (e.g. inflectional affixes) as independent syntactic elements, that is, as syntactic primitives on a par with words.[5]

In contrast to the transformational tradition, there is another approach to generative grammar, equally committed (if not more so) to the original goal of developing precisely formulated grammars. This tradition has two distinctive properties:

[3]In some versions of this approach, lexical insertion into the structures generated by the CFG is accomplished by a separate specialized mechanism.

[4]There are exceptions, though, e.g. Koster 1987 and Brody 1995. Chomsky has always maintained that it is 'not easy' to provide empirical (or theory-independent) evidence that would lead one to prefer a transformational theory over simpler alternatives. Despite this repeated claim (Chomsky (1981: 90f; 1995: 223f)), Chomsky has included transformational operations in every version of grammatical theory he has developed since the 1950s.

[5]Hence Chomsky's introduction of the term (syntactic) 'formative' to encompass stems, noninflecting words, and inflectional affixes.

(i) **Constraint-Based Architecture**: Grammars are based on the notion of constraint satisfaction, rather than transformational derivation.

(ii) **Strict Lexicalism**: Words, formed in accordance with an independent lexical theory (or 'module'), are the atoms of the syntax. Their internal structure is invisible to syntactic constraints.

These two design properties together form the basis of the 'constraint-based lexicalist' (CBL) approach to generative grammar. In CBL approaches, surface structures are generated directly, though ancillary kinds of syntactic representation may be cogenerated (see below).

The principle of strict lexicalism has its origin in the pioneering work of Chomsky (1970), who challenged previous attempts to derive nominalizations (e.g. *the enemy's destruction of the city*) from clauses (e.g. *the enemy destroyed the city*) via syntactic transformations. In the mid- and late 1970s, many other alternatives to transformational analyses were developed. There are two particularly significant developments in this period. The first is Bresnan's 'Realistic' Transformational Grammar (widely circulated in unpublished form; a version was published as Bresnan (1978)), which for the first time provided a cogent treatment of numerous phenomena (e.g. passivization) in lexical rather than transformational terms. Bresnan's dramatic first step inspired a number of people, notably Brame (1979) and Gazdar (1981) [first drafted in 1980], to take the further step of purging transformations from syntactic theory altogether. Second, the emergence of the framework of Montague Grammar provided new techniques for characterizing meanings directly in terms of surface structure, thereby eventually eliminating any semantic motivation for syntactic transformations. In many versions of transformational grammar, active and passive sentences were derived from a common underlying structure, leading to the (controversial) suggestion that many aspects of meaning are preserved by transformational derivations. With the advent of more sophisticated methods of semantic analysis, distinct surface structures could be assigned formally distinct but equivalent semantic interpretations, thus accounting for the semantics in a principled fashion without appeal to transformations.

'Realistic' Transformational Grammar and Montague Grammar together set the stage for the emergence of fully nontransformational generative frameworks in the late 1970s and early 1980s. Most notable among these are Lexical Functional Grammar, Generalized Phrase Structure Grammar, Categorial Grammar, and Dependency Grammar, each of which we summarize below. The subsequent history of CBL generative grammar witnessed not only considerable development in each of

these frameworks, but also the introduction of such new approaches as Construction Grammar. Of immediate relevance also is the evolution of Generalized Phrase Structure Grammar, through the integration of ideas from various other frameworks, into the framework of HPSG, from which are directly drawn many analyses and the general orientation of the present text.

The remainder of this appendix briefly sketches a number of the most influential theories of grammar developed over the past forty years. We do not pretend that this is a comprehensive listing of approaches to syntactic description that have been explored in that period; our main purpose here is to situate the theory developed in this text in a wider intellectual landscape, indicating the many conceptual strands that have contributed to our approach. At the same time, we hope to convey a general sense of the historical development of generative grammar and its various schools.

Transformational Grammar

Early Transformational Grammar (1955–1964) An early version of transformational generative grammar was presented in Chomsky's 1957 book, *Syntactic Structures*. The analyses presented there and in other transformational works of the period included explicit formal statements of rules intended to license all and only the well-formed sentences of the language under discussion (usually English). This emphasis on the precise formulation of hypotheses is perhaps the greatest influence of early transformational grammar on the approach presented here.

As noted above, a key claim of transformational grammar (in all its versions) is that an empirically adequate grammar requires that sentences be associated not with a single tree structure, but with a sequence of trees, each related to the next by a transformation. The initial trees in Chomsky's *Syntactic Structures* theory were to be generated by a CFG. For example, passive sentences (such as *The cat was chased by the dog*) were derived from the same underlying structures as their active counterparts (*The dog chased the cat*) by means of a passivization transformation that permuted the order of the two NPs and inserted the words *be* and *by* in the appropriate places.

The most celebrated analysis in this theory is its treatment of the English auxiliary system (roughly, the material covered in Chapter 13 of this text). Chomsky (1957) proposed that tense was, in the underlying syntactic structure, a 'formative' separate from the verb on which it ultimately appears. A movement transformation was posited to account for inversion in questions (deriving, e.g. *Is the sun shining?* from

the same underlying structure as *The sun is shining*); and an insertion transformation placed *not* in the appropriate position for sentence negation. Both these transformations in some instances have the effect of stranding tense – that is, leaving it in a position not adjacent to any verb. For these cases, Chomsky posited a transformation to insert *do* as a carrier of tense. Several other uses of auxiliary *do* (e.g. in ellipsis) were also treated on this view as instances of tense stranding. This unified account of apparently disparate uses of *do*, together with the formal explicitness of the presentation, won many converts to transformational grammar.

The Standard Theory (1965–1970) Katz and Postal (1964) and Chomsky (1965) introduced a number of major changes into transformational grammar, and Chomsky dubbed the resulting theory 'the Standard Theory'. It differed from early transformational grammar in several ways, some rather technical. Among the important innovations of this theory were the use of recursive phrase structure rules (allowing for the elimination of transformations that combined multiple trees into one) and the introduction of syntactic features to account for subcategorization (valence).

Perhaps the most important conceptual change was the addition of a semantic component to the theory of transformational grammar. In this theory, the initial tree in each sentence's derivation, known as its 'deep structure', transparently represented all the information necessary for semantic interpretation. In particular, it was claimed that there is a simple mapping between the semantic roles played by arguments to a verb (intuitively, who did what to whom) and the deep structure grammatical relations (subject, object, etc.). In the final tree of the derivation (the surface structure), the words and phrases were arranged as the sentence would actually be pronounced. On this theory, then, transformations were thought to be the primary link between sound and meaning in natural language.

The Standard Theory had great intuitive appeal and attracted much attention from neighboring disciplines. In particular, many philosophers were attracted by the idea that deep structures might provide something very much like the 'logical form' of sentences needed for precise analysis of their role in inference. Likewise, psychologists hoped that the transformational derivations were a first approximation to the mental processes involved in the production and comprehension of utterances. Initial experiments gave credibility to this idea, in that they showed a correlation between the psychological complexity of a sentence and the number of transformations posited in its derivation. Further research

on this idea (known as the 'derivational theory of complexity') failed to support it, however, and by the early 70s it had been largely abandoned (Fodor, Bever and Garrett (1974)).

Most contemporary grammatical theories have preserved the most important innovations of the Standard Theory, namely, syntactic features, recursive phrase structure, and some sort of semantic component. On the other hand, no current theory maintains the centrality of transformations in mediating between sound and meaning.

Generative Semantics (1967–1974) Generative Semantics was the first major challenge to Chomsky's views within the generative paradigm; its leading figures included George Lakoff, James McCawley, Paul Postal, and John ('Haj') Ross. They carried the central idea of the Standard Theory to its logical conclusion, claiming that deep structures should themselves be viewed as representations of meaning, and denying that syntactic and semantic rules should be considered distinct components of a grammar. That is, on the Generative Semantics view, something was considered a possible input to the transformational rules just in case it represented a proposition that made sense. Hence all languages could be derived from the same underlying source, differing only in how the underlying representations get transformed into sounds.

The underlying trees of Generative Semantics were far larger and more elaborate than those of the Standard Theory (though the inventory of grammatical categories was much reduced). Virtually all the work involved in describing the relationships between form and meaning in language was done in this theory by transformations, though these rules were rarely formulated explicitly.

Generative Semantics enjoyed wide currency for a few years and served as the vehicle for the exploration of a wide range of fascinating phenomena in many languages. Although the theory itself had a short life span (for reasons that have been debated by historians of linguistics), many of the constructions first discovered by generative semanticists continue to figure prominently in theoretical discussions. Moreover, some recent analyses have borne striking resemblances to earlier Generative Semantics proposals; see Harris (1993) and Huck and Goldsmith (1995) for discussion.

The Extended Standard Theory (1967-1980) Unlike the generative semanticists, Chomsky and some others (notably, Ray Jackendoff) quickly abandoned the idea that pairs of sentences with identical deep structures must be synonymous. In particular, they argued that transformations that reordered quantified NPs could change the scopes of the quantifiers (e.g. *Many people read few books* was claimed to have

a range of interpretations different from *Few books are read by many people*). Hence they claimed that structures other than deep structures must play a role in semantic interpretation.

Instead of the complex underlying trees and elaborate transformational derivations of Generative Semantics, the framework that Chomsky dubbed the 'Extended Standard Theory' (EST) posited a relatively impoverished theory of transformations; instead, it enriched other components of the theory to carry much of the descriptive burden. In addition to the new types of semantic rules alluded to above, schematization over phrase structure rules and an enriched conception of the lexicon – including lexical rules – were introduced. These innovations have been carried over into much contemporary work, including the theory developed in this text. The approach of EST led to a highly 'modularized' theory of grammar, with a variety of distinct types of mechanisms to account for different kinds of empirical phenomena.

EST also saw the introduction of 'empty categories' – that is, elements that occupy positions in a tree but which have no phonetic realization. These included a type of null pronoun used in control constructions (e.g. the subject of *leave* in *We tried to leave*) and 'traces' of elements that have been moved.

A central concern of EST and much subsequent work has been to constrain the power of the theory – that is, to restrict the class of grammars that the theory makes available. The primary rationale for seeking such constraints has been to account for the possibility of language acquisition, which (as noted in Chapters 1 and 9) Chomsky regards as the central question for linguistics.

Government and Binding Theory (1980 to the present) Government and Binding theory, or GB, as it is generally called, was first laid out in Chomsky's (1981) book, *Lectures on Government and Binding*.[6] It develops the modular style of EST, dividing the theory of grammar into a set of subtheories, each with its own set of universal principles. Although GB still uses transformational derivations to analyze sentences, it reduces the transformational component to a single rule (referred to as 'Move α'), which can move anything anywhere. The idea is that general principles will filter out most derivations, preventing the massive overgeneration that might be expected.

Elaborating on earlier work in EST, GB analyses posit a rich array of empty categories. Binding theory, which has been a major topic of research within GB, is applied not only to overt pronouns, but also to empty categories. Movement leaves behind traces (a kind of empty cat-

[6]For an introductory presentation, see Haegeman (1994).

egory), which are bound by the moved element. Binding theory thus relates constraints on movement to possible pronoun-antecedent relations. Since movement is used to deal with a wide range of phenomena (including filler-gap dependencies, the active-passive relation, raising, extraposition, and auxiliary inversion), linking all of these to the binding principles yields a richly interconnected system.

The primary focus in GB has been the development of a theory of universal grammar. GB claims that many of the principles that make up the theory of grammar are parameterized, in the sense that they vary within a narrow range. (Another name for this approach to syntax is 'Principles and Parameters'). Learning a language, on this view, consists of fixing a small set of parameters (plus learning vocabulary). That is, GB claims that all languages are essentially alike, with only a few restricted parameters of possible variation. Working out the details of this ambitious program has been the most active area of syntactic research since the early 1980s.

The literature within this framework is massive, and it represents a much wider range of analyses than any of the other theories listed here, though these analyses are seldom formulated with a precision comparable to that assumed in this text and hence are often difficult to evaluate. However, they tend to share certain noteworthy characteristics, including the following:

- Highly articulated phrase structures (linguistically significant distinctions and relations are encoded into tree configurations);
- Use of movement (that is, the transformation Move α);
- Extensive use of empty categories;
- A rich set of universal principles, some of which are parameterized;
- Avoidance of language-particular rules (properties specific to a language are to be expressed in terms of values of universally available parameters);
- Deductive structure (small changes in a grammar should have far-reaching consequences for the language, so that stipulation is minimized).

The theory we have presented here has been influenced by GB in a number of ways. These include very general goals, such as striving for deductive structure in the theory. They also include more specific design features, such as the general form of the Binding Theory (though not the detailed statement of the binding principles). Finally, there are specific points of our grammar that were first proposed within GB analyses, such as treating complementizers as heads that could take sentences as their complements.

The Minimalist Program As of this writing, the most recent descendant of GB is the 'Minimalist Program'. As its name implies, MP is a program for research, rather than a theory of syntax.

MP explores the idea that instead of generating sentences directly, what grammars do is pick out the best expressions from some candidate set. MP grew out of Government and Binding theory, but it represents a significant departure from earlier work in that framework. Its goal is to explain linguistic structure in terms of intuitively natural 'economy conditions' on grammars and their operations. Analyses are most highly valued if they minimize the amount of structure and the length of derivations posited. In his seminal work on MP, Chomsky (1995) lays out some principles that embody this general idea. These include 'GREED', which says that a constituent will move only to satisfy its own licensing requirements, and 'PROCRASTINATE', which says that movement will take place as late as possible. More recent work has sought to localize such effects, applying economy constraints at each step of a transformational derivation. This conception of grammar, in which the properties of competing transformational derivations are crucial in determining sentence well-formedness,[7] represents a radical departure from the original goals and methodology of generative grammar and has no direct connection with the theoretical orientation of the present work. For an extensive comparison of MP with CBL approaches in general, see Johnson and Lappin (in press).

The work of filling in the details of MP is still in the early stages, but it has stimulated a great deal of interest within the field. For an elementary exposition of MP, see Radford (1997); also useful is Webelhuth, ed. (1995).

Constraint-Based Lexicalist Grammar

Categorial Grammar (1974 to the present) Categorial Grammar (CG) has a long history dating back to the 1930s, but it was developed primarily by mathematical logicians before the early 1970s. It came to the attention of linguists when the logician Richard Montague used it as the syntactic framework to go with his new approach to analyzing natural language semantics.[8]

The central idea of CG is that an enriched conception of grammatical

[7]Generative semanticists in the early 1970s briefly discussed the need for what were then termed 'transderivational constraints', but the idea was not pursued for long.

[8]Montague's intensional logic and his precise framework for studying the relation between expressions and their meaning had a considerable influence on work in linguistics. His famous remark (Montague, 1970:373) that 'there is in my opinion no important theoretical difference between natural languages and the artificial lan-

categories can eliminate the need for many of the constructs (e.g. transformations) found in other theories of grammar. CG makes no formal distinction between lexical and nonlexical categories. Hence, for example, an intransitive verb like *sleeps* is treated as belonging to the same category as a phrase consisting of a transitive verb plus a direct object, such as *gets some rest*. Categories are defined in terms of their members' potential for combining with other constituents (and thus CG is often seen as a variety of Dependency Grammar - q.v.). So, for example, verb phrases (and intransitive verbs) can be characterized as those elements which, when combined with a noun phrase on their left, form sentences; one notation for this is NP\S. A transitive verb like *gets* belongs to the category of elements that take an NP on their right to form a NP\S; this can be written (NP\S)/NP. A categorial grammar consists simply of a lexicon together with a few simple rules for how categories can be combined.

The primary attractions of CG have always been its conceptual simplicity and the fact that it is well suited to the formulation of tightly linked syntactic and semantic analyses. It also provides particularly elegant and appealing accounts of coordination. For a general introduction to categorial grammars, see Wood (1993) and Carpenter (1997). Steedman (1996) summarizes a tradition within CG that makes use of logical 'combinators'. For the competing 'type-logical' approach, see Morrill (1994) and Moortgat (1997).

The influence of CG on the theory developed in this text is quite clear. The valence features of HPSG do much the same work as complex category labels do in CG. The nodes of trees in HPSG are labeled with feature structures that contain all the information in CG category labels (plus quite a bit more). Our grammar rules cancel elements off our valence lists in a way analogous to the combinatory rules of CG. Hence many CG analyses can be translated fairly straightforwardly into HPSG.

Construction Grammar (1988 to the present) Construction Grammar (CnG) represents a related yet interestingly different perspective from other CBL frameworks. Linguistic objects (sentences, phrases, and words), viewed formally, are taken to be constituent structure trees whose nodes are occupied by feature structures. The feature structures contain additional syntactic as well as semantic and phonological information. The constructions that license linguistic objects are partial descriptions of such feature structure trees, much as in the presentation of this text.

---

guages of logicians' is sometimes referred to (following Bach 1989) as Montague's Hypothesis.

A construction grammar is a set of constructions that can be unified to create the set of sentences of a language (and no sentence objects not in the language). A CnG feature structure tree can be translated into either an HPSG feature structure of type *sign*, or the analogue of this in Lexical Functional Grammar (q.v.) (insofar as the detailed kinds of information contained in the particular implementations of these theories at hand are comparable).[9]

There are two principle substantive motivations for CnG. The first is to account for grammatically determined, non-truth-conditional aspects of meaning – including such 'pragmatic' factors as conventional implicature and presupposition – in the formal representations of sentences and of the grammatical constructions that license them. The second is to account for the full range of idiomatic and semi-idiomatic constructions that occur in every language. Although every generative approach to the grammar of natural languages is interested in full coverage of the facts of every language, as well as in the extraction of intralanguage and interlanguage generalizations – the latter usually considered to constitute the stuff of universal grammar – varying approaches differ in their relative emphasis on the full coverage of language facts versus the development of a parsimonious theory of universal grammar. CnG falls at the end of this scale concerned primarily with empirical coverage. For more on CnG, see Fillmore and Kay (forthcoming), Fillmore et al. (1988), and Goldberg (1995).

Dependency Grammar (1959 to the present) Work on transformational grammar rests on two crucial (but controversial) assumptions about sentence structure: that it is organized hierarchically into 'phrases' (hence 'phrase structure'), and that grammatical relations such as 'subject' and 'object' are redundant. The assumption of phrase structure is a distinctively American contribution to linguistics, having been suggested by Leonard Bloomfield (1933). Bloomfield suggested that sentences should be analyzed by a process of segmentation and classification: segment the sentence into its main parts, classify these parts, then repeat the process for each part, and so on until the parts are 'morphemes', the indivisible atoms of grammar. Thus *Cool students write short essays* divides into the noun phrase *cool students* plus the verb phrase *write short essays*, which in turn divides into the verb *write* plus *short essays*, and so on. This contrasts with the European tradition (which dates back to classical Greece) in which the focus is on individ-

[9]Recent developments in HPSG have reduced many of the analytic differences between CnG and HPSG. See Sag (1997) and the discussion of phrase hierarchies in Chapter 16.

ual words and their relationships - for example, *cool* is an 'attributive modifier' of *students*, and *students* is the subject of *write*.

The attraction of phrase structure analysis is its formal clarity, which is revealed by the familiar phrase structure trees. Various linguists (mainly European) have attempted to develop the traditional approach in the same way, with the emphasis on the relationships among words rather than on the groupings of words. One of the characteristics of these relationships is that the words concerned are generally not equal, in that one serves to modify the meaning of the other; so *cool students* denote certain students, and *students writing essays* denotes a kind of writing. The relationships are called 'dependencies', with the modifying word depending on the modified (so *cool* depends on *students*, and *students* on *write*), and the approach is called 'Dependency Grammar' (DG) to contrast it with phrase structure grammar.

There are several ways to represent DG analyses diagrammatically, including a system that has been widely used in American schools since the nineteenth century which is often called simply 'sentence diagramming'. The first real attempt to build a theory of DG analysis was Tesnière (1959), but since then developments in PSG theory have been paralleled in DG theories. One of these which is particularly close in other respects to HPSG is 'Word Grammar' (Hudson 1984, 1990, 1998). In some respects, HPSG bridges the gap between DG and PSG, for in HPSG all the parts of a phrase depend directly on its head word - the phrase is 'head-driven', just as in DG. On the other hand, in HPSG the dependent parts are themselves phrases with their own internal structure consisting of a head word and its dependents.

Generalized Phrase Structure Grammar (1979–1987) Generalized Phrase Structure Grammar, or GPSG, as it was known, was initiated by Gerald Gazdar in a pair of papers (Gazdar 1981, 1982) that attracted the attention of numerous researchers in the field of syntax. The theory was further developed by him and a number of colleagues in the early 1980s and was codified in the 1985 book, *Generalized Phrase Structure Grammar* (Gazdar et al. 1985), which provides a detailed exposition of the theory.

The central idea of GPSG is that standard context-free phrase structure grammars can be enhanced in ways that do not enrich their generative capacity, but which do make them suitable for the description of natural language syntax. The implicit claim of work in GPSG was that the tenable arguments against CFG as a theory of syntax were arguments about efficiency or elegance of notation, and not about coverage in principle.

Among the important ideas that originated in GPSG are the separation of CFG rules into (i) rules of immediate dominance ('ID rules'), which specify only which phrases can appear as daughters in a local syntactic tree, and (ii) rules of linear precedence ('LP rules'), which specify general constraints determining the order of daughters in any local tree. This factorization of the two functions of traditional CFG rules is preserved in HPSG, though we have not employed it in the formulation of grammar rules in this text.

A second idea stemming from work in GPSG is the treatment of long-distance dependency constructions, including filler-gap constructions (such as topicalization, *wh*-questions, and relative clauses). The GPSG treatment of these phenomena involved locally encoding the absence of a given constituent via a feature specification. The remarkable result of the transformationless GPSG analysis of filler-gap dependencies was that it succeeded where transformational theories had failed, namely in deriving the Coordinate Structure Constraint and its 'across-the-board' exceptions (see Chapter 15). This feature-based analysis of filler-gap dependencies is preserved in HPSG, and we have carried it over virtually intact to the current text.

Head-Driven Phrase Structure Grammar (1984 to the present)
HPSG evolved directly from attempts to modify GPSG in the interdisciplinary environment of Stanford's Center for the Study of Language and Information (CSLI), the site of several experimental computational systems for language processing. From its inception, HPSG has been developed as a conscious effort to synthesize ideas from a variety of perspectives, including those of Situation Semantics (which originated at CSLI at about the same time as HPSG), data type theory, and a variety of other linguistic frameworks under development in the early and mid-1980s. The name 'Head-driven Phrase Structure Grammar' was chosen to reflect the increasingly recognized importance (as compared with, say, GPSG) of information encoded in the lexical heads of syntactic phrases. Dependency relations are lexically encoded, as they are in Dependency Grammar (q.v.), Categorial Grammar (q.v.) and LFG (q.v.). The theoretical aspects of HPSG have been developed in considerable detail in two books (Pollard and Sag 1987, 1994) and a number of major articles.

Some of the key ideas of work in HPSG are: (1) a sign-based architecture (see Chapter 16); (2) the organization of linguistic information via types, type hierarchies, and constraint inheritance; (3) the projection of phrases via general principles from rich lexical information; (4) the organization of such lexical information via a system of lexical types; and (5) the factorization of phrasal properties into construction-specific

and more general constraints. These properties have all been discussed at various places in this text.

Since the inception of HPSG, researchers have been involved with its computational implementations. From 1980 until 1991, Hewlett-Packard Laboratories in Palo Alto, California supported one such project, which involved the authors of this text and a number of colleagues and students. It was with this project that many of us learned for the first time how far the rhetoric of theoretical linguistics can be from the reality of working grammars. At the time of this writing, implementations of HPSG and HPSG-like grammars are being developed at numerous universities and industrial research laboratories around the world, including sites in Canada, the United States, western and eastern Europe, Japan, Korea, Taiwan, and Australia.

Lexical Functional Grammar (1979 to the present) The theory of Lexical Functional Grammar, commonly referred to as 'LFG' (Bresnan (1982); Dalrymple et al. (1995); Bresnan (to appear)) shares with Relational Grammar (q.v.) the idea that relational concepts like 'subject' are of central importance and cannot be defined in terms of tree structures. But it also treats phrase structure as an essential part of grammatical description and has focussed on the development of a universal theory of how constituent structures are associated with grammatical relations.

In LFG, each phrase is associated with multiple structures of distinct types, with each structure expressing a different sort of information about the phrase. The two representations that have been the center of attention in most LFG literature are: the 'functional structure', which expresses the relational information that is analogous in certain respects to our ARG-ST and in other respects to the valence features SPR and COMPS; and the 'constituent structure', which is a tree diagram very similar to the surface structures of the Standard Theory. General principles and construction-specific constraints define the possible pairings of functional and constituent structures. In addition, LFG recognizes a number of further levels of representation. Perhaps most notable among these are σ-structure, which represents linguistically significant aspects of meaning, and a-structure, which serves to link syntactic arguments with aspects of their meanings. Thus the analogue of the HPSG sign presented in Chapter 16 is a tuple of LFG structures, possibly the four-tuple consisting of a sentence's c-structure, f-structure, a-structure, and σ-structure.

There are no transformations in LFG. Much of the descriptive work done in earlier theories by transformations is handled by an enriched lexicon, an idea pioneered by LFG researchers. For example, the active-passive relation in English is treated as a lexical relation between two

forms of verbs. In early LFG, this was codified in terms of lexical rules similar to those presented in this text. Subsequent work has sought to develop a more abstract conception of lexical relations in terms of 'lexical mapping theory'. LMT provides for constraints on the relation between f-structures and a-structures, that is, constraints associated with particular arguments that partially determine their grammatical function. It also contains mechanisms whereby arguments can be suppressed in the course of lexical derivation. In LFG, information from lexical entries and phrasal annotations is unified to produce the functional structures of complex expressions.

As the above description makes clear, LFG and HPSG bear many resemblances. In particular, HPSG has been able to incorporate many insights from work in LFG, most notably: the significant use of lexical analyses, well-developed dependency-based analyses of numerous phenomena (rather than accounts based on constituent structure), and the general constraint-based approach to grammatical description. There are crucial differences between LFG and HPSG as well (e.g. the use of types and type-based inheritance, which plays no role in the LFG literature, and HPSG's use of defaults). The differences in practice between the HPSG and LFG communities often lead to rather different analyses of the same phenomena; yet these analyses are often compatible with either framework. For an overview of current developments in LFG, see Bresnan (forthcoming).

Two Other Grammatical Frameworks

Relational Grammar (1974 to the present) We can now return to the second controversial claim of PSG, namely, that grammatical relations are redundant. In early theories of generative grammar, transformations were defined in terms of structural properties of tree diagrams. To the extent that traditional notions like 'subject' and 'direct object' were employed in these theories, they were regarded simply as shorthand for relations between linguistic elements definable in terms of the geometry of trees. Relational Grammar (RG), developed by Paul Postal, David Perlmutter, David Johnson and others, adopts primitives that are conceptually very close to the traditional relational notions of subject, direct object, and indirect object. In this respect there is a strong affinity between RG and Dependency Grammar (q.v.). The grammatical rules of RG are formulated in relational terms, replacing the earlier formulations based on tree configurations. For example, the passive rule is stated in terms of promoting the direct object to subject, rather than as a structural rearrangement of NPs.

This approach allows rules to be given very general formulations

that apply across languages. The characterization of passivization as promotion of the object does not depend on whether subjecthood and objecthood are indicated by word order or by other means, such as case marking on the nouns or some marking on the verb.

Although the influence of RG on the theory presented here may not be obvious, it is real. The notions of 'specifier' and 'complement' employed in this text are generalizations of 'subject' and 'object'. Languages use different grammatical devices to mark these relations (word order, case marking, agreement, etc.), and so a theory whose primitives are too closely linked to these devices would be unable to express cross-linguistic similarities. A number of contemporary theories, including LFG (q.v.) and HPSG, have adopted this central insight of RG.

The RG framework was applied to the description of a much wider variety of languages than were earlier generative theories (which tended to concentrate on the familiar European languages, East Asian languages, and a few others). Various results of work in this framework are anthologized in Perlmutter, ed. (1983) and Postal and Joseph, eds. (1990). 'Arc Pair Grammar' (Johnson and Postal (1980)) is an axiomatization and elaboration of many of the central ideas of Relational Grammar. The graph-theoretic foundations of HPSG are essentially the same as those first developed for Arc Pair Grammar by Johnson and Postal.

Optimality Theory Optimality Theory, or OT was first developed as a phonological framework (Prince and Smolensky (forthcoming)), and has recently been adapted to syntactic analysis (see Barbosa et al. (1998)). OT posits a universal set of defeasible constraints. The grammar of a language consists of a ranking of the constraints. Determining whether a given string of words is a well-formed sentence involves comparing it with other candidate expressions of the same proposition. The candidate whose highest-ranking constraint violation is lower than any other candidate's is grammatical. For example, if constraint A outranks constraint B, which outranks constraint C, and if candidate sentence 1 violates A whereas candidate sentence 2 violates B and C, then sentence 2 is preferred over sentence 1, and sentence 1 is ungrammatical. If no other candidate sentence wins such a competition against sentence 2, then sentence 2 is licensed by the grammar.

The idea of constraints that can be violated is also incorporated in the theory presented in this book, since default constraints specified in type hierarchies can be overridden. Moreover, a hierarchy of types with defeasible constraints defines a partial ordering on those constraints, with those introduced lower in the hierarchy taking precedence over those introduced at higher levels. Although there are substantive differ-

ences, certain central properties of OT can also be found in inheritance hierarchies with default constraints.[10]

OT follows much earlier work in generative grammar in positing rich systems of universal grammar. However, the idea that determinations of well-formedness necessarily involve comparing structures or derivations is a break with past views, as we already noted in discussing the Minimalist Program (q.v.). Another common characteristic of MP and OT is the use of defeasible constraints. As noted above, such constraint mechanisms of various sorts have been proposed from time to time within some theories, including the theory presented in this book. This is not surprising, since idiosyncratic exceptions to general patterns are commonplace in natural languages. Defeasible constraint mechanisms are now accepted fairly widely in various theories of syntax. It remains to be seen whether a similar consensus will arise concerning the idea of defining well-formedness in terms of the outcome of some sort of competition.

[10]Such hierarchies are explored in some detail in the artificial intelligence literature of the 1970s.

Glossary

This glossary contains linguistic terms that either play a direct role in or are presupposed by this book. For further information, there are a number of dictionaries of linguistics, including Crystal (1985) and Trask (1993).

AAVE In informal speech, many African Americans use a distinctive variety of English known as African American Vernacular English, or AAVE for short. Also known as Black English, African American English, and Ebonics, both the phonology and aspects of the syntax of AAVE have been extensively studied, largely by sociolinguists.

active A verb form or clause that is not in the passive is referred to as active. *See also* passive; voice.

affix An affix is a morphological element added to a stem to form another stem or a word. Two common types of affix are prefixes (e.g. *re-*, as in *reread*; *out-*, as in *outperform*) and suffixes (*-ed*, as in *visited*; *-s*, as in *visits*. Many languages exhibit other types of affix as well, including infixes (an element inserted into a stem) and circumfixes (e.g. a pair of elements wrapped around a stem).

agreement In many languages, the forms of certain elements can vary to indicate such properties as person [q.v.], number [q.v.], animacy, gender [q.v.], etc. Often, these variations are marked with affixes. Some grammatical relationships between pairs of linguistic elements require that they agree on these properties. In English, for example, present tense verbs are marked to indicate whether their subjects are third-person singular (with the suffix *-s*), and nouns indicate plurality (also with a suffix *-s*). The systematic covariation of the forms of the subject and verb is called 'subject-verb agreement.' Similarly, third-person pronouns must agree with their antecedents in person, number, and gender. *See also* inflection.

anaphor *See* anaphora.

anaphora Certain expressions depend for their interpretation on their association with some other element in the discourse (usually earlier). Paradigm examples are pronouns like *he, her*, and *itself*; other examples include *do so* and verb phrase ellipsis. 'Anaphora' is the term for the relationship between such elements and their antecedents. The term 'anaphor' is sometimes used for all anaphoric elements and is sometimes reserved for only certain kinds (primarily reflexives [q.v.] and reciprocals [q.v.]). *See also* antecedent; binding; ellipsis.

antecedent This term is used for a linguistic element that is the basis for interpreting some anaphoric element occurring (typically later) in the sentence or discourse. In particular, pronouns are often described as referring back to their antecedents (or to the referents of their antecedents). *See also* anaphora.

argument (or **syntactic argument)** This is a general term for any phrase that is selected for by a lexical head, such as a complement or specifier. This usage derives from the semantic term 'argument', which refers to a component of a proposition that must occur with a given predicate or relation. For example, the meaning of the verb *wash* has two semantic arguments (the washer and the washee) that correspond to the two syntactic arguments of the verb *wash* (the subject and the object) – as in *Alex washed the car*. The simplest examples of (syntactic) arguments are noun phrases, but prepositional phrases and subordinate clauses can also function as arguments. *See also* complement; specifier; subcategorization; valence; argument structure.

argument-marking preposition English prepositions serve two distinct functions. In some cases, a preposition is used to indicate the role its object NP plays within the semantic structure of some predicate. In other cases, the preposition itself functions as a predicate, and its object is one of its arguments. In this text, the first kind of use is called an 'argument-marking preposition,' or just a 'marking preposition.' An example is the preposition *on* in *They rely on us*. We call the second kind of preposition 'predicational,' illustrated by the use of *on* in *They are on the porch*.

argument structure In the theory developed in this text, the phrases that serve as arguments of a given head are listed in the value of a feature called ARGUMENT-STRUCTURE. This term is also sometimes used in a less technical sense to designate the semantic relations between a head [q.v.] and its arguments. *See also* argument.

aspect Many languages have special grammatical elements for locating in time the situations referred to. Among the temporal notions often expressed are whether situations are in process or completed and whether they occur repeatedly. These notions are often called 'aspect,' and words or affixes whose function is to express aspect are called 'aspectual markers.' *See also* perfective, progressive.

aspectual marker *See* aspect.

auxiliary This term refers to elements found in many languages that share the following semantic and syntactic characteristics: (i) they express such notions as time (past, present, future; continuation, completion), necessity, possibility, obligation, permission, negation, or questioning; and (ii) they occur in fixed positions in sentences, usually at or near the beginning or end. English auxiliaries are a special kind of verb. It is the auxiliary verb that is inverted with the subject in yes/no questions (e.g. *Did she fall?*) and that carries the negative suffix in contractions [q.v.] (e.g. *can't, won't*).

binding Pronouns are often said to be 'bound' by their antecedents [q.v.], and the term 'binding' is used to refer to the relationship between pronouns and their antecedents. The study of the principles governing pronominal anaphora [q.v.] is called 'binding theory.' *See also* reciprocal; reflexive.

case Certain words – particularly nouns and pronouns – can appear in different forms depending on their relationship to other elements in the sentence in which they appear. In English, for example, personal pronouns exhibit nominative case (e.g. *I, they*) or accusative case (e.g. *me, them*), depending on whether they are subjects or objects. In many languages, case is the primary way of indicating the roles of the noun phrases in the clause – that is, who did what to whom. Among the names of cases commonly used across languages are 'nominative,' 'accusative,' 'dative,' 'genitive,' 'ergative,' and 'absolutive.' *See also* inflection.

clause A clause is a phrase that includes a predicate and all of its arguments and modifiers. The term is sometimes limited to phrases headed by a verb.

common noun Nouns are often divided into two kinds: proper and common. Proper nouns are names, denoting individual things by themselves, and do not normally take determiners, complements, or modifiers. Common nouns stand for kinds of things and take determiners [q.v.], modifiers [q.v.], and (sometimes) complements [q.v.]. In English orthography, proper nouns are conventionally capitalized, but common nouns are not.

competence In a number of works, Chomsky has distinguished between the (largely unconscious) knowledge of language that people have and how they put that knowledge to use in speech and writing. The former is called 'competence'; the latter 'performance.' The study of linguistic competence abstracts away from such factors as memory limitations, disfluencies, and speech errors. Work in generative grammar [q.v.] has concentrated largely on developing models of competence, though there has been much discussion of how such models relate to what is known about performance.

complement The lexical head [q.v.] of a phrase characteristically selects which arguments co-occur with it, and some of these are referred to as 'complements.' When the phrase's head is a verb, the complements include what are traditionally called direct and indirect objects, as well as some prepositional phrases and subordinate clauses [q.v.]. Subjects – and determiners of NPs – are arguments that are not complements, but specifiers. Complements occur as sisters to the lexical head in syntactic structure and, in English, follow the head. For example, a verb such as *hit* takes one complement, namely, an NP (e.g. *hit the ball*); *rely* takes a PP complement (e.g. *rely on Sandy*). A preposition such as *in* also takes a single NP complement (e.g. *in the box*). Some nouns can also take complements, such as *picture* which takes an optional PP complement (e.g. *picture of Kim*). *See also* argument; specifier.

complementizer Words whose sole function is to introduce certain kinds of complements are known as 'complementizers.' This label is normally reserved for elements introducing clausal complements headed by a verb. In English, the use of *that* to introduce subordinate clauses [q.v.] (as in *It is remarkable that ice floats*) is the clearest example of a complementizer. In this text, the *to* that introduces VPs (as in *Everybody wants to win*) is also analyzed as a complementizer. *See also* complement.

conjunction (conjunct) Traditional grammarians use the term 'conjunction' to refer to words that connect two linguistic units in some way. In this text, we use it only for what are traditionally called 'coordinate conjunctions,' that is, words connecting elements that are, intuitively, of equal status. In English, the paradigmatic coordinate conjunctions are *and* and *or*, though *but* and *nor* can also function in this way. The individual words or phrases that have been conjoined can be referred to as conjuncts. *See also* coordination.

constituent The term 'constituent' is used by linguists as a near synonym for 'phrase,' meaning a part of a sentence that functions syntactically as a single unit. The difference is that 'constituent' is usually limited to phrases that are proper parts of larger expressions.

context-free grammar (CFG) A context-free grammar is a particular type of formal system that has proved very useful in the precise characterization of computer languages and also serves as the starting point for much work in syntactic theory. CFGs consist of an initial symbol [q.v.], a finite lexicon with words classified into grammatical categories [q.v.], and a finite collection of rules of the form $A \rightarrow \omega$, where A is a single symbol (representing a type of phrase), and ω is a finite string of lexical and/or phrasal categories.

contraction Reduced forms of words are sometimes combined with other words (that would typically occur adjacent to the reduced words) to form a new word; these are referred to as 'contractions.' English examples include combinations of a finite auxiliary [q.v.] verb with a reduced form of *not* to produce such words as *isn't* and *can't*, as well as simple contraction of finite auxiliaries, e.g. *They're arriving tomorrow* and *Kim's here.*

control Some complements have no overt specifier, but are interpreted as if they had subjects with the same reference as (i.e. coindexed with) another complement to the same predicate. For example, in both *Pat tries to be on time* and *We urged Pat to be on time* the individual Pat is understood as the person who is meant to be on time. This relationship (between two noun phrases, the second typically an unexpressed subject) is referred to as 'control'; in this case with the NP *Pat* being the 'controller' of the unexpressed subject of the infinitival phrase. Control predicates are not to be confused with raising [q.v.] predicates (like *continue* and *expect*), one of whose arguments actually is simultaneously an argument of another complement. A number of syntactic and semantic diagnostics distinguish these two types of predicates. *See also* raising.

coordination This term refers to the linking of two words or phrases of equivalent syntactic status (i.e. neither conjoined element is subordinate to the other). An example of a coordinated clause is *Kim cooked the food and Sandy washed the dishes. See also* conjunction.

copula This term is used by traditional grammarians to refer to verbs with little semantic content, which serve to link a subject with a predicate. In English, the copula is *be* and its various inflections.

coreference (or**) coreferential** Two referring expressions that refer to the same entity are called 'coreferential,' and the relationship between them is called 'coreference.' *See also* anaphora.

count noun Nouns are divided into two subclasses known as 'count' and 'mass' according to which determiners they can combine with. Count nouns co-occur with *a(n), several, few*, etc; mass nouns co-occur with *much* and can occur in the singular with no determiner. This distinction is correlated with a semantic distinction: mass nouns usually are used to refer to substances and count nouns to (sets of) entities. A portion of a substance (e.g. helium) is still the same substance, whereas a portion of an entity (e.g. a bicycle) is not usually an entity of the same kind. This correlation is not perfect, however, as evidenced by the mass noun *furniture* and by minimal pairs like *cabbage* (which can be either count or mass) vs. *lettuce* (which, for many speakers, must be mass).

declarative/interrogative/imperative These are terms used in the classification of sentence types. Declarative sentences are used to make a statement (or – equivalently for our purposes – to assert the truth of a proposition), as in *The mayor is reading a book.* Interrogative sentences are used to ask questions, as in *What are they doing?* Imperative sentences are used to give orders (or to issue 'directives'), as in *Read a book!*

defeasible A constraint is said to be 'defeasible' if it can be overridden – that is, if it allows for the existence of exceptions.

demonstrative Expressions used for referring through direct indication (often accompanied by pointing) are called 'demonstratives.' The best examples in English are *this, that, these*, and *those.*

descriptive grammar *See* prescriptive grammar.

determiner The sorts of specifiers [q.v.] that nouns take are called 'determiners.' These include articles (*a, the*, etc.), quantifiers [q.v.] (*some, every, many, two*, etc.), and possessives [q.v.] (*my, Kim's*, etc.). *See also* specifier.

discourse This term refers to units of language longer than a sentence – for example, dialogues or paragraphs.

distribution Linguists use this term to refer to the set of total environments – or contexts – in which some linguistic unit can occur.

ditransitive verb Verbs that take two NP objects are called 'ditransitive.' The standard example is *give*, in examples like *The teacher gave the students an exam. See also* intransitive verb; transitive verb; valence.

dummy Words that evidently have no meaning and serve only to fill some grammatical function are sometimes called 'dummies.' The paradigm examples in English are the *there* that occurs in existential sentences (e.g. *There is a seat available* and the *it* of extraposition [q.v.] (e.g. *It is fortunate that you have a seat*). Other terms used for these are 'expletives' and 'pleonastic' elements.

ellipsis Ellipsis means 'leaving out' or 'omitting': in certain contexts, parts of a sentence can be omitted if their interpretation is reconstructable. An example is the following case of verb phrase ellipsis, where the bracketed material may be left out:

(107) Pat won't taste the soup, but Chris will [taste the soup].

See also anaphora.

existential *be*/**existential** *there* English has a special construction for expressing existence, involving the dummy *there* as subject and forms of the verb *be*. These are called 'existential.' *See also* dummy.

extraposition Predicates that can take complementizer [q.v.] phrases (i.e. *that*-clauses or *to*-infinitive phrases) as subjects can also occur with a dummy *it* as subject and the CP as the last complement. The latter construction is called 'extraposition.' The term is also sometimes used for expressions in which a complement or modifier is separated from its head by intervening material, as in *A review appeared of Lee's latest book*. *See also* dummy.

feature structure A standard way of representing linguistic information is in terms of complexes of features and values. A feature can be thought of as a dimension along which different linguistic entities (such as words, phrases, or sentences) may differ, and values identify locations on those dimensions. A feature-value pair models a property of an expression that distinguishes it in a linguistically interesting way from some other expressions. For example, the feature PERSON in English has three possible values, namely '1st,' '2nd,' and '3rd.' It is a property of the word *you* that it is second person, and we represent that with the feature-value pair [PERSON 2nd]. A feature structure can thus be treated as a set of such feature-value pairs, in which no feature is paired with more than one value. Values of features in our theory may themselves be feature structures, or even lists of feature structures. A variant of this idea formalizes feature structures as directed graphs, where the arcs are labeled by feature names. Feature structure descriptions are standardly given in terms of matrices, listing feature

names paired with their values, also known as 'feature specifications.' *See also* inheritance hierarchy; type.

finite-state grammar finite-state grammars are a type of formal system sometimes used to describe certain rather simple artificial languages. They are mathematically equivalent to regular expressions. *See also* context-free grammar; regular expression.

finite verb A finite verb is one that is marked for tense [q.v.] (present or past, in English).

gap *See* long-distance dependency.

gender The nouns in many languages divide into classes, differing in their patterns of inflection and agreement. In a number of languages (e.g. French and German), these noun classes are referred to as 'genders,' because nouns used to refer to males or females (of any species) are generally (though not invariably) grammatically masculine or feminine, respectively. In English, gender is marked grammatically only on third-person singular pronouns (*he, she*, and *it*,) and is virtually always predictable from the actual gender of the referent.

generative grammar Chomsky introduced this term based on the idea that a grammar is a formal system for generating the sentences of a language. The term is now used somewhat ambiguously, either to mean work in the Chomskyan tradition (fairly broadly conceived), or to mean an explicit system of rules, principles, and/or constraints that characterizes all and only the well-formed sentences of a language.

grammatical category Words and phrases can be classified in various ways, any of which can be called a 'grammatical category.' The term is usually used to refer to parts of speech [q.v.], such as noun, verb, etc., as well as types of phrases, such as noun phrase, verb phrase, and sentence. But it can also be used for less traditional classifications, such as the various types of lexemes [q.v.] presented in this text.

head The constituent [q.v.] of a phrase that is grammatically the most important constituent of that phrase is called the 'head' of the phrase. The head usually determines the category of the phrase, as well as many of its other properties. Thus noun phrases have nouns as heads, verb phrases have verbs as heads, etc. The term is used ambiguously to refer to the word that functions as head of the phrase and any subphrase containing that word. For example, in *the destruction of the city*, both *destruction* and *destruction of the city* can be called heads of the phrase.

idiom Some combinations of words have interpretations that are not fully predictable from the meanings that those same words have in other contexts. These are known as 'idioms.' English examples include *take advantage* to mean (roughly) 'exploit,' *keep tabs on* for 'monitor,' and *kick the bucket* for 'die.' Parts of an idiom are sometimes called 'idiom chunks,' e.g. *advantage* in *take advantage*. Idiom chunks play a central role in one of the diagnostics for distinguishing raising [q.v.] predicates from control [q.v.] predicates.

imperative *See* declarative

infinitive A certain class of nonfinite verbs is referred to as 'infinitives.' English infinitives are uninflected, but in many other languages, there are special infinitive affixes. Use of the term in English grammar is sometimes limited to verbs preceded by 'to,' but we also use it elsewhere, e.g. for verbs following modals [q.v.].

inflection Languages often add affixes to words to mark the syntactic function or relationships of the word in the sentence. For example, present tense verbs in English are usually inflected with the suffix *-s* when the subject is third person singular, and past tense verbs are inflected with *-ed*. The term may also be used to refer to the affix itself. Among the common uses of inflectional affixes are to indicate tense [q.v.], agreement [q.v.], number [q.v.] (singular or plural), and case [q.v.].

inheritance hierarchy The elements of some domains of study can naturally be organized into classes, based on shared properties. Some classes can be further subdivided into subclasses, with additional shared properties. The organization of such domains can be thought of as a hierarchy, with the most inclusive class (encompassing the entire domain) at the top, and the individual members of the domain at the bottom. In between are the various classes of interest, with classes above their subclasses. The properties associated with particular classes are inherited by their subclasses, and ultimately by their individual members. Domains organized in this way are referred to as 'inheritance hierarchies.' In linguistics, inheritance hierarchies have been used to organize lexical information, among other things. *See also* type.

initial symbol Grammars characterize languages. But languages can be conceived of in a variety of ways: as consisting of sentences, of phrases, of any expressions that can serve as stand-alone utterances, etc. A formal theory of grammar must include a specification of which of the expressions it characterizes are to be regarded as those that constitute the language. The initial symbols of a for-

mal theory are precise statements of what is to count as an element of the language. In this book, the initial symbol definition specifies conditions that phrases must satisfy if they can stand alone, i.e. be used in isolation to communicate a message.

interrogative *See* declarative.

intonation This term is used to refer to the patterns of pitch in speech.

intransitive verb A verb that does not take any NP objects is referred to as 'intransitive.' A standard example in English is *die. See also* ditransitive verb; transitive verb; valence.

inversion Grammarians use this term fairly generally to refer to any construction in which two elements appear with their typical ordering reversed. In this text, it is used in particular for sentences (mostly questions) in which a finite auxiliary [q.v.] verb precedes the subject, as in *Are you sleeping?*

island constraint While long-distance dependencies can, in principle, stretch over arbitrary distances, there are some pairings of filler and gap positions that are not possible. For example, a gap inside a CP subject cannot, in general, be paired with a filler outside that CP, as in **Which candidate did [that I voted for __] bother you.* The constraints on possible filler-gap pairings are known as 'island constraints.' *See also* long-distance dependency.

Kleene star It is useful in the formal representation of languages (both natural and artificial) to allow certain patterns to be repeated any finite number of times (including zero). The standard notation for this is a superscripted asterisk, known as the 'Kleene star' (after the mathematician Stephen Kleene). For example, ab^*c is shorthand for the infinite set of strings: *ac, abc, abbc, abbbc, ...* 'Kleene plus,' denoted by a superscripted plus sign, means any nonzero number of repetitions. *See also* regular expression.

lexeme The term 'word' is used ambiguously to mean either a particular form, such as *sees*, or a set of related forms such as *see, sees, saw, seen*, and *seeing.* To avoid this ambiguity, linguists sometimes posit an abstract entity called a 'lexeme' that gives rise to a family of related words. *See also* word.

lexicalism Lexicalism often refers to the doctrine that (1) the internal structure of words is independent of how words are put together to make sentences, and (2) words are the atoms of syntactic combination. For example, in a lexicalist theory, bound morphemes (inflectional affixes that must be attached to a word) are not treated as independent syntactic elements, as they are in most (nonlexicalist) versions of Transformational Grammar (see Appendix B).

Theories of grammar also differ in their organization and in where they locate syntactic information. Some theories (e.g. Generalized Phrase Structure Grammar) have rich systems of rules and relatively impoverished lexical entries. Others (e.g. Categorial Grammar or Lexical Functional Grammar) have highly structured lexical entries and a small number of very general rule schemata. 'Lexicalism' is sometimes also used to distinguish the latter sort of theory.

lexicon The list of all words [q.v.] (or lexemes [q.v.]) of a language is called its 'lexicon.' The lexicon is the repository of all idiosyncratic information about particular words, including syntactic, semantic, and phonological information. In some theories of grammar, the lexicon can also contain a great deal more systematic information, organized by a type hierarchy [q.v.] and/or lexical rules.

long-distance dependency Certain constructions, including *wh*-questions, relative clauses, and topicalization, permit an element in one position to fill the grammatical role associated with another position. The two positions can be arbitrarily far apart. For example, in *Which student did the principal say that the teacher thought was responsible?* the NP *which student* functions as the subject of *was responsible*, although they are separated by most of the sentence. Such constructions are called 'long-distance dependencies' (LDDs). Elements like *which student* in the above example are called 'fillers,' and the position normally associated with the filler's role (in this case, immediately preceding *was responsible*) is called the 'gap.' *See also* island constraints.

main clause *See* root sentence.

modal The English verbs *can, could, may, might, must, shall, should, will*, and *would*, along with their negated forms (*can't*, etc.) are referred to as 'modals' or 'modal verbs.' They share the following properties: they function only as finite verbs [q.v.]; they exhibit auxiliary behavior (negation, inversion, contraction, and ellipsis); they take infinitival [q.v.] VP complements; and they show no agreement [q.v.] (i.e. no third-person singular *-s* suffix). Some other languages have similar syntactically distinctive classes of words expressing necessity, possibility, obligation, and permission; these are also known as modals. *See also* auxiliary.

mass noun *See* count noun.

modifier Most phrases consist of a head [q.v.], together with that head's arguments [q.v.]. Semantically, the head typically denotes either a situation or an individual, and the arguments denote es-

sential associated entities. In addition, phrases may contain modifiers, which serve to place further restrictions on the situation or individual picked out by the phrase as a whole. Modifiers can take a wide variety of syntactic forms, including adjectives and adjective phrases, adverbs and adverbial phrases, prepositional phrases, and modifying clauses (such as relative clauses). *See also* argument structure.

morphology This term refers ambiguously to the study of word structure – how words are put together out of stems and affixes – or to word structure itself.

negation Languages include devices for reversing or contradicting the meaning or truth conditions of expressions, a semantic effect known as 'negation.' In English, the paradigmatic element expressing negation is the word *not*.

nominalization Nominalizations are nouns constructed out of words of other categories, usually through affixation. An example is *destruction*, derived from the verb *destroy* through the affixation of *-tion* (together with some other modifications). The term 'nominalization' is also used to refer to a process of turning verbs and adjectives into nouns.

number Most English nouns take different forms depending on whether they refer to single entities or multiple entities, e.g. *dog/dogs, man/men*. Similarly, present tense [q.v.] verbs with third-person subjects have different forms depending on whether the subjects are singular or plural. The term 'number' is used for such distinctions. Some languages also mark number on other types of words, e.g. adjectives may be marked for the number of the noun they modify. There are also languages that make finer number distinctions than just singular vs. plural, notably languages that have special 'dual' forms for expressions referring to sets with exactly two members.

orthography This term refers to written representations of language. For example, the plural of the noun *doe* and the present tense form of the verb *do* that goes with a third-person singular subject share the same orthography (namely, 'does'), although their pronunciations (and almost everything else about them) are different.

paradigm Certain words have multiple inflectional forms. For example, verbs in English typically change their form depending on whether they are past or present tense, and their present-tense forms depend on the person and number of the subject. They also have a

variety of nonfinite forms. The full array of inflectional forms of a word is known as its 'paradigm.' *See also* inflection.

parsing This term refers to the process of assigning a structural description to a sentence. Many computer systems designed to process natural language include components for parsing, and much psycholinguistic research is concerned with discovering what (if any) parsing strategies humans use in language comprehension.

part of speech This is the traditional term for lexical categories (i.e. categories of words), based on a combination of semantic and distributional criteria. Among the standard parts of speech are noun, verb, adjective, preposition, adverb, and conjunction. *See also* grammatical category.

participle Certain nonfinite verbs – usually ones that share some properties with adjectives – are referred to as 'participles.' English has three types of participles: present participles, which end in *-ing* and usually follow some form of *be*; past participles, which usually end in *-ed* or *-en* and follow some form of *have*; and passive participles, which look exactly like past participles but indicate the passive voice [q.v.]. The three participles of *eat* are illustrated in the following sentences:

a. Termites are eating the house.
b. Termites have eaten the house.
c. The house was eaten by termites.

passive Many languages have a construction in which the grammatical subject of a verb plays the same semantic role that the object plays when the verb in question appears elsewhere (in active [q.v.] forms). The term 'passive' is used to refer both to this construction, and to the verb whose arguments' roles are at issue. In English, the passive form of the verb looks exactly like the past participle and is usually preceded by a form of *be*; a prepositional phrase headed by *by* is also common, and is used for marking what would be the subject if the verb were not passive. An example is *The dog was attacked (by wombats)*. *See also* participle; voice.

perfective Many languages have special verb forms or constructions used to indicate that the event denoted by the verb is completed. These are referred to as 'perfective' (or just 'perfect') in aspect. The English perfective involves the combination of *have* with a past participle [q.v.], as in *The dog has eaten the cake*. *See also* aspect.

performance *See* competence.

person Many languages distinguish grammatically among expressions referring to the speaker, to the hearer, and to third parties. This is called the expression of 'person.' Reference to the speaker or a set including the speaker is called 'first person'; reference to (sets including) the adressee(s) is called 'second person'; and everything else is called 'third person.' Person distinctions are clearest with pronouns, since these are the most common forms used to refer to the speaker and hearer. But in some languages nouns also show person marking, and verbs and adjectives may then agree with the noun in person.

phonetics Phonetics is the study of the acoustic or articulatory properties of speech sounds.

phonology Phonology is the study of the sound systems of languages, i.e. the systematic grammatical patterns in the distribution [q.v.] of speech sounds.

possessive Many languages have grammatical mechanisms for indicating a relation of possession between the referents of two NPs. When one noun or NP is marked as the possessor of another, this marking is referred to as the 'possessive.' In English, the possessive is marked by *'s* attached at the end of the noun phrase functioning as the 'possessor.'

pragmatics The information conveyed by a linguistic expression in a particular instance of use is typically much more than just its literal (or 'linguistic') meaning. The study of how linguistic meaning contributes more generally to communication is called (linguistic) 'pragmatics.'

predicational preposition *See* argument-marking preposition.

prescriptive grammar Much of traditional work in grammar is concerned with setting norms – that is, dictating that some usages are 'incorrect' and hence to be avoided. Modern linguists refer to this as 'prescriptive grammar' (or just 'prescriptivism') and argue that it has no scientific basis. Instead, they advocate describing languages as they are ('descriptive grammar') and seeking scientific explanations for what they find.

productive A relationship between two linguistic forms is said to be 'productive' if it generalizes to novel forms. For example, the use of the suffix *-ing* to mark the present participle form of a verb is productive, since it gets applied to new coinages (as in *faxing*). Productivity is usually thought of as a matter of degree, with exceptionless relationships counting as more productive than those with exceptions.

progressive Special verb forms or constructions used to indicate that the event denoted by the verb is in progress are referred to as 'progressive' aspect. The English progressive involves combination of *be* with a present participle [q.v.], as in *The dog is eating the cake*. *see also* aspect.

proper noun *See* common noun.

quantifier Words or phrases used to restrict the number or amount of some referent are called 'quantifiers.' In English, these include such expressions as *all, each, some, many, few, two, more than half*, etc.

raising Some predicates take one more syntactic argument than semantic argument. In these cases, the extra syntactic argument functions as the subject of another complement and must obey any special co-occurrence restrictions imposed by that complement. These predicates are called 'raising' predicates. Raising is exemplified by the sentences *Pat continues to be on time* and *We expected Pat to be on time*. In these examples, *Pat*, though a syntactic argument of *seem* and *expect*, is semantically an argument only of *be on time*. A semantically empty dummy [q.v.] is possible with raising predicates, where it would not be possible in the corresponding positions with control predicates: *There continued/*tried to be demonstrations on campus*. *See also* control.

reciprocal A reciprocal pronoun is one that expresses a mutual relationship, such as the English pronoun *each other*. *See also* anaphora; reflexive.

referent This term is used for the entity (e.g. a person, object, notion, or situation) that is denoted by (a use of) a linguistic expression.

reflexive Many languages use special forms of pronouns when the subject and object refer to the same individual or individuals, e.g. the English forms ending in *-self* or *-selves*. These are called 'reflexives' or 'reflexive pronouns.' It is common for these pronouns also to be acceptable in some other environments, but those environments differ from language to language. *See also* anaphora; binding; reciprocals.

regular expression It is possible to characterize the well-formed expressions of some simple formal languages by means of a few abbreviatory devices. One system that has proved very useful in some contexts involves templates, made up of words and/or categories of words, together with parentheses (to indicate optionality), a disjunction symbol (to indicate alternatives), and Kleene star [q.v.] (and/or Kleene plus), to indicate arbitrary numbers of repetitions

of a sequence. Such templates are called 'regular expressions.' *See also* finite-state grammar.

relative clause These are clauses that are used to modify nouns or noun phrases. A relative clause characteristically contains either a gap or a pronoun understood to be coreferential with the noun or NP they modify.

root sentence The traditional distinction between main clause and subordinate clause is motivated in part by the fact that certain phenomena seem to be restricted to main clauses, e.g. the inversion of finite auxiliaries [q.v.] in English questions (compare: *Will I win?* vs. *I wonder whether I will win*). Consequently, some version of this distinction has been maintained in most formal theories of grammar. The term 'root sentence' is sometimes used for main clauses, or, more technically, a phrase of category S that is not dominated by anything else. *See also* subordinate clause.

saturated In the system of grammar developed in this book, a saturated phrase is one that is specified as [SPR $\langle$ *rangle*] (and [COMPS $\langle$ *rangle*]). The intuition behind this is that headed phrases can be thought of as being generated bottom-up, starting from the lexical head, via a regime of cancelling elements from the head's valence specifications. For example, a verb combines first with however many complements are on its COMPS list to build a VP (a verbal phrase that is [COMPS $\langle$ *rangle*] but [SPR $\langle$*NPrangle*]); the resulting (SPR-)unsaturated phrase then combines with the subject NP to build a saturated phrase, i.e. an S.

semantics Semantics is the branch of linguistics concerned with the study of linguistic meaning. Linguists also use the locution 'the semantics of' some expression as a way of talking about the literal interpretation of that expression. Not all information that is conveyed by the utterance of an expression is part of its semantics, but the line between literal meaning and what is conveyed can be hard to draw. At a minimum, the semantics of a (declarative) sentence is usually taken to include a specification of the conditions under which it would be true. *See also* pragmatics.

specifier We use the term 'specifier' to cover subjects of clauses, determiners of noun phrases, and certain other constituents that are neither heads of the phrases they appear in nor complements to the heads. In English, the specifier of a phrase precedes its head [q.v.] and complements [q.v.]. *See also* determiner.

structural description (SD) In this text, phrases are modeled by phrase structures – trees consisting of subtrees whose nodes are fully specified feature structures. Phrase structures are informally described by means of 'structural descriptions,' which share the same tree configuration as the phrase structures themselves, but can leave some features underspecified. Likewise, a word is modeled by a word structure (tree), consisting of a fully specified feature structure that exhaustively dominates the word. Partial descriptions of word structures (lexical SDs) are also structural descriptions. *See also* feature structure.

subcategorization Lexical heads differ according to how many and what types of things they must combine with in order to make complete phrases. Each grammatical category [q.v.] (that is, part of speech [q.v.]) can be divided into subcategories, based on the valence, or combinatory potential, of the particular words. When we talk of the subcategorization of a verb (or other type of head), we mean the restrictions on which sorts of phrases it can combine with. Another common locution is to say that a given verb 'subcategorizes for' a certain phrase, meaning that it combines with such a phrase. *See also* valence

subordinate clause A subordinate clause is one that is dependent on, and usually a constituent [q.v.] of, another clause [q.v.]. An example of a subordinate clause is *when Kim went* in *Sandy came when Kim went*. *See also* root sentence.

tense Finite verbs come in different forms depending on the time they denote; these forms are called 'tenses.' English has present and past tense, exemplified by the present tense forms *walk* and *walks*, and by the past tense form *walked*. Some languages also have future tenses, but English uses other means (e.g. the modal [q.v.] *will*) to express future time. *See also* aspect; finite verb.

transitive verb Verbs that take an NP object are called 'transitive.' The term can also be used for other parts of speech that can take objects, e.g. prepositions. It is sometimes taken to encompass both simple transitive verbs (that is, those taking a single object) and ditransitive verbs. A standard example of a transitive verb is *hit*. *See also* ditransitive verb; intransitive verb; valence.

type Elements of any collection can be sorted into types, based on similarities of properties. In the theory presented in this text, linguistic entities (e.g. words and phrases) are described by means of feature structures [q.v.]. Particular features are appropriate only

to certain types of entities, and constraints on possible feature-value pairings are also associated with some types. The types of linguistic entities are arranged in an inheritance hierarchy. The type hierarchy is especially important for capturing regularities in the lexicon. *See also* feature structure; inheritance hierarchy; lexicon.

unification The operation of unification merges two feature structure descriptions into a description that contains all the information in both. Two feature structure descriptions can unify so long as the information in them is consistent – that is, so long as there is no feature for which they have conflicting values. The unification simply consists of all of the features and values specified in the two feature structure descriptions. If description D_1 is satisfied by a family of feature structures F_1 and description D_2 is satisfied by a family of feature structures F_2, then the unification of D_1 and D_2 (written $D_1 \sqcup D_2$) is satisfied by all feature structures in the intersection of F_1 and F_2.

universal grammar Many linguists claim that there is a great deal in common among the grammars of the world's languages. Most advocates of this position believe that the commonalities exist because linguistic structure is largely determined by human biology. The term 'universal grammar' is used to mean three subtly different things: (i) what is common to the world's languages; (ii) linguists' representations of these commonalities; and (iii) the biological structures that are claimed to underlie the common features.

unsaturated *See* saturated.

valence This term is used (by analogy with the chemical term) to refer to the combinatoric potential of words and phrases. In this text, the VALENCE features are those that specify this kind of information. Specifically, the VALENCE features SPR and COMPS for the verb *put* specify that it requires a subject NP, an object NP, and a PP in order to form a clause. *See also* argument; argument structure; ditransitive verb; intransitive verb; transitive verb.

voice This term refers to the way the semantic arguments of the verb are expressed grammatically. The term is used in English primarily to distinguish active voice and passive voice, but some other languages have far richer systems of voice. *See also* active; passive.

word This term is used in many different ways. In this text, a word is a particular form whose lexical entry is derived from a lexeme by some inflectional rule. *See also* lexeme, lexicon.