Toshiaki Nishihara, Katsuaki Nakanishi, Rachel Nordlinger, Stephan Oepen, Michael Orme, Gerald Penn, Paul Postal, Geoffrey Pullum, Emily Reyna, John Rickford, Susanne Riehemann, Sarah Risken, Jed Rose, Scott Schwenter, Jeramie Scott, Rylan Sekiguchi, Peter Sells, Yael Shrager, Luke Swartz, Ryosuke Takahashi, Stuart Tannock, Shiao Wei Tham, Ida Toivonen, Judith Tonhauser, Louise Vigeant, Rick Warren, Gert Webelhuth, Steve Wechsler, Bill Weigel, and Corinne Yates. We would also like to thank Dikran Karagueuzian, Director of CSLI Publications, for his continued support and patience, as well as Lauri Kanerva, Anubha Kothari, Michael Frank, Christine Sosa, Maureen Burke, and Tony Gee for their help in matters of production.

This book was written at Stanford's Center for the Study of Language and Information – an ideal environment for thought and writing. Thanks to Emma Pease for sustained help with all matters technological. Some of the material in this text is based on research conducted under the auspices of CSLI's Linguistic Grammars Online (LINGO) project. In that connection, we gratefully acknowledge contracts from the Bundesministerium für Bildung, Wissenschaft, Forschung, und Technologie (BMBF), who supported LINGO'S participation in the VERBMOBIL project. This material is also based in part upon work supported by the National Science Foundation under Grant No. BCS-0094638. Work on the second edition was completed while Sag was a Fellow at the Center for Advanced Study in the Behavioral Sciences, in which connection we acknowledge the support of a grant (# 2000-5633) to CASBS from The William and Flora Hewlett Foundation.

Finally, we express our deep appreciation to our spouses, Penny Eckert, Judith Wasow, and Vijay Menon, for putting up with us during the long and often tedious revision process.

Sag, I, Wasow, T & E. Bender 2003

Syntactic Theory 2nd edition

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 our concerns are very different. We view human language as a natural phenomenon amenable to scientific investigation, rather than something to be regulated by the decrees of authorities. Your seventh grade math teacher might have told you the (apocryphal) story about how the Indiana legislature almost passed a bill establishing the value of π as 3, and everybody in class no doubt laughed at such foolishness. Most linguists regard prescriptive grammar as silly in much the same way: natural phenomena simply cannot be legislated.

Of course, unlike the value of π, the structure of language is a product of human activity, and that can be legislated. And we do not deny the existence of powerful social and economic reasons for learning the grammatical norms of educated people.[1] But how these norms get established and influence the evolution of languages is a (fascinating) question for sociolinguistics and/or historical linguistics, not for syntactic theory. Hence, it is beyond the scope of this book. Similarly, we will 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 and Cameron 1995 for insightful discussion of these issues.) Our concern instead is with language as it is used in everyday communication; and the rules of prescriptive grammar are of little help in describing actual usage.

[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).

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. We employ this technique throughout this book, using our own intuitions about English as our principal source of data. In keeping with standard linguistic practice, we will use an asterisk to mark an expression that is not well-formed – that is, an expression that doesn't 'sound good' to our ears. 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*?

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 prescriptive grammarians 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 objects 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

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

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', on the other hand, refers only to the ways in which words combine into phrases, and phrases into sentences – the form or structure of well-formed expressions.

Linguists divide grammar into 'syntax', 'semantics' (the study of linguistic meaning), 'morphology' (the study of word structure), and 'phonology' (the study of the sound patterns of language). Although these distinctions are conceptually clear, many phenomena in natural languages involve more than one of these components of grammar.

1.2 An Extended Example: Reflexive and Nonreflexive Pronouns

To get a feel for the sort of research syntacticians conduct, consider the following question:[4]

> 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 ≠ her]
d. She likes herself.
e. Nobody likes us.
f. *Leslie 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 referent (i.e. a preceding COREFERENTIAL expression); a *nonreflexive* pronoun cannot appear in a sentence that contains such an expression.

[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 serious 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.

[4] The presentation in this section owes much to the pedagogy of David Perlmutter; see Perlmutter and Soames (1979: chapters 2 and 3).

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 # 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.*Leslie told us about us.
l. Leslie told us about ourselves.
m.*Leslie told ourselves about us.
n.*Leslie 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 Leslie likes us.
b.*We think that Leslie 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 (1) and (2) contain only simple sentences.

Exercise 1: Some Other Subordinate Clauses

> Throughout the book we have provided exercises designed to allow you to test your understanding of the material being presented. Answers to these exercises can be found beginning on page 543.

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 # 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 your letter to us in the trash.
f.*We found your 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 refer to the participants in that action or state. These are often called 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 *your 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 referent. 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 our*selves),* 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,[5] 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 $them_1$ to vote for $them_2$. [$them_1 \neq them_2$]
b. We appealed to them to vote for themselves.
c. We appealed to them to vote for us.

(7) a. We appeared to them to vote for them.
b.*We appeared to them to vote for themselves.
c. We appeared to them to vote for ourselves.

In (6), the pronouns indicate that them 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 not directly observable. We will return to the analysis of such cases in later chapters.

You can see that things get quite complex quite fast, requiring abstract notions like 'coreference', being 'arguments of the same verb', and 'phantom arguments' that the rules for pronoun type must make reference to. 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 an intricate system, which is not directly accessible to conscious reflection.
- 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.

[5]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.

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 languages in between).[6] 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 systematic 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'[7]) study to the 'synchronic'[8] 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

[6]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.

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

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

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. 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 to argue that 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, and 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.[9]

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.[10]

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 certain types of languages.

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

[10]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.

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 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 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 often uses the term 'Universal Grammar' to mean the innate endowment that makes language acquisition possible. 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. We need to *pen* the pigs to keep them from getting into the corn. ('to put in a fenced enclosure')
d. They should *pen* the letter quickly. ('to write')
e. 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')
1. 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.

(1la) 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 woth or good is modifying (the verb or the noun).

These two types of 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.

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* lzeutenant *governor/50 states/oxygen*-based life here). Moreover, although certain proper names like Sally Ride, Sandra Day *O'Connor*, or *Condoleezza* Rice 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, 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.

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 literally said. Other examples of such inference are similar, but perhaps more subtle. A famous example[11] 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

[11]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.

another with considerable precision – far more precision than the language itself would seem to allow. 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.

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) rnade 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 finish reading this sentence, you do not need to go back and think about whether to interpret on as in (14a) or (14b). The decision about how to construe on is made by the time the word atom is understood.

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 now facing the study of language.

Syntax plays a crucial role in all this. It imposes constraints on how sentences can or cannot be construed. 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 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[12] 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 people with certain disabilities 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, these programs 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, there are many potential uses for software to reduce the number of keystrokes needed to input text, 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

[12] That is, English, Japanese, Swahili, etc. in contrast to programming languages or the languages of mathematical logic.

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 testified against defendants who incriminated them?
     b. Which witnesses testified against defendants who incriminated 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 practical 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 these 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 Phenomena Addressed

Over the next fifteen chapters, we develop theoretical apparatus to provide precise syntactic descriptions. We motivate our formal machinery by examining various phenomena in English. We also address the applicability of our theory to other languages, particularly in some of the problems.

The following is a brief overview of the most important phenomena of English that we deal with. We omit many subtleties in this preliminary survey, but this should give readers a rough sense of what is to come.

- Languages are infinite. That is, there is no limit to the length of sentences, and most utterances have never been uttered before.
- There are different types of words – such as nouns, verbs, etc. – which occur in different linguistic environments.
- There are many constraints on word order in English. For example, we would say *Pat writes books,* not **Writes Pat books, *Books writes Pat,* or **Pat books writes.*
- Some verbs require objects, some disallow them, and some take them optionally. So we get: *Pat devoured the steak,* but not **Pat devoured; Pat dined,* but not **Pat dined the steak*; and both *Pat ate the steak,* and *Pat ate.*
- Verbs agree with their subjects, so (in standard English) we wouldn't say **Pat write books* or **Books is interesting.*

- There is also a kind of agreement within noun phrases; for example, *this bird* but not **this birds*; *these birds* but not **these bird*; and *much water* but not **much bird* or **much birds.*
- Some pronouns have a different form depending on whether they are the subject of the verb or the object: *I saw them* vs. **Me saw them* or **I saw they.*
- As was discussed in Section 1.2, reflexive and nonreflexive pronouns have different distributions, based on the location of their antecedent.
- Commands are usually expressed by sentences without subjects, whose verbs show no agreement or tense marking, such as *Be careful!*
- Verbs come in a variety of forms, depending on their tense and on properties of their subject. Nouns usually have two forms: singular and plural. There are also cases of nouns and verbs that are morphologically and semantically related, such as *drive* and *driver.*
- Sentences with transitive verbs typically have counterparts in the passive voice, e.g. *The dog chased the cat* and *The cat was chased by the dog.*
- The word *there* often occurs as the subject of sentences expressing existential statements, as in *There is a unicorn in the garden.*
- The word *it* in sentences like *It is clear that syntax is difficult* does not refer to anything. This sentence is synonymous with *That syntax is difficult is clear,* where the word *it* doesn't even appear.
- Certain combinations of words, known as idioms, have conventional meanings, not straightforwardly inferable from the meanings of the words within them. Idioms vary in their syntactic versatility. Examples of idioms are *keep tabs on* and *take advantage of.*
- Pairs of sentences like *Pat seems to be helpful* and *Pat tries to be helpful,* though superficially similar, are very different in the semantic relationship between the subject and the main verb. This difference is reflected in the syntax in several ways; for example, *seems* but not *tries* can have the existential *there* as a subject: *There seems to be a unicorn in the garden* vs. **There tries to be a unicorn in the garden.*
- There is a similar contrast between the superficially similar verbs *expect* and *persuade: We expected several students to be at the talk* and *We persuaded several students to be at the talk* vs. *We expected there to be several students at the talk* but **We persuaded there to be several students at the talk.*
- Auxiliary ('helping') verbs in English (like *can, is, have,* and *do)* have a number of special properties, notably:
  - fixed ordering *(They have been sleeping* vs **They are having slept)*
  - occurring at the beginning of yes-no questions *(Are they sleeping?)*
  - occuring immediately before *not (They are not sleeping)*
  - taking the contracted form of *not,* written *n't (They aren't sleeping)*
  - occurring before elliptical (missing) verb phrases (We *aren't sleeping, but they are)*

- There is considerable dialectal variation in the English auxiliary system, notably British/American differences in the use of auxiliary *have* (Have you the time?) and the existence of a silent version of *as* in African American Vernacular English *(She the teacher)*.
- A number of constructions (such as '*wh*-questions') involve pairing a phrase at the beginning of a sentence with a 'gap' – that is, a missing element later in the sentence. For example, in What are you talking about? what functions as the object of the preposition about, even though it doesn't appear where the object of a preposition normally does.

These are some of the kinds of facts that a complete grammar of English should account for. We want our grammar to be precise and detailed enough to make claims about the structure and meanings of as many types of sentence as possible. We also want these descriptions to be psychologically realistic and computationally tractable. Finally, despite our focus on English, our descriptive vocabulary and formalization should be applicable to all natural languages.

1.6 Summary

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. 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.

1.7 Further Reading

An entertaining (but by no means unbiased) 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, Cameron 1995, 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 1986b. There have been few recent attempts at surveying work in (human or machine) sentence processing. Fodor et al. 1974 is a comprehensive review of early psycholinguistic work within the Chomskyan paradigm, but it is now quite dated. Garrett 1990 and Fodor 1995 are more recent, but much more limited in scope. For a readable, linguistically oriented, general introduction to computational linguistics, see Jurafsky and Martin 2000.

1.8 Problems

⚠ *This symbol before a problem indicates that it should not be skipped. The problem either deals with material that is of central importance in the chapter, or it introduces something that will be discussed or used in subsequent chapters.*

⚠ Problem 1: Judging **Examples**

For each of the following examples, indicate whether it is acceptable or unacceptable. (Don't worry about what prescriptivists might say: we want native speaker intuitions of what sounds right). 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 (for any variety of English, to the best of your knowledge),
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 (your own or another); 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) Terry left tomorrow.
(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.

Problem 2: 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. Is there some general property that all antecedents of reciprocals have that not all antecedents of reflexives have? Give both grammatical and ungrammatical examples to make your point.

B. Aside from the difference noted in part (A), do reciprocals behave like reflexives with respect to Hypothesis III? Provide evidence for your answer, including both acceptable and unacceptable examples, illustrating the full range of types of configurations we considered in motivating Hypothesis III.

C. Is the behavior of reciprocals similar to that of reflexives in imperative sentences and in sentences containing *appeal* and *appear?* Again, support your answer with both positive and negative evidence.

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.]

Problem 3: Ambiguity

Give a brief description of each ambiguity illustrated in (12) on page 11, saying what the source of ambiguity is – that is, whether it is lexical, structural (modificational), or both.

2

Some Simple Theories of Grammar

2.1 Introduction

Among the key points in the previous chapter were the following:

- Language is rule-governed.
- The rules aren't the ones we were taught in school.
- Much of our linguistic knowledge is unconscious, so we have to get at it indirectly; one way of doing this is to consult intuitions of what sounds natural.'

In this text, we have a number of objectives. First, we will work toward developing a set of rules that will correctly predict the acceptability of (a large subset of) English sentences. The ultimate goal is a grammar that can tell us for any arbitrary string of English words whether or not it is a well-formed sentence. Thus we will again and again be engaged in the exercise of formulating a grammar that generates a certain set of word strings – the sentences predicted to be grammatical according to that grammar. We will then examine particular members of that set and ask ourselves: 'Is this example acceptable?' The goal then reduces to trying to make the set of sentences generated by our grammar match the set of sentences that we intuitively judge to be acceptable.[1]

A second of our objectives is to consider how the grammar of English differs from the grammar of other languages (or how the grammar of standard American English differs from those of other varieties of English). The conception of grammar we develop will involve general principles that are just as applicable (as we will see in various exercises) to superficially different languages as they are to English. Ultimately, much of the outward differences among languages can be viewed as differences in vocabulary.

This leads directly to our final goal: to consider what our findings might tell us about human linguistic abilities in general. As we develop grammars that include principles of considerable generality, we will begin to see constructs that may have universal applicability to human language. Explicit formulation of such constructs will help us evaluate Chomsky's idea, discussed briefly in Chapter 1, that humans' innate linguistic endowment is a kind of 'Universal Grammar'.

[1] Of course there may be other interacting factors that cause grammatical sentences to sound less than fully acceptable – see Chapter 9 for further discussion. In addition, we don't all speak exactly the same variety of English, though we will assume that existing varieties are sufficiently similar for us to engage in a meaningful discussion of quite a bit of English grammar; see Chapter 15 for more discussion.

In developing the informal rules for reflexive and nonreflexive pronouns in Chapter 1, we assumed that we already knew a lot about the structure of the sentences we were looking at – that is, we talked about subjects, objects, clauses, and so forth. In fact, a fully worked out theory of reflexive and nonreflexive pronouns is going to require that many other aspects of syntactic theory get worked out first. We begin this grammar development process in the present chapter.

We will consider several candidates for theories of English grammar. We begin by quickly dismissing certain simple-minded approaches. We spend more time on a formalism known as 'context-free grammar', which serves as a starting point for most modern theories of syntax. Appendix B includes a brief overview of some of the most important schools of thought within the paradigm of generative grammar, situating the approach developed in this text with respect to some alternatives.

2.2 Two Simplistic Syntactic Theories

2.2.1 Lists as Grammars

The simplest imaginable syntactic theory asserts that a grammar consists of a list of all the well-formed sentences in the language. The most obvious problem with such a proposal is that the list would have to be too long. There is no fixed finite bound on the length of English sentences, as can be seen from the following sequence:

(1) Some sentences go on and on.
Some sentences go on and on and on.
Some sentences go on and on and on and on.
Some sentences go on and on and on and on and on.

Every example in this sequence is an acceptable English sentence. Since there is no bound on their size, it follows that the number of sentences in the list must be infinite. Hence there are infinitely many sentences of English. Since human brains are finite, they cannot store infinite lists. Consequently, there must be some more compact way of encoding the grammatical knowledge that speakers of English possess.

Moreover, there are generalizations about the structure of English that an adequate grammar should express. For example, consider a hypothetical language consisting of infinitely many sentences similar to those in (1), except that every other sentence reversed the order of the words some and sentences[2]

(2) An Impossible Hypothetical Language:

Some sentences go on and on.
Sentences some go on and on and on.
Some sentences go on and on and on and on.
Sentences some go on and on and on and on and on.
*Sentences some go on and on.
*Some sentences go on and on and on.

[2]The asterisks in **(2)** are intended to indicate the ungrammaticality of the strings in the hypothetical language under discussion, not in normal English.

*Sentences some go on and on and on and on.
*Some sentences go on and on and on and on and on.
...

Of course, none of these sentences[3] where the word sentences precedes the word some is a well-formed English sentence. Moreover, no natural language exhibits patterns of that sort – in this case, having word order depend on whether the length of the sentence is divisible by 4. A syntactic theory that sheds light on human linguistic abilities ought to explain why such patterns do not occur in human languages. But a theory that said that grammars consisted only of lists of sentences could not do that. If grammars were just lists, then there would be no patterns that would be excluded – and none that would be expected, either.

This form of argument – that a certain theory of grammar fails to 'capture a linguistically significant generalization' – is very common in generative grammar. It takes for granted the idea that language is 'rule governed', that is, that language is a combinatoric system whose operations are 'out there' to be discovered by empirical investigation. If a particular characterization of the way a language works fails to distinguish in a principled way between naturally occurring types of patterns and those that do not occur then it's assumed to be the wrong characterization of the grammar of that language. Likewise, if a theory of grammar cannot describe some phenomenon without excessive redundancy and complications, we assume something is wrong with it. We will see this kind of argumentation again, in connection with proposals that are more plausible than the 'grammars-as-lists' idea. In Chapter 9, we will argue that (perhaps surprisingly), a grammar motivated largely on the basis of considerations of parsimony seems to be a good candidate for a psychological model of the knowledge of language that is employed in speaking and understanding.

2.2.2 Regular Expressions

A natural first step toward allowing grammars to capture generalizations is to classify words into what are often called 'parts of speech' or 'grammatical categories'. There are large numbers of words that behave in similar ways syntactically. For example, the words apple, book, color, and dog all can appear in roughly the same contexts, such as the following:

(3) a. That __ surprised me.
b. I noticed the __ .
c. They were interested in his __ .
d. This is my favorite __ .

Moreover, they all have plural forms that can be constructed in similar ways (orthographically, simply by adding an *-s*).

Traditionally, the vocabulary of a language is sorted into nouns, verbs, etc. based on loose semantic characterizations (e.g. 'a noun is a word that refers to a person, place, or thing'). While there is undoubtedly a grain of insight at the heart of such definitions,

[3]Note that we are already slipping into a common, but imprecise, way of talking about unacceptable strings of words as 'sentences'.

we can make use of this division into grammatical categories without committing ourselves to any semantic basis for them. For our purposes, it is sufficient that there are classes of words that may occur grammatically in the same environments. Our theory of grammar can capture their common behavior by formulating patterns or rules in terms of categories, not individual words.

Someone might, then, propose that the grammar of English is a list of patterns, stated in terms of grammatical categories, together with a lexicon – that is, a list of words and their categories. For example, the patterns could include (among many others):

(4) a. ARTICLE NOUN VERB
 b. ARTICLE NOUN VERB ARTICLE NOUN

And the lexicon could include (likewise, among many others):

(5) a. Articles: a, the
 b. Nouns: cat, dog
 c. Verbs: attacked, scratched

This mini-grammar licenses forty well-formed English sentences, and captures a few generalizations. However, a grammar that consists of a list of patterns still suffers from the first drawback of the theory of grammars as lists of sentences: it can only account for a finite number of sentences, while a natural language is an infinite set of sentences. For example, such a grammar will still be incapable of dealing with all of the sentences in the infinite sequence illustrated in (1).

We can enhance our theory of grammar so as to permit infinite numbers of sentences by introducing a device that extends its descriptive power. In particular, the problem associated with (1) can be handled using what is known as the 'Kleene star'.[4] Notated as a superscripted asterisk, the Kleene star is interpreted to mean that the expression it is attached to can be repeated any finite number of times (including zero). Thus, the examples in (1) could be abbreviated as follows:

(6) Some sentences go on and on [and on]*.

A closely related notation is a superscripted plus sign (called the Kleene plus), meaning that one or more occurrences of the expression it is attached to are permissible. Hence, another way of expressing the same pattern would be:

(7) Some sentences go on [and on]$^{+}$.

We shall employ these, as well as two common abbreviatory devices. The first is simply to put parentheses around material that is optional. For example, the two sentence patterns in (4) could be collapsed into: ARTICLE NOUN VERB (ARTICLE NOUN). The second abbreviatory device is a vertical bar, which is used to separate alternatives.[5] For example, if we wished to expand the mini-grammar in (4) to include sentences like *The dog looked big*, we could add the pattern ARTICLE NOUN VERB ADJECTIVE and collapse it with the previous patterns as: ARTICLE NOUN VERB (ARTICLE NOUN)|ADJECTIVE. Of

[4]Named after the mathematician Stephen Kleene.
[5]This is the notation standardly used in computer science and in the study of mathematical properties of grammatical systems. Descriptive linguists tend lo use curly brackets to annotate alternatives.

course, we would also have to add the verb *looked* and the adjective *big* to the lexicon.[6]

Patterns making use of the devices just described – Kleene star, Kleene plus, parentheses for optionality, and the vertical bar for alternatives – are known as 'regular expressions'.[7] A great deal is known about what sorts of patterns can and cannot be represented with regular expressions (see, for example, Hopcroft et al. 2001, chaps. 2 and 3), and a number of scholars have argued that natural languages in fact exhibit patterns that are beyond the descriptive capacity of regular expressions (see Bar-Hillel and Shamir 1960, secs. 5 and 6). The most convincing arguments for employing a grammatical formalism richer than regular expressions, however, have to do with the need to capture generalizations.

In (4), the string ARTICLE NOUN occurs twice, once before the verb and once after it. Notice that there are other options possible in both of these positions:

(8) a. *Dogs* chase *cats.*
 b. *A large dog* chased *a small cat.*
 c. *A dog with brown spots* chased *a cat with no tail.*

Moreover, these are not the only positions in which the same strings can occur:

(9) a. Some people yell at *(the) (noisy) dogs (in my nezghborhood).*
 b. Some people consider *(the) (noisy) dogs (in my neighborhood)* dangerous.

Even with the abbreviatory devices available in regular expressions, the same lengthy string of symbols – something like (ARTICLE) (ADJECTIVE) NOUN (PREPOSITION ARTICLE NOUN) – will have to appear over and over again in the patterns that constitute the grammar. Moreover, the recurring patterns are in fact considerably more complicated than those illustrated so far. Strings of other forms, such as *the nozsy annoying dogs, the dogs that live in my neighborhood,* or *Rover, Fido, and Lassie* can all occur in just the same positions. It would clearly siniplify the grammar if we could give this apparently infinite set of strings a name and say that any string from the set can appear in certain positions in a sentence.

Furthermore, as we have already seen, an adequate theory of syntax must somehow account for the fact that a given string of words can sometimes be put together in more than one way. If there is no more to grammar than lists of recurring patterns, where these are defined in terms of parts of speech, then there is no apparent way to talk about the ambiguity of sentences like those in (10).

(10) a. We enjoyed the movie with Cher.
 b. The room was filled with noisy children and animals.
 c. People with children who use drugs should be locked up.
 d. I saw the astronomer with a telescope,

[6]This extension of the grammar would license some unacceptable strings, e.g. **The cat scratched big.* Overgeneration is always a danger when extending a grammar, as we will see in subsequent chapters.
[7]This is not intended as a rigorous definition of regular expressions. A precise definition would include the requirement that the empty string is a regular expression, and would probably omit some of the devices mentioned in the text (because they can be defined in terms of others). Incidentally, readers who use computers with the UNIX operating system may be familiar with the command 'grep'. This stands for 'Global Regular Expression Printer'.

In the first sentence, it can be us or the movie that is 'with Cher'; in the second, it can be either just the children or both the children and the animals that are noisy; in the third, it can be the children or their parents who use drugs, and so forth. None of these ambiguities can be plausibly attributed to a lexical ambiguity. Rather, they seem to result from different ways of grouping the words.

In short, the fundamental defect of regular expressions as a theory of grammar is that they provide no means for representing the fact that a string of several words may constitute a unit. The same holds true of several other formalisms that are provably equivalent to regular expressions (including what is known as 'finite-state grammar').

The recurrent strings we have been seeing are usually called 'phrases' or '(syntactic) constituents'.[8] Phrases, like words, come in different types. All of the italicized phrases in (8)–(9) above obligatorily include a noun, so they are called 'Noun Phrases'. The next natural enrichment of our theory of grammar is to permit our regular expressions to include not only words and parts of speech, but also phrase types. Then we also need to provide (similarly enriched) regular expressions to provide the patterns for each type of phrase. The technical name for this theory of grammar is 'context-free phrase structure grammar' or simply 'context-free grammar', sometimes abbreviated as CFG. CFGs, which will also let us begin to talk about structural ambiguity like that illustrated in (10), form the starting point for most serious attempts to develop formal grammars for natural languages.

2.3 Context-Free Phrase Structure Grammar

The term 'grammatical category' now covers not only the parts of speech, but also types of phrase, such as noun phrase and prepositional phrase. To distinguish the two types, we will sometimes use the terms 'lexical category' (for parts of speech) and 'nonlexical category' or 'phrasal category' to mean types of phrase. For convenience, we will abbreviate them, so that 'NOUN' becomes 'N', 'NOUN PHRASE' becomes 'NP', etc.

A coiltext-free phrase structure grammar has two parts:

- A LEXICON, consisting of a list of words, with their associated grammatical categories.[9]
- A set of RULES of the form A $\rightarrow \varphi$ where A is a nonlexical category, and 'φ' stands for a regular expression formed from lexical and/or nonlexical categories; the arrow is to be interpreted as meaning, roughly, 'can consist of'. These rules are called 'phrase structure rules'.

The left-hand side of each rule specifies a phrase type (including the sentence as a type of phrase), and the right-hand side gives a possible pattern for that type of phrase. Because

[8] There is a minor difference in the way these terms are used: linguists often use 'phrase' in contrast to 'word' to mean something longer, whereas words are always treated as a species of constituent.

[9] This conception of a lexicon leaves out some crucial information. In particular, it leaves out information about the meanings and uses of words, except what might be generally associated with the grammatical categories. While this impoverished conception is standard in the formal theory of CFG, attempts to use CFG to describe natural languages have made use of lexicons that also included semantic information. The lexicon we develop in subsequent chapters will be quite rich in structure.

phrasal categories can appear on the right-hand sides of rules, it is possible to have phrases embedded within other phrases. This permits CFGs to express regularities that seem like accidents when only regular expressions are permitted.

A CFG has a designated 'initial symbol', usually notated 'S' (for 'sentence'). Any string of words that can be derived from the initial symbol by means of a sequence of applications of the rules of the grammar is licensed (or, as linguists like to say, 'generated') by the grammar. The language a grammar generates is simply the collection of all of the sentences it generates.[10]

To illustrate how a CFG works, consider the following grammar: (We use 'D' for 'Determiner', which includes what we have up to now been calling 'articles': but will eventually also be used to cover some other things, such as *two* and my; 'A' stands for 'Adjective'; 'P' stands for 'Preposition'.)

(11) a. Rules:

S → NP VP
NP → (D) A* N PP*
VP → V (NP) (PP)
PP → P NP

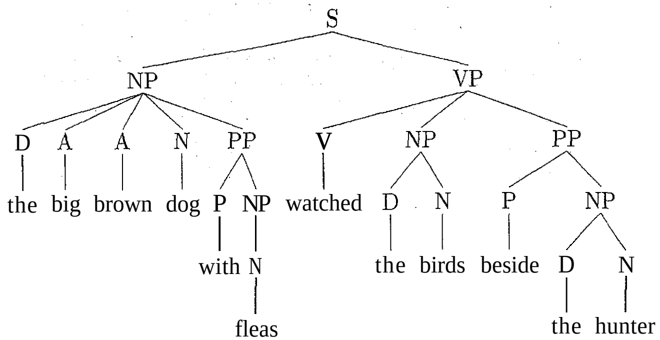
b. Lexicon:

D: the, some
A: big, brown, old
N: birds, fleas, dog, hunter
V: attack, ate, watched
P: for, beside, with

This grammar generates infinitely many English sentences. Let us look in detail at how it generates one sentence: *The big brown dog with fleas watched the birds beside the hunter*. We start with the symbol S, for 'Sentence'. This must consist of the sequence NP VP, since the first rule is the only one with S on the left-hand side. The second rule allows a wide range of possibilities for the NP, one of which is D A A N PP. This PP must consist of a P followed by an NP, by the fourth rule, and the NP so introduced may consist of just an N. The third rule allows VP to consist of V NP PP, and this NP can consist of a D followed by an N. Lastly, the final PP again consists of a P followed by an NP, and this NP also consists of a D followed by an N. Putting these steps together the S may consist of the string D A A N P N V D N P D N, which can be converted into the desired sentence by inserting appropriate words in place of their lexical categories. All of this can be summarized in the following figure (called a 'tree diagram'):

[10] Our definition of CFG differs slightly from the standard ones found in textbooks on formal language theory. Those definitions restrict the right-hand side of rules to finite strings of categories, whereas we allow any regular expression, including those containing the Kleene operators. This difference does not affect the languages that can be generated, although the trees associated with those sentences (see the next section) will be different in some cases.

(12)

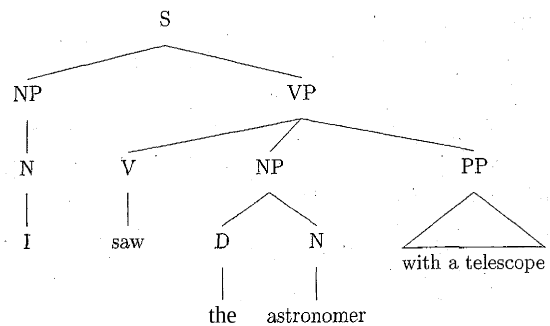


Note that certain sentences generated by this grammar can be associated with more than one tree. (Indeed, the example just given is one such sentence, but finding the other tree will be left as an exercise.) This illustrates how CFGs can overcome the second defect of regular expressions pointed out at the end of the previous section. Recall the ambiguity of (13):

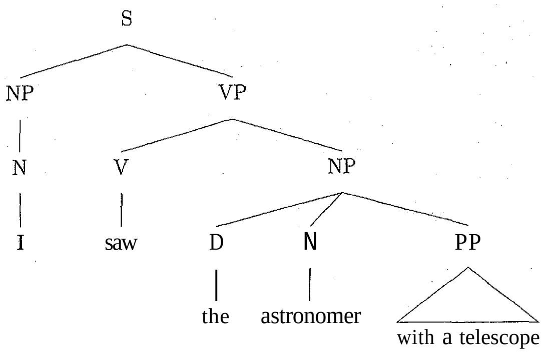
(13) I saw the astronomer with a telescope.

The distinct interpretations of this sentence ('I used the telescope to see the astronomer'; 'I saw the astronomer who had a telescope') correspond to distinct tree structures that our grammar will assign to this string of words. The first interpretation corresponds to (14a) and the latter to (14b):

(14) a.



b.



CFG thus provides us with a straightforward mechanism for expressing such ambiguities, whereas grammars that use only regular expressions don't.

The normal way of talking about words and phrases is to say that certain sequences of words (or categories) 'form a constituent'. What this means is that these strings function as units for some purpose (for example, the interpretation of modifiers) within the sentences in which they appear. So in (12), the sequence *with fleas* forms a PP constituent (as does the sequence P NP), *the big brown dog with fleas* forms an NP, and the sequence *dog with fleas* forms no constituent. Structural ambiguity arises whenever a string of words can form constituents in more than one way.

Exercise 1: Practice with CFG

Assume the CFG grammar given in (11). Draw the tree structure for the other interpretation (i.e. not the one shown in (12)) of *The big brown dog with fleas watched the birds beside the hunter*.

2.4 Applying Context-Free Grammar

In the previous sections, we introduced the formalism of context-free grammar and showed how it allows us to generate infinite collections of English sentences with simple rules. We also showed how it can provide a rather natural representation of certain ambiguities we find in natural languages. But the grammar we presented was just a teaching tool, designed to illustrate certain properties of the formalism; it was not intended to be taken seriously as an attempt to analyze the structure of English. In this section, we begin by motivating some phrase structure rules for English. In the course of doing this, we develop a new test for determining which strings of words are constituents. We also introduce a new abbreviatory convention that permits us to collapse many of our phrase structure rules into rule schemas.

2.4.1 Some Phrase Structure Rules for English

For the most part, we will use the traditional parts of speech, such as noun, verb, adjective, and preposition. In some cases, we will find it useful to introduce grammatical categories that might be new to readers, and we may apply the traditional labels somewhat differently than in traditional grammar books. But the traditional classification of words into types has proved to be an extremely useful categorization over the past two millennia, and we see no reason to abandon it wholesale.

We turn now to phrases, beginning with noun phrases.

Noun Phrases

Nouns can appear in a number of positions, e.g. those occupied by the three nouns in *Dogs give people fleas*. These same positions also allow sequences of an article followed by a noun, as in *The child gave the dog a bath*. Since the place of the article can also be filled by demonstratives (e.g. *this, these*), possessives (e.g. *my, their*), or quantifiers (e.g. *each, some, many*), we use the more general term 'determiner' (abbreviated D) for this category. We can capture these facts by positing a type of phrase we'll call NP (for 'noun phrase'), and the rule NP → (D) N. As we saw earlier in this chapter, this rule will need to be elaborated later to include adjectives and other modifiers. First, however, we should consider a type of construction we have not yet discussed.

Coordination

To account for examples like *A dog, a cat, and a wombat fought,* we want a rule that allows sequences of NPs, with *and* before the last one, to appear where simple NPs can occur. A rule that does this is NP → NP^+ CONJ NP. (Recall that NP^+ means a string of one or more NPs).

Whole sentences can also be conjoined, as in *The dog barked, the donkey brayed, and the pig squealed.*[11] Again, we could posit a rule like S → S^+ CONJ S. But now we have two rules that look an awful lot alike. We can collapse them into one rule schema as follows, where the variable 'X' can be replaced by any grammatical category name (and 'CONJ' is the category of conjunctions like *and* and *or,* which will have to be listed in the lexicon):

(15) X → X^+ CONJ X.

Now we have made a claim that goes well beyond the data that motivated the rule, namely, that elements of any category can be conjoined in the same way. If this is correct, then we can use it as a test to see whether a particular string of words should be treated as a phrase. In fact, coordinate conjunction is widely used as a test for constituency – that is, as a test for which strings of words form phrases. Though it is not an infallible diagnostic, we will use it as one of our sources of evidence for constituent structure.

Verb Phrases

Consider (16):

(16) A neighbor yelled, chased the cat, and gave the dog a bone.

(16) contains the coordination of strings consisting of V, V NP, and V NP NP. According to (15), this means that all three strings are constituents of the same type. Hence, we posit a constituent which we'll call VP, described by the rule VP → V (NP) (NP). VP is introduced by the rule S → NP VP. A tree structure for the coordinate VP in (16) would be the following:

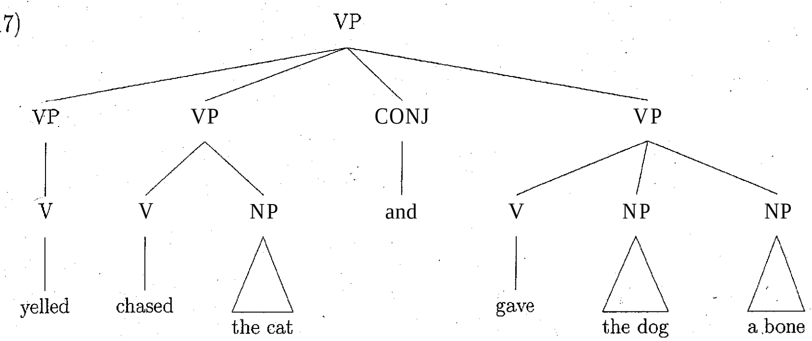
[11]There are other kinds of coordinate sentences that we are leaving aside here – in particular, elliptical sentences that involve coordination of nonconstituent sequences:

(i) Chris likes blue and Pat green.

(ii) Leslie wants to go home tomorrow, and Terry. too.

Notice that this kind of sentence, which will not be treated by the coordination rule discussed in the text, has a characteristic intonation pattern – the elements after the conjunction form separate intonational units separated by pauses.

(17)



Prepositional Phrases

Expressions like *in Rome* or *at noon* that denote places or times ('locative' and 'temporal' expressions, as linguists would say) can be added to almost any sentence, and to NPs, too. For example:

(18) a. The fool yelled at noon.
b. This disease gave Leslie a fever in Rome.
c. A tourist in Rome laughed.

These are constituents, as indicated by examples like A *tourist yelled at noon and at midnight, in Rome and in Paris.* We can get lots of them in one sentence, for example, A *tourist laughed on the street in Rome at noon on Tuesday.* These facts can be incorporated into the grammar in terms of the phrasal category PP (for 'prepositional phrase'), and the rules:

(19) a. PP → P NP
b. VP → VP PP

Since the second rule has VP on both the right and left sides of the arrow, it can apply to its own output. (Such a rule is known as a RECURSIVE rule).[12] Each time it applies, it adds a PP to the tree structure. Thus, this recursive rule permits arbitrary numbers of PPs within a VP.

As mentioned earlier, locative and temporal PPs can also occur in NPs, for example, *A protest on the street in Rome on Tuesday at noon disrupted traffic.* The most obvious analysis to consider for this would be a rule that said: NP → NP PP. However, we're going to adopt a slightly more complex analysis. We posit a new nonlexical category, which we'll call NOM (for 'nominal'), and we replace our old rule: NP → (D) N with the following:

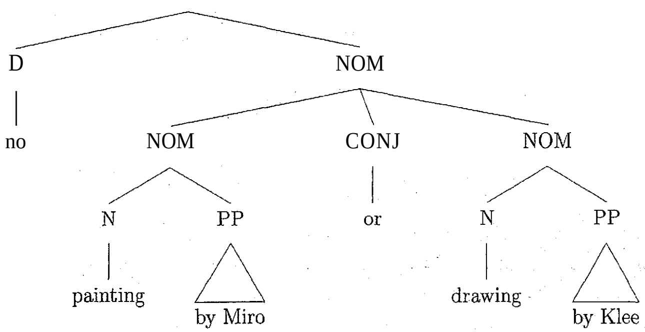
(20) a. NP → (D) NOM
b. NOM → N
c. NOM → NOM PP

[12]More generally, we use the term RECURSION whenever rules permit a constituent to occur within a larger constituent of the same type.

The category NOM will be very useful later in the text. For now, we will justify it with the following *sentences:*

(21) a. The love of my life and mother of my children would never do such a thing.
b. The museum displayed no painting by Miro or drawing by Klee.

(21b) means that the museum displayed neither paintings by Miro nor drawings by Klee. That is, the determiner *no* must be understood as 'having scope' over both *painting by Miro* and *drawing by Klee* – it applies to both phrases. The most natural noun phrase structure to associate with this interpretation is:

(22)


This, in turn, is possible with our current rules if *painting by Miro or drawing by Klee* is a conjoined NOM. It would not be possible without NOM.

Similarly, for (21a), *the* has scope over both *love of my life* and *mother of my children* and hence provides motivation for an analysis involving coordination of NOM constituents.

2.4.2 Summary of Grammar Rules

Our grammar now has the following rules:

(23)
S → NP VP
NP → (D) NOM
VP → V (NP) (NP)
NOM → N
NOM → NOM PP
VP → VP PP
PP → P NP
X → X^+ CONJ X

In motivating this grammar, we used three types of evidence for deciding how to divide sentences up into constituents:

- In ambiguous sentences, a particular division into constituents sometimes can provide an account of the ambiguity in terms of where some constituent is attached (as in (14)).

- Coordinate conjunction usually combines constituents, so strings that can serve as coordinate conjuncts are probably constituents (as we' argued for VPs, PPs, and NOMs in the last few pages).
- Strings that can appear in multiple environments are typically constituents.

We actually used this last type of argument' for constituent structure only once. That was when we motivated the constituent NP by observing that pretty-much the same strings could appear as subject, object, or object of a preposition. In fact, variants of this type of evidence are commonly used in linguistics to motivate particular choices about phrase structure. In particular, there are certain environments that linguists use as diagnostics for constituency – that is, as a way of testing whether a given string is a constituent.

Probably the most common such diagnostic is occurrence. before the subject of a sentence. In the appropriate contexts, various types of phrases are acceptable at the beginning of a sentence. This is illustrated in the following sentences, with the constituent in question italicized, and its label indicated in parentheses after the example:

(24) a. Most elections are quickly forgotten, but *the election of* ,2000, everyone will remember for a long time. (NP)
b. You asked me to fix the drain, and *fix the drain*, I shall. (VP)
c. In *the mornzng,* they drink tea. (PP)

Another environment that is frequently used as a diagnostic for constituency is what is sometimes called the 'cleft' construction. It has the following form: *It is* (or *was)* __ *that* ... For example:

(25) a. It was *a book about syntax* that she was reading. (NP)
b. It is *study for the exam* that *I* urgently need to do. (VP)
c. It is *after lunch* that they always fall asleep. (PP)

Such diagnostics can be very useful in deciding how to divide up sentences into phrases. However, some caution in their use is advisable. Some diagnostics work only for some kinds of constituents. For example, while coordination provided some motivation for positing NOM as a constituent (see (21)), NOM cannot appear at the beginning of a sentence or in a cleft:

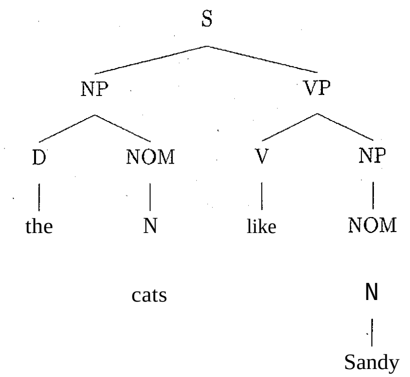
(26) a.*Many artists were represented, but painting by Klee or drawing by Miro' the museum displayed no.
b.*It is painting by Klee or drawing by Miro that the museum displays no.

More generally, these tests should be regarded only as heuristics, for there may be cases where they give conflicting or questionable results. Nevertheless, they can be very useful in deciding how to analyze particular sentences, and we will make use of them in the chapters to come.

2.5 Trees Revisited

In grouping words into phrases and smaller phrases into larger ones, we are assigning internal structure to sentences. As noted earlier, this structure can be represented as a tree diagram. For example, our grammar so far generates the following tree:

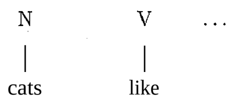
(27)



A tree is said to consist of NODES, connected by BRANCHES. A node above another on a branch is said to DOMINATE it. The nodes at the bottom of the tree – that is, those that do not dominate anything else – are referred to as TERMINAL (or LEAF) nodes. A node right above another node on a tree is said to be its MOTHER and to IMMEDIATELY DOMINATE it. A node right below another on a branch is said to be its DAUGHTER. Two daughters of the same mother node are, naturally, referred to as SISTERS.

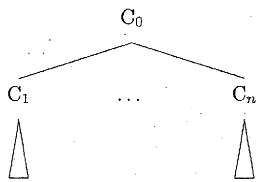
One way to think of the way in which a grammar of this kind defines (or generates) trees is as follows. First, we appeal to the lexicon (still conceived of as just a list of words paired with their grammatical categories) to tell us which lexical trees are well-formed. (By 'lexical tree', we simply mean a tree consisting of a word immediately dominated by its grammatical category.) So if cats is listed in the lexicon as belonging to the category N, and *like* is listed as a V, and so forth, then lexical structures like the following are well-formed:

(28)



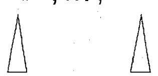
And the grammar rules'are equally straightforward. They simply tell us how well-formed trees (some of which may be lexical) can be combined into bigger ones:

(29)



is a well-formed nonlexical tree if (and only if)

C_1 , . . . , C_n are well-formed trees, and



$C_0 \rightarrow C_1 \dots C_n$ is a grammar rule.

So we can think of our grammar as generating sentences in a 'bottom-up' fashion – starting with lexical trees, and then using these to build bigger and bigger phrasal trees, until we build one whose top node is S. The set of all sentences that can be built that have S as their top node is the set of sentences the grammar generates. But note that our grammar could just as well have been used to generate sentences in a 'top-down' manner, starting with S. The set of sentences generated in this way is exactly the same. A CFG is completely neutral with respect to top-down and bottom-up perspectives on analyzing sentence structure. There is also no particular bias toward thinking of the grammar in terms of generating sentences or in terms of parsing. Instead, the grammar can be thought of as constraining the set of all possible phrase structure trees, defining a particular subset as well-formed.

Direction neutrality and process neutrality are consequences of the fact that the rules and lexical entries simply provide constraints on well-formed structure. As we will suggest in Chapter 9, these are in fact important design features of this theory (and of those we will develop that are based on it), as they facilitate the direct embedding of the abstract grammar within a model of language processing.

The lexicon and grammar rules together thus constitute a system for defining not only well-formed word strings (i.e. sentences), but also well-formed tree structures. Our statement of the relationship between the grammar rules and the well-formedness of trees is at present rather trivial, and our lexical entries still consist simply of pairings of words with parts of speech. As we modify our theory of grammar and enrich our lexicon, however, our attention will increasingly turn to a more refined characterization of which trees are well-formed.

2.6 CFG as a Theory of Natural Language Grammar

As was the case with regular expressions, the formal properties of CFG are extremely well studied (see Hopcroft et al. 2001, chaps. 4–6 for a summary). In the early 1960s, several scholars published arguments purporting to show that natural languages exhibit properties beyond the descriptive capacity of CFGs. The pioneering work in the first two decades of generative grammar was based on the assumption that these arguments were sound. Most of that work can be viewed as the development of extensions to CFG designed to deal with the richness and complexity of natural languages. Similarly, the theory we develop in this book is in essence an extended version of CFG, although our extensions are rather different in kind from some of the earlier ones.

In 1982, Geoffrey Pullum and Gerald Gazdar published a paper showing that the earlier arguments against the adequacy of CFG as a theory of natural language structure all contained empirical or mathematical flaws (or both). This led to a flurry of new work on the issue, culminating in new arguments that natural languages were not describable by CFGs. The mathematical and empirical work that resulted from this controversy substantially influenced the theory of grammar presented in this text. Many of the central papers in this debate were collected together by Savitch et al. (1987); of particular interest are Pullum and Gazdar's paper and Shieber's paper in that volume.

While the question of whether natural languages are in principle beyond the generative capacity of CFGs is of some intellectual interest, working linguists tend to be more concerned with determining what sort of formalisms can provide elegant and enlightening accounts of linguistic phenomena in practice. Hence the arguments that tend to carry the most weight are ones about what formal devices are needed to capture linguistically significant generalizations. In the next section and later chapters, we will consider some phenomena in English that suggest that the simple version of CFG introduced above needs to be extended.

Accompanying the 1980s revival of interest in the mathematical properties of natural languages, considerable attention was given to the idea that, with an appropriately designed theory of syntactic features and general principles, context-free phrase structure grammar could serve as an empirically adequate theory of natural language syntax. This proposition was explored in great detail by Gazdar et al. (1985), who developed the theory known as 'Generalized Phrase Structure Grammar' (or GPSG). Work in phrase structure grammar advanced rapidly, and GPSG quickly evolved into a new framework, now known as 'Head-driven Phrase Structure Grammar' (HPSG), whose name reflects the increased importance of information encoded in the lexical heads[13] of syntactic phrases. The theory of grammar developed in this text is most closely related to current HPSG. See Appendix B for discussion of these and other modern theories of grammar.

2.7 Problems with CFG

Two of our arguments against overly simple theories of grammar at the beginning of this chapter were that we wanted to be able to account for the infinity of language, and that we wanted to be able to account for structural ambiguity. CFG addresses these problems, but, as indicated in the previous section, simple CFGs like the ones we have seen so far are not adequate to account for the full richness of natural language syntax. This section introduces some of the problems that arise in trying to construct a CFG of English.

2.7.1 Heads

As we have seen, CFGs can provide successful analyses of quite a bit of natural language. But if our theory of natural language syntax were nothing more than CFG, our theory would fail to predict the fact that certain kinds of CF rules are much more natural than others. For example, as far as we are aware, no linguist has ever wanted to write rules like those in (30) in describing any human language:

(30) Unnatural Hypothetical Phrase Structure Rules
VP → P NP
NP → PP S

What is it that is unnatural about the rules in (30)? An intuitive answer is that the categories on the left of the rules don't seem appropriate for the sequences on the right. For example, a VP should have a verb in it. This then leads us to consider why we named NP, VP, and PP after the lexical categories N, V, and P. In each case, the phrasal category was named after a lexical category that is an obligatory part of that kind of phrase. At least in the case of NP and VP, all other parts of the phrase may sometimes be absent (e.g. Dogs *bark)*.

[13] The notion of 'head' will be discussed in Section 2.7.1 below.

The lexical category that a phrasal category derives its name from is called the HEAD of the phrase. This notion of 'headedness' plays a crucial role in all human languages and this fact points out a way in which natural language grammars differ from some kinds of CFG. The formalism of CFG, in and of itself, treats category names as arbitrary: our choice of pairs like 'N' and 'NP', etc., serves only a mnemonic function in simple CFGs. But we want our theory to do more. Many phrase structures of natural languages are headed structures, a fact we will build into the architecture of our grammatical theory. To do this, we will enrich the way we represent grammatical categories, so that we can express directly what a phrase and its head have in common. This will lead eventually to a dramatic reduction in the number of grammar rules required.

The notion of headedness is a problem for CFG because it cuts across many different phrase types, suggesting that the rules are too fine-grained. The next two subsections discuss problems of the opposite type – that is, ways in which the syntax of English is sensitive to finer-grained distinctions among grammatical categories than a simple CFG can encode.

2.7.2 Subcategorization

The few grammar rules we have so far cover only a small fragment of English. What might not be so obvious, however, is that they also overgenerate – that is, they generate strings that are not well-formed English sentences. Both *denied* and *disappeared* would be listed in the lexicon as members of the category V. This classification is necessary to account for sentences like (31):

(31) a. The defendant denied the accusation.
b. The problem disappeared.

But this classification would also permit the generation of the ungrammmatical examples in (32):

(32) a.*The defendant denied.
b.*The teacher disappeared the problem.

Similarly, the verb *handed* must be followed by two NPs, but our rules allow a VP to be expanded in such a way that any V can be followed by only one NP, or no NPs at all. That is, our current grammar fails to distinguish among the following:

(33) a. The teacher handed the student a book!
b.*The teacher handed the student.
c.*The teacher handed a book.
d.*The teacher handed.

To rule out the ungrammatical examples in (33), we need to distinguish among verbs that cannot be followed by an NP, those that must be followed by one NP, and those that must be followed by two NPs. These classes are often referred to as INTRANSITIVE, TRANSITIVE, and DITRANSITIVE verbs, respectively. In short, we need to distinguish subcategories of the category V.

One possible approach to this problem is simply to conclude that the traditional category of 'verb' is too coarse-grained for generative grammar, and that it must be replaced by at least three distinct categories, which we can call IV, TV, and DTV. We can then replace our earlier phrase structure rule

VP → V (NP) (NP)

with the following three rules:

(34) a. VP → IV
b. VP → TV NP
c. VP → DTV NP NP

2.7.3 Transitivity and Agreement

Most nouns and verbs in English have both singular and plural forms. In the case of nouns, the distinction between, say, bird and bzrds indicates whether the word is being used to refer to just one fowl or a multiplicity of them. In the case of verbs, distinctions like the one between sing and *sings* indicate whether the verb's subject refers to one or many individuals. In present tense English sentences, the plurality marking on the head noun of the subject NP and that on the verb must be consistent with each other. This is referred to as SUBJECT-VERB AGREEMENT (or sometimes just 'agreement' for short). It is illustrated in (35):

(35) a. The bird sings.
b. Birds sing.
c.*The bird sing.[14]
d.*Birds sings.

Perhaps the most obvious strategy for dealing with agreement is the one considered in the previous section. That is, we could divide our grammatical categories into smaller categories, distinguishing singular and plural forms. We could then replace the relevant phrase structure rules with more specific ones. In examples like (35), we could distinguish lexical categories of N-SG and N-PL, as well as IV-SG and IV-PL. Then we could replace the rule

S → NP VP

with two rules:

S → NP-SG VP-SG

and

S → NP-PL VP-PL

But since the marking for number appears on the head noun and head verb, other rules would also have to be changed. Specifically, the rules expanding NP and VP all would have to be divided into pairs of rules expanding NP-SG, NP-PL, VP-SG, and VP-PL. Hence, we would need all of the following:

(36) a. NP-SG → (D) NOM-SG
b. NP-PL → (D) NQM-PL
c. NOM-SG → NOM-SG PP

[14]There are dialects of English in which this is grammatical, but we will be analyzing the more standard dialect in which agreement marking is obligatory.

d. NOM-PL → NOM-PL PP
e. NOM-SG → N-SG
f. NOM-PL → N-PL
g. VP-SG → IV-SG
h. VP-PL → IV-PL
i. VP-SG → VP-SG PP
j. VP-PL → VP-PL PP

This set of rules is cumbersome, and clearly misses linguistically significant generalizations. The rules in this set come in pairs, differing only in whether the category names end in '-SG' or '-PL'. Nothing in the formalism or in the theory predicts this pairing. The rules would look no less natural if, for example, the rules expanding -PL categories had their right-hand sides in the reverse order from those expanding -SG categories. But languages exhibiting this sort of variation in word order do not seem to exist.

Things get even messier when we consider transitive and ditransitive verbs. Agreement is required regardless of whether the verb is intransitive, transitive, or ditransitive. Thus, along with (35), we have (37) and (38):

(37) a. The bird devours the worm.
b. The birds devour the worm.
c.*The bird devour the worm.
d.*The birds devours the worm.

(38) a. The bird gives the worm a tug.
b. The birds give the worm a tug.
c.*The bird give the worm a tug.
d.*The birds gives the worm a tug.

If agreement is to be handled by the rules in (39):

(39) a. S → NP-SG VP-SG
b. S → NP-PL VP-PL

then we will now need to introduce lexical categories TV-SG, TV-PL, DTV-SG, and DTV-PL, along with the necessary VP-SG and VP-PL expansion rules (as well as the two rules in (39)). What are the rules for VP-SG and VP-PL when the verb is transitive or ditransitive? For simplicity, we will look only at the case of VP-SG with a transitive verb. Since the object of the verb can be either singular or plural, we need two rules:

(40) a. VP-SG → TV-SG NP-SG
b. VP-SG → TV-SG NP-PL

Similarly, we need two rules for expanding VP-PL when the verb is transitive, and four rules each for expanding VP-SG and VP-PL when the verb is ditransitive (since each object can be either singular or plural). Alternatively, we could make all objects of category NP and introduce the following two rules:

(41) a. NP → NP-SG
b. NP → NP-PL

This would keep the number of VP-SG and VP-PL rules down to three each (rather than seven each), but it introduces extra noun phrase categories. Either way, the rules are full of undesirable redundancy.

Matters would get even worse when we examine a wider range of verb types. So far, we have only considered how many NPs must follow each verb. But there are verbs that only appear in other environments; for example, some verbs require following PPs or Ss, as in (42).

(42) a. Terry wallowed in self-pity.
b.*Terry wallowed.
c.*Terry wallowed the self-pity.
d. Kerry remarked (that) it was late.
e.*Kerry remarked.
f.*Kerry remarked the time.

Exercise 2: Wallowing in **Categories**

A. Provide examples showing that the verbs *wallow* and *remark* exhibit the same agreement patterns as the other types of verbs we have been discussing.

B. What additional categories and rules would be required to handle these verbs?

When a broader range of data is considered, it is evident that the transitivity distinctions we have been assuming are simply special cases of a more general phenomenon. Some verbs (and, as we will see later, some other types of words as well) occur only in the environment of particular kinds of constituents. In English, these constituents characteristically occur after the verb, and syntacticians call them COMPLEMENTS. Complements will be discussed in greater detail in Chapter 4.

It should be clear by now that as additional coverage is incorporated – such as adjectives modifying nouns – the redundancies will proliferate. The problem is that we want to be able to talk about nouns and verbs as general classes, but we have now divided nouns into (at least) two categories (N-SG and N-PL) and verbs into six categories (IV-SG, IV-PL, TV-SG, TV-PL, DTV-SG, and DTV-PL). To make agreement work, this multiplication of categories has to be propagated up through at least some of the phrasal categories. The result is a very long and repetitive list of phrase structure rules.

What we need is a way to talk about subclasses of categories, without giving up the commonality of the original categories. That is, we need a formalism that permits us to refer straightforwardly to, for example, all verbs, all singular verbs, all ditransitive verbs, or all singular ditransitive verbs. In the next chapter, we introduce a device that will permit us to do this.

2.8 Transformational Grammar

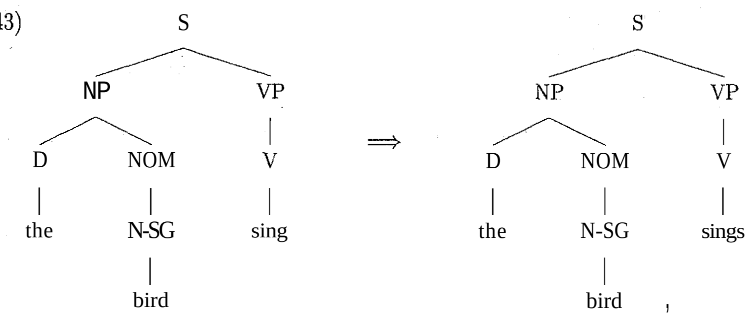
As noted in Section 2.6, much of the work in generative grammar (including this book) has involved developing extensions of Context Free Grammar to make it better adapted to the task of describing natural languages. The most celebrated proposed extension

was a kind of rule called a 'transformation', as introduced into the field of generative grammar by Noam Chomsky.[15] Transformations are mappings from phrase structure representations to phrase structure representations (from trees to trees, in our terms) that can copy, delete, and permute parts of trees, as well as insert specified new material into them. The initial trees were to be generated by a CFG. For example, in early work on transformations, it was claimed that declarative and interrogative sentence pairs (such as *The* sun is shining and Is *the* sun *shining?*) were to be derived from the same underlying phrase structure by a transformation that moved certain verbs to the front of the sentence. Likewise, 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. The name 'transformational grammar' is sometimes used for theories positing rules of this sort.[16]

In a transformational grammar, then, each sentence is associated not with a single tree structure, but with a sequence of such structures. This greatly enriches the formal options for describing particular linguistic phenomena.

For example, subject-verb agreement can be handled in transformational terms by assuming that number (that is, being singular or plural) is an intrinsic property of nouns, but not of verbs. Hence, in the initial tree structures for sentences, the verbs have no number associated with them. Subsequently, a transformation changes the form of the verb to the one that agrees with the subject NP. Such an analysis avoids the proliferation of phrase structure rules described in the preceding section, but at the cost of adding an agreement transformation.

As an illustration of how this would work, consider again the contrast in (35)[17] Instead of creating separate singular and plural versions of NP, VP, NOM, N, and V (with the corresponding phrase structure rules in (36)), a transformational analysis could limit this bifurcation of categories to N-SG and N-PL (with the rules NOM → N-SG and NOM → N-PL). In addition, an agreement transformation (which we will not try to formalize here) would give the verb the correct form, roughly as follows:

(43)


[15]The original conception of a transformation, as developed in the early 1950s by Zellig Harris, was intended somewhat differently – as a way of regularizing the information content of texts, rather than as a system for generating sentences.

[16]See Appendix B for more discussion of varieties of transformational grammar.

[17]The analysis sketched in this paragraph is a simplified version of the one developed by Chomsky (1957). It has long since been superceded by other analyses. In presenting it here (for pedagogical purposes) we do not mean to suggest that contemporary transformationalists would advocate it.

Notice that in a theory that posits a passivization transformation (which, among other things, would move the object NP into subject position), something like the agreement transformation described in the previous paragraph would be required. To make this more concrete, consider examples like (44):

(44) a. Everyone loves puppies.
b. Puppies are loved by everyone.

Substituting the the singular form of the verb in (44b) results in ill-formedness:

(45) *Puppies is loved by everyone.

In a transformational analysis, puppies only becomes the subject of the sentence following application of the passivization transformation. Since agreement (in English) is consistently with the subject NP, if transformations are permitted to change which NP is the subject, agreement cannot be determined until after such transformations have applied.

In general, transformational analyses involve such rule interactions. Many transformational derivations involve highly abstract underlying structures with complex sequences of transformations deriving the observable forms.

Because versions of transformational grammar have been so influential throughout the history of generative grammar, many of the phenomena to be discussed have come to be labeled with names that suggest transformational analyses (e.g. "raising", discussed in Chapter 12).

This influence is also evident in work on the psychology of language. In contemplating the mental processes underlying language use, linguists naturally make reference to their theories of language structure, and there have been repeated efforts over the years to find evidence that transformational derivations play a role in at least some aspects of language processing.

In later chapters, we will on occasion be comparing our (nontransformational) analyses with transformational alternatives. We make no pretense of doing justice to all varieties of transformational grammar in this text. Our concern is to develop a theory that can provide rigorous and insightful analyses of a wide range of the structures found in natural languages. From time to time, it will be convenient to be able to consider alternative approaches, and these will often be transformational.

2.9 What Are Grammars Theories Of?

In the opening paragraphs of Chapter 1, we said that linguists try to study language scientifically. We then went on to describe some of the grammatical phenomena that we would be investigating in this book. In this chapter, we have taken the first steps towards formulating a precise theory of grammar, and we have presented evidence for particular formulations over others.

We have not, however, said much about what a grammar is taken to be a theory of. Chapter 1 discussed the view, articulated most forcefully by Chomsky, that one reason for studying language is to gain insight into the workings of the human mind. On this view – which is shared by many but by no means all linguists – choosing one form of grammar over another constitutes a psychological hypothesis. That is, a grammar is a theory about the mental representation of linguistic knowledge.

As we noted, there are other views. Some linguists point out that communicating through language requires that different people share a common set of conventions. Any approach to language that seeks to represent only what is in the mind of an individual speaker necessarily gives short shrift to this social aspect of language.

To begin to get a handle on these issues, consider a concrete example: Pat says, "What time is it?" and Chris answers, "It's noon". The two utterances are physical events that are directly observable. But each of them is an instance of a sentence,' and both of these sentences have been uttered many times. As syntacticians, we are interested in only some properties of these utterances; other properties, such as where they were uttered and by whom, are not relevant to our concerns. Moreover, there are many other English sentences that have never been spoken (or written), but they still have properties that our grammar should characterize. In short, the subject matter of our theory is sentences, which are abstractions, rather than observable physical events. We are interested in particular utterances only as evidence of something more abstract and general, just as a biologist is only interested in particular organisms as instances of something more abstract and general, such as a species.

A grammar of English should characterize the structure and meaning of both-Pat's utterance and Chris's. So we need to abstract across different speakers, too. This raises some difficult issues, because no two speakers have exactly the same linguistic knowledge. In fact, linguistic differences among individuals and groups of individuals make it notoriously difficult to draw boundaries between languages. The conventional labels applied to languages (such as English, Chinese, or Arabic) are determined as much by political facts as by linguistic ones.[18] It is largely for this reason that Chomsky and many other linguists say that their object of study is the mental representations of individual speakers.

Of course, similar difficulties arise in drawing boundaries between species, but few biologists would say on those grounds that biology should only be concerned with the DNA of individual organisms. Just as biologists seek to generalize across populations of heterogeneous individuals, we want our grammar to characterize something more general than what is in one person's mind. Occasionally, we will deal with phenomena which are not uniform across all varieties of English (see especially Chapter 15).

In short, we want our grammar to characterize the syntax of English. This involves multiple levels of abstraction from what is directly observable, as well as some attention to variation among speakers. Our object of study is not purely a matter of individual psychology, nor is it exclusively a social phenomenon. There are some aspects of language that are primarily manifestations of individual speakers' mental representations and others that critically involve the interactions of multiple language users. Just as molecular biology and population biology both contribute to our understanding of species, linguists need not make an exclusive choice between an internal and an external perspective.

2.10 Summary

In this chapter, we began our search for an adequate model of the grammar of one natural language: English. We considered and rejected two simple approaches to grammar,

[18]Linguists sometimes joke that a 'language' is simply a 'dialect' with an army and a navy

including a theory based on regular expressions ('finite-state grammar'). The theory of context-free grammars, by contrast, solves the obvious defects of these simple approaches and provides an appropriate starting point for the grammatical description of natural language. However, we isolated two ways in which context-free grammars are inadequate as a theory of natural language:

- CFGs are arbitrary. They fail to capture the 'headedness' that is characteristic of many types of phrase in natural language.
- CFGs are redundant. Without some way to refer to kinds of categories rather than just individual categories, there is no way to eliminate the massive redundancy that will be required in order to analyze the agreement and subcategorization patterns of natural languages.

For these reasons, we cannot accept CFG alone as a theory of grammar. As we will show in the next few chapters, however, it is possible to retain much of the character of CFG as we seek to remedy its defects.

2.11 Further Reading

The standard reference work for the basic mathematical results on formal languages (including regular expressions and context-free languages) is Hopcroft et al. 2001. Partee et al. 1990 covers much of the same material from a more linguistic perspective. Classic works arguing against the use of context-free grammars for natural languages include Chomsky 1963 and Postal 1964. Papers questioning these arguments, and other papers presenting new arguments for the same conclusion are collected in Savitch et al. 1987. For (somewhat dated) surveys of theories of grammar, see Sells 1985 and Wasow 1989. A more detailed presentation of GPSG is Gazdar et al. 1985. The history of generative grammar is presented from different perspectives by Matthews (1993), Newmeyer (1986), Harris (1993), and Huck and Goldsmith (1995).

Perhaps the best discussions of the basic phrase structures of English are to be found in good descriptive grammars, such as Quirk et al. 1972, 1985, Huddleston and Pullum 2002, or Greenbaum 1996. Important discussions of the notion of 'head' and its role in phrase structure can be found in Chomsky 1970 and Gazdar and Pullum 1981. A detailed taxonomy of the subcategories of English verbs is provided by Levin (1993).

2.12 Problems

⚠ Problem 1: More Practice with CFG

Assume the grammar rules given in (23), but with the following lexicon:

D:	a, the
V:	admired, disappeared, put, relied
N:	cat, dog, hat, man, woman, roof
P:	in, on, with
CONJ:	and, or

A. Give a well-formed English sentence that this grammar sanctions and assigns only one structure to. Draw the tree structure that the grammar assigns to it. ,

B. Give a well-formed English sentence that is structurally ambiguous according to this grammar. Draw two distinct tree structures for it. Discuss whether the English sentence has two distinct interpretations corresponding to the two trees.

C. Give a sentence (using only the words from this grammar) that is not covered by this grammar but which is nonetheless well-formed in English.

D. Explain what prevents the example in (C) from being covered.

E. Give a sentence sanctioned by this grammar that is not a well-formed English sentence

F. Discuss how the grammar might be revised to correctly exclude your example in (E), without simultaneously excluding good sentences. Be explicit about how you would change the rules and/or the lexicon.

G. How many sentences does this grammar admit?

H. How many would it admit if it didn't have the last rule (the coordination schema)?

Problem 2: Structural Ambiguity

Show that the grammar in (23) can account for the ambiguity of each of the following sentences by providing at least two trees licensed for each one, and explain briefly which interpretation goes with which tree:

(i) Bo saw the group with the telescope.

(ii) Most dogs and cats with fleas live in this neighborhood.

(iii) The pictures show Superman and Lois Lane and Wonder Woman.

[Note: We haven't provided a lexicon, so technically, (23) doesn't generate any of these. You can assume, however, that all the words in them are in the lexicon, with the *obvious* category assignments.]

Problem 3: Infinity

The grammar in (23) has two mechanisms, each of which permits us to have infinitely many sentences: the Kleene operators (plus and star), and recursion (categories that can 'dominate themselves'). Construct arguments for why we need both of them. That is, why not use recursion to account for the unboundedness of coordination or use Kleene star to account for the possibility of arbitrary numbers of PPs?

[*Hint*: Consider the *different groupings* into phrases – that is, the different tree structures – provided by the two mechanisms. Then look for English data supporting one choice of structure over another.]

Problem 4: CFG for Japanese

Examples (i)–(x) give examples of grammatical Japanese sentences and strings made up of the same words which are not grammatical Japanese sentences.

(i) Suzuki-san-ga sono eiga-wo mita.
Suzuki-NOM that movie-ACC saw
'Suzuki saw that movie.'

(ii)*Mita Suzuki-san-ga sono eiga-wo.
Saw Suzuki-NOM that movie-ACC

(iii)*Suzuki-san-ga mita sono eiga-wo.
Suzuki-NOM saw that movie-ACC

(iv)*Suzuki-san-ga eiga-wo sono mita.
Suzuki-NOM movie-ACC that saw.

(v) Suzuki-san-ga sono omoshiroi eiga-wo mita.
Suzuki-NOM that interesting movie-ACC saw
'Suzuki saw that interesting movie.'

(vi)*Suzuki-san-ga sono eiga-wo omoshiroi mita.
Suzuki-NOM that movie-ACC interesting saw

(vii)*Suzuki-san-ga omoshiroi sono eiga-wo mita.
Suzuki-NOM interesting that movie-ACC saw

(viii) Suzuki-san-ga Toukyou e itta.
Suzuki-NOM Tokyo to went.
'Suzuki went to Tokyo.'

(ix)*Suzuki-san-ga e Toukyou itta.
Suzuki-NOM to Tokyo went.

(x)*Suzuki-san-ga itta Toukyou e.
Suzuki-NOM went Tokyo to.

A. Using the lexicon in (xi), write phrase structure rules that will generate the grammatical examples and correctly rule out the ungrammatical examples.

[Notes: The data presented represent only a very small fragment of Japanese, and are consistent with many different CFGs. While some of those CFGs would fare better than others when further data are considered, any answer that accounts for the data presented here is acceptable. The abbreviations 'NOM' and 'ACC' in these examples stand for nominative and accusative case, which you may ignore for the purposes of this problem.]

(xi) N: Suzuki-san-ga, eiga-wo, Toukyou
D: sono
P: e
A: omoshiroi
V: mita, itta

B. Draw the trees that your grammar assigns to (i), (v), and (viii).

Problem 5: Properties Common to Verbs

The rules in (34) embody the claim that IVs, TVs, and DTVs are entirely different categories. Hence, the rules provide no reason to expect that these categories would have more in common than any other collection of three lexical categories, say, N, P, and D. But these three types of verbs do behave alike in a number of ways. For example, they all exhibit agreement with the subject of the sentence, as discussed in Section 2.7.3. List at least three other properties that are shared by intransitive, transitive, and ditransitive verbs.

Problem 6: Pronoun Case

There are some differences between the noun phrases that can appear in different positions. In particular, pronouns in subject position have one form (referred to as NOMINATIVE, and including the pronouns I, he, she, we, and *they*), whereas pronouns in other positions take another form (called ACCUSATIVE, and including me, *him*, her, us, and *them*). So, for example, we say He saw her, not *Him saw she.

A. How would the category of NP have to be further subdivided (that is, beyond NP-SG and NP-PL) in order to account for the difference between nominative and accusative pronouns?

B. How would the rules for S and the various kinds of VPs have to be modified in order to account for the differences between where nominative and accusative pronouns occur?

3

Analyzing Features of Grammatical Categories

3.1 Introduction

In the last chapter, we saw that there are constraints on which words can go together (what linguists call CO-OCCURRENCE RESTRICTIONS) that are not adequately described using the standard formalism of context-free grammar. Some verbs must take an object; others can never take an object; still others (e.g. put, hand) require both an object and another phrase of a particular kind. These co-occurrence restrictions, as we have seen, give rise to a great deal of redundancy in CFGs. In addition, different forms of a given verb impose different conditions on what kind of NP can precede them (i.e. on what kind of subject they co-occur with). For example, walks requires a third-person singular NP as its subject; walk requires a plural subject, or else one that is first- or second-person singular. As we saw in the last chapter, if we try to deal with this complex array of data by dividing the category V into more specific categories, each with its unique co-occurrence restrictions, we end up with a massively redundant grammar that fails to capture linguistically significant generalizations.

We also isolated a second defect of CFGs, namely that they allow rules that are arbitrary. Nothing in the theory of CFG reflects the headedness of phrases in human language – that is, the fact that phrases usually share certain key properties (nounhood, verbhood, prepositionhood, etc.) with a particular daughter within them. We must somehow modify the theory of CFG to allow us to express the property of headedness.

Our solution to the problem of redundancy is to make grammatical categories decomposable into component parts. CFG as presented so far treats each grammatical category symbol as ATOMIC – that is, without internal structure. Two categories are either identical or different; there is no mechanism for saying that two categories are alike in some ways, but different in others. However, words and phrases in natural languages typically behave alike in certain respects, but not in others. For example, the two words deny and *denies* are alike in requiring an NP object (both being forms of a transitive verb). But they differ in terms of the kind of subject NP they take: denies requires a third-person-singular subject like *Kim* or she, while deny accepts almost any NP subject except the third-person-singular kind. On the other hand, denies and disappears both take a singular subject NP, but only the former can co-occur with a following object NP. In other words, the property of taking a third-person-singular subject is independent of the property of taking a direct object NP. This is illustrated in the following table:

(1)

	3rd singular subject	plural subject
direct object NP	*denaes*	*deny*
no direct object NP	*disappears*	*disappear*

The table in (1) illustrates only two of the cross-cutting properties of verbs. There are many more. For example, the properties of forming the third-person-singular form with -s, the past tense form with *-ed,* and the present participle with *-ing* are all orthogonal to the property of taking a direct object NP. In Chapter 8, we will see how to write rules for generating these INFLECTIONAL forms of verbs. In order to write such rules with maximal generality, we need to be able to refer to the class of all verbs, regardless of whether they take a direct object NP. More generally, an adequate theory of grammar needs to be able to categorize words into classes defined in terms of cross-cutting properties. In Chapter 2, we showed CFG to be inadequate as a theory of grammar, because it provides no means to represent cross-cutting properties. Instead, it ends up proliferating atomic categories and missing generalizations.

To accommodate these observations, we will develop the view that grammatical categories are not atomic, but rather are COMPLEXES of grammatical properties. In some ways, this innovation is similar to the periodic table of the elements in chemistry, which represents the elements as complexes of physical properties. The rows and columns of the table represent classes of elements that have properties in common, and the classes interesect: each element belongs to more than one class, and shares only some of its properties with the other elements in each of the classes it belongs to. Treating grammatical categories as complexes of grammatical properties will also pave the way for a solution to the second defect of CFGs, by allowing us to express the property of headedness.

3.2 Feature Structures

This section introduces the formal mechanism we will use for representing grammatical categories as complexes of grammatical properties. But let us first review the grammatical properties we have covered so far. We have seen that verbs differ in their transitivity. In fact, this kind of variation is not restricted to verbs. More generally, linguists talk about elements that have different combinatoric potential in terms of differing VALENCE.[1] Likewise, we talk of the NUMBER (singular or plural) of a noun, the PART OF SPEECH of a word (whether it's a noun, verb, etc.), and the FORM of a verb (e.g. whether it is a present participle, an infinitive, etc.). Previously we have been associating each word in the lexicon with a single atomic category (such a P, N-SG, etc.). Now, in order to model grammatical categories as complexes of information, we will use FEATURE STRUCTURES instead of atomic labels.

A feature structure is a way of representing grammatical information. Formally, a feature structure consists of a specification of a set of features (which we will write in upper case), each of which is paired with a particular value. Feature structures can be

[1] This term, borrowed from chemistry, refers to the capacity to combine with atoms, ions, and the like

thought of in at least two roughly equivalent ways. For example, they may be conceived of as functions (in the mathematicians' sense of the word)[2] specifying a value for each of a set of features, or else as directed graphs where feature names label arcs that point to appropriately labeled nodes. For grammatical purposes, however, it will be most useful for us to focus on DESCRIPTIONS of feature structures, which we will write in a square bracket notation, as shown in (2):

(2) $\begin{bmatrix} \text{FEATURE}_1 & \text{VALUE}_1 \\ \text{FEATURE}_2 & \text{VALUE}_2 \\ \ldots & \\ \text{FEATURE}_n & \text{VALUE}_n \end{bmatrix}$

For example, we might treat the category of the word *bird* in terms of a feature structure that specifies just its part of speech and number. We may assume such a category includes appropriate specifications for two appropriately named features: its part of speech (POS) is noun, and its number (NUM) is singular (sg). Under these assumptions, the lexical entry for *bird* would be a pair consisting of a form and a feature structure description, roughly as shown in (3):[3]

(3) $\left\langle \text{bird}, \begin{bmatrix} \text{POS} & \text{noun} \\ \text{NUM} & \text{sg} \end{bmatrix} \right\rangle$

One of the first things we will want to do in developing a theory of grammar is to classify linguistic entities in various ways. To this end, it is particularly useful to introduce the notion of TYPE. This concept is really quite simple: if we think of a language as a system of linguistic entities (words, phrases, categories, sounds, and other more abstract entities that we will introduce as we go along), then types are just classes of those entities. We assign entities to these classes on the basis of certain properties that they share. Naturally, the properties we employ in our type classification will be those that we wish to refer to in our descriptions of the entities. Thus each grammatical type will be associated with particular features and sometimes with particular values for those features. As we develop our theory of grammatical types, we will in fact be developing a theory of what kinds of linguistic entities there are, and what kinds of generalizations hold of those entities.

Let us make this very abstract discussion more concrete by considering the use of feature structures to elucidate a simple nonlinguistic domain: universities and the people who are associated with them. We'll start from the assumption that the people and the other entities are really 'out there' in the real world. Our first step then in constructing a theory of this part of the world is to develop a model. A simple model will be a set of mathematical entities that we assume to correspond to the real ones. Our theory will be successful to the extent that we can show that the properties that our theory ascribes to our modeling entities (through stipulation or deduction from the stipulations) also hold

[2] A function in this sense is a set of ordered pairs such that no two ordered pairs in the set share the same first element. What this means for feature structures is that each feature in a feature structure must have a unique value.

[3] Throughout this book, we will describe linguistic forms in terms of standard English orthography. In fact, a lexical entry such as this should contain a phonological description that will play a role in the word's phonological realization, a topic we will not consider in detail here

of the real world entities that they correspond to.

The domain at hand includes entities such as universities, departments and individuals (people). We might want to talk about certain properties of these entities, for example their names or telephone numbers. In order to do so, we will start' to build our model by declaring the existence of a general type called entity and say that the features NAME and TEL(EPHONE) are appropriate features for all entities (tokens) of this type. So for each university, department, or person in this university world, we would hypothesize a distinct feature structure model that we could describe as follows:

(4) a. $\begin{bmatrix} \textit{entity} & \\ \text{NAME} & \text{Stanford University} \\ \text{TEL} & \text{650-723-2300} \end{bmatrix}$

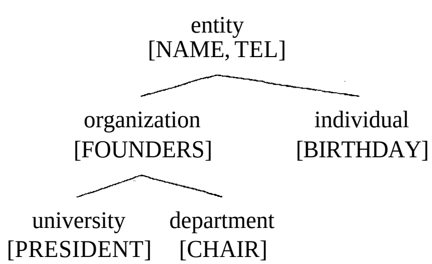
b. $\begin{bmatrix} \textit{entity} & \\ \text{NAME} & \text{John Hennessy} \\ \text{TEL} & \text{650-723-2481} \end{bmatrix}$

c. $\begin{bmatrix} \textit{entity} & \\ \text{NAME} & \text{Stanford Linguistics} \\ \text{TEL} & \text{650-723-4284} \end{bmatrix}$

Note that we use type names (in this case entity), written in italics, **as** labels on the top line within feature structures.

Of course 'entity' is a very general classification – our theory would not have progressed far if it recognized no more specific kinds of things. So in fact, we would want our theory to include the fact that there are different subtypes of the type entity. Let's call these new types university, department, and individual. Entities belonging to each of these types have their own special properties. For example, individual people have birthdays, but universities and departments don't (or not in the same sense). Similarly, departments have chairs (or 'heads of department'), but neither universities nor individuals do. And only universities have presidents. Finally, universities and deparments, but not individuals, have founders, a fact that will motivate grouping these two types together under a common intermediate-level type which we will call organization. We can then accommodate all these facts by declaring each of the relevant features (BIRTHDAY, CHAIR, PRESIDENT, FOUNDERS) to be appropriate for entities of the appropriate subtype. This organization of the types of entity and the features that 'are appropriate for each of them results in the TYPE HIERARCHY shown in (5):

(5)



Each type of entity has its own constellation of features – some of them are declared appropriate for entities of the indicated type; others are sanctioned by one of the supertypes: entity or organization. This is a simple illustration of how a hierarchical classification system works. A given feature structure contains only those features that are declared appropriate by one of its types, that is, by its LEAF type[4] or one of its supertypes in a hierarchy like (5). This formal declaration is just a precise way of saying that the members of the relevant subclasses have certain properties that distinguish them from other entities in the system, as well as certain properties that they share with other entities.

Now that we've extended the model by adding types and features, the resulting de-, scriptions that we write will be appropriately more specific, as in (6):

(6) a. $\begin{bmatrix} \textit{university} & \\ \text{NAME} & \text{Stanford University} \\ \text{FOUNDERS} & \left\langle \text{Leland Stanford , Jane Stanford} \right\rangle \\ \text{PRESIDENT} & \text{John Hennessy} \\ \text{TEL} & \text{650-723-2300} \end{bmatrix}$

b. $\begin{bmatrix} \textit{individual} & \\ \text{NAME} & \text{John Hennessy} \\ \text{BIRTHDAY} & \text{9-22-52} \\ \text{TEL} & \text{650-723-2481} \end{bmatrix}$

c. $\begin{bmatrix} \textit{department} & \\ \text{NAME} & \text{Stanford Linguistics} \\ \text{FOUNDERS} & \left\langle \text{Joseph Greenberg , Charles Ferguson} \right\rangle \\ \text{CHAIR} & \text{Eve Clark} \\ \text{TEL} & \text{650-723-4284} \end{bmatrix}$

Note that we also need to specify what kind of value is appropriate for each feature. Here we've used angled brackets ('⟨' and '⟩') to construct a list as the value of the feature FOUNDERS. As we will see, a feature structure also inherits any type constraints, (that is, potentially complex constraints on feature values) that are associated with its supertypes. Articulating a type hierarchy and the constraints associated with each type in the hierarchy is an important component of a theory that uses typed feature structures as its models.

Let us reconsider the feature structures in (6). These structures, as explicated above, aren't yet expressing the proper information about the objects they are trying to model. In particular, the value of features like PRESIDENT and CHAIR are atomic, i.e. the names John Hennessy and Eve Clark. But this isn't right – the president of Stanford University is the individual John Hennessy, not his name. The same goes for Eve Clark, who gives more to being chair of the Stanford Linguistics Department than just her name.

[4] The leaf types are the basic or bottom-level types in a hierarchy, i.e. the types that have no subtypes. These are also referred to in the literature (somewhat counterintuitively) as 'maximal' types.

Similarly, the value of the FOUNDERS feature should be a list of individuals, not a list of names. To reflect these observations, we now introduce complex feature structures, those whose features may have nonatomic feature structures (or lists of feature structures) as their value, where appropriate. This modification leads to the following more accurate models of Stanford and its Linguistics Department (the model of John Hennessy remains unaffected by this change):

(7) a.

$$\begin{bmatrix} \textit{university} & \\ \text{NAME} & \text{Stanford University} \\ \text{FOUNDERS} & \left\langle \begin{bmatrix} \textit{individual} & \\ \text{NAME} & \text{Leland Stanford} \end{bmatrix}, \begin{bmatrix} \textit{individual} & \\ \text{NAME} & \text{Jane Stanford} \end{bmatrix} \right\rangle \\ \text{PRESIDENT} & \begin{bmatrix} \textit{individual} & \\ \text{NAME} & \text{John Hennessy} \\ \text{BIRTHDAY} & \text{9-22-52} \\ \text{TEL} & \text{650-723-2481} \end{bmatrix} \\ \text{TEL} & \text{650-723-2300} \end{bmatrix}$$

b.

$$\begin{bmatrix} \textit{department} & \\ \text{NAME} & \text{Stanford Linguistics} \\ \text{FOUNDERS} & \left\langle \begin{bmatrix} \textit{individual} & \\ \text{NAME} & \text{Joseph Greenberg} \\ \text{BIRTHDAY} & \text{5-28-15} \end{bmatrix}, \begin{bmatrix} \textit{indzvidual} & \\ \text{NAME} & \text{Charles Ferguson} \\ \text{BIRTHDAY} & \text{7-6-21} \end{bmatrix} \right\rangle \\ \text{CHAIR} & \begin{bmatrix} \textit{individual} & \\ \text{NAME} & \text{Eve Clark} \\ \text{BIRTHDAY} & \text{7-26-42} \\ \text{TEL} & \text{650-723-4284} \end{bmatrix} \\ \text{TEL} & \text{650-723-4284} \end{bmatrix}$$

When we model some empirical problem in this way, it is important to distinguish the modeling objects (the typed feature structures) from the statements we make about them. The objects in our model are meant to be simplified analogs of objects in the real world (if they weren't simplified, it wouldn't be a model). The statements we make about the modeling objects – our constraints – constitute our theory of the domain we are investigating. The system of types we set up of course is the first step in developing such a theory:

- It states what kinds of objects we claim exist (the types).
- It organizes the objects hierarchically into classes with shared properties (the type hierarchy).
- It states what general properties each kind of object has (the feature and feature value declarations).

We could summarize the beginnings of our theory of universities in terms of the following table (where 'IST' stands for 'immediate supertype'):[5]

(8)

TYPE	FEATURES/VALUES	IST
entity	[NAME *string*, TEL *number*]	
organization	[FOUNDERS *list(individual)*]	*entity*
university	[PRESIDENT *individual*]	*organization*
department	[CHAIR *individual*]	*organization*
individual	[BIRTHDAY *date*]	*entity*

Against this background, it is the particular constraints we write that fill in the details. Type constraints specify properties that relevant classes of objects have (e.g. that universities have presidents who are individuals) and other constraints characterize properties of certain idiosyncratic entities that we find it necessary to recognize (e.g. that Stanford's president is John Hennessy). We then make the standard assumption that our modeling objects are in correspondence with the real world. In so doing, our constraints are making claims about reality in ways that distinguish our theory of the relevant empirical domain from many others that could be formulated.

Our (admittedly somewhat artificial) theory of Stanford University then consists of a set of constraints that reflect our claims about the way Stanford is, some of which may reflect the way all universities are. Those constraints are meant to describe (or be SATISFIED by) the objects in our model of Stanford – the feature structures assigned to appropriate types, exhibiting the relevant properties. And if we've modeled things correctly, our feature structures will reflect the reality of Stanford and we will view our theory as making correct predictions.

Theories often include constraints requiring two things to be identical. For example, suppose we wanted to state the hypothesis that the phone number of a department chair was always the same as the department's phone number. This somewhat trivial (yet precise) claim might be formulated as follows:

[5] Note that this table assumes the types *number*, *string* and *date*. These three types would also need to be incorporated into the type hierarchy.

(9) $\text{department}: \begin{bmatrix} \text{TEL} & \boxed{1} \\ \text{CHAIR} & \begin{bmatrix} \text{TEL} & \boxed{1} \end{bmatrix} \end{bmatrix}$

The colon here denotes a conditional ('if–then') relation between a type and a claim being made about the instances of that 'type. The boxed numerals in (9) are called 'tags'. They function like variables in algebra, logic, or programming languages. That is, they indicate that two values within a given feature structure are identical. What the constraint in (9) is saying then is that for any feature structure of type department, if you start at the outside and follow the feature path CHAIR|TEL, you'll arrive at the same value that you find when you start at the outside again and follow the (single-feature) path TEL.

Of course, it's easy to test the predictions of a one-sentence theory like (9). The feature structure models of type department that satisfy (9) have a clear and simple property and the relevant objects out in the real world are all listed in the Stanford Directory with their phone numbers. It's presumably not hard to verify whether (9) is true or not.[6] But science is full of theories whose predictions are much harder to test. Indeed, we'll see that evaluating the predictions of a theory of language based on feature structure models can sometimes be quite a subtle matter.

Interesting theories involve a number of different claims that interact. For this reason, it's essential that we have a way of combining constraints and determining which models satisfy the resulting combinations, however complex they might be. We will in fact use a simple method for combining (conjoining) constraints – one that we'll sometimes write with the symbol '&', as in (10a). Quite often, however, we will simply combine two constraints into a bigger one like (10b):[7]

(10) a. $\begin{bmatrix} \text{TEL} & \text{650-723-4284} \end{bmatrix} \ \& \ \begin{bmatrix} \text{NAME} & \text{Stanford Linguistics} \end{bmatrix}$

b. $\begin{bmatrix} \text{NAME} & \text{Stanford Linguistics} \\ \text{TEL} & \text{650-723-4284} \end{bmatrix}$

Notice how our constraints relate to their models. The first conjunct (the bracketed constraint before the '&' in (10a)) is satisfied by a set of feature structures (in our current model, it's the set that contains the feature structure we used to model the Stanford Linguistics Department and the one we used to model its chair). The second conjunct in (10a) is also satisfied by a set of feature structures, but this set has only one member: the feature structure serving as our model of the Stanford Linguistics Department. And the constraint in (10), whether we formulate it as in (10a) or as in (10b), is satisfied by the intersection of the two other sets, i.e. by the (singleton) set that contains just the feature structure we used to model the Stanford Linguistics Department.

[6] In fact, this theory of Stanford department phone numbers is easily falsified.

[7] The process of combining constraints in the fashion of (10b) is often called 'unification'. Theories of the sort we describe in this book are sometimes called 'unification-based', but this term is misleading. Unification is a method (i.e. a procedure) for solving sets of identity constraints. But it is the constraints themselves that constitute the theory, not any procedure we might use with them. Hence, we will refer to the theory of grammar we devlop, and the class of related theories, as 'constraint-based', rather than 'unification-based'.

Note that the constraints in (11) are incompatible because they differ in the value they assign to the feature NAME:

(11) a. $\begin{bmatrix} \textit{university} & \\ \text{NAME} & \text{Stanford University} \end{bmatrix}$

b. $\begin{bmatrix} \textit{university} & \\ \text{NAME} & \text{Harvard University} \end{bmatrix}$

And because (11a) and (11b) are incompatible, they couldn't be used to describe the same entity.

Similarly, the constraints in (12) cannot be combined:

(12) a. $\begin{bmatrix} \textit{individual} & \\ \text{TEL} & \text{650-555-4284} \end{bmatrix}$

b. $\begin{bmatrix} \textit{department} & \\ \text{TEL} & \text{650-555-4284} \end{bmatrix}$

In this case, the problem is that (12a) and (12b) specify incompatible types, namely, *individual* and department. Hence (12a) and (12b) must be describing distinct entities. But the constraint in (13) is compatible with any of those in (14a)-(14c):

(13) $\begin{bmatrix} \text{TEL} & \text{888-234-5789} \end{bmatrix}$

(14) a. $\begin{bmatrix} \textit{university} \end{bmatrix}$

b. $\begin{bmatrix} \textit{individual} & \\ \text{NAME} & \text{Sailor Moon} \end{bmatrix}$

c. $\begin{bmatrix} \textit{department} & \\ \text{NAME} & \text{Metaphysics} \\ \text{CHAIR} & \text{Alexius Meinong, Jr.} \end{bmatrix}$

For example, the combination of (13) and (14b), shown in (15), is satisfied by those objects (in our model) that satisfy both (13) and (14b):

(15) $\begin{bmatrix} \textit{individual} & \\ \text{NAME} & \text{Sailor Moon} \\ \text{TEL} & \text{888-234-5789} \end{bmatrix}$

Finally, the constraints in (16) cannot be combined:

(16) a. $\begin{bmatrix} \text{BIRTHDAY} & \text{10-10-1973} \end{bmatrix}$

b. $\begin{bmatrix} \text{PRESIDENT} & \begin{bmatrix} \text{individual} & \\ \text{NAME} & \text{Sailor Moon} \end{bmatrix} \end{bmatrix}$

In this case, the constraints cannot be combined because there is no type for which the features BIRTHDAY and PRESIDENT are appropriate. Since all of the modeling objects must belong to some type, there will be none that satisfy both (16a) and (16b).

When our feature structure constraints get a bit more complicated, we will sometimes want to indicate simultaneously the value of a particular feature and the fact that that value is identical with the value of another feature (or feature path), as shown in (17):

(17) $$\begin{bmatrix} \text{TEL} & \boxed{1}\,650\text{-}723\text{-}4284 \\ \text{CHAIR} & [\text{TEL} \quad \boxed{1}] \end{bmatrix}$$

But it would make no difference if we wrote the phone number after the other occurrence of $\boxed{1}$ in (17):

(18) $$\begin{bmatrix} \text{TEL} & \boxed{1} \\ \text{CHAIR} & [\text{TEL} \quad \boxed{1}\,650\text{-}723\text{-}4284] \end{bmatrix}$$

The intended interpretation would be exactly the same. It also makes no difference what order we write the features in. For example, (17) and (18) are both equivalent to either of the following:

(19) a. $$\begin{bmatrix} \text{CHAIR} & [\text{TEL} \quad \boxed{1}\,650\text{-}723\text{-}4284] \\ \text{TEL} & \square \end{bmatrix}$$

b. $$\begin{bmatrix} \text{CHAIR} & [\text{TEL} \quad \boxed{1}] \\ [\text{TEL} & \boxed{1}\,650\text{-}723\text{-}4284 \end{bmatrix}$$

Finally, it should be noticed that the choice of a particular tag is also completely arbitrary. The following constraints are also equivalent to the ones in (17)–(19):

(20) a. $$\begin{bmatrix} \text{CHAIR} & [\text{TEL} \quad \boxed{279}\,650\text{-}723\text{-}4284] \\ \text{TEL} & \boxed{279} \end{bmatrix}$$

b. $$\begin{bmatrix} \text{TEL} & \boxed{N} \\ \text{CHAIR} & [\text{TEL} \quad \boxed{N}\,650\text{-}723\text{-}4284] \end{bmatrix}$$

These are still simple examples. In the chapters that follow: we will have occasion to combine the various tools introduced here into fairly complex constraints.

Exercise **1:** Practice with Combining Constraints

Are the following pairs of constraints compatible? If so, what does the combined constraint look like?

A. $$\begin{bmatrix} \text{TEL} & 650\text{-}723\text{-}4284 \end{bmatrix} \& \begin{bmatrix} \textit{department} \\ \text{NAME Metaphysics} \end{bmatrix}$$

B. $$\begin{bmatrix} \text{TEL} & 650\text{-}723\text{-}4284 \end{bmatrix} \& \begin{bmatrix} \text{TEL} & \\ \text{CHAIR} & [\text{TEL}\ \boxed{23}] \end{bmatrix}$$

C. $$\begin{bmatrix} \text{PRESIDENT} & \square \\ [\text{FOUNDERS} & \langle \boxed{1} \rangle \end{bmatrix} \& \begin{bmatrix} \textit{individual} & \\ [\text{NAME} & \text{John Hennessy} \end{bmatrix}$$

3.3 The Linguistic Application of Feature Structures

3.3.1 Feature Structure Categories

So how do typed feature structures help us with our linguistic concerns? Instead of saying that there is just one kind of linguistic entity, which must bear a value for every feature we recognize in our feature structures, we will often want to say that a given entity is of a certain type for which only certain features are appropriate. In fact, we will use typing in many ways: for example, to ensure that [NUM sg] (or [NUM pl]) can only be specified for certain kinds of words (for example, nouns, pronouns, and verbs), but not for prepositions or adjectives.[8] Likewise, we will eventually introduce a feature AUX to distinguish auxiliaries (helping verbs like *will* and have) from all other verbs, but we won't want to say that nouns are all redundantly specified as [AUX –]. Rather, the idea that we'll want our grammar to incorporate is that the feature AUX just isn't appropriate for nouns. We can use types as a basis for classifying the feature structures we introduce and the constraints we place on them. In so doing, we provide an easy way of saying that particular features only go with certain types of feature structure. This amounts to the beginnings of a linguistic ontology: the types lay out what kinds of linguistic entities exist in our theory, and the features associated with those types tell us what general properties each kind of entity exhibits.[9]

In addition, the organization of linguistic objects in terms of a type hierarchy with intermediate types (analogous to *organization* in the university example) is significant. Partial generalizations – generalizations that hold of many but not all entities – are very common in the domain of natural language. Intermediate types allow us to state those generalizations. This feature of our theory will become particularly prominent when we organize the lexical entries into a hierarchy in Chapter 8.

In this chapter, we will develop a feature-based grammar that incorporates key ideas from the CFG we used in Chapter 2. We will show how feature structures can solve some of the problems we raised in our critical discussion of that grammar. As we do so, we will gradually replace all the atomic category names used in the CFG (S, NP, V, etc.) by typed feature structures. Since the grammar presented in this chapter is modeled on the CFG of Chapter 2, it is just an intermediate step in our exposition. In Chapter 4, we will refine the Chapter 3 grammar so that in the chapters to come we can systematically expand its coverage to include a much wider set of data.

3.3.2 Words and Phrases

To start with, let us draw a very intuitive distinction between two types: word and phrase. Our grammar rules (i.e. our phrase structure rules) all specify the properties of phrases;

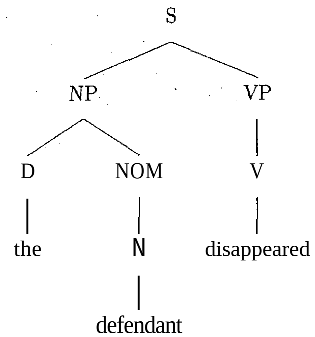
[8] Many such restrictions are language-particular. For example, adjectives are distinguished according to number (agreeing with the noun they modify) in many languages. Even prepositions exhibit agreement inflection in some languages (e.g. modern Irish) and need to be classified in similar terms.

[9] We might instead introduce some mechanism for directly stipulating dependencies between values of different features – such as a statement that the existence of a value for AUX implies that the value of POS is 'verb'. (For a theory that incorporates a mechanism like this, see Gazdar et al. 1985.) But mechanisms of this kind are unnecessary, given the availability of types in our theory

the lexicon provides a theory of words.

Consider the CFG tree in *(21):*

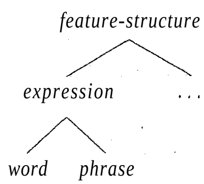
(21)



In this tree, the nodes S, NP, NOM, and VP are all *phrases.* The nodes *D,* N and V are all *words.* Both of these statements may seem unintuitive at first, because the words *word* and *phrase* are used in various ways. Sometimes a particular form, e.g. *the, defendant* or *disappeared,* is referred to as a word and certain sequences of forms, e.g. *the defendant* are called phrases. In the sense we intend here, however, 'word' refers to the category that the lexicon associates with a given form like *disappeared* and 'phrase' refers to the category that the grammar associates with a sequence of such forms.

Although there is an intuitive contrast between *words* and *phrases,* they also have some properties in common, especially in contrast to the more abstract grammatical types we will be positing below. We will therefore create our type hierarchy so that *word* and *phrase* are both subtypes of *expression*:[10]

(22)



One property that words and phrases have in common is part of speech. In the CFG of Chapter 2, this similarity was represented mnemonically (although not formally) in the atomic labels we choose for the categories: NP and N have in common that they are essentially nominal, VP and V that they are essentially verbal, etc. With feature structures, we can represent this formally. We will assume that all *expressions* specify values for a feature we will call HEAD. The value of HEAD will indicate the expression's part of speech. This feature is called HEAD because the part of speech of a phrase depends on the part of speech of one particular daughter, called the head daughter. That is, an NP structure is nominal because it has an N inside of it. That N is the head daughter of the NP structure.

[10]Note that the most general type in our theory will be called *feature-structure.* All of the types we introduce will be subtypes of *feature-structure.* If we were to fully flesh out the university example, something similar would have to be done there.

So far, then, our feature structure representation of the category NP looks like this:

(23) $\begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \textit{noun} \end{bmatrix}$

and our feature structure representation of the lexical entry for a noun, say *bird,* looks like this:

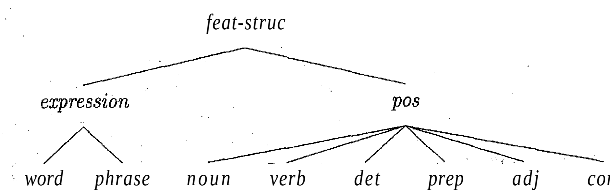
(24) $\left\langle \text{bird}\ ,\ \begin{bmatrix} \textit{word} & \\ \text{HEAD} & \textit{noun} \end{bmatrix} \right\rangle$

3.3.3 Parts of Speech

Let us reflect for a moment on parts of speech. There are certain features that are appropriate for certain parts of speech, but not others. We proposed above to distinguish helping verbs from all others in terms of the feature AUX(ILIARY), which will be appropriate only for verbs. Likewise, we will use the feature AGR(EEMENT) only for nouns, verbs, and determiners. To guarantee that only the right features go with the right parts of speech, we will introduce a set of types. Then we can declare feature appropriateness for each part of speech type, just as we did in our type hierarchy for Stanford University.

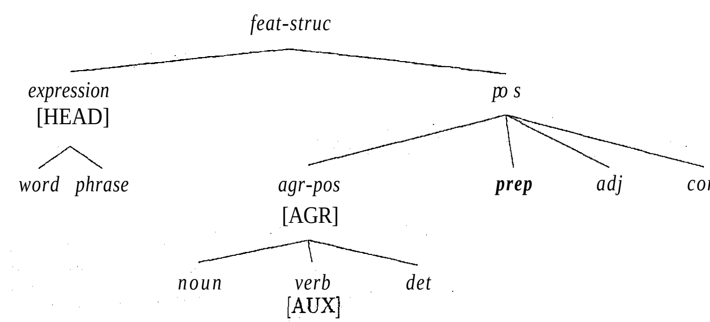
We therefore introduce the types *noun, verb, adj, prep, det,* and *conj* for the six lexical categories we have so far considered. We then make all of these subtypes of a type called *part-of-speech* (*pos*), which is itself a subtype of *feat(ure)-struc(ture).* The resulting type organization is as shown in **(25):**

(25)



But in fact, if we want to introduce features only once in a given type hierarchy, then we will have to modify this picture slightly. That's because there are three parts of speech that take the feature *AGR.*[11] We will thus modify the type hierarchy to give these three types a common supertype where the feature AGR is introduced, as shown in (26):

(26)



[11]There will be a few more as we expand the coverage of our grammar in later chapters.

In this way, determiners and nouns will both specify values for AGR and verbs will specify values for both AGR and AUX. Notice, however, that it is not the words themselves that specify values for these features – rather, it is the feature structures of type noun, verb or det. Individual words (and phrases) get associated with this information because they have a feature HEAD whose value is always a feature structure that belongs to some subtype of pos.

So far, we have motivated distinguishing the different subtypes of pos as a way of making sure that words only bear features that are appropriate for their part of speech. There is, however, another benefit. As discussed in Section 3.3.2 above, the value of the HEAD feature represents information that a phrase (more precisely, the mother nodes of a phrase structure) shares with its head daughter. (We will see how the grammar enforces this identity in Section 3.3.5 below.) The features we posit for the pos types (so far, AGR and AUX) also encode information that phrases share with their head daughters. This is particularly clear in the case of agreement: just as an NP is only nominal because it has an N inside of it, a singular NP is only singular because it has a singular N inside of it. By making AGR a feature of (the relevant subtypes of) pos, we can represent this very efficiently: we identify the HEAD value of the mother (say, NP) and that of its head daughter (N). In doing so, we identify not only the mother and head daughter's part of speech, but also any other associated information, for example, their number.[12] In refining our account of the feature structures of type pos, we will thus be formulating a general theory of what features the head daughter shares with its mother in a headed phrase.

3.3.4 Valence Features

The approach we are developing also provides a more satisfying analysis of our earlier categories IV, TV, and DTV. Instead of treating these as unanalyzable (i.e. as atoms), we now decompose these as feature structures. To do this, we introduce a new feature VAL (for 'valence'). The value of VAL is a feature structure (of type val-cat) representing the combinatoric potential of the word or phrase. The first feature we will posit under VAL is COMPS (for 'complements' – see Chapter 2, Section 2.7), which we use to indicate what the required following environment is for each type of verb: (For now, we assume that the possible values of COMPS are itr = intransitive, str = strict-transitive, and dtr = ditransitive, though we will revise this in the next chapter.)

(27)

$$\text{IV} = \begin{bmatrix} \textit{word} & \\ \text{HEAD} & \textit{verb} \\ \text{VAL} & \begin{bmatrix} \textit{val-cat} & \\ \text{COMPS} & \text{itr} \end{bmatrix} \end{bmatrix} \qquad \text{TV} = \begin{bmatrix} \textit{word} & \\ \text{HEAD} & \textit{verb} \\ \text{VAL} & \begin{bmatrix} \textit{val-cat} & \\ \text{COMPS} & \text{str} \end{bmatrix} \end{bmatrix}$$

$$\text{DTV} = \begin{bmatrix} \textit{word} & \\ \text{HEAD} & \textit{verb} \\ \text{VAL} & \begin{bmatrix} \textit{val-cat} & \\ \text{COMPS} & \text{dtr} \end{bmatrix} \end{bmatrix}$$

[12] We will return to the feature AGR and describe what kinds of things it takes as its value in Section 3.3.6 below.

The three categories described in (27) all share the type word and the feature specification [HEAD *verb*]. This is just the combination of types and features that we would naturally identify with the category V. And by analyzing categories in terms of types and features, we can distinguish between the different valence possibilities for verbs, while still recognizing that all verbs fall under a general category. The general category V is obtained by leaving the value of the VAL feature unspecified, as in (28):

(28) $$\text{V} = \begin{bmatrix} \textit{word} & \\ \text{HEAD} & \textit{verb} \end{bmatrix}$$

The term UNDERSPECIFICATION is commonly used in linguistics to indicate a less specific linguistic description. Given our modeling assumptions, underspecification has a precise interpretation: an underspecified description (or constraint) always picks out a larger class of feature structures than a fully specified one. In general, the less information given in a description (i.e. the more underspecified it is), the more models (feature structures) there are that will satisfy that description.

In the grammar so far, the category VP differs from the category V only with respect to its type assignment.[13] So VP is recast as the following description:

(29) $$\text{VP} = \begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \textit{verb} \end{bmatrix}$$

And the class of grammatical categories that includes just verbs and verb phrases is defined precisely by the underspecification in (30):

(30) $$\begin{bmatrix} \text{HEAD} & \textit{verb} \end{bmatrix}$$

Similarly, we can reanalyze the categories N and NP as follows:

(31) $$\text{N} = \begin{bmatrix} \textit{word} & \\ \text{HEAD} & \text{noun} \end{bmatrix} \qquad \text{NP} = \begin{bmatrix} \textit{phrase} & \\ \textbf{HEAD} & \text{noun} \end{bmatrix}$$

Within this general approach, we can retain all our previous categories (V, S, NP, etc.) as convenient abbreviations.

Underspecification allows us to provide compact descriptions for the sets of categories that our grammar will actually need to refer to, what linguists usually call 'natural classes'. For example, while we couldn't even talk about IV, DTV, and TV as one class in CFG, we can now refer to them together as words that are [HEAD *verb*]. We will use the symbol V as an abbreviation for this feature structure description, but it should now be regarded merely as an abbreviated description of the class of typed feature structures just described. The same is true for N, NP, VP, etc.

Observe that the feature analysis we have just sketched does not yet accommodate the category NOM. NP and NOM are both [HEAD *noun*]. And since the COMPS value is used to indicate what the following environment must be, it is not appropriate for the distinction between NP and NOM. Recall that NOM differs from NP in that it

[13] Additional differences with respect to their VAL values will be discussed shortly. A more sweeping reanalysis of the feature composition of these categories is introduced in Chapter 4 and carried on to subsequent chapters.

does not include the determiner, which is at the beginning of the phrase. In fact, it is a straightforward matter to use features to model our three-level distinction among N, NOM, and NP. NOM is the category that includes everything in the NP except the determiner, e.g. *picture of Yosemite* in *that picture of Yosemite*. We can distinguish NOM and NP using features in much the same way that we distinguished transitive and intransitive verbs – that is, by introducing a valence feature that indicates a restriction on the possible contexts in which the category in question can appear. In this case, the feature will specify whether or not a determiner is needed. We call this feature SPR (SPECIFIER). Just as we introduced 'complement' as a generalization of the notion of object, we are now introducing 'specifier' as a generalization of the notion of determiner.

For now, we will treat SPR as having two values: [SPR –] categories need a specifier on their left; [SPR +] categories do not, either because they label structures that already contain a specifier or that just don't need one. Note that like COMPS, SPR encodes an aspect of an expression's combinatoric potential. NP and NOM are thus defined as in (32):

(32)

$$NP = \begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \textit{noun} \\ \text{VAL} & \begin{bmatrix} \textit{val-cat} & \\ \text{COMPS} & \text{itr} \\ \text{SPR} & + \end{bmatrix} \end{bmatrix} \quad \text{NOM} = \begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \textit{noun} \\ \text{VAL} & \begin{bmatrix} \textit{val-cat} & \\ \text{COMPS} & \text{itr} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix}$$

We can also use the feature SPR to distinguish between VP and S, by treating a subject NP as the VP's specifier. That is, VP and S can be distinguished as follows:

(33)

$$\text{S} = \begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \textit{verb} \\ \text{VAL} & \begin{bmatrix} \textit{val-cat} & \\ \text{COMPS} & \text{itr} \\ \text{SPR} & + \end{bmatrix} \end{bmatrix} \quad \text{VP} = \begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \textit{verb} \\ \text{VAL} & \begin{bmatrix} \textit{val-cat} & \\ \text{COMPS} & \text{itr} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix}$$

In calling both determiners and subject NPs specifiers, we are claiming that the relationship between subject and VP is in important respects parallel to the relationship between determiner and NOM. The intuition behind this claim is that specifiers (subject NPs and determiners) serve to complete the phrases they are in. S and NP are fully formed categories, while NOM and VP are still incomplete. The idea that subjects and determiners play parallel roles seems particularly intuitive when we consider examples like (34).

(34) a. We created a monster.
  b. our creation of a monster

We will have more to say about the feature SPR in the next chapter.

Returning to (32), notice that we have extended the intuitive meaning of the specification [COMPS itr] so that it applies to phrases as well as to words. This is a natural

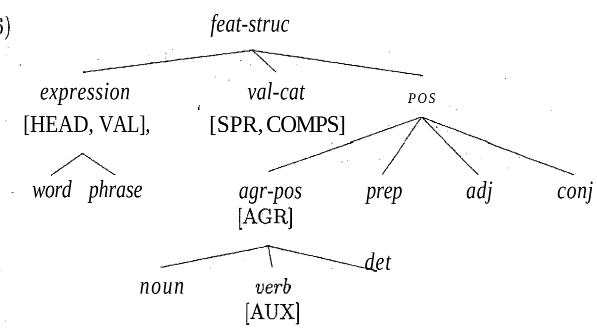
extension, as phrases (whether NP, S, VP or NOM) are like strictly intransitive verbs in that they cannot combine with complements. (Recall that a phrase contains its head's complement(s), so it can't combine with any more). Notice also that under this conception, the abbreviations NP and S both include the following feature specifications:

(35)

$$\begin{bmatrix} \text{VAL} & \begin{bmatrix} \textit{val-cat} & \\ \text{COMPS} & \text{itr} \\ \text{SPR} & + \end{bmatrix} \end{bmatrix}$$

As *words* and *phrases* both need to be specified for the valence features, we declare VAL to be appropriate for the type *expression*. The value of VAL is a *val-cat*, and COMPS and SPR are both features of *val-cat*.[14] Our type hierarchy now looks like this:

(36)



3.3.5 Reformulating the Grammar Rules

Turning now to the phrase structure rules considered in Chapter 2, we can reformulate our VP rules in terms of our new feature structure categories. Consider the following way of stating these rules:

(37) a.
$$\begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \boxed{1} \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix} \rightarrow \begin{bmatrix} \textit{word} & \\ \text{HEAD} & \boxed{1} \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix}$$

b.
$$\begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \boxed{1} \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix} \rightarrow \begin{bmatrix} \textit{word} & \\ \text{HEAD} & \boxed{1} \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{str} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix} \text{NP}$$

c.
$$\begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \boxed{1} \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix} \rightarrow \begin{bmatrix} \textit{word} & \\ \text{HEAD} & \boxed{1} \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{dtr} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix} \text{NP NP}$$

[14]In Chapter 5, we will add a further feature, MOD, to val-cat.

The two occurrences of $\boxed{1}$ in each of these rules tell us that the HEAD value of the mother and that of the first daughter must be identified. Since the rules in (37) were introduced as VP rules, the obvious value to assign to $\boxed{1}$ is verb. But, by stating the rules in this underspecified way, we can use them to cover some other structures as well. The first rule, for intransitives, can be used to introduce nouns, which can never take NP complements (in English). This is done simply by instantiating $\boxed{1}$ as noun, which will in turn cause the mother to be a NOM. To make this work right, we will have to specify that lexical nouns, like intransitive verbs, must be [COMPS itr]:

(38)
$$\left\langle \text{bird}, \begin{bmatrix} \text{word} & \\ \text{HEAD} & \text{noun} \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix} \right\rangle$$

Note that both verbs and nouns are lexically specified as [SPR −], i.e. as having not (yet) combined with a specifier.

We can now recast the CFG rules in (39):

(39) a. S → NP VP
b. NP → (D) NOM

Assuming, as we did above, that S is related to VP and V in just the same way that NP is related to NOM and N, the rules in (39) may be reformulated as (40a) and (40b), respectively:

(40) a.
$$\begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \boxed{1}\textit{verb} \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & + \end{bmatrix} \end{bmatrix} \rightarrow \text{NP} \begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \boxed{1} \\ \text{VAL} & \begin{bmatrix} \text{SPR} & - \end{bmatrix} \end{bmatrix}$$

b.
$$\begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \boxed{1}\textit{noun} \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & + \end{bmatrix} \end{bmatrix} \rightarrow \text{D} \begin{bmatrix} \text{phrase} & \\ \text{HEAD} & \boxed{1} \\ \text{VAL} & \begin{bmatrix} \text{SPR} & - \end{bmatrix} \end{bmatrix}$$

In these rules, 'NP' and 'D' are abbreviations for feature structure descriptions. NP was defined in (32) above. We'll assume that 'D' is interpreted as follows:

(41)
$$\text{D} = \begin{bmatrix} \text{D} & \text{det} \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & + \end{bmatrix} \end{bmatrix}$$

Note that the feature structure rule in (40b) differs from the CFG NP rule in (39b) in that the former makes the determiner obligatory. In fact, the optionality in the CFG rule caused it to overgenerate: while some nouns (like information or facts) can appear with

or without a determiner, others (like fact) require a determiner, and still others (like you or *Alex*) never take a determiner:

(42) a. I have the information.
b. I have information.
c. I was already aware of that fact.
d.*I was already aware of fact.
e. I know you.
f.*I know the you.

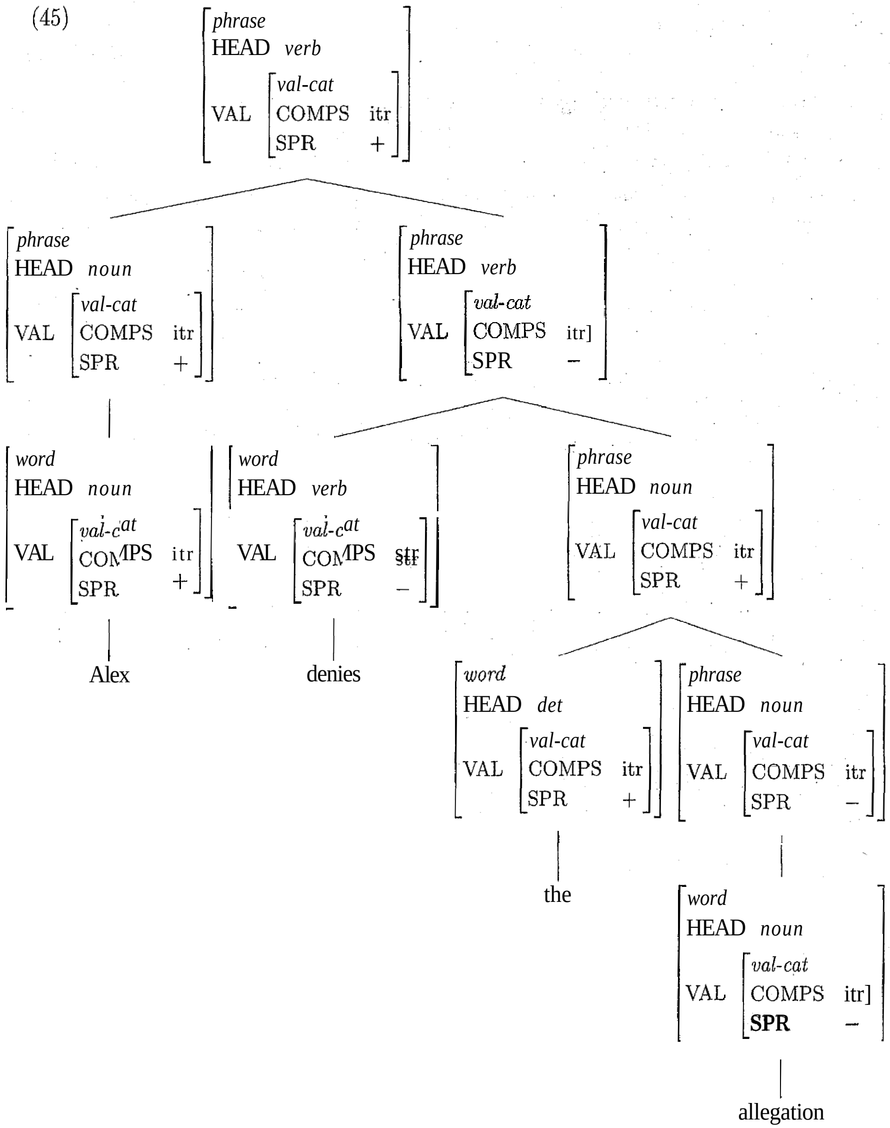
Since the CFG rule in (39b) doesn't distinguish between different kinds of Ns, it in fact licenses all of the NPs in (42). We will return to the problem of nouns whose determiners are truly optional (like information) in Chapter 8. The thing to note here is that the feature SPR allows us to distinguish nouns that require determiners (like fact or bird) from those that refuse determiners (like you or Alex). The former are specified as [SPR −], and build NPs with the rule in (40b). The latter are [SPR +] (see (43)), and require a new rule, given in (44):

(43)
$$\left\langle \text{Alex}, \begin{bmatrix} \text{word} & \\ \text{HEAD} & \text{noun} \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & + \end{bmatrix} \end{bmatrix} \right\rangle$$

(44)
$$\begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \boxed{1}\textit{noun} \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & + \end{bmatrix} \end{bmatrix} \rightarrow \begin{bmatrix} \textit{word} & \\ \text{HEAD} & \boxed{1} \\ \text{VAL} & \begin{bmatrix} \text{SPR} & + \end{bmatrix} \end{bmatrix}$$

Given the rules and categories just sketched, it is important to see that our grammar now licenses trees like the one shown in (45):

(45)



Exercise 2: Understanding Tree (45)

A. For each node in (45) other than the preterminal nodes, identify the rule that licensed it.

B. Find the right abbreviation (e.g. NP, S, ...) for each node in (45).

Two rules we haven't yet reconsidered are the ones that introduce PP modifiers, repeated in (46):

(46) a. VP $\rightarrow$ VP PP

b. NOM $\rightarrow$ NOM PP

Although we will have nothing to say about the internal structure of PPs in this chapter, we would like to point out the potential for underspecification to simplify these rules, as well. Once categories are modeled as feature structures, we can replace the two CFG rules in (46) with one grammar rule, which will look something like (47):

(47) $$\begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \boxed{2} \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix} \rightarrow \begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \boxed{2} \\ \text{VAL} & \begin{bmatrix} \text{SPR} & - \end{bmatrix} \end{bmatrix} \text{PP}$$

Note that the head daughter of this rule is unspecified for COMPS. In fact, all of the categories of type phrase licensed by our grammar are [COMPS itr], so specifying a COMPS value on the head daughter in addition to giving its type as phrase would be redundant.

Exercise 3: COMPS Value of Phrases

Look at the grammar summary in Section 3.6 and verify that this last claim is true.

In the next chapter, we will carry the collapsing of phrase structure rules even further. First, however, let us examine how features can be used in the analysis of agreement.

3.3.6 Representing Agreement with Features

In Section **3.3.3** above, we stated that the types noun, verb and det bear a feature AGR. In this section, we will consider what the value of that feature should be and how it can help us model subject-verb agreement.[15]

Agreement in English involves more than one kind of information. For subject-verb agreement, both the person and the number of the subject are relevant. Therefore, we want the value of AGR to be a feature structure that includes (at least) these two kinds of information, i.e. bears at least the features PER(SON) and NUM(BER). We will call the type of feature structure that has these features an agr-cat (agreement-category). The type agr-cat is a subtype of *feature-structure*.[16] The values of PER and NUM are atomic. The values of PER are drawn from the set {1st, 2nd, 3rd} and the values for NUM from the set {sg, pl}. The result is that instances of the type agr-cat will look like (48):

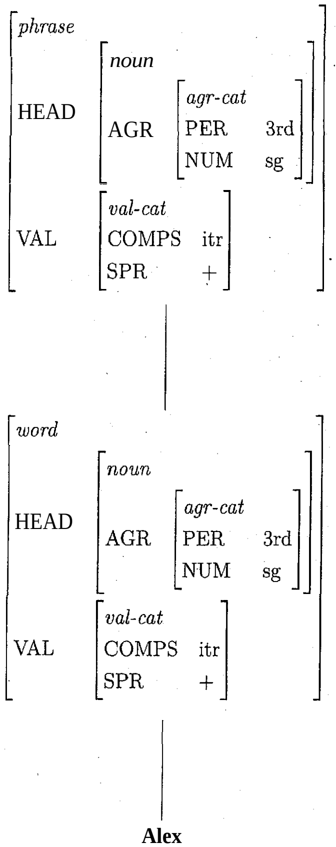
(48) $$\begin{bmatrix} \textit{agr-cat} & \\ \text{PER} & \text{3rd} \\ \text{NUM} & \text{sg} \end{bmatrix}$$

[15] Determiner-noun agreement will be addressed in Problem **3** and then brought up again in Chapter 4.

[16] See the grammar summary in Section 3.6 for how this addition affects the type hierarchy.

AGR is a feature of (certain) subtypes of *pos.* This means that it is a HEAD FEATURE, i.e. one of the features that appears inside the HEAD value. Consequently, AGR-specifications get passed up from words to phrases and then to larger phrases. For example, the mother node in (49) will have the same specification for AGR as its head daughter:

(49)

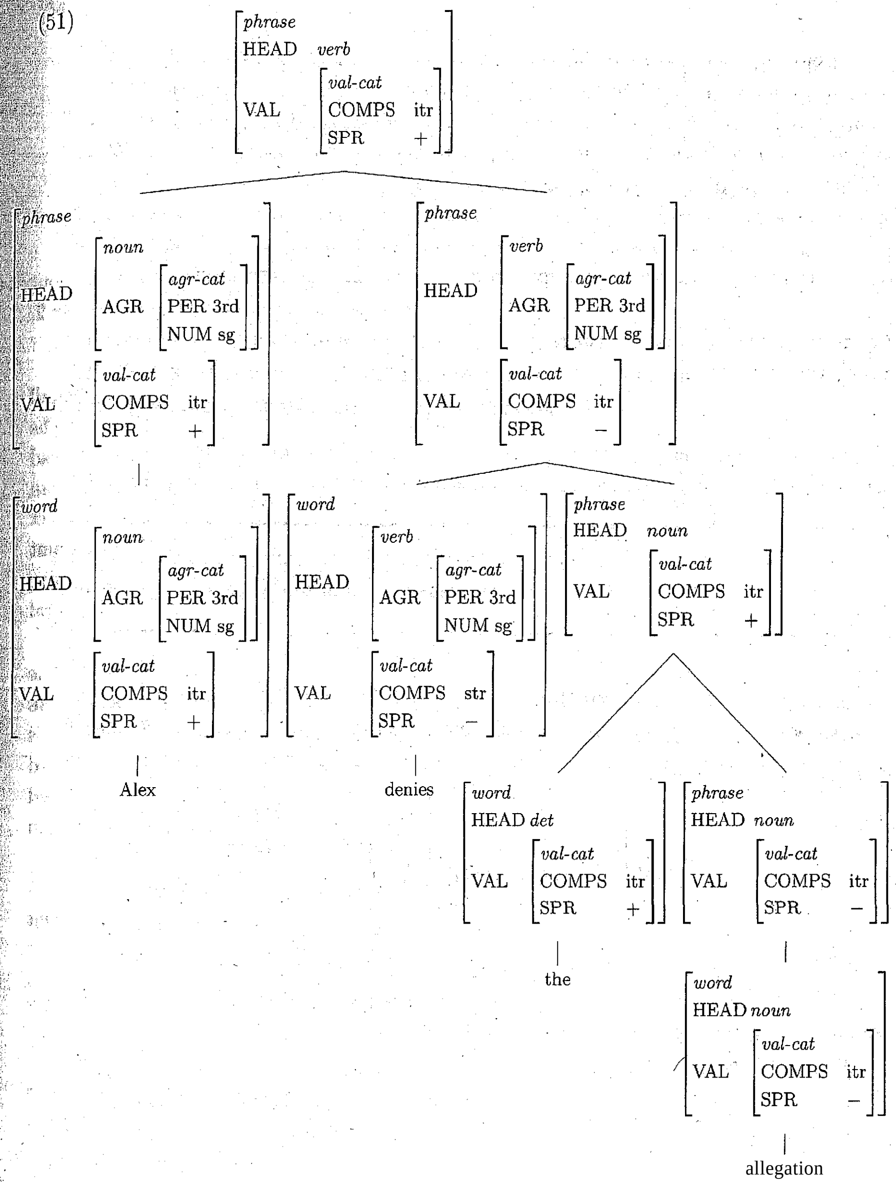


We want AGR information to be part of a phrase like this, because it is the kind of phrase that can be the subject of a simple sentence. If the verb within the VP and the noun that is the head of the subject NP both pass up their AGR specifications in this way, it is a simple matter to account for subject-verb agreement by revising our rule (40a) for combining NP and VP into an S. This revision may take the following form:

(50)
$$\begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \boxed{1}\textit{verb} \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & + \end{bmatrix} \end{bmatrix} \rightarrow \overset{\text{NP}}{\begin{bmatrix} \text{HEAD} & \begin{bmatrix}\text{AGR} & \boxed{2}\end{bmatrix} \end{bmatrix}} \begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \boxed{1}\begin{bmatrix}\text{AGR} & \boxed{2}\end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix}$$

And in consequence of the revision in (50), AGR values are constrained as illustrated in (51):[17]

(51)



[17] In this tree, we omit the AGR specifications on the object NP and the root node, even though the grammar will provide them.

More generally, assuming the appropriate lexical entries, the revised analysis correctly accounts for all the contrasts in (52):

(52) a. The defendant denies the allegation.
b. *The defendant deny the allegation.
c. The defendants deny the allegation.
d. *The defendants denies the allegation.
e. The defendant walks.
f. *The defendant walk.
g. The defendants walk.
h. *The defendants walks.

Representing categories as complexes of features enables us to capture these facts without proliferating grammar rules. This is a distinct improvement over the CFG of Chapter 2.

3.3.7 The **Head Feature Principle**

The grammar rules proposed in the previous sections ((37a–c), (40), and (47)) have all identified the mother's HEAD value with the HEAD value of one of the daughters. The relevant HEAD-sharing daughter is always the one we have been referring to as the head daughter: the N in a NOM phrase, the NOM in an NP, the V in a VP, the VP in an S, and the VP or NOM that co-occurs with a PP modifier. But our theory does not yet include any notion of head daughter. If it did, we could factor out a general constraint about identity of HEAD values, instead of stating the same constraint in each of our five rules (with possibly more to come). The purpose of this section is to propose a general principle with this effect.

Rather than stipulating identity of features in an ad hoc manner on both sides of the rules, our analysis will recognize that in a certain kind of phrase – a HEADED PHRASE – one daughter is assigned special status as the HEAD DAUGHTER. Once this notion is incorporated into our theory (thus providing a remedy for the second defect of standard CFGs noted in the last chapter), we can factor out the identity constraint that we need for all the headed phrases, making it a general principle. We will call this generalization the Head Feature Principle (HFP).

Certain rules introduce an element that functions as the head of the phrase characterized by the rule. We will call such rules HEADED RULES. To indicate which element introduced in a headed rule is the head daughter, we will label one element on the right hand side of the rule with the letter '**H**'. So a headed rule will have the following general form:[18]

(53) $[\mathit{phrase}] \rightarrow \ \ldots \ \mathbf{H}[\ ] \ \ldots$

So far, we have done two things: (i) we have identified the head daughter in a headed rule and (ii) we have bundled together (within the HEAD value) all the feature specifications that the head daughter must share with its mother. With these two adjustments in place, we are now in a position to simplify the grammar of headed phrases.

[18] Note that 'H', unlike the other shorthand symbols we use occasionally (e.g. 'V' and 'NP'), does not abbreviate a feature structure in a grammar rule. Rather, it merely indicates which feature structure in the rule corresponds to the phrase's head daughter.

First we simplify all the headed rules: they no longer mention anything about identity of HEAD values:

(54) a. $\begin{bmatrix}\mathit{phrase} & \\ \text{VAL} & \begin{bmatrix}\text{COMPS} & \text{itr}\\ \text{SPR} & -\end{bmatrix}\end{bmatrix} \rightarrow \mathbf{H}\begin{bmatrix}\mathit{word} & \\ \text{VAL} & \begin{bmatrix}\text{COMPS} & \text{itr}\\ \text{SPR} & -\end{bmatrix}\end{bmatrix}$

b. $\begin{bmatrix}\mathit{phrase} & \\ \text{VAL} & \begin{bmatrix}\text{COMPS} & \text{itr}\\ \text{SPR} & -\end{bmatrix}\end{bmatrix} \rightarrow \mathbf{H}\begin{bmatrix}\mathit{word} & \\ \text{VAL} & \begin{bmatrix}\text{COMPS} & \text{str}\\ \text{SPR} & -\end{bmatrix}\end{bmatrix} \ \text{NP}$

c. $\begin{bmatrix}\mathit{phrase} & \\ \text{VAL} & \begin{bmatrix}\text{COMPS} & \text{itr}\\ \text{SPR} & -\end{bmatrix}\end{bmatrix} \rightarrow \mathbf{H}\begin{bmatrix}\mathit{word} & \\ \text{VAL} & \begin{bmatrix}\text{COMPS} & \text{dtr}\\ \text{SPR} & -\end{bmatrix}\end{bmatrix} \ \text{NP} \ \text{NP}$

d. $\begin{bmatrix}\mathit{phrase} & \\ \text{VAL} & \begin{bmatrix}\text{COMPS} & \text{itr}\\ \text{SPR} & +\end{bmatrix}\end{bmatrix} \rightarrow \overset{\text{NP}}{\begin{bmatrix}\text{HEAD} & \begin{bmatrix}\text{AGR} & \boxed{2}\end{bmatrix}\end{bmatrix}} \ \mathbf{H}\begin{bmatrix}\mathit{phrase} & \\ \text{HEAD} & \begin{bmatrix}\mathit{verb} & \\ \text{AGR} & \boxed{2}\end{bmatrix}\\ \text{VAL} & \begin{bmatrix}\text{SPR} & -\end{bmatrix}\end{bmatrix}$

e. $\begin{bmatrix}\mathit{phrase} & \\ \text{VAL} & \begin{bmatrix}\text{COMPS} & \text{itr}\\ \text{SPR} & +\end{bmatrix}\end{bmatrix} \rightarrow \text{D} \ \mathbf{H}\begin{bmatrix}\mathit{phrase} & \\ \text{HEAD} & \mathit{noun}\\ \text{VAL} & \begin{bmatrix}\text{SPR} & -\end{bmatrix}\end{bmatrix}$

f. $\begin{bmatrix}\mathit{phrase} & \\ \text{VAL} & \begin{bmatrix}\text{COMPS} & \text{itr}\\ \text{SPR} & +\end{bmatrix}\end{bmatrix} \rightarrow \mathbf{H}\begin{bmatrix}\mathit{word} & \\ \text{HEAD} & \mathit{noun}\\ \text{VAL} & \begin{bmatrix}\text{SPR} & +\end{bmatrix}\end{bmatrix}$

g. $\begin{bmatrix}\mathit{phrase} & \\ \text{VAL} & \begin{bmatrix}\text{COMPS} & \text{itr}\\ \text{SPR} & -\end{bmatrix}\end{bmatrix} \rightarrow \mathbf{H}\begin{bmatrix}\mathit{phrase} & \\ \text{VAL} & \begin{bmatrix}\text{SPR} & -\end{bmatrix}\end{bmatrix} \ \text{PP}$

The element labeled '**H**' in the above rules is the head daughter.

Second, we state the Head Feature Principle as a general constraint governing all trees built by headed rules.

(55) Head Feature Principle (HFP)
In any headed phrase, the HEAD value of the mother and the HEAD value of the head daughter must be identical.

The HFP makes our rules simpler by factoring out those properties common to all headed phrases, and making them conditions that will quite generally be part of the trees defined by our grammar. By formulating the HFP in terms of HEAD value identity, we allow information specified by the rule, information present on the daughter or the mother, or information required by some other constraint all to be amalgamated, as long as that information is compatible.[19]

3.4 Phrase Structure Trees

At this point, we must address the general question of how rules, lexical entries and principles like the HFP interact to define linguistic structures. Our earlier discussion of this question in Chapter 2 requires some revision, now that we have introduced feature structures and types. In the case of simple context-free grammars, descriptions and structures are in simple correspondence: in CFG, each local subtree (that is, a mother node with its daughters) corresponds in a straightforward fashion to a rule of the grammar. All of the information in that local subtree comes directly from the rule. There is no reason to draw a distinction between the linguistic objects and the grammar's descriptions of them. But now that rules, lexical entries and principles like the HFP all contribute constraints (of varying degrees of specificity) that linguistic tokens must satisfy, we must take care to specify how these constraints are amalgamated and how the grammar specifies which expressions are grammatical.

3.4.1 The Formal System: an Informal Account

The distinction between descriptions and the structures they describe is fundamental. We use feature structures in our models of linguistic entities. Consider what this meant for the feature structures we used to model universities, departments and individuals. Each feature structure model was assumed to have all the properties relevant to understanding the university system; in our example, this included (for individuals) a name, a birthday, and a telephone number. The objects we took as models were thus complete in relevant respects.[20] Contrast this with descriptions of university individuals. These come in varying degrees of completeness. A description may be partial in not specifying values for every feature, in specifying only part of the (complex) value of a feature, in failing to specify a type, or in specifying nothing at all. A complete description of some entity will presumably be satisfied by only one thing – the entity in question. An empty description is satisfied by all the entities in the modeling domain. Any nonempty partial description is satisfied by some things in the modeling domain, and not by others.

Our theory of language works the same way. We use trees to model phrases and we use feature structures to model the grammatical categories that label the nodes in those trees. These models are complete (or RESOLVED) with respect to all linguistically relevant properties.[21] On the other hand, the lexical entries, grammar rules and principles are not models but rather partial descriptions of models. They thus need not be (and in

[19]The Head Feature Principle is sometimes formulated as 'percolation' of properties of lexical heads to the phrases that they 'project'. While it is often helpful to think of information as propagating up or down through a tree, this is just a metaphor. Our formulation of the generalization avoids attributing directionality of causation in the sharing of properties between phrases and their heads.

[20]Of course, a model and the thing it is a model of differ with respect to certain irrelevant properties. Our models of university individuals should omit any irrelevant properties that all such individuals presumably have, ranging from hair color to grandmothers' middle names to disposition with respect to Indian food.

[21]'Resolvedness' is a direct consequence of our decision to define complex feature structures as total functions over a given domain of features.

fact usually aren't) fully resolved. For example, since the English word you is ambiguous between singular and plural, we might want to posit a lexical entry for it like the following:

(56)
$$\left\langle \text{you}\ ,\ \begin{bmatrix} \textit{word} & \\ \text{HEAD} & \begin{bmatrix} \textit{noun} & \\ \text{AGR} & [\text{PER}\ \ \text{2nd}] \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & + \end{bmatrix} \end{bmatrix} \right\rangle$$

This lexical entry is not complete in that it does not provide a specification for the feature NUM.[22]

Because the lexical entry is underspecified, it licenses two distinct WORD STRUCTURES (local, non-branching subtrees whose mother is of type word). These are shown in (57) and (58):

(57)
$$\begin{array}{c} \begin{bmatrix} \textit{word} & \\ \text{HEAD} & \begin{bmatrix} \text{noun} & \\ \text{AGR} & \begin{bmatrix} \textit{agr-cat} & \\ \text{PER} & \text{2nd} \\ \text{NUM} & \text{sg} \end{bmatrix} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \textit{val-cat} & \\ \text{COMPS} & \text{itr} \\ \text{SPR} & + \end{bmatrix} \end{bmatrix} \\ | \\ \text{you} \end{array}$$

(58)
$$\begin{array}{c} \begin{bmatrix} \textit{word} & \\ \text{HEAD} & \begin{bmatrix} \text{noun} & \\ \text{AGR} & \begin{bmatrix} \textit{agr-cat} & \\ \text{PER} & \text{2nd} \\ \text{NUM} & \text{pl} \end{bmatrix} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \textit{val-cat} & \\ \text{COMPS} & \text{itr} \\ \text{SPR} & + \end{bmatrix} \end{bmatrix} \\ | \\ \text{you} \end{array}$$

Here all the appropriate features are present (the mothers' feature structures are 'totally well-typed') and each feature has a completely resolved value.[23]

[22]This analysis of the ambiguity of *you* won't work in later versions of our grammar, and is presented here by way of illustration only.

[23]Again, this follows from defining feature structures in terms of total functions.

The relationship of the models to the grammar becomes more intricate when we consider not only lexical entries, but also grammar rules and the one general principle we have so far. These can all be thought of as constraints. Together, they serve to delimit the class of tree structures licensed by the grammar. For example, the grammar rule in (54b) above, repeated here as (59), is a constraint that can be satisfied by a large number of local subtrees. One such subtree is given in (60):

(59)
$$\begin{bmatrix} \textit{phrase} & \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix} \rightarrow \mathbf{H} \begin{bmatrix} \text{word} & \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{str} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix} \ \text{NP}$$

(60)
$$\begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \begin{bmatrix} \textit{verb} & \\ \text{AGR} & \begin{bmatrix} \textit{agr-cat} & \\ \text{PER} & \text{2nd} \\ \text{NUM} & \text{pl} \end{bmatrix} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \textit{val-cat} & \\ \text{COMPS} & \text{itr} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix}$$

$$\begin{bmatrix} \textit{word} & \\ \text{HEAD} & \begin{bmatrix} \textit{verb} & \\ \text{AGR} & \begin{bmatrix} \textit{agr-cat} & \\ \text{PER} & \text{2nd} \\ \text{NUM} & \text{pl} \end{bmatrix} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \textit{val-cat} & \\ \text{COMPS} & \text{str} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix} \quad \begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \begin{bmatrix} \textit{noun} & \\ \text{AGR} & \begin{bmatrix} \textit{agr-cat} & \\ \text{PER} & \text{3rd} \\ \text{NUM} & \text{pl} \end{bmatrix} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \textit{val-cat} & \\ \text{COMPS} & \text{itr} \\ \text{SPR} & + \end{bmatrix} \end{bmatrix}$$

How many local subtrees are there that satisfy rule (59)? The answer to this question breaks down into a number of subquestions:

(61) a. How many feature structure categories can label the mother node?
b. How many feature structures categories can label the first daughter?
c. How many feature structures categories can label the second daughter?

The number of models satisfying (59) will be obtained by multiplying the answer to (61a) times the answer to (61b) times the answer to (61c), because, in the absence of other constraints, these choices are independent of one another.

Let us consider the mother node first. Here the types of the mother's and head daughter's feature structures are fixed by the rule, as are the SPR and COMPS values,

but the HEAD value is left unconstrained. In the grammar developed in this chapter, we have six parts of speech. This means that there are six options for the type of the HEAD value. If we pick noun, det, or verb, however, we have more options, depending on the values of AGR. Given that PER has three distinct values and NUM has two, there are six possible AGR values. Hence there are six distinct HEAD values of type noun, six of type det and six of type verb. Given that there is only one HEAD value of type *adj*, one of type prep and one of type *conj*, it follows that there are exactly 21 (= **(3** x 6) **+ 3)** possible HEAD values for the mother. Since all other feature values are fixed by the rule, there are then 21 possible feature structures that could label the mother node.

By similar reasoning, there are exactly 21 possible feature structures that could label the first (head) daughter in a local subtree satisfying rule (59). As for the second daughter, which is constrained to be an NP, there are only 6 possibilities – those determined by varying AGR values. Thus, there are 2646 (21 x 21 x 6) local subtrees satisfying rule (59), given the grammar developed in this chapter.

Note that one of these is the local subtree shown in (62), where the mother and the head daughter have divergent HEAD values:

(62) A Tree Not Licensed by the Grammar

$$\begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \begin{bmatrix} \textit{noun} & \\ \text{AGR} & \begin{bmatrix} \textit{agr-cat} & \\ \text{PER} & \text{3rd} \\ \text{NUM} & \text{pl} \end{bmatrix} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \textit{val-cat} & \\ \text{COMPS} & \text{itr} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix}$$

$$\begin{bmatrix} \textit{word} & \\ \text{HEAD} & \begin{bmatrix} \textit{verb} & \\ \text{AGR} & \begin{bmatrix} \textit{agr-cat} & \\ \text{PER} & \text{2nd} \\ \text{NUM} & \text{pl} \end{bmatrix} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \textit{val-cat} & \\ \text{COMPS} & \text{str} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix} \quad \begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \begin{bmatrix} \textit{noun} & \\ \text{AGR} & \begin{bmatrix} \textit{agr-cat} & \\ \text{PER} & \text{3rd} \\ \text{NUM} & \text{pl} \end{bmatrix} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \textit{val-cat} & \\ \text{COMPS} & \text{itr} \\ \text{SPR} & + \end{bmatrix} \end{bmatrix}$$

It is subtrees like this that are ruled out by the HFP, because the HFP requires that the HEAD value of the mother be identical to that of the head daughter. Hence, by incorporating the HFP into our theory, we vastly reduce the number of well-formed local subtrees licensed by any headed rule. The number of local subtrees satisfying both (59) and the HFP is just 126 (21 x 6). And in fact only 42 ((6 + 1) x 6) of these will ever be used in trees for complete sentences licensed by our grammar: in such trees, a word structure must be compatible with the head daughter, but only word structures for verbs or prepositions are ever specified as [COMPS str].

We complete the picture in much the same way as we did for CFGs. A phrase structure tree Φ is licensed by a grammar G if and only if:

- Φ is terminated (i.e. the nodes at the bottom of the tree are all labeled by lexical forms),
- the mother of Φ is labeled by S,[24]
- each local subtree within Φ is licensed by a grammar rule of G or a lexical entry of G, and
- each local subtree within Φ obeys all relevant principles of G.

A grammar is successful to the extent that it can be shown that the tree structures it licenses – its models – have properties that correspond to our observations about how the language really is. Recall from our discussion in Section 2.9 of Chapter 2 that what we are taking to be the reality of language involves aspects of both the mental representations of individual speakers and the social interactions among speakers. Thus, we're idealizing a fair bit when we talk about the sentences of the language being 'out there' in the world. In particular, we're abstracting away from variation across utterances and systematic variation across speakers. But we will have plenty to talk about before this idealization gets in our way, and we will have many observations and intuitions to draw from in evaluating the claims our models make about the external reality of language.

3.4.2 An Example

Consider the sentence They *swim*. Let's suppose that the lexical entries for *they* and swim are as shown in (63). Note that lexical entry for the plural form *swim* is underspecified for person.

(63) a.

$$\left\langle \text{they}\ ,\ \begin{bmatrix} \textit{word} & \\ \text{HEAD} & \begin{bmatrix} \textit{noun} & \\ \text{AGR} & \begin{bmatrix} \text{PER} & \text{3rd} \\ \text{NUM} & \text{pl} \end{bmatrix} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & + \end{bmatrix} \end{bmatrix} \right\rangle$$

[24]Remember that S is now an abbreviation defined in **(33)** above.

b.

$$\left\langle \text{swim}\ ,\ \begin{bmatrix} \textit{word} & \\ \text{HEAD} & \begin{bmatrix} \textit{verb} & \\ \text{AGR} & \begin{bmatrix} \text{NUM} & \text{pl} \end{bmatrix} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix} \right\rangle$$

Given these two lexical entries, the following are both well-formed local subtrees, according to our theory:

(64) a.

$$\begin{array}{c} \begin{bmatrix} \textit{word} & \\ \text{HEAD} & \begin{bmatrix} \textit{noun} & \\ \text{AGR} & \begin{bmatrix} \textit{agr-cat} & \\ \text{PER} & \text{3rd} \\ \text{NUM} & \text{pl} \end{bmatrix} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \textit{val-cat} & \\ \text{COMPS} & \text{itr} \\ \text{SPR} & + \end{bmatrix} \end{bmatrix} \\ | \\ \text{they} \end{array}$$

b.

$$\begin{array}{c} \begin{bmatrix} \textit{word} & \\ \text{HEAD} & \begin{bmatrix} \textit{verb} & \\ \text{AGR} & \begin{bmatrix} \textit{agr-cat} & \\ \text{PER} & \text{3rd} \\ \text{NUM} & \text{pl} \end{bmatrix} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \textit{val-cat} & \\ \text{COMPS} & \text{itr} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix} \\ | \\ \text{swim} \end{array}$$

Observe that these word structures contain only fully resolved feature structures. Furthermore, the structure in (64b) contains a specification for the feature PER that will make the relevant tree structure compatible with the structure over *they* when we combine them to build a sentence.

These lexical structures can now be embedded within larger structures sanctioned by the rules in (54f,a) and the HFP, as illustrated in (65a,b):

(65) a.

$$\begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \begin{bmatrix} \textit{noun} & \\ \text{AGR} & \begin{bmatrix} \textit{agr-cat} & \\ \text{PER} & \text{3rd} \\ \text{NUM} & \text{pl} \end{bmatrix} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \textit{val-cat} & \\ \text{COMPS} & \text{itr} \\ \text{SPR} & + \end{bmatrix} \end{bmatrix}$$

$$|$$

$$\begin{bmatrix} \textit{word} & \\ \text{HEAD} & \begin{bmatrix} \textit{noun} & \\ \text{AGR} & \begin{bmatrix} \textit{agr-cat} & \\ \text{PER} & \text{3rd} \\ \text{NUM} & \text{pl} \end{bmatrix} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \textit{val-cat} & \\ \text{COMPS} & \text{itr} \\ \text{SPR} & + \end{bmatrix} \end{bmatrix}$$

$$|$$

they

b.

$$\begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \begin{bmatrix} \textit{verb} & \\ \text{AGR} & \begin{bmatrix} \textit{agr-cat} & \\ \text{PER} & \text{3rd} \\ \text{NUM} & \text{pl} \end{bmatrix} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \textit{val-cat} & \\ \text{COMPS} & \text{itr} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix}$$

$$|$$

$$\begin{bmatrix} \textit{word} & \\ \text{HEAD} & \begin{bmatrix} \textit{verb} & \\ \text{AGR} & \begin{bmatrix} \textit{agr-cat} & \\ \text{PER} & \text{3rd} \\ \text{NUM} & \text{pl} \end{bmatrix} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \textit{val-cat} & \\ \text{COMPS} & \text{itr} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix}$$

$$|$$

swim

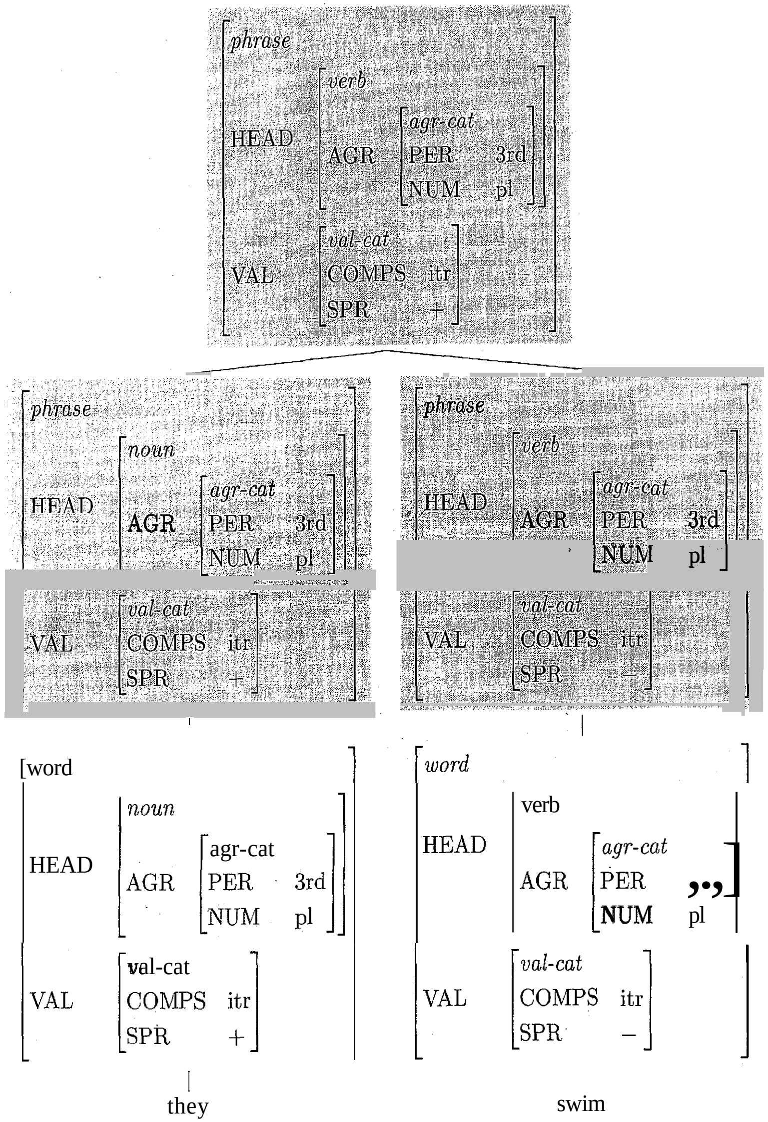
The shading in these and subsequent trees indicates the portion of the tree that is licensed by the rule in question (together with the HFP).

And finally, we can use rule (54d), repeated here as (66) to build a sentential phrase structure that combines the two previous structures. This is shown in (67):

(66)

$$\begin{bmatrix} \textit{phrase} & \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & + \end{bmatrix} \end{bmatrix} \rightarrow \overset{\text{NP}}{\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{AGR} & \boxed{2} \end{bmatrix} \end{bmatrix}} \ \mathbf{H}\begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \begin{bmatrix} \textit{verb} \\ \text{AGR} & \boxed{2} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SPR} & - \end{bmatrix} \end{bmatrix}$$

(67)

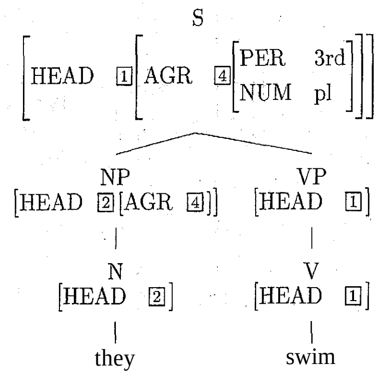


The nodes of the local subtree licensed by the'rule in (66) (and the HFP) are again indicated by shading.

We will display phrase structure trees throughout this book, usually to illustrate the effect of particular constraints that are under discussion. Though the feature structures in the trees licensed by our grammar are always total functions, we will often display tree diagrams that contain defined abbreviations (e.g. NP or S) or which omit irrelevant feature specifications (or both). Similarly, we may want to illustrate particular identities

within phrase structure trees that have been enforced by linguistic constraints. To this end, we will sometimes include tags (e.g. 3) in our tree digrams to indicate identities induced by linguistic constraints. To illustrate the effect of the HFP, for example, we might replace the tree diagram in (67) with one like (68):

(68)



A diagram like (68) always abbreviates a phrase structure tree whose nodes are labeled by fully determinate, resolved feature structures.

3.5 Summary

The introduction of features has given us a formal mechanism for talking about ways in which sets of words (and phrases) behave both alike and differently. By allowing embedded feature structures, underspecifying categories, and formulating general constraints stating identities that must hold in well-formed trees, we have been able to generalize our phrase structure rules and,reduce their number. This in turn has led us to carefully distinguish between our grammar rules and the fully determinate ('resolved') structures that satisfy them, and further between the models licensed by our grammar and the abbreviated representations of those models such as (68) that we will often use to focus our discussions throughout the remainder of this book.

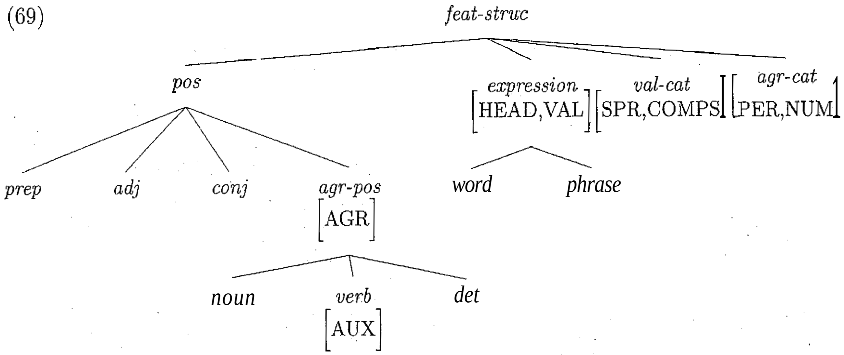
The theory we are developing is still closely related to standard CFG, yet it is somewhat more abstract. We no longer think of our phrase structure rules as specifying all the information that labels the nodes of trees. Rather, the rules, the lexicon, and some general principles – of which the HFP is the first example – all place certain constraints on trees, and any imaginable tree is well-formed so long as it conforms to these constraints. In this way, our grammar continues to be constraint-based, with the rules, lexical entries, and general principles all working together to define the well-formed structures of the language.

But the changes introduced in this chapter are not yet sufficient. They still leave us with three rules introducing complements that have too much in common and should be collapsed, and two rules introducing specifiers that similarly need to be collapsed. Moreover, as we will see in the next chapter, we have simplified the facts of agreement too much. The grammar we develop there will allow the more complex facts to be systematized, while at the same time eliminating further redundancy from the phrase structure rules of our grammar.

3.6 The Chapter 3 Grammar

3.6.1 The Type Hierarchy

(69)



3.6.2 Feature Declarations and Type Constraints

TYPE	FEATURES/CONSTRAINTS	IST
feat-struc		
pos		*feat-struc*
agr-pos	$\begin{bmatrix} \text{AGR} & \textit{agr-cat} \end{bmatrix}$	*pos*
noun		*agr-pos*
det		*agr-pos*
verb	$\begin{bmatrix} \text{AUX} & \{+, -\} \end{bmatrix}$	*agr-pos*
prep		*pos*
adj		*pos*
conj		*pos*
expression	$\begin{bmatrix} \text{HEAD} & \textit{pos} \\ \text{VAL} & \textit{val-cat} \end{bmatrix}$	*feat-struc*
word		*expression*
phrase		*expression*
val-cat	$\begin{bmatrix} \text{COMPS} & \{\text{itr, str, dtr}\} \\ \text{SPR} & \{+, -\} \end{bmatrix}$	*feat-struc*
agr-cat	$\begin{bmatrix} \text{PER} & \{\text{1st, 2nd, 3rd}\} \\ \text{NUM} & \{\text{sg, pl}\} \end{bmatrix}$	*feat-struc*

3.6.3 Abbreviations

(70)

$$\text{S} = \begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \textit{verb} \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & + \end{bmatrix} \end{bmatrix} \qquad \text{NP} = \begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \textit{noun} \\ & \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & + \end{bmatrix} \end{bmatrix}$$

$$\text{VP} = \begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \textit{verb} \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix} \qquad \text{NOM} = \begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \textit{noun} \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix}$$

$$\text{V} = \begin{bmatrix} \textit{word} & \\ \text{HEAD} & \textit{verb} \end{bmatrix} \qquad = \begin{bmatrix} \textit{word} & \\ \text{HEAD} & \textit{noun} \end{bmatrix}$$

$$\text{D} = \begin{bmatrix} \textit{word} & \\ \text{HEAD} & \textit{det} \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & + \end{bmatrix} \end{bmatrix}$$

3.6.4 The Grammar Rules

(71) Head-Complement Rule 1:

$$\begin{bmatrix} \textit{phrase} & \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix} \rightarrow \mathbf{H} \begin{bmatrix} \textit{word} & \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix}$$

(72) Head-Complement Rule 2:

$$\begin{bmatrix} \textit{phrase} & \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix} \rightarrow \text{H} \begin{bmatrix} \textit{word} & \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{str} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix} \ \text{NP}$$

(73) Head-Complement Rule 3:

$$\begin{bmatrix} \textit{phrase} & \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & - \end{bmatrix} \end{bmatrix} \rightarrow \mathbf{H} \begin{bmatrix} \textit{word} & \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{dtr} \\ \textbf{SPR} & - \end{bmatrix} \end{bmatrix} \ \text{NP} \ \ \text{NP}$$

(74) Head-Specifier Rule 1:

$$\begin{bmatrix} \textit{phrase} & \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & + \end{bmatrix} \end{bmatrix} \rightarrow \overset{\text{NP}}{\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{AGR} & \boxed{1} \end{bmatrix} \end{bmatrix}} \ \mathbf{H} \begin{bmatrix} \textit{phrase} & \\ \text{HEAD} & \begin{bmatrix} \textit{verb} & \\ \text{AGR} & \boxed{1} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SPR} & - \end{bmatrix} \end{bmatrix}$$

(75) Head-Specifier Rule 2:[25]

$$\begin{bmatrix} \textit{phrase} & \\ \text{VAL} & \begin{bmatrix}\text{COMPS} & \text{itr}\\ \text{SPR} & +\end{bmatrix}\end{bmatrix} \rightarrow \text{D} \quad \mathbf{H}\begin{bmatrix}\textit{phrase} & \\ \text{HEAD} & \textit{noun} \\ \text{VAL} & \begin{bmatrix}\text{SPR} & -\end{bmatrix}\end{bmatrix}$$

(76) Non-Branching NP Rule:

$$\begin{bmatrix} \textit{phrase} & \\ \text{VAL} & \begin{bmatrix}\text{COMPS} & \\ \text{SPR} & +\end{bmatrix}\end{bmatrix} \rightarrow \mathbf{H}\begin{bmatrix}\textit{word} & \\ \text{HEAD} & \textit{noun} \\ \text{VAL} & \begin{bmatrix}\text{SPR} & +\end{bmatrix}\end{bmatrix}$$

(77) Head-Modifier Rule:

$$\begin{bmatrix} \textit{phrase} & \\ \text{VAL} & \begin{bmatrix}\text{COMPS} & \text{itr}\\ \text{SPR} & -\end{bmatrix}\end{bmatrix} \rightarrow \mathbf{H}\begin{bmatrix}\textit{phrase} & \\ \text{VAL} & \begin{bmatrix}\text{SPR} & -\end{bmatrix}\end{bmatrix} \text{ PP}$$

(78) Coordination Rule:

$$\boxed{1} \rightarrow \boxed{1}^{+} \begin{bmatrix}\textit{word} & \\ \text{HEAD} & \textit{conj}\end{bmatrix} \boxed{1}$$

3.6.5 The Head Feature Principle (HFP)

In any headed phrase, the HEAD value of the mother and the HEAD value of the head daughter must be identical.

3.6.6 Sample Lexical Entries

(79)

$$\left\langle \text{walks}, \begin{bmatrix}\textit{word} & \\ \text{HEAD} & \begin{bmatrix}\textit{verb} & \\ \text{AGR} & \begin{bmatrix}\text{NUM} & \text{sg}\\ \text{PER} & \text{3rd}\end{bmatrix}\end{bmatrix} \\ \text{VAL} & \begin{bmatrix}\text{COMPS} & \text{itr}\\ \text{SPR} & -\end{bmatrix}\end{bmatrix} \right\rangle$$

$$\left\langle \text{walk}, \begin{bmatrix}\textit{word} & \\ \text{HEAD} & \begin{bmatrix}\textit{verb} & \\ \text{AGR} & \begin{bmatrix}\text{NUM} & \text{pl}\end{bmatrix}\end{bmatrix} \\ \text{VAL} & \begin{bmatrix}\text{COMPS} & \text{itr}\\ \text{SPR} & -\end{bmatrix}\end{bmatrix} \right\rangle$$

[25]See Problem **3** for more on this rule.

$$\left\langle \text{denies}, \begin{bmatrix}\textit{word} & \\ \text{HEAD} & \begin{bmatrix}\textit{verb} & \\ \text{AGR} & \begin{bmatrix}\text{NUM} & \text{sg}\\ \text{PER} & \text{3rd}\end{bmatrix}\end{bmatrix} \\ \text{VAL} & \begin{bmatrix}\text{COMPS} & \text{str}\\ \text{SPR} & -\end{bmatrix}\end{bmatrix} \right\rangle$$

$$\left\langle \text{deny}, \begin{bmatrix}\textit{word} & \\ \text{HEAD} & \begin{bmatrix}\textit{verb} & \\ \text{AGR} & \begin{bmatrix}\text{NUM} & \text{pl}\end{bmatrix}\end{bmatrix} \\ & \begin{bmatrix}\text{COMPS} & \text{str}\\ \text{SPR} & -\end{bmatrix}\end{bmatrix} \right\rangle$$

$$\left\langle \text{defendant}, \begin{bmatrix}\textit{word} & \\ \text{HEAD} & \begin{bmatrix}\textit{noun} & \\ \text{AGR} & \begin{bmatrix}\text{PER} & \text{3rd}\\ \text{NUM} & \text{sg}\end{bmatrix}\end{bmatrix} \\ \text{VAL} & \begin{bmatrix}\text{COMPS} & \text{itr}\\ \text{SPR} & -\end{bmatrix}\end{bmatrix} \right\rangle$$

$$\left\langle \text{the}, \begin{bmatrix}\textit{word} & \\ \text{HEAD} & \textit{det} \\ \text{VAL} & \begin{bmatrix}\text{COMPS} & \text{itr}\\ \text{SPR} & +\end{bmatrix}\end{bmatrix} \right\rangle$$

$$\left\langle \text{Chris}, \begin{bmatrix}\textit{word} & \\ \text{HEAD} & \begin{bmatrix}\textit{noun} & \\ \text{AGR} & \begin{bmatrix}\text{PER} & \text{3rd}\\ \text{NUM} & \text{sg}\end{bmatrix}\end{bmatrix} \\ \text{VAL} & \begin{bmatrix}\text{COMPS} & \text{itr}\\ \text{SPR} & +\end{bmatrix}\end{bmatrix} \right\rangle$$

$$\left\langle \text{YOU}, \begin{bmatrix}\textit{word} & \\ \text{HEAD} & \begin{bmatrix}\textit{noun} & \\ \text{AGR} & \begin{bmatrix}\text{PER} & \text{2nd}\end{bmatrix}\end{bmatrix} \\ \text{VAL} & \begin{bmatrix}\text{COMPS} & \text{itr}\\ \text{SPR} & +\end{bmatrix}\end{bmatrix} \right\rangle$$

$$\left\langle \text{and}\ ,\ \begin{bmatrix} \textit{word} & \\ \text{HEAD} & \textit{conj} \\ \text{VAL} & \begin{bmatrix} \text{COMPS} & \text{itr} \\ \text{SPR} & + \end{bmatrix} \end{bmatrix} \right\rangle$$

3.7 Further Reading

One of the earliest (but often ignored) demonstrations of the descriptive power of feature structures is Harman 1963. Chomsky (1965) provides one of the earliest explicit discussions of syntactic features in generative grammar. The modern tradition of using complex feature structures (that is, features with feature structures as their values) begins with Kay 1979, Bear 1981, Bresnan 1982b, and Gazdar 1981 (see also Kaplan 1975 and Gazdar et al. 1985). For an elementary discussion of the formal properties of unification and its use in grammatical description, see Shieber 1986. For differing and more detailed technical presentations of the logic of typed feature structures, see King 1989, Carpenter 1992, Richter 1999, 2000, and Penn 2000.

3.8 Problems

⚠ Problem 1: Applying the Chapter 3 Grammar

A. Formulate a lexical entry for the word *defendants*.

B. Draw a tree for the sentence *The defendants walk.* Show the values for all of the features on every node and use tags to indicate the effects of any identities that the grammar requires.

C. Explain how your lexical entry for *defendants* interacts with the Chapter 3 grammar to rule out **The defendants walks.* Your explanation should make reference to grammar rules, lexical entries and the HFP.

Problem 2: 1st Singular and 2nd Singular Forms of Verbs

The sample lexical entry for *walk* given in (79) is specified as [AGR [NUM pl]]. This accounts for (i)–(iii), but not (iv) and (v):

(i) They walk.

(ii) We walk.

(iii) You (pl) walk. (cf. You yourselves walk.)

(iv) You (sg) walk. (cf. You yourself walk.)

(v) I walk.

Formulate lexical entries for *walk* in (iv) and (v). Be sure that those lexical entries don't license (vi):

(vi)*Dana walk.

⚠ **Problem 3**: Determiner-Noun Agreement

The Chapter 3 grammar declares AGR to be a feature appropriate for the types *noun*, *verb*, and *det*, but so far we haven't discussed agreement involving determiners. Unlike the determiner *the*, most other English determiners do show agreement with the nouns they combine with:

(i) a bird/*a birds

(ii) this bird/*this birds

(iii) that bird/*that birds

(iv) these birds/*these bird

(v) those birds/*those bird

(vi) many birds/*many bird

A. Formulate lexical entries for *this* and *these*.

B. Modify Head-Specifier Rule 2 so that it enforces agreement between the noun and the determiner just like Head-Specifier Rule 1 enforces agreement between the NP and the VP.

C. Draw a tree for the NP *these birds*. Show the value for all features of every node and use tags to indicate the effects of any identities that the grammar (including your modified HSR2) the Head Feature Principle requires.

Problem 4: Coordination and Modification

The Chapter 3 Grammar includes a coordination rule that is very similar to the coordination rule from the context-free grammar in (23) in Chapter 2 (see page 32).[26] The only difference is notational: Now that we have a more general kind of notation – tags – for representing identity, we can replace the 'X's in the Chapter 2 version of the rule with tags.

The Chapter 3 Grammar also includes a Head-Modifier Rule. This rule corresponds to the two rules that introduced PPs in the Chapter 2 CFG:

(i) NOM → NOM PP

(ii) VP → VP PP

The first thing to notice about these rules is that they allow PPs to modify coordinate structures.[27] That is, the head daughter in the Head-Modifier Rule can be the entire italicized phrases in sentences like (iii) and (iv).

(iii) Alex *walks and reads books* without difficulty.

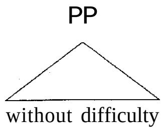
(iv) Terry likes the *poetry and* muszc on this program.

Of course, (iii) and (iv) are ambiguous: The PP can also be modifying just the rightmost conjunct within the coordinate structures.

[26] We will in fact revise this coordination rule in subsequent chapters.

[27] This was also true of the rules in the Chapter 2 grammar.

A. Draw the two trees for (iii) using the Chapter **3** grammar, and indicate which interpretation goes with which tree. [Notes: You may use abbreviations for the feature structures at the nodes. Since we haven't given any sample lexical entries for prepositions, abbreviate the structure under the PP node with a triangle like this:



The node above *and* may be abbreviated as *CONJ.*]

The Chapter 3 grammar, in its present form, doesn't allow PPs to modify Ss or NPs (which are both [SPR +]). Is this prediction correct? Consider the examples in (v) and (vi):

(v) Alex walks without difficulty.
(vi) Terry likes the music on the program.

In these examples, it is hard to tell which constituents the PPs *without difficulty* and *on the program* modify. Whether they attach low (modifying VP and NOM respectively, as currently permitted by the Chapter 3 grammar) or high (modifying S and NP, respectively, not currently permitted by the Chapter 3 grammar), we get the same string of words, and it's difficult to tell what the semantic differences between the two possible attachment sites would be. This question cannot be resolved just by considering simple examples like (v) and (vi).

B. Use coordination to resolve this question. That is, provide an argument USING EXAMPLE WITH COORDINATION to show that the prediction of the Chapter 3 grammar is incorrect: PPs must be able to modify S and NP as well as VP and NOM. [Hint: Your argument should make reference to the different meanings associated with the different tree structures, depending on where the PP attaches.]

Problem 5: Identifying the Head of a Phrase

The head of a phrase is the element inside the phrase whose properties determine the distribution of that phrase, i.e. the environments in which it can occur. We say that nouns head noun phrases, since (ii)-(v) can all show up in the same environments as (i): e.g. as the specifier of a verb, as a complement of a transitive verb and as the complement of prepositions like of or *on*.

(i) giraffes
(ii) tall giraffes
(iii) giraffes with long necks
(iv) all giraffes
(v) all tall giraffes with long necks

On the other hand (vi)–(ix) do not have the same distribution as the phrases in (i)–(v).

(vi) tall

(vii) with long necks
(viii) all
(ix) all tall

Thus it appears to be the noun in (i)–(v) that defines the distributional properties of the whole phrase, and it is the noun that we call the head.

In this problem we apply this criterion for identifying heads to a domain that is off the beaten path of grammatical analysis: English number names.[28] The goal of this problem is to identify the head in expressions like *two hundred* and *three hundred*. That is, which is the head of *two hundred: two* or *hundred?* In order to answer this, we are going to compare the distribution of *two hundred* with that of two minimally different phrases: *three hundred* and *two thousand.*

Now, many environments that allow *two hundred* also allow *three hundred* and *two thousand*:

(x) There were two hundred/three hundred/two thousand.
(xi) Two hundred/three hundred/two thousand penguins waddled by.

Some environments do distinguish between them, however. One such environment is the environment to the right of the word *thousand*

(xii) four thousand two hundred
(xiii) four thousand three hundred
(xiv) *four thousand two thousand

A. Based on the data in (xii)–(xiv), which phrase has the same distribution as *two hundred: three hundred* or *two thousand?*

B. Does your answer to part (A) support treating *two* or *hundred* as the head of *two hundred?* Explain your answer in a sentence or two.

Similarly, we can compare the distribution of *two hundred five* to the two minimally different phrases *two hundred* six and *two thousand five.* Once again, the environment to the right of *thousand* will do:

(xv) four thousand two hundred five
(xvi) four thousand two hundred six
(xvii) *four thousand two thousand five

C. Based on the data in (xv)–(xvii), which phrase has the same distribution as *two hundred five: two hundred six or two thousand five?*

D. Does your answer to part (C) support treating *two hundred* or *five* as the head of *two hundred five?* Briefly explain why.

---

[28]This problem is based on the analysis of English number names in Smith 1999.