Chapter 2

Rules, Constituents, and Fragments

Only connect! That was the whole of her sermon. Only connect the prose and the passion .... Live in fragments no longer.
E. M. Forster, *Howards End*

The artificial languages that we design ourselves, such as logics or programming languages, exhibit a very strong form of the rule-to-rule relation between their semantics and the syntax as it is defined in the textbook or reference manual. This condition in its most general form means simply that there is a functional relation mapping semantic rules and interpretations to syntactic rules and constituents. We will return to the nature of the mapping and its consequences below, but the function of syntax as a vehicle to convey semantics makes the requirement of simplicity in the rule-to-rule mapping seem so reasonable and desirable that it might be expected to transcend all particulars of function and content.

When we finally come to understand the natural system, we must therefore expect to find a similarly direct relation between syntax and semantics, for it is hard to imagine any evolutionary pressure that would force it to be otherwise.

Indeed, there is at least one identifiable force that can be expected to work positively to keep them in line. It arises from the fact that children have to learn languages, apparently on the basis of rather unsystematic presentation of positive instances alone. Since under the simplest assumptions even such trivial classes of grammars as finite-state grammars are not learnable from mere exposure to positive instances of the strings of the language (Gold 1967), and since there appears to be little evidence that any more explicit guidance is provided by adults (Brown and Hanlon 1970 and much subsequent work), some other source of information, "innate" in the weak sense that it is available to children prelinguistically, must guide them. While statistical approximation and information-theoretic analysis using unsupervised machine learning techniques over large volumes of linguistic material remains a theoretically interesting alternative, the most psychologically plausible source for the information children actually use is semantic interpretation or the related conceptual representation.[1]

In the context of modern linguistics, the suggestion that children learn language by hanging language-specific rules and lexical categories on semantic or conceptual representations goes back at least to Chomsky (1965, 29, 59) and Miller (1967). Of course, the idea is much older. See Pinker 1979, 1994 and Fisher et al. 1994 for reviews of some proposed mechanisms, and see Gleitman 1990 for some cogent warnings against the assumption that such semantic representations have their origin solely in the results of direct perception of the material world in any simple sense of that term.

However inadequate our formal (and even informal) grasp on children's prelinguistic conceptualization of their situation, we can be in no doubt that they have one. If so, it is likely that this cognitive representation includes such grammatically relevant prelinguistic notions as actual and potential participants in, properties of, and causal relations among, events; probable attitudes and attentional focus of other conversational participants; and representations of more obviously material aspects of the instantaneously perceivable world.

2.1 Constituents

Three consequences follow from assuming a rule-to-rule relation between syntax and semantics. The first, which follows from the expectation of transparency between syntax and semantics, is so strong and so uncontentious that no theory of grammar has failed to observe it in spirit, though it is probably true to say that none has so far succeeded in following it to the letter. To say that syntax and semantics are related rule-to-rule is to say no more than that every syntactic rule has a semantic interpretation. However, it immediately follows that the syntactic entities that are combined by a syntactic rule must also be semantically interpretable. (Otherwise, they could not be combined by the semantic interpretation of the rule.) It follows that syntactic rules can only combine or yield *constituents*.

This condition, which has been called "The Constituent Condition on Rules," has been a central feature of Generative Grammar from its earliest moments. It frequently surfaces in that literature in the guise of "structure dependency" of grammatical rules. It is also the notion that is embodied in the "proper analysis" condition on transformations proposed in Chomsky 1975a (chapters written in 1955). Perhaps the most illuminating and ambitious early endorsement of this principle is to be found in Chomsky 1975a (210–211, chapters written in 1956), where the following four "criteria" (the scare quotes are Chomsky's) are offered as tests for grammatical constituents and constituent boundaries:

(1) a. The rule for conjunction
    b. Intrusion of parenthetical expressions
    c. Ability to enter into transformations
    d. Certain intonational features

These criteria are very cautiously advanced and carefully surrounded with qualifications, and the subsequent discussion is deliberately designed to demonstrate that some of them raise as many questions as they answer. Nevertheless, there is an implicit claim of great boldness, however programmatically stated. If these operations are tests for constituency, it can only be because they are rules of grammar, subject to the Constituent Condition on Rules. The bulk of Chomsky 1975a, and most work in Generative Grammar since, mainly bears on the claim relative to the third criterion, concerning transformational rules of movement (and their modern equivalents and alternatives), which it has overwhelmingly supported.

It has proved much more difficult to make good on the implicit claim with respect to the remaining three phenomena. Theories of coordination, intonation, and (insofar as there are any) parentheticalization have generally been forced at some point to compromise the Constituent Condition on Rules. The present work should be viewed as an attempt to bring Chomsky's original program nearer completion.

The second consequence of assuming a rule-to-rule relation between syntax and semantics is to imply that the *only* grammatical entities that have interpretations are constituents. This consequence is again entirely uncontentious, and virtually all theories of competence grammar have adhered to it (insofar as they have involved an explicit semantics at all). However, it will be relevant to the discussion of processing in part III.

Finally, the rule-to-rule hypothesis, and its justification in terms of its parsimony with respect to the theory of language learning and evolution, imply that syntactic and semantic rules should have the property of monotonicity. That is, there should be no rules like certain old-style transformations which convert structures that are ill formed (and hence uninterpretable) into structures which are well formed, and vice versa.

To claim that syntax is monotonic is not of course to deny that theories of language need different levels of rules, such as phonology, morphology, syntax, and semantics, or to deny the modular nature of the grammar. However,

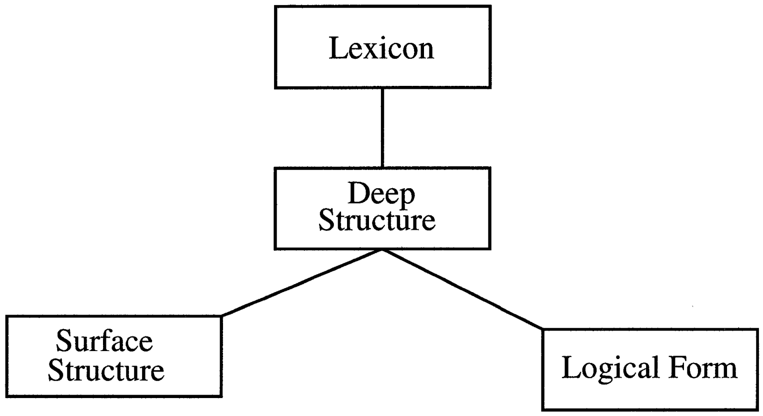


Figure 2.1
Architecture of a standard theory of grammar

It does imply that those levels too should be monotonically related, a point to which I return below.

2.2 Fragments

To what extent do the observed regularities of natural language syntax conform to the expectations set out above? As noted earlier, the generative theoretical tradition has had considerable success in accounting for many constructions involving discontiguities between elements that are semantically dependent upon one another. Many such constructions were originally brought within the fold of the Constituent Condition on Rules by the introduction of transformational rules of "movement" of constituents, relating an underlying level or levels of representation at which predicate-argument relations relevant to semantic interpretation were contiguously related, and from which a Logical Form representing further distinctions such as quantifier scope could be directly computed, to a surface structural level at which the discontiguities were manifest. The introduction of these rules gave rise to the familiar "Y-diagram" architecture of first-generation generative theories, shown in figure 2.1.[2] Theories within the generative tradition differ with respect to how many underlying levels they postulate (one in the case of *Aspects*-style Transformational Grammar "Deep Structures" and Lexical-Functional Grammar (LFG) "f-structure," two in the case of Government-Binding Theory (GB) "S-Structure" and "D-Structure," and even more in some cases). These theories also differ in how they interpret the notion of "movement." Nevertheless, they can all be seen as modifications of this architecture, and the metaphor of movement is so persuasive and revealing that I will use it freely to describe syntactic phenomena, even though the present theory eschews the notion of movement as a syntactic operation.

Movement transformations and the constructions that they capture fall naturally into two groups. The first group includes phenomena that can be accounted for entirely in terms of "bounded" dependencies—roughly, dependencies between items that occur within the same tensed clause, like those between *Dexter* and the verb *win* in the following examples:

(2) a. Dexter expects to win.
 b. Dexter is expected to win by Warren.

As Brame (1976, 1978) and Bresnan (1978) were among the first to point out, the clause-bounded nature of these dependencies means that they can be base-generated or (equivalently) specified in the lexicon, thus bringing them within the domain of the Constituent Condition on Rules without the use of movement as such, and explaining a number of "structure preserving" constraints upon such constructions (Emonds 1976).

This and much subsequent work has shown that the bounded constructions are subject to a number of apparently universal constraints upon such dependencies which reflect dominance and command and an "obliqueness" ordering of arguments of predicates, according to which subjects are ordered with respect to objects and other arguments. An important example for present purposes is afforded by asymmetries in binding possibilities for reflexives and reciprocals like the following, which Keenan (1988) shows to be independent of basic word order across languages:

(3) a. The dogs like each other./*Each other like the dogs.
 b. I showed the dogs to each other/*each other to the dogs.
 c. Sid wants Nancy to like herself/*himself.

I will return to the question of the source of such asymmetries.

The generative approach has also proved extremely successful in accounting for the phenomenon of unbounded dependency exhibited in relative clauses and topicalizations such as the following, again in terms of movement:

(4) a. a book which *I expect I will find*
 b. These articles, *I think that you must have read without understanding.*

In such constructions, elements that are related in the interpretation of the construction, such as the topicalized or relativized NPs and the verb(s) of which they are arguments, can be separated by arbitrarily long substrings and unbounded embedding. Although the residue of topicalization or relativization at first glance looks like a nonconstituent fragment of a sentence, it can be regarded as a constituent of type S, with a special kind of invisible coindexed or "moved" argument, and can thereby be brought under the Constituent Condition on Rules.

Examples like the following suggest that unbounded dependencies are also subject to constraints reflecting conditions of dominance and command involving obliqueness among arguments:

(5) a. *a man whom he thinks that Mary likes
b. *a woman whom I persuaded to like

(The first cannot describe a man who thinks that Mary likes him, and the second cannot describe a woman whom I persuaded to like herself. Again, I will return to this question.)

It has proved much more difficult to account for coordination, parentheticalization, and phrasal intonation within the confines of the Constituent Condition on Rules. It is worth looking at some data in this connection.

2.2.1 Coordination and Parentheticals

At first glance, there is a striking overlap between the kinds of fragments that result from relativization and the related topicalizing construction, and those that can coordinate. In particular, practically everything that can occur as the residue of leftward movement can be coordinated, as in examples like the following:

(6) a. a book which *I expect I will find*, and *I think that you must have read without really understanding*
b. *I expect I will find*, but *I think that you must have read without really understanding*, that novel about the secret life of legumes.

The second, (6b), involves rightward movement (again, the term is used descriptively).

There is a similar (though less complete) conspiracy between the residues of leftward and rightward movement. That is, most residues that arise from leftward movement can also arise from rightward movement. Moreover, both kinds of extraction are subject to very similar "island constraints:"[3]

(7) a. *a book which *I hope that I will meet the woman who wrote*
b. **I hope that I will meet the woman who wrote*, and *you expect to interview the consortium that published*, that novel about the secret life of legumes.

However, the fragments that result from coordination are much more diverse than those that result from (leftward and rightward) movement. For example:

(8) a. I *want to try to write*, and hope to see produced, a musical about the life of Sir Stafford Cripps.
b. Give Deadeye Dick a sugar-stick, and *Mexican Pete, a bun*.
c. I want to try to write a novel, and *you, a screenplay*.

The study of such constructions has revealed that they too are subject to some very strong crosslinguistic constraints, many of which were first discussed by Ross (1970), and which will be discussed at length in chapter 7. These can be summarized as reflecting an "order-preserving" property of coordination, whereby (in configurational languages, at least) if a leftmost constituent is moved, is raised, or otherwise "goes missing" from one conjunct, then it shows up to the left of the entire coordinate structure, whereas a missing rightmost constituent turns up on the right. Thus, in a language like English whose basic word order is Subject-Verb-Object, coordinate sentences like the following are prohibited:[4]

(9) a. *A musical *want to try to write*, and hope to see produced, I.
b. *Deadeye Dick a sugar-stick, and *Mexican Pete, a bun* give.
c. *I a novel, and *you want to try to write a screenplay*

(I will show later in what sense example (9c) is an instance of the same universal.)

Although considerably less attention has been devoted to parenthetical utterances (but see Emonds 1976, II.9, and 1979, McCawley 1982, 1989, Levelt 1989, Espinal 1991, Croft 1995, Taglicht 1998 and Doran 1998), some similarly unconstrained fragments arise from their intrusion, as in (10):

(10) a. *Have you ever been*, they asked, a member of the Friends of the Ukraine Film Society?
b. *You could give*, suggested Dexter, a policeman a flower.

The result has been that, although linguistic theories have had some success in accounting for the relative-clause construction in terms of devices that reinstate the Constituent Condition on Rules by deriving such fragments from traditional constituents such as S via devices like movement (Chomsky 1965), indexed "traces" (Chomsky 1975b), and feature passing (Gazdar 1981), they have been much less successful in showing that the same devices will account for coordination. Instead, coordination has led to the introduction of rules of deletion to supplement rules of movement. Such rules again attempt to reinstate the Constituent Condition over the rule of coordination, by deriving the fragments from underlying constituents of type *S*. However, the deletion rules themselves have generally failed to adhere strictly to that condition. For example, (8b) appears to require either the movement or the deletion of a nonconstituent, and (8c) appears to offer no alternative to the deletion of the nonconstituent *want to try to write*. More worrying still, this fragment looks suspiciously like the kind of fragment that is the surface-structural *result* of deletion or movement, as in (8a).

These complications are surprising, because intuitively all of these constructions appear to be related to the semantic notion of *abstraction*, or definition of a *property*. Most obviously, a restrictive relative clause like (11a) seems to correspond to a predicate or property of *being married by Anna*. Formally, such properties, concepts, or abstractions can be conveniently and transparently represented by terms in the λ-calculus like (11b):

(11) a. (the man that) Anna married
b. $\lambda x.marry'x\ anna'$

Those who are unfamiliar with the λ-calculus are referred to Partee, ter Meulen and Wall 1990, chap. 13, for a full exposition. However, it will be sufficient for present purposes to note that the operator λ declares the symbol x to be a variable local to the expression that follows, distinct from any other variable elsewhere that happens to have the same name. The variable is thus in every way comparable to a parameter or formal variable of a subroutine or function in a computer programming language, which, when instantiated with a value, passes that value to the occurrences of the variable in the expression. The λ-term (11b) can therefore be thought of as defining a function in such a language that maps entities onto truth-values according to whether Anna married them or not. Here as elsewhere in the book, constants like $marry'$, distinguished from variables by primes, are used to identify semantic interpretations whose details are not of immediate interest. Application of a function f to an argument a is simply written fa, and a convention of "left associativity" of function application is observed, so that the above formula is equivalent to the following:

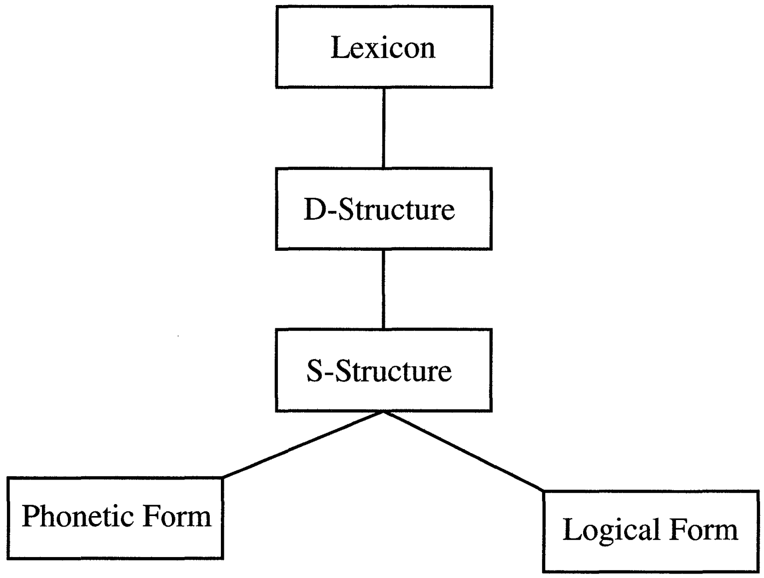


Figure 2.2
Architecture of a government-binding theory of grammar

(12) $\lambda x.(marry' x)anna'$

Most current theories of natural language grammar since "standard" Transformational Grammar (Chomsky 1965), including the present one, more or less explicitly embody the analogy between relativization and abstraction over a variable.

Nevertheless, in the case of coordination and these other unruly phenomena, their apparent freedom from the Constituent Condition on Rules has led to a situation in which, despite the apparently close relation between coordination and relativization, the responsibility for the former phenomenon, together with parentheticalization, and on occasion phenomena like "scrambling" in other languages, has been delegated to a separate, more surface-oriented domain of "stylistic" rules. This led directly to the distinction in GB between the level of S-Structure, at which relativization or *wh*-movement is defined in terms of traces closely resembling syntactic realizations of the bound variables of the λ-calculus, and the level of Phonetic or Phonological Form (PF), at which, contrary to what its name might suggest, significant syntactic work is done. (Chomsky 1986 refers to PF as "Surface Structure.") The result is the theoretical architecture shown in figure 2.2.

2.2.2 Intonation Structure

In a similar apparent contradiction to the Constituent Condition on Rules, some closely related fragments abound in spoken language, arising from phenomena associated with prosody and intonation, as well as less well behaved phenomena like restarts and the parentheticals discussed earlier. For example, one quite normal prosody for an answer to question (13a), involving stress on the word *Anna* and a rise in pitch at the end of the word *married*, imposes an Intonation Structure which is orthogonal to the traditional syntactic structure of the sentence, as informally indicated in (13b) by the parentheses (stress, marked in this case by pitch maxima, is indicated by capitals):

(13) a. I know that Alice married ALAN. But who did ANNA marry?
b. (ANNA married)(MANNY).

We will of course need more empirical evidence and more formal notations to define this phenomenon more precisely in the chapters that follow. But the effect is very strong. (It is ironic that one of the difficulties in teaching introductory syntax is to persuade students that this is *not* the notion of structure that is relevant to the study of syntax. One conclusion that can be drawn from the argument in this book is that the students are, in an important sense, quite right.)

As is often the case with informally controlled contexts like this, other intonation contours are possible. In particular, because the context (13a) is compatible with an interpretation under which *Anna* is the topic or theme of the utterance, a contour with an intonational boundary separating the subject and predicate, *(ANNA)(married MANNY)*, is also possible. (For further discussion, see Jackendoff 1972; Steedman 1991a; and chapter 5 for further discussion.)

Intonation Structure nevertheless remains strongly constrained by meaning. For example, contours imposing bracketings like the following do not seem to be allowed, as Selkirk (1984) has pointed out:

(14) #(Three CATS)(in ten prefer CORDUROY).

Halliday (1967a) observes that this constraint, which Selkirk calls the "Sense Unit Condition," seems to follow from the *function* of phrasal intonation, which in English is in part to convey what Halliday called "Information Structure"—that is, distinctions of focus, presupposition, and propositional attitude toward entities in the discourse model. These discourse entities are more diverse than mere NP or propositional referents, but they do not seem to include such nonconcepts as "in ten prefer corduroy."

The question in (13), *Who did Anna marry?*, appears to introduce a new "theme" or topic of conversation, corresponding like the relative clause in (11a) to the concept of *someone such that Anna married them.* As Jackendoff (1972) points out, it is once again natural to think of this theme as a functional *abstraction* and to express it using exactly the same expression of the λ-calculus as was used in (11b) for the relative clause, repeated here:[5]

(15) $\lambda x.marry'x\ anna'$

When this function or concept is supplied with an argument *manny'*, it *reduces* to give a proposition, with the same function-argument relations as the canonical sentence (Again, function application associates to the left.)

(16) $marry'manny'anna'$

It is the presence of the abstraction (15) rather than some other that makes the intonation contour in (13b) felicitous. (That is not to say that its presence uniquely *determines* this response, or that its explicit mention is necessary for interpreting the response.)

These observations have led linguists such as Selkirk to postulate a level of Intonation Structure, independent of syntactic structure and related to an Information Structural component of LF, implying an architecture something like the one in figure 2.3 for the theory of grammar as a whole (see Selkirk 1984, 205, and cf. Chomsky 1971).[6]

The involvement of two apparently uncoupled levels of underlying structure on the way from sound to meaning in natural language grammar appears to complicate the path from speech to interpretation unreasonably, and to thereby threaten the entire theory of grammar (not to mention its worrying implications for the feasibility of a number of applications in computational speech recognition and speech synthesis).

In the light of the increasing complexity of the mainstream theories of grammar in the face of these less well behaved constructions, it is interesting to observe that the coordinate constructions considered in section 2.2.1, whose semantics also seems to be reminiscent of functional abstraction, are also subject to something like the Sense Unit Condition that limits intonational phrases. For example, strings like *in ten prefer corduroy* seem to be as reluctant to take part in coordination as they are to be treated as intonational phrases:

(17) *Three cats in twenty like velvet, and in ten prefer corduroy.

Parentheticalization is similarly bad at such junctures:

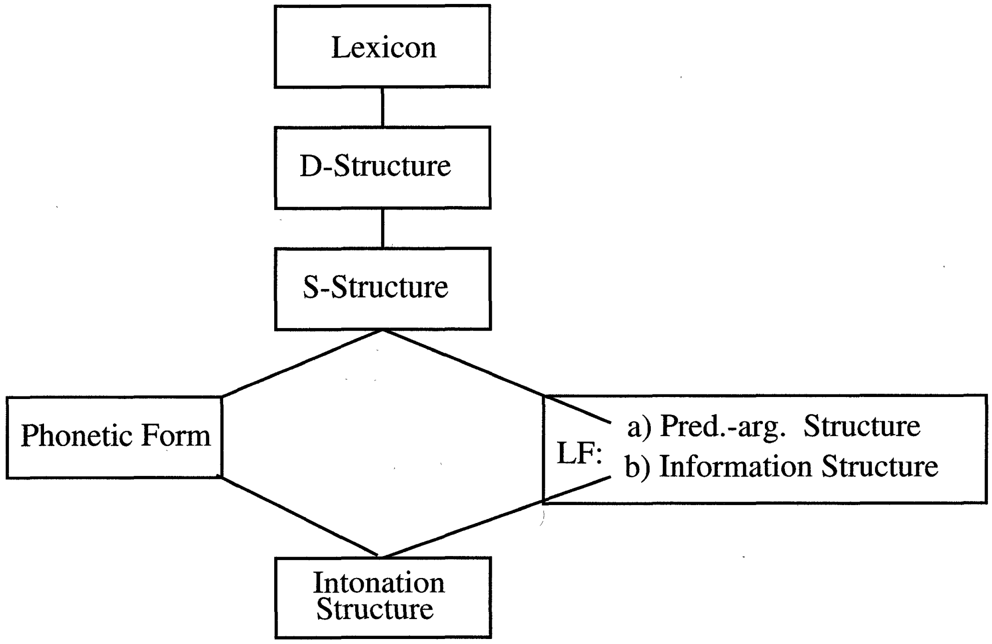


Figure 2.3
Architecture of a government-binding prosody

(18) *"Three cats," ejaculated Tom, prematurely, "in ten prefer corduroy."

Since coordinate constructions and parentheticals constitute major sources of complexity for current theories of natural language grammar and also offer serious obstacles to computational applications, it is tempting to suspect that this conspiracy between syntax and prosody might point to a unified notion of syntactic constituent structure that is somewhat different from the traditional one, but that unifies the GB notions of S-Structure, PF and Intonation Structure under a single notion that I shall call Information Structure, to avoid confusion with other notions of Surface Structure that are in play.

2.3 Issues of Power and Explanation

To cut ourselves adrift from traditional linguistic notions of constituency is a frightening prospect. Fortunately, there are other principles of linguistic investigation to guide the search for an alternative.

The first among these principles is Occam's razor, which says that we should capture the phenomena of language in theories with as few degrees of freedom as possible. The strength of a theory lies in the extent to which (a) it captures all and only the phenomena that can occur, and (b) could not in principle capture phenomena that we have good reason to believe could not occur, and therefore has no need to explicitly exclude them by stipulation.

It follows that a theory should involve as few levels of representation as possible, consistent with the main goal of capturing generalizations. Ideally, following the Montagovian tradition, we would like to assume no more than an interpretation related in a strictly monotonic fashion to syntactic derivation.

We should also try to minimize power in the modules of the theory, consistent with the primary goal of building interpretations that capture predicate-argument relations correctly. If we can do so with grammars that are provably low on the Chomsky hierarchy of formal languages and automata-theoretic power, then we are on stronger theoretical ground than if we adopted theories that achieve the same coverage at the expense of greater power, because greater automata-theoretic power increases the variety of alternative constructions and phenomena that we *could* capture beyond the point where there seems to be any motivation from empirical or imaginable semantic dependencies. Here I follow the tradition of Generalized Phrase Structure Grammar (GPSG; Gazdar 1982) and mildly context-sensitive theories of grammar such as Tree-Adjoining Grammar (TAG; Joshi, Levy and Takahashi 1975).

Of course, to achieve parsimony in automata-theoretic power is possible only to the extent that we have good information about the real space of possible natural languages and grammars. Fortunately there is a certain amount of information available on this question.

In asking what is the least powerful class of grammars on some scale such as the Chomsky hierarchy that includes all natural grammars, we must distinguish between "strong" and "weak" adequacy of grammars or sets of rules for capturing languages in the formal sense of sets of strings of words or other terminal vocabulary symbols. A grammar or set of rules that merely generatively specifies all and only the strings of a language may be only weakly adequate as a grammar for the language. To be strongly adequate, it must also assign a correct structural description to the string, where "correct" structural descriptions are the ones that support the semantics of the language. For example, a language whose semantics demands a context-free grammar whose rules happen to permit embedding only on rightmost elements has a weakly adequate finite-state grammar that generates all and only the same strings. However, the weakly adequate finite-state grammar does not directly express the embedding that supports the semantics. It is only the strongly adequate context-free grammar that does that.

However strong our intuitions may be concerning some aspects of meaning in natural language, we do not have access to natural semantics in the direct sense in which we have access to its strings. It follows that the only formal proofs that we can construct concerning lower bounds on the power implicit in natural languages involve weak adequacy. Of course, if we can show that some level of the automata-theoretic power hierarchy is not even weakly adequate to capture all natural languages, then it immediately follows that the level in question is not strongly adequate either. However, finding a lower bound on the power of strongly adequate grammars via a proof of weak adequacy depends on finding a construction in the language that not only has an intuitive semantics that demands that power to produce correct structural descriptions, but that also happens not to admit a weakly adequate grammar of lower power.

Partly because of the widespread presence of lexical ambiguity in natural languages, the search for a formal proof of a lower bound on weak adequacy was extremely protracted. Chomsky (1957) gave an early argument that nothing lower than context-free grammars—that is, grammars that can be written with production rules expanding a single nonterminal or phrasal type—could be weakly adequate. However, many of the earliest examples of constructions that were claimed to prove that the lower bound on power was strictly greater than context-free were flawed, either because they confounded extragrammatical factors with grammatical ones (like the argument from the English *respectively* construction) or because they depended on intuitive assumptions about strong adequacy and failed to exclude the possibility of a weakly equivalent grammar of lower power. These arguments are helpfully reviewed by Pullum and Gazdar (1982) and by Pullum (1987).

This curious delay in proving any lower bound on power greater than context-free should be kept in perspective. Since the late 1970s, there has been very little doubt that (a) the competence grammars implicated by the semantics or predicate-argument relations for natural languages were strictly greater than context-free in power, and (b) that power was not very much greater than context-free.

One reason for believing natural grammars not to greatly exceed context-free power, and in particular not to come anywhere near the power of context-sensitive grammars, is that most phenomena of natural languages, including those involving unbounded dependencies, can be captured by context-free rules. Although I will not at this point go into the way in which certain kinds of unbounded dependencies can be captured using context-free or near-context-free rules, the possibility is strongly suggested by the observation that the two dependencies in (19a) must nest, rather than intercalate, as they would have to for (19b) to have a meaning to do with playing sonatas on violins (the asterisk here means "not allowed with the intended reading.")

(19) a. a violin which$_i$ [this sonata]$_j$ is hard to play$_j$ upon$_i$

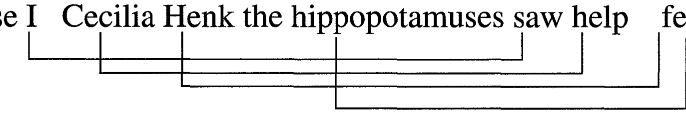
b. *a sonata which$_i$ [this violin]$_j$ is hard to play$_i$ upon$_j$

To nest dependencies is of course the characteristic behavior of a pushdown automaton, which is the automaton for recognizing context-free languages, so it seems likely that there is a strongly adequate context-free grammar for this construction. GPSG (Gazdar 1981, 1982; Gazdar et al. 1985) extensively explored the extent to which all grammatical phenomena can be captured in context-free terms.

The reason for believing that natural grammars are of strictly greater than context-free power lies in the fact that, although nonnesting or intercalating dependencies are rare, strong adequacy will undoubtedly require them to be captured in some, and perhaps all, natural languages. The most convincing observation of this kind came from Dutch, in work by Huybregts (1976), although it was some time before a formal proof in the form of a proof of weak inadequacy of context-free grammars was forthcoming on the basis of a similar phenomenon in related Germanic dialects (Huybregts 1984; Shieber 1985).

In Dutch there is a strong tendency for all verbs in subordinate clauses to be clause-final and for matrix verbs to appear to the left of the verbs in their complements. This means that certain sentences with embedded infinitival complements that in English embed right-peripherally involve crossing or intercalating dependencies between predicates and arguments, as indicated by connecting lines in (20):[7]

(20) ... omdat ik Cecilia Henk de nijlpaarden zag helpen voeren.
... because I Cecilia Henk the hippopotamuses saw help feed



'... because I saw Cecilia help Henk feed the hippopotamuses.'

The important property of this construction is that the semantic dependencies between nouns and verbs do not nest in Dutch, as they do in English. They intercalate, or interleave. This means that any strongly adequate grammar for Dutch—that is, one that maintains a rule-to-rule relation between syntax and semantics and captures the dependencies correctly—is likely to be non-context-free (Wall 1972).

The challenge that is offered by the contrast between nesting examples like

(19) and intercalating examples like (20) is to find a strongly adequate class of grammars that is "mildly" context-sensitive, allowing these examples and the previous ones that are characteristic of coordination and prosody, without allowing all sorts of illicit phenomena.

There is a third source of information that we can draw upon in our search for such a class. There are a number of known crosslinguistic constraints on grammar that are so strong as to apparently constitute universal limitations on natural grammars. We will be concerned with many such universal constraints below, but two in particular will play a central role in the argument.

The first is due to Ross (1970), who pointed out that the construction that shows up in English as (medial) gapping, in sentences like (21), shows a strong crosslinguistic regularity concerning which of the two conjuncts undergoes "deletion" or otherwise has the verb go missing:

(21) Dexter likes cats, and Warren, dogs.

The pattern is that in languages whose basic clause constituent order subject-verb-object (SVO), the verb or verb group that goes missing is the one in the right conjunct, and not the one in the left conjunct. The same asymmetry holds for VSO languages like Irish. However, SOV languages like Japanese show the opposite asymmetry: the missing verb is in the *left* conjunct.[8] The pattern can be summarized as follows for the three dominant constituent orders (asterisks indicate the excluded cases):[9]

(22) SVO: *SO and SVO SVO and SO
VSO: *SO and VSO VSO and SO
SOV: SO and SOV *SOV and SO

This observation can be generalized to individual constructions within a language: just about any construction in which an element apparently goes missing preserves canonical word order in an analogous fashion. For example, English ditransitive verbs subcategorize for two complements on their right, like VSO verbs. In the following "argument cluster" coordination, it is indeed in the right conjunct that the verb goes missing:

(23) Give a policeman a flower, and a dog a bone.

The second phenomenon identified above, the crosslinguistic dependency of binding of reflexives and anaphors upon Jackendoff's Jackendoff (1972) obliqueness hierarchy is discussed by Keenan (1988) and Clark (1985, 1991), among others. Regardless of basic word order (here there are data from OS languages and constructions within languages), or indeed of configurationality itself, anaphoric pronouns like *themselves* and *each other* may corefer with another argument of the verb just in case that other argument is less oblique—that is, earlier in a series that places subject first, object next, and more oblique arguments later. In English this shows up in the fact that a subject may bind an object, but not vice versa:

(24) a. Dexter and Warren like each other.
b. *Each other like Dexter and Warren.

In the case of the ditransitive verbs, it shows up in the fact that the first object can bind the second, but not vice versa (see Barss and Lasnik 1986 for discussion):

(25) a. I introduced Dexter and Warren to each other.
b. *I introduced each other to Dexter and Warren.

This is not an idiosyncratic fact about English or SVO languages. Keenan shows that in VSO and even VOS and mixed VSO/VOS languages, less oblique arguments such as subjects bind more oblique arguments such as objects, and not vice versa.

2.4 Grammar as an Applicative System

The two universals singled out in the last section, both of which can conveniently be illustrated using the same English ditransitive construction, induce opposite tensions in the theory of grammar. For reasons that will be developed below, the dependency of gapping upon canonical word order suggests that directionality under concatenation is directly projected from the lexicon to the string by strictly concatenative rules. On the other hand, binding suggests the existence of a level of representation at which obliqueness is represented independently of surface order, or that projection is not strictly concatenative, or both.

It is interesting in this connection to note that there are alternative systems to the λ-calculus for capturing the notion of abstraction, and that these systems entirely avoid the use of bound variables. They are the combinatory systems invented by Schönfinkel (1924) and Curry and Feys (1958) as a formal foundation for the semantics of the λ-calculus. They are entertainingly discussed in Smullyan 1985, where the combinatory operators take the form of birds, and from which a number of the epigraphs to the present chapters are taken. In such systems, which I discuss in detail in chapter 8, terms equivalent to abstractions are built up using a few simple operations for combining functions, such as functional composition. Systems using quite small numbers of combinators can be shown to be equivalent in expressive power to the λ-calculi. The existence of these systems raises the possibility that alternative theories of grammar can be developed based as directly upon the combinatory applicative systems as the traditional ones implicitly are upon the λ-calculus. The significance of this possibility is that the different form that syntactic rules take in combinatory systems may lead us to look at the kinds of phenomena discussed above in a new way.

Because combinators are operations that map *functions* onto other functions, and because the categorial grammars that were originally developed in their "pure" context-free form by Ajdukiewicz (1935) and Bar-Hillel (1953) provide a notation in which functional type is made explicit, this insight has led in recent years to an explosion of research in frameworks that are collectively known as "flexible" categorial grammars (see, e.g., Lambek 1958; Montague 1970; Geach 1972; Cresswell 1973; Karlgren 1974; Bach 1976, 1979, 1980; Shaumyan 1977; Keenan and Faltz 1978; von von Stechow 1979; Levin 1982; Ades and Steedman 1982; Dowty 1982; Hausser 1984; van Benthem 1986; Flynn 1983; Huck 1985; Zwarts 1986; Uszkoreit 1986; Wittenburg 1986; Desclés, Guentchéva and Shaumyan 1986; Oehrle 1987, 1988; Zeevat, Klein and Calder 1987; Bouma 1987; Szabolcsi 1989; Moortgat 1988a; Hoeksema 1989; Carpenter 1989; Hepple 1990; Jacobson 1990; Segond 1990; Karttunen 1989; Hepple 1990; Jowsey 1989; Steele 1990; Reape 1996; Wood 1993; van der Linden 1993; Potts 1994; Houtman 1994; Ranta 1994; Morrill 1994; Hendriks 1995; and Aarts 1995.[10] One of the interesting properties of combinatory applicative systems is that in general they offer many equivalent combinatory expressions for a given normalized λ-term. This property is reflected in another distinguishing feature of certain of the above theories. The use of rules related to combinators encourages a property of "associativity" in linguistic derivations. That is, for any unambiguous sentence, there are typically several distinct categorial derivations, all of which are semantically equivalent in the sense that they deliver the same function-argument relations. The notion "constituent of a derivation" is correspondingly generalized to cover many of the puzzling fragments discussed above.

I will not attempt to review this very diverse literature here.[11] However, many of the ideas explored below draw upon these works, and I will try to make their parentage clear as we go. This book is most closely related to the subgroup of these theories developed by Ades, Oehrle, Jacobson, Szabolcsi, and Hepple, among other cited above, although they should not be assumed to endorse the present theory in all (or even many) respects. In these theories, syntactic rules corresponding to simple individual combinators such as functional composition are used to lend such fragments as *want to try to write* and even *a policeman a flower* the status of constituents, without the use of movement or deletion. Such grammars will be shown to provide a unified treatment of a wide variety of syntactic phenomena in natural language and to explain phenomena of long-distance dependency (including relativization), coordination, parentheticalization, and intonation, within the confines of the Constituent Condition on Rules and in terms of a single principle. That principle is that the predicate-argument relations that hold in sentences of natural languages are projected onto long-range syntactic dependencies from the relations defined locally in the lexicon by syntactic operations corresponding to combinators, rather than by syntactic operations involving empty categories or traces corresponding to syntactically realized bound variables.

In order to demonstrate that these novel grammars deliver the correct interpretations, we will need a semantic notation. Although we could use combinators for the purpose, λ-calculus is far more readable and in every way equivalent. The appearance of variables in the semantic notation might give the impression that traces have been reintroduced. However, these variables are no more than a readable notation for a uniform mechanism whereby *all* arguments, whether "extracted" or "in situ," get semantically bound to predicates.

BLANK PAGE

Chapter 3

Intuitive Basis of Combinatory Categorial Grammars

Given any birds A, B, and C, the bird C is said to compose *A with B if for every bird x the following condition holds:* $Cx = A(Bx)$
Raymond Smullyan, *To Mock a Mockingbird*

Because combinators are operations on functions, we will need a notation for grammars that makes prominent the functional type or "category" of linguistic entities—that is, a notation that specifies the kinds of things that a linguistic entity combines with and the kind of thing that results. Categorial Grammar (CG), invented by Ajdukiewicz, Bar-Hillel, and others, which in its pure form is (weakly) equivalent to other context-free grammars, provides such a notation, and it is there that we will begin.[1]

3.1 Pure Categorial Grammar

Categorial grammars put into the lexicon most of the information that is standardly captured in context-free phrase structure (PS) rules. For example, consider the following PS rules, which capture some basic syntactic facts concerning English transitive sentences:

(1)
$$\begin{array}{lll} S & \rightarrow & NP\ \ VP \\ VP & \rightarrow & TV\ \ NP \\ TV & \rightarrow & \{\textit{married}, \textit{finds}, \ldots\} \end{array}$$

In a categorial grammar, all constituents—and in particular the lexical elements, such as verbs and nouns—are associated with a syntactic "category" that identifies them as either *functions* or *arguments*. In the case of functions, the category specifies the type and directionality of their argument(s) and the type of their result. The present work uses a notation in which the argument or domain category always appears to the right of the slash, and the result or range category to the left. A forward slash / means that an argument of the appropriate type must appear to the right of the functor; a backward slash \ means that the argument must appear to the left.[2] All functions are "Curried,"