Introduction

In Information-Based Syntax and Semantics, Volume 1 (Pollard and Sag 1987; henceforth P&S-87), we sketched the fundamentals of head-driven phrase structure grammar (HPSG), an integrated theory of natural language syntax and semantics. There our focus was on foundational issues, such as the nature of features, categories, lexical entries, rules, and principles; the relationship between syntax and semantics; and technical aspects of formalization. We were at pains to emphasize the extent to which HPSG was intellectually indebted to a wide range of recent research traditions in syntax (principally nonderivational approaches such as categorial grammar (CG), generalized phrase structure grammar (GPSG), arc pair grammar (APG), and lexical-functional grammar (LFG)), semantics (especially situation semantics), and computer science (data type theory, knowledge representation, unification-based formalisms).

But the range of linguistic facts that we considered was quite narrow, being limited for the most part to matters such as subcategorization, the distinction between complements and adjuncts, constituent ordering, and inflection. In this volume, our goal is to demonstrate the applicability of a theory like the one described in P&S-87 to a wide range of empirical problems. The phenomena with which we will be concerned are among those that have occupied center stage within syntactic theory for well over thirty years: the control of 'understood' subjects, long-distance dependencies conventionally treated in terms of wh-movement, and syntactic constraints on the relationship between various kinds of pronouns and their antecedents.

Within that time period, detailed accounts of these phenomena—and of the relationships among them—have been developed within the research framework established by Noam Chomsky and known in its successive stages as the 'standard' theory, the 'extended standard' theory, the 'revised extended standard' theory, and 'government-binding' theory (GB, or the 'principles-and-parameters' approach). But given the widespread acceptance of that framework as a standard in recent years, especially among an extensive community of syntacticians in the United States and much of continental western Europe, it is incumbent on the proponents of a competing framework to explicate the sense in and extent to which the proposed alternative addresses the concerns of
that community. For that reason, we will be relatively less concerned in this volume with paying our intellectual debts and correspondingly more concerned with making our proposals intelligible to workers within Chomsky's paradigm. More than that, we will try to make clear in what respects our accounts resemble those provided within GB theory and—more importantly—in what respects they differ.

A number of similarities between GB theory and the theory advocated here will be apparent. For example, in both theories, structure is determined chiefly by the interaction between highly articulated lexical entries and parametrized universal principles of grammatical well-formedness, with rules reduced to a handful of highly general and universally available phrase structure (or immediate dominance) schemata. In both theories, phonetically empty constituents (GB's variables, HPSG's traces) are central to the account of unbounded dependency (wh-movement) phenomena. A number of key GB principles (such as Principles A, B, and C of the binding theory, subjacency, and the empty category principle) have more or less direct analogs in HPSG; and two other HPSG principles (the head feature principle and the subcategorization principle) play a role in the theory roughly comparable to that of the projection principle in GB. Moreover, in both GB and HPSG, there are assumed to be several distinct 'levels' (or, as we will call them, attributes or features) of linguistic structure.

At the same time, however, there are a great many differences between the two theories, with respect to both global theory architecture and matters of technical detail. One key architectural difference is the absence from HPSG of any notion of transformation. Unlike GB levels (at least as they are most commonly explicated1), the attributes of linguistic structure in HPSG are related not by movement but rather by structure sharing, that is, token identity between substructures of a given structure in accordance with lexical specifications or grammatical principles (or complex interactions between the two).2 In common, then, with a number of linguistic theories (including those commonly referred to as 'unification-based'), HPSG is nonderivational, in contradistinction to nearly all variants of GB and its forebears, wherein distinct levels of syntactic structure are sequentially derived by means of transformational op-


2. The notion of structure sharing has a somewhat obscure origin in modern linguistics. As noted by Johnson and Postal (1980: 479–483), it has played a central role (under various names, e.g., 'loops,' 'vines,' 'multiattachment,' and 'overlapping arcs') in various theoretical frameworks and a body of research in relational grammar and arc pair grammar (see especially the formulation in Johnson and Postal (1980) and the references cited therein).

3. Exceptional in this respect is the nonderivational GB model developed by Koster (1987).

4. This inventory of attributes differs from the one assumed in the version of HPSG presented in P&S-87. This and other differences between the two versions will be discussed in full in Chapter 1.

5. For expository simplicity, we are ignoring the distinction within SYNSEM values between LOCAL (LOC) and NONLOCAL (NONLOC) features, which will be discussed in Chapter 1.

6. The CONTEXT attribute contains linguistic information that bears on certain context-dependent aspects of semantic interpretation. As we will explain below, the CONTEXT attribute supersedes the INDICES attribute of P&S-87.
least oblique complement of the relevant head, where relative obliqueness is modelled by position on the list that forms the SUBCATEGORIZATION (SUBCAT) value of that head. To give another example, in HPSG, Principle A (which constrains the possible antecedents of anaphors) makes no reference to co-command or government, merely requiring that an anaphor be coindexed with some less oblique argument (provided such exists). We will try to show that such nonconfigurational formulations are not only conceivable alternatives, perhaps to be preferred on grounds of simplicity and conceptual clarity, but are also superior with respect to conformity with the facts.

As mentioned above, although HPSG does not employ movement, the account that we propose for phenomena traditionally treated under the rubric of wh-movement does resemble the GB account inasmuch as phonetically null constituents—traces—are assumed to occupy the ‘gap’ position; however, we will argue that the relationship between the gap and its ‘filler’ is more clearly understood as a matter of structure sharing than as one of movement.

To put it another way, we deny that transformations themselves model anything in the empirical domain (and therefore HPSG shares the property of ‘non-derivationality’ with CG, GPSG, APG, and LFG, in contradistinction to GB and its derivational kin). Similarly, raising will be treated in terms of structure sharing between a matrix argument and the complement’s SUBCAT specification corresponding to the complement subject. In this case, however, there is no need to posit an actual constituent (e.g. NP-trace) corresponding to that specification, and hence the complement will simply be a VP, not an S. Thus HPSG has no analog of GB’s ‘extended’ projection principle, which appears to us to have been introduced by Chomsky (1982) essentially without argument: lexical requirements (as expressed in SUBCAT lists) do not always have to be satisfied on the surface (i.e. in the DAUGHTERS attribute).

Another GB assumption explicitly denied in HPSG is the principle, proposed by Chomsky (1981), that every (nonsubject) subcategorized element must be assigned a semantic role. Thus there is no obstacle to a ‘raising-to-object’ analysis of sentences like *Kim believes Sandy to be happy*. In HPSG, this amounts to structure sharing between the matrix object and the subject specification on the complement’s SUBCAT list. Thus raising to subject and raising to object are handled in entirely parallel fashion: by sharing of structure between the complement subject and the matrix controller at the ‘level’ of subcategorization.

As we have seen, the closest HPSG analog of movement is structure sharing either with a phonetically null constituent (unbounded dependencies) or with a SUBCAT element that is not realized as a constituent at all (raising). But not all instances of movement in GB correspond to structure sharing in HPSG; passive, for example, as mentioned above, is not treated in the syntax at all but rather by lexical rule. Another case in which movement in GB has a ‘non-movement’ (i.e. non-structure-sharing) account in HPSG is that of ‘head movement,’ as manifested, for example, in VSO word order or in English ‘subject-auxiliary inversion.’ On our account, such structures simply arise from the existence of a phrase-structure schema, utilized (like all schemata) to different extents by different languages, that permits the realization of all complements (including the subject) as sisters of the lexical head (P&S-87, sec. 6.2); the orderings are the consequence of independently motivated language-specific constituent ordering principles (P&S-87, sec. 7.2). But in Chapter 9, we will consider the possibility that inverted structures are licensed by the same schema as typical head-complement structures (i.e. that inverted ‘subjects’ are actually nonsubject complements).

The other core case of head movement in GB, namely, movement of the head of VP into INFL, does not require any treatment at all in HPSG, for HPSG does not posit an independent category INFL to serve as a repository of tense and subject agreement features. Instead, subject agreement features (like object agreement features, in languages that have object agreement) occur within the corresponding SUBCAT element of the verb; and the role of the tense element of INFL is taken over by the head feature VERB-INFLATIONAL-FORM (VFORM). Thus whether the verb is tensed is simply a question of whether the VFORM value is finite (fin) or some other (nonfinite) value; and the independent question of whether the verb is an auxiliary (and therefore can license VP deletion, contracted negation, etc.) is treated in terms of another (binary) head feature AUXILIARY (AUX).

7. But we will propose an alternative, traceless analysis in Chapter 9.

8. The proposal to treat extraction phenomena in terms of structure sharing (or ‘overlapping arcs,’ in their terms) was first made, we believe, by Johnson and Postal (1980). Our proposals for the analysis of extraction, coreference, and a variety of other linguistic phenomena, though differing in many points of detail from those of Johnson and Postal, nonetheless share the important feature of being based on structure sharing rather than derivational processes.

9. If one adopts the distinction between level and stratum proposed by Ladusaw (1988), then HPSG would be said to be a multilevel theory, in that it posits more than one kind of representation, but monosratial (like GPSG and CG), in that there is only one representation of each such kind.

10. Moreover, since passive is handled by lexical rule rather than within the syntax, the necessity for an analog of NP-trace is obviated altogether. See P&S-87, Chapter 8.

11. Postal and Pullum (1988) argue persuasively that this assumption, though conventional, is justified by neither empirical nor GB-internal theoretical considerations. We return to these issues in Chapter 3.
Indeed, from the point of view of HPSG, Chomsky's rule move-α must be seen as a kind of procrustean bed. On our account, the phenomena that have been relegated to it are a heterogeneous assemblage, each of which deserves a more comfortable resting place of its own, be it in the lexicon (passive and verb inflection), in the phrase structure schemata (verb-object nonadjacency), or in structure sharings that accord with different kinds of interactions between lexical specifications and universal principles (raising and unbounded dependencies).  

The Nature of Linguistic Theory  

Let us begin by making explicit some methodological assumptions. In any mathematical theory about an empirical domain, the phenomena of interest are modelled by mathematical structures, certain aspects of which are conventionally understood as corresponding to observables of the domain. The theory itself does not talk directly about the empirical phenomena; instead, it talks about, or is interpreted by, the modelling structures. Thus the predictive power of the theory arises from the conventional correspondence between the model and the empirical domain.  

An informal theory is one that talks about the model in natural language, say a technical dialect of English, German, or Japanese. But as theories become more complicated and their empirical consequences less straightforwardly apparent, the need for formalization arises. In cases of extreme formalization, of course, the empirical hypotheses are cast as a set of axioms in a logical language, where the modelling structures serve as the intended interpretations of expressions in the logic.  

For example, in one kind of standard model of celestial mechanics, the positions and velocities of bodies subject to mutual gravitation are represented by vectors in a higher-dimensional Euclidean space ("phase space"), the masses of the bodies by positive real numbers, and their motions by paths along certain smooth vector fields ("flows") on the space. Of course such a model is not the same thing as what it models (e.g. the solar system), but certain formal properties of such a model may represent aspects of the solar system of interest to a physicist. In a formal theory based on such a model, the underlying logic is just a standard first-order language (e.g. the language of Zermelo-Fraenkel set theory), and the axioms are certain systems of differential equations (e.g. Hamiltonian systems) that the flows are required to satisfy. An observed motion of the solar system is then predicted by the theory insofar as it agrees—under the conventional correspondence—with an admissible flow (i.e. one that satisfies the equations). This state of affairs is summarized in (1):  

1. Phenomena (possible motions of n-body systems)  
2. Model (Hamiltonian vector fields)  
3. Formal theory (1st-order predicate logic)  
4. Model-theoretic interpretation (Zermelo-Fraenkel set theory)  
5. Certain differential equations  

In our view, a linguistic theory should bear exactly the same relation to the empirical domain of natural language, namely, the universe of possible linguistic objects, as a mathematical theory of celestial mechanics bears to the possible motions of n-body systems. Thus we insist on being explicit as to what sorts of constructs are assumed (i.e. what ontological categories of linguistic objects we suppose to populate the empirical domain) and on being mathematically rigorous as to what structures are used to model them. Moreover, we require that the theory itself act count as a theory in the technical sense of precisely characterizing those modelling structures that are regarded as admissible or well-formed (i.e. corresponding to those imaginable linguistic objects that are actually predicted to be possible ones). This does not mean that the empirical hypotheses must be rendered in a formal logic as long as their content can be made clear and unambiguous in natural language (the same holds true in mathematical physics), but in principle they must be capable of being so rendered. Unless these criteria are satisfied, an enterprise purporting to be a theory cannot have any determinate empirical consequences. Thus we emphatically reject the currently widespread view which holds that linguistic theory need not be formalized. Rather, our position is the same as the one advocated by Chomsky (1957: 5):  

Precisely constructed models for linguistic structure can play an important role, both negative and positive, in the process of discovery itself. By pushing
a precise but inadequate formulation to an unacceptable conclusion, we can often expose the exact source of this inadequacy and, consequently, gain a deeper understanding of the linguistic data. More positively, a formalized theory may automatically provide solutions for many problems other than those for which it was explicitly designed. Obscure and intuition-bound notions can neither lead to absurd conclusions nor provide new and correct ones, and hence they fail to be useful in two important respects. I think that some of those linguists who have questioned the value of precise and technical development of linguistic theory have failed to recognize the productive potential in the method of rigorously stating a proposed theory and applying it strictly to linguistic material with no attempt to avoid unacceptable conclusions by ad hoc adjustments or loose formulation.

In HPSG, the modelling domain—the analog of the physicist’s flows—is a system of sorted feature structures (Moshier 1988; Pollard and Moshier 1990), that are intended to stand in a one-to-one relation with types of natural language expressions and their subparts. The role of the linguistic theory is to give a precise specification of which feature structures are to be considered admissible; the types of linguistic entities that correspond to the admissible feature structures constitute the predictions of the theory.

Just as in other empirical domains, linguistic theory has become sufficiently modular, complex, and deductive that a need for formalization has become apparent, especially to researchers concerned with the computational implementation of current theories. Thus in the past few years, a number of specialized “feature logics” have been proposed (Kasper and Rounds 1986; Johnson 1988, 1991; Gazdar et al. 1988) for specifying constraints on the feature structures used in linguistic analysis. At the same time, theoretical computer scientists have proposed various constraint languages for programming language description (e.g. Moshier 1988) and knowledge representation (e.g. Höhfeld and Smolka 1988; Ait-Kaci and Nasr 1986) that are easily adapted to this end. A very recent integration of these lines of work is the development of feature logics appropriate for the formalization of linguistic theories, languages whose formulas serve as the linguist’s analog of the space physicist’s differential equations (see Carpenter 1990; Carpenter et al. 1991; Carpenter and Pollard 1991; King 1989; Pollard, n.d.; Pollard and Carpenter 1990; and Carpenter 1992).

The last-mentioned work in particular sets forth a logic very close to the one that we will assume will underlie a fully formalized version of our theory. In very general terms, this can be characterized as a sorted variant of Kasper and Rounds’s (1986) logic augmented with path inequalities, definite relations, and set values.

Our research program thus assumes the following three-way relation connecting theory, model, and the empirical domain of language:

Since our principal goal in this volume is to propose analyses of linguistic phenomena and make them intelligible to the linguistic community, we eschew extreme formalization here and thereby avoid the many technical design decisions associated with the problem of choosing a feature logic; our rules and principles, in the form of feature structure constraints, will be expressed—clearly and unambiguously, we hope—variously in English or in a quasi-formal description language that we turn to directly. We doubt that the relative informality of analyses such as those presented here is likely to impede understanding; if anything, probably the reverse is true. Nevertheless we consider formalization of the theory an important goal, indeed a necessary one for the proof of computational properties (see below).

A further methodological principle, shared by the scientific community at large, is that of ontological parsimony: insofar as it is possible without doing violence to the simplicity and elegance of the theory, we do not posit constructs that do not correspond to observables of the empirical domain. Of course, all scientific theories contain such constructs. An obsolete example is the phlogiston that used to form the basis for the theory of combustion; a contemporary one is the quarks that are postulated to account for the observed variety of subatomic particles. But with respect to nonobservable constructs the parsimony principle dictates: use only as needed. Perhaps phrase structure itself (variously manifested as, e.g., GB’s S-structure, LFG’s c-structure, and HPSG’s daughters attribute) is the nonobservable linguistic construct that enjoys the widest acceptance in current theoretical work. Surely the evidence for it is far less direct, robust, and compelling than that for phonological structure (e.g. GB’s
bounded information-processing organisms that successfully employ them in a communicative function; second, that language users are able to render judgments as to the well-formedness of candidate expressions (generally taken as the primary data to be accounted for by the theory). On the other hand, in order to circumscribe our task, we do not charge our theory with providing a specific algorithm, though we would expect an adequate theory of language use to provide one.

Of course, decidability of this sort, in and of itself, is a modest criterion to impose on a linguistic theory. If the grammars offered by a linguistic theory are to be embedded into a theory of human language processing, then there is a variety of properties of language processing that might be expected to inform the design of grammar. For example, even the most superficial observation of actual language use makes plain the fact that language processing is typically highly incremental: speakers are able to assign partial interpretations to partial utterances (and quite rapidly, in fact). Thus, other things being equal, a theory of grammar that provides linguistic descriptions that can be shown to be incrementally processable should be regarded as superior to one that does not.

Similarly, we know that language processing is highly integrative—information about the world, the context, and the topic at hand is skillfully woven together with linguistic information whenever utterances are successfully decoded. For example, it is the encyclopedic fact that books don’t fit on atoms—integrated mid-sentence—that allows the correct attachment of the prepositional phrase on the atom to be determined well before word-by-word processing of a sentence like (3) is complete:

(3) After finding the book on the atom, Kim decided that the library really wasn’t as bad as people had been claiming.

Without such nonlinguistic sources of constraint, the interpretation of even the most mundane of utterances can become highly indeterminate. So profound, in fact, is this indeterminacy (and the concomitant reliance of language on situational information) that the very fact that communication is possible using natural language acquires an air of considerable mystery. Although we lack at present any well-developed scientific theory of how linguistic and nonlinguistic information are brought together to resolve such indeterminacy, it is nonetheless clear that we must prefer a linguistic theory whose grammars provide partial linguistic descriptions of a sort that can be flexibly integrated with nonlinguistic information in a model of language processing.

In addition to the incremental and integrative nature of human language processing, we may also observe that there is no one order in which information is consulted that can be fixed for all language use situations. In fact, an even stronger claim can be justified. In examples like (4), early accessing of mor-
phological information allows the cardinality of the set of sheep under discussion to be determined incrementally, and well before the world knowledge necessary to select the 'fenced enclosure' sense of pen rather than its 'writing implement' sense:’

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

In (5), on the other hand, the relevant information about the world (the information, however represented, that allows a hearer to determine that sheep might fit inside a fenced enclosure but not inside a writing implement) seems to be accessed well before the relevant morphological information constraining the cardinality of the set of sheep:

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

What contrasts like these suggest is that the order in which information is accessed in language understanding, linguistic or otherwise, is tied fairly directly to the order of the words being processed. Assuming then that it is the particular language process that will in general dictate the order in which linguistic (and other) information is consulted, a grammar—if it is to play the role, as we assume, of information that fits directly into a model of processing—should be unbiased as to order. Grammars that are to fit into realistic models of processing should be completely order-independent.

Finally, we know that linguistic information, in the main, functions with like effect in many diverse kinds of processing activity, including comprehension, production, translation, playing language games, and the like. By ‘like effect,’ we mean, for example, that the set of sentences potentially producible by a given speaker-hearer is quite similar to, in fact bears a natural relation (presumably proper inclusion) to, the set of sentences that that speaker-hearer can comprehend. This might well have been otherwise. The fact that there is so close and predictable a relation between the production activity and the comprehension activity of any given speaker of a natural language argues strongly against any theory where production grammars are independent from comprehension grammars, for instance. Rather, this simple observation suggests that the differences between, say, comprehension and production should be explained by a theory that posits different kinds of processing regimes based on a single linguistic description—a process-neutral grammar of the language that is consulted by the various processors that function in linguistic activity. The fact that production is more restricted than comprehension can then be explained within a theory of comprehension that allows certain kinds of linguistic constraints to be relaxed, or even word-by-word processing to be suspended, when situational information is sufficient to signal partial communicative intent. Suspension of word-by-word processing clearly cannot enter into production in the

same way (though incomplete sentences sometimes achieve communicative success). Hence, if we appeal to differences of process—not differences of grammar—there is at least the beginning of a natural account for why production should lag behind comprehension. Speakers who stray very far from the grammar of their language run a serious risk of not being understood; yet hearers who allow grammatical principles to relax when necessary may understand more than those who do not. There is thus a deep functional motivation for why the two kinds of processing might differ as they appear to.

Observations of this sort about real language use and language processing are quite robust. Yet, given our current understanding, it is not completely clear how to convert such intuitive observations into criteria for evaluating linguistic theories. The problem is in essence that our understanding of language processing lags well behind our understanding of linguistic structure. Whereas it is reasonable to expect that further research into human language processing will produce specific results that inform the minute details of future linguistic theories, we do not yet know how to bring these considerations to bear.

Despite this uncertainty, the foregoing observations about human language processing suggest certain conclusions about the design of grammar. Grammars whose constructs are truly process-neutral, for example, hold the best hope for the development of processing models. And the best known way to ensure process-neutrality is to formulate a grammar as a declarative system of constraints. Such systems of constraints fit well into models of processing precisely because all the information they provide is on an equal footing. To see this, consider a theory of grammar that does not meet this criterion. A grammar of the sort proposed by Chomsky (1965), for example, embodies transformational rules whose application is order-dependent. The fixed order imposed on such rules is one that is more compatible with models of production than models of comprehension. This is so because production models may plausibly be closely associated with the application of transformations, and the information that must be accessible to determine transformational applicability is localized within a single structural description (a phrase marker) at some level in the transformational derivation. Comprehension models based on transformational grammar, by contrast, seem ineluctably saddled with the problem of systematically applying transformations in reverse, and this is a problem that no one, to our knowledge, has ever solved.

Declaratively formulated grammars like those we develop in this book exhibit no biases toward one mode of processing rather than another. Because each partial linguistic description is to be viewed denotatively, that is, as being satisfied by a certain set of linguistic structures (see above), the constructs of such grammars (e.g. words, rules, or principles) can be consulted in whatever

17. We owe this sort of example to Martin Kay.

18. A similar point is made by Bresnan and Kaplan (1982). See also Halvorsen 1983; Sag et al. 1985; and Fenstad et al. 1987.
order a process may dictate—the constructs are all constraints that, by their very nature, are order-independent and that allow themselves to be processed in a monotonic fashion. Given the current state of our knowledge of language use, a constraint-based architecture of this sort would seem to be the most plausible choice for the design of the theory of language, at least if the goal of embedding that theory within a model of language processing is ever to be realized.

In our concern for processing issues like those we have touched on briefly here, we have accepted the conventional wisdom that linguistic theory must account for linguistic knowledge (a recursively definable system of linguistic types) but not necessarily for processes by which that knowledge is brought to bear in the case of individual linguistic tokens. Indeed, we take it to be the central goal of linguistic theory to characterize what it is that every linguistically mature human being knows by virtue of being a linguistic creature, namely, universal grammar. And a theory of a particular language—a grammar—characterizes what linguistic knowledge (beyond universal grammar) is shared by the community of speakers of that language. Indeed, from the linguist’s point of view, that is what the language is.

But what does language consist of? One thing that it certainly does not consist of is individual linguistic events or utterance tokens, for knowledge of these is not what is shared among the members of a linguistic community. Instead, what is known in common, that makes communication possible, is the system of linguistic types. For example, the type of the sentence I’m sleepy is part of that system, but no individual token of it is.

Just what sorts of things these linguistic types are is another question. Indeed, the precise ontological status of linguistic types is the subject of a very long-standing debate among various schools of conceptualists (e.g. Ferdinand de Saussure, Noam Chomsky), who take them to be mental objects, and realists (e.g. Leonard Bloomfield, Jerrold Katz, Paul Postal, Jon Barwise), who consider them to belong to extramental reality. Thus we might identify linguistic types with such psychological entities as Saussure’s signs or with certain presumably nonmental objects of situation theory (situation types or perhaps infons). For our part, we doubt that the question of whether objects of knowledge are mental or extramental is an empirical one. Fortunately, as Rich Thomason has pointed out, a successful science does not have to have solved its foundational problems: the interminable philosophical debate over the meaning of quantum mechanics has failed to diminish its predictive power. Our concern in this book will be with the internal architecture of the system that the linguistic types form, not with that system’s ultimate ontological status.

19. For a recent, slightly different assessment of the possible stances on the foundations of linguistics, see Katz and Postal 1991.
20. For further discussion of this point, see P&S-87, secs. 1.1–2.

HPSG: A System of Signs

1.1 The Structure of the Sign

In the early days of generative grammar (e.g. Chomsky 1957), the linguistic types singled out for attention were the sentences, considered as strings of phonetic shapes. Correspondingly, a grammar was just a computational device, for example, a context-free grammar or a transformational grammar, that enumerated a set of strings. Most current linguistic theories, of course, are much more demanding: the linguistic types par excellence, the expressions—or signs (in roughly the Saussurean sense)—include not only sentences, but also the words and subsentential phrases, even multisentence discourses. And a sign is taken to consist not only of a phonetic form, but of other attributes or features as well. That is, we conceive of signs as structured complexes of phonological, syntactic, semantic, discourse, and phrase-structural information.

As we noted in the introduction, most current syntactic theories posit two or more levels of representation. In the present formulation of HPSG theory, we assume that all signs at minimum possess the two attributes PHON and SYNSEM. Here the value of the PHON attribute is assumed to be some kind of feature representation of the sign’s sound content that serves as the basis for phonological and phonetic interpretation. We will have nothing to say about the nature of PHON in this book, and we will content ourselves with glossing PHON values as lists of phoneme strings, or often, to enhance readability, lists of English orthographies. The SYNSEM attribute includes a complex of linguistic information that was distributed among the two attributes SYNTAX and SEMANTICS employed in P&S-87. As we will see, this complex of information is more or less analogous to the information that is distributed between the levels of D-structure and LF in current transformational models. Thus the top-
level structure of a sign posited in the present theory is as shown in (1):³

(1)

```
PHON → (a list of phoneme strings)
```

```
sign
```

```
SYNSEM → (an object of type synsem)
```

```
(an object of type syntactic-category)
```

```
(an object of type semantic-object)
```

The information brought together within the SYNSEM attribute forms a natural class in the sense that it is precisely this information that has the potential of being subcategorized for by other signs; in addition it is SYNSEM information that is shared by the complement subject and the controller in raising (Chapter 3). It is these facts that justify the localization of this particular complex of information into a single structure.

Thus the value of the SYNSEM attribute is another structured object, of a type that we will call a synsem object, with attributes of its own called LOCAL (LOC) and NONLOCAL (NONLOC). NONLOC information figures centrally in the analysis of unbounded dependency phenomena (Chapter 4). LOC information in turn is divided into CATEGORY, CONTENT, and CONTEXT attributes. What these three pieces of information have in common, which justifies their being viewed as attributes of a single structure, is that they and they alone are shared between a trace and its filler in an unbounded dependency; we return to this topic in Chapter 4.

Before explicated these notions, we digress briefly to explain a few technical points about feature structures as employed in our theory. Some readers, particularly those steeped in the tradition of generative grammar, may prefer to proceed directly to section 1.4, where we sketch the nature of phrase structure schemata, referring back to sections 1.2 and 1.3 as needed. A summary of sorts

3. This should be compared with the following structure, which was assumed in P&S-87:

```
PHON → (a list of phoneme strings)
```

```
sign
```

```
SYNTAX → (an object of type synsem)
```

```
(an object of type syntactic-category)
```

```
(an object of type semantic-object)
```

3A. Some Formal Properties of HPSG Feature Structures

of objects, appropriate attributes, rule schemata, and universal principles can also be found in the appendix.

1.2 Some Formal Properties of HPSG Feature Structures

The structure for the English pronoun she is shown in (2):⁴

(2)

```
FIRST
```

```
neist
```

```
est
```

```
LIST
```

```
elist
```

```
CATEGORY
```

```
cat
```

```
SUBCAT
```

```
word
```

```
SYNSEM
```

```
LOCAL
```

```
loc
```

```
CONTENT
```

```
npsd
```

```
BACKGROUND
```

```
nest
```

```
RELN
```

```
female
```

```
INSTANCE
```

```
3rd
```

A few purely formal properties of feature structures such as the one depicted in (2) should be noted; the linguistic significance of this structure will be explained in the next section. We presuppose some familiarity with the use of feature structures in linguistics; for introductory accounts, see Shieber 1986 or Pereira and Shieber 1987; for a full formal account of feature structures and feature logic as employed here, see Carpenter 1992.

The first formal point to note is that the feature structures employed in HPSG are sorted. This means simply that each node is labelled with a sort symbol that tells what type of object the structure is modelling; that is, there is one sort symbol for each basic type (ontological category) of construct. It is the labels on the nodes in (2) that specify such sort assignments (for notational ease, the sort symbol labelling a node in a feature structure will often be deleted when it can be recovered from the context). The (finite) set of all sort symbols is as-

4. For simplicity, QSTORE and SYNSEM | NONLOCAL values are omitted here.
a list or a set of linguistic entities of a certain kind. For example, the value of
the SUBCAT feature of a sign is always a list each of whose members is an
entity of sort synsem. And the value of the BACKGROUND feature is a set of
parametrized states of affairs (psosas). For the sake of specificity and
familiarity, we have chosen to model nonempty lists as a sort of feature structure—
nelist—specified for the two attributes FIRST and REST; the empty list is
modelled by another sort—elist—for which no attribute labels are appropriate.
Set values are modelled by nodes with arcs labelled ‘e’ leading to the nodes of
the graph that model the members of the set. For notational convenience, such
nodes bear the label neset (if there are outgoing e arcs) or eset (otherwise), but
technically speaking neset and eset are not sort symbols.

It is crucially important that two distinct paths in a feature structure can
lead to one and the same node: for example, the paths synsem | loc | content
and synsem | loc | content | background | e | instance
in (2). In such cases, we speak of structure sharing: two paths share the same
structure as their common value. Informally (but not quite correctly), the
values of the two paths are often said to be unified. It is important to be clear
that structure sharing involves token identity of values, not just values that are
structurally identical feature structures; identity of the latter kind is referred to
as type identity or structural identity. It is unification in this sense of structure
sharing that gives its name to the family of ‘unification-based’ linguistic fram-
eworks, of which HPSG is an exemplar. It is not going too far to say that in
HPSG structure sharing is the central explanatory mechanism, much as move-
α is the central explanatory mechanism in GB theory; indeed, the relationships
between fillers and traces, between ‘understood’ subjects and their controllers,
between pronouns and their antecedents, between ‘agreement sources’ and
‘agreement targets,’ and between the category of a word and the category of its
phrasal projections will all be analyzed as instances of structure sharing.

A final formal point to note is that for some sorts, no attribute labels are
appropriate. Such sorts are called atoms. Examples are case and its subsorts
nom and acc; number (num) and its subsorts singular (sing) and plural (plu);
and elist. A feature structure consisting of a single node labelled by an atom
will sometimes be called an atom itself, though ‘atomic feature structure’ is
more technically correct.

As noted above, we often formulate constraints on modelling structures
like (2) in terms of an appropriately designed description language. In one
easily imagined such language, essentially a quantifier-free predicate calculus
with sorts and features represented as unary and binary predicates, respec-
tively, a reasonably complete description of a word like (2) might be given as
(3):

(3) rooted-at(X1) & word(X1) & PHON(X1,X2) & nelist(X2) &
    FIRST(X2,X3) & she(X3) & REST(X2,X4) & elist(X4) &
SYNSEM(X1,X5) & synem(X5) & LOCAL(X5,X6) & local(X6) & CAT(X6,X7) & cat(X7) & HEAD(X7,X8) & noun(X8) & CASE(X8,X9) & nom(X9) & SUBCAT(X7,X10) & elist(X10) & CONTENT(X6,X11) & ppro(X11) & INDEX(X11,X12) & ref(X12) & PERSON(X12,X14) & 3rd(X14) & NUMBER(X12,X15) & sing(X15) & GENDER(X12,X16) & fem(X16) & RESTRICTION(X11,X17) & eset(X17) & CONTEXT(X6,X18) & context(X18) & BACKGROUND(X8,X19) & SINGLETON-OFO(X19,X20) & psoo(X20) & INSTANCE(X20,X12) & RELATION(X20,X21) & female(X21)

We say that (3) describes (2), or that (2) satisfies (3) relative to some assignment of the variables in (3) to nodes of (2). Note that the conjuncts INDEX(X11,X12) and INSTANCE(X20,X12) clearly express the structure sharing appropriate to (2).

In a proper formalization of linguistic theory, grammars, universal and language-particular principles, and lexical entries will all be given as logical formulas that are interpreted as denoting the set of feature structures that satisfy those formulas. Throughout this volume, however, we will describe feature structures using attribute-value matrix (AVM) diagrams. The perspicuity gained by this choice of description language can readily be appreciated by comparing (3) with the AVM description of (2) shown in (4):

\[
\begin{align*}
\text{SYNSEM} \quad \text{LOCAL} \\
\text{word} \quad \text{synem} \\
\text{CATEGORY} \quad \text{HEAD noun\{CASE nom\}} \quad \text{CAT} \quad \text{SUBCAT} \quad \text{PER} \\
\text{INDEX} \quad \text{NUM} \quad \text{sing} \quad \text{GEND fem} \\
\text{PRO} \quad \text{RESTR} \quad \text{INST} \\
\text{BACKGR} \quad \text{psoo} \\
\text{CONTEXT} \quad \text{RELN female} \\
\end{align*}
\]

Note that descriptions of (or, equivalently, constraints on) feature structures, be they first-order formulas like (3) or AVM diagrams like (4), need not be complete. For example, although satisfied by the feature structure (2), the AVM diagram

\[
\begin{align*}
\text{SIGN} \\
\text{SYNSEM} \quad \text{LOCAL} \\
\text{cat} \quad \text{HEAD noun\{CASE nom\}} \quad \text{SUBCAT} \quad \text{PER} \\
\text{INDEX} \quad \text{NUM} \quad \text{sing} \quad \text{GEND fem} \\
\text{PRO} \quad \text{RESTR} \quad \text{INST} \\
\text{BACKGR} \quad \text{psoo} \\
\end{align*}
\]

provides only a partial description of (2). For instance, the sort is given not as word, but as the less specific supersort sign; the value of the path SYNSEM | LOC | CAT | HEAD is specified only as noun, with no indication of the CASE value; and only the CATEGORY value is described for the path SYNSEM | LOC, not the CONTENT or CONTEXT value. Of course the description (4') is also satisfied by many feature structures in addition to the one in (2); in general, the more specific (or explicit, or informative) a description, the fewer feature structures satisfy it.

A common source of confusion is that feature structures themselves can be used as descriptions of other feature structures. Since feature structures that are not totally well-typed and sort-resolved can be arranged into a partial ordering relation (called the subsumption relation; see Shieber 1986 or P&S-87, Chapter 2), any feature structure can be thought of as partially describing any of the feature structures that it subsumes. Similarly, a feature structure can be taken as a partial description of any of the well-typed (or totally well-typed, or totally well-typed and sort-resolved) feature structures that it subsumes. Next, we choose to eliminate this possible source of confusion by using only totally well-typed, sort-resolved feature structures as (total) models of linguistic entities and AVM diagrams (not feature structures) as descriptions.

The following matters of notation should be observed with respect to AVM descriptions like (4). First, sort assignments are indicated by left subscripts (e.g. word) if the object in question has one or more of its attributes specified, and by an atomic symbol (e.g. feminine (fem)) otherwise. Second, structure sharing is indicated by multiple occurrences of boxed numerals called tags, for example, \{'\}. Third, descriptions of sets are given within curly braces, with \{'\} describing the empty set. And fourth, descriptions of lists are usually abbreviated by the use of angle-bracket notation (instead of the attribute labels FIRST and REST), with \{\} describing the empty list. These notations are modelled after those of Shieber 1986 and those of P&S-87. AVM descriptions will also employ functional or relational symbols such as append, union (\cup), and \& all of which we consider to be necessary in a linguistically adequate description language; the use and interpretation of such symbols will be explained as they are introduced.

1.3 The Linguistic Application of Feature Structure Descriptions

We turn now to the linguistic significance of the various substructures contained within the feature structure (2) described in (4). First, note that the indicated sort of the whole structure is word. Now let us consider the structure of sort loc that lies at the end of the path SYNSEM | LOC. As noted above, this structure has the three attributes CATEGORY (cat), CONTENT (cont), and CONTEXT (contx). Here the CATEGORY value includes not only what would be regarded by most syntacticians as the syntactic category of the word in question, but also the grammatical arguments it requires. The CONTENT value

7. More precisely, feature structures can be viewed as representing logical equivalence classes of nondisjunctive formulas in certain feature logics.
constitutes the word’s contribution to (context-independent) aspects of the semantic interpretation of any phrase that contains it. And the CONTEXT value contains certain context-dependent linguistic information usually discussed under such rubrics as indexicality, presupposition, and/or conventional implicature. Let us examine these three structures more closely.

The CATEGORY value is an object of sort category (cat), and it contains the two attributes HEAD and SUBCAT.8 Roughly speaking, the HEAD value of a sign is its part of speech, analogous to the information contained in an X-theory category stripped of bar level information. As we will discuss in the following section, a principle of universal grammar (the Head Feature Principle) essentially stipulates that the HEAD value of any sign is always structure-shared with that of its phrasal projections. The appropriate values for HEAD are divided into the two sorts substantive (subst) and functional (funct). Subsorts of the sort substantive are noun, verb, adjective (adj), and preposition (prep), whereas determiner (det) and marker (mark) (e.g., complementizers) are the two subsorts of the sort functional that we will deal with here. For the sake of familiarity, we will refer to all these sorts as parts of speech.9 This list is not intended to be exhaustive; we leave open the question of the precise inventory of the parts of speech. Some parts of speech have attributes of their own.10 For example, noun has the feature CASE, prepositions have the attribute PREPOSITION-FORM (PFORM), and verbs have the attribute VFORM as well as the boolean features AUX and INVERTED (INV). In addition, the boolean feature PREDICATIVE (PRD) is appropriate for the sort subst (i.e., for all the parts of speech noun, adj, verb, and prep).11 And the feature SPECIFIED (SPEC) is appropriate only for the sort funct.12 In the present case, the HEAD value is specified as noun [CASE nom].

The SUBCAT value of a sign is in essence the sign’s valence, that is, a specification of what other signs the sign in question must combine with in order to become saturated. More precisely, the SUBCAT value is a list of synsem objects, corresponding to the SYNSEM values of the other signs selected as complements by the sign in question. Here the notion complement is broadly construed to include only not sisters of lexical heads but also certain dependent elements classified as specifiers in GB theory (i.e., subjects, including determiner subjects of NPs). The condition that the SYNSEM value of a complement is structure-shared with the corresponding specification within the SUBCAT value of the sign that selects the complement is required by another universal principle to be discussed in the following section (the Subcategorization Principle). In the present case the SUBCAT value is ( ) because pronouns do not require any complements in order to become saturated; more complex cases will be considered below. It is noteworthy that only the SYNSEM values of complements are selected, not any of their other attributes (e.g., PHON and DTRS).13 We thus eliminate a defect of the account of subcategorization given in P&S-87, wherein the SUBCAT value was assumed to be a list of signs. Because that account did not constrain what attributes of complements can be subcategorized for, a Locality Principle had to be introduced in order to disallow the possibility of heads imposing conditions on internal constituent structure of their complements (P&S-87, pp. 143–144). In the current theory, the essential content of the Locality Principle follows immediately from the internal structure of signs. That is, because of the particular information that is included in objects of the sort synsem (e.g., information about part of speech, agreement features, case and referential index), we derive the effect that category selection, head-dependent agreement, case assignment, and semantic role assignment are all strictly local. Looking at this from a somewhat different perspective, precisely because such information about subconstituents of a

8. In P&S-87, a sort syntactic-category was employed for values of the SYNTAX attribute. This sort had the two attributes BINDING (similar to the NONLOCAL attribute of SYNSEM values in the present arrangement) and LOCAL (essentially the same as the present attribute CAT). The revision of the top-level structure of signs can be summarized as in P&S-87 (i) and this volume (ii), respectively:

(i)

(ii)

A third LOCAL attribute, LEXICAL (LEX), was employed in Chapter 7 of P&S-87, in the account of English constituent ordering principles. In order to carry that account over into the present system, LEX should be added to the inventory of CAT attributes; but since we will not touch on questions of constituent ordering in this volume, the LEX feature will not be employed in this volume.

9. The part of speech marker is introduced for the analysis of complementizers and certain other ‘minor’ words, discussed below.

10. Such attributes correspond for the most part to the HEAD features of P&S-87. But note that the HEAD feature MAJOR (MAJ) of P&S-87 is superseded here by the sort of the HEAD value; also, as explained below, the function of the old HEAD feature NOUN-FORM (NFORM) for nouns is taken over by the sort of the INDEX value.

11. The motivation for these features and inventories of their possible values are given in P&S-87, sec. 3.1.

12. The feature SPEC will play a crucial role in our analysis (section 1.6) of complementizers as markers that select the kind of phrase they ‘mark’ and our treatment of noun phrases (sections 1.7 and 1.8), wherein determiners select the kind of nominal they combine with.

13. The DAUGHTERS attribute, which is defined for phrases but not for words, will be introduced in the following section.
complement is absent from synsem objects, it follows that no head can select for, agree with, or assign a role to such a subconstituent.

As justified at length in P&S-87, we assume that the ordering of elements on the SUBCAT list corresponds not to surface order, but rather to a version of the traditional obliqueness hierarchy. Thus subjects appear first (leftmost), followed by other complements (if any) in the order primary object, secondary object, then oblique PP and verbal and/or predicative complements.14 As explained in P&S-87 (see also Sag and Pollard 1989), this ordering is broadly similar to other proposed hierarchies of grammatical relations such as the Keenan-Conrie accessibility hierarchy, the 1-2-3-oblique ordering of relational grammar, and the SUBJ-OBJ-OBJ hierarchy employed in the LFG Lexical Rule of Functional Control. Our obliqueness order also corresponds closely to the semantic order of arguments assumed in categorial grammar accounts such as that of Dowty 1982a, 1982b. However, under the set of semantic assumptions that we adopt here, where the contents of grammatical arguments are structure-shared with substructures of the head’s content (rather than being taken as arguments), the categorial notion of semantic argument order does not have any analog.

We now turn to the CONTENT value, which (together with CONTEXT) specifies the sign’s contribution to semantic interpretation, especially with respect to matters of reference.15

For the CONTENT value of nominals (i.e. lexical nouns and their phrasal projections), we will employ a feature structure sort called nominal-object (ncom-obj), which in turn bears an attribute called INDEX (IND). The INDEX value, a structure of sort index, should be thought of as the HPSG analog of a reference marker in discourse representation theory (DRT: Kamp 1981) or of a parameter introduced by an NP use in situational semantics. As we will soon see, it is to indices that semantic (or thematic) roles are assigned.16 Indices themselves are classified into subsets, according to what might be called their mode of referring. Thus index has the three subsets referential (ref), there, and it. Indices of the latter two sorts are used only for the expletive (dummy or pleonastic) pronouns there and it, respectively. Referential indices are used for semantically contentful nouns as well as for nonpredicative PPs (i.e. PPs in which the head preposition functions analogously to a case marking, so that the whole PP gets its content from the prepositional object NP).

Nominal-objects are further divided into the two subsets nonpronom (npro) and pronoun (pron), with the latter sort itself subdivided into the subsets personal-pronom (ppro) and anaphor (ana). Finally, anaphor has the two (maximal) subsets reflexive (refl) and reciprocal (recp). The three subsets ppro, ana, and npro correspond roughly to the classification within GB theory of NPs as pure pronomininals (+p, -a), pure anaphors (-p, +a), or R-expressions (-p, -a). In the present example, the CONTENT value of sort ppro, which entails that she will be subject to a certain principle of the HPSG binding theory (Chapter 6) analogous to GB’s Principle B.

Indices in turn have the three agreement features PERSON (PER), NUMBER (NUM), and GENDER (GEND). Indices play a role in our theory analogous to that of NP indices in GB theory: two nominals are said to be coindexed if their indices are token-identical (structure-shared). For example, in the sentence he shaved himself the indices of he and himself will be structure-shared. Their CONTENT values, however, will not be; indeed, the CONTENT value of he is of sort ppro, while that of himself is of sort refl. Since agreement features are internal to indices, it follows that coindexed elements always agree.17 This in turn has important consequences for the HPSG theory of agreement, presented in Chapter 2.

We also allow for the possibility that the content of a (nonexpletive) nominal introduces a semantic restriction on its index, which, when present, will be the value of the RESTRICTION attribute of the nominal-object. The value of the RESTRICTION attribute is a set of parametrized states-of-affairs (posas).18 Expletive pronouns, which are always nonreferential, are specified as [RESTRICTION {}].

Within a posa, the values of the attributes other than the RELATION attribute (i.e. the arguments of the posa) can be either referential indices or posa (which may in turn contain further indices of their own). In general the referential index arguments will originate in the contents of NPs or nonpredicative PPs, while the posa arguments will arise from predicative phrases (such as sentential...

14. Unlike P&S-87, however, left-to-right order on the SUBCAT list corresponds to increasing, rather than decreasing, obliqueness. This is more mnemonic, since it corresponds roughly to English surface order (rather than its reverse).

15. The way that the CONTENT-related notions presented here figure in a situation-based account of semantic interpretation will be discussed at length in Chapter 8. For background on relevant situation-semantic notions, see Gawron and Peters 1990, Devlin 1991, Cooper 1990, and Ginzburg 1992.

16. Our approach thus differs in detail from most work in situation semantics, where the objects to which roles are assigned are restricted parameters, which are roughly analogous to our nominal-objects.

17. At present, we have no account of such colloquialisms as everyone has behaved themselves (pointed out to us by Tom Wasow), which are counterexamples to all theories of agreement that we are familiar with.

18. States-of-affairs (soa) were called ‘circumstances’ in P&S-87. In recent situation semantics literature, the term ‘infon’ is also frequently employed in this connection (see Devlin 1991). A (posa) is represented by a feature structure that specifies a relation (for us, the value of the RELATION (RELN) attribute) together with values for the (argument) roles of that relation (for us, represented by the other attributes). In situation theory, a soa also specifies a polarity (either 1 or 0), but since we will not be concerned with the analysis of negation, we suppress the POLARITY attribute, which can always be assumed to have the value 1.
and VP complements, or predicative APs and PP's). Each psoa in the RESTRICTION value is interpreted as placing semantic conditions on the entities that the indices appearing in them can be anchored to in a given context (or range over, in case an index is quantified over). In the present example, the set of restrictions is empty; but nonempty restrictions are introduced in the case of common nouns. For example, the CONTENT value of the common noun book would be as in (5):

\[
\text{INDEX} \begin{bmatrix}
\text{PER} & \text{3rd} \\
\text{NUM} & \text{sing} \\
\text{INDEX} & \text{GEN} & \text{neut}
\end{bmatrix}
\]

\[\text{RESTRICTION} \{ \text{psoa} \begin{bmatrix}
\text{RELATION} & \text{book} \\
\text{INSTANCE} & 1
\end{bmatrix} \} \]

The significance of the RESTRICTION value is that when the word book is used referentially (e.g. in a referential use of the phrase a book), the index 1 introduced by that use must be anchored to an entity that renders each psoa in the set (in this case, a single psoa) factual; that is, the index must be anchored to a book. And when the index is bound by quantification (e.g. in a use of the expression every book), the quantification is constrained to range over a set of entities that render the psoa factual; that is, it will range over some contextually salient set of books.

19. When an NP is used referentially, its referent is just the object that its index is anchored to. If two NPs are coindexed (i.e. they share a common index) and one of the NPs refers to some entity X (which is then the anchor of that NP's index), then the other NP necessarily also refers to X. Similarly, if some index is bound by a quantifier, then any other occurrence of the same index must also be bound by the same quantifier. For further discussion of these and related issues, see Chapter 8.

20. The present terminology and structuring of information within CONTENT values differ in numerous respects from those employed in P&S-87. The differences can be summarized by comparing the CONTENT value for book in (5) with the diagram in (i), which shows how (roughly) the same information would have been structured in P&S-87:

\[
\text{INDEX} \begin{bmatrix}
\text{VARIABLE} & \text{1} \\
\text{NUM} & \text{sing} \\
\text{INDEX} & \text{GEN} & \text{neut}
\end{bmatrix}
\]

\[\text{RESTRICTION} \{ \text{psoa} \begin{bmatrix}
\text{RELATION} & \text{book} \\
\text{INSTANCE} & 1
\end{bmatrix} \} \]

Restructuring along these lines leads to numerous simplifications in the theory with respect to expletive pronouns (Chapter 3), binding (Chapter 6), control (Chapters 3 and 7), and quantification (Chapter 8).

Finally, we consider the CONTEXT value, which is a structure of sort context. For the moment we consider a single context attribute called BACKGROUND (BACKGR), whose value is a set of psoas. Like the psoa in RESTRICTION values, each of these background psoas restricts the possible anchors of indices. Unlike RESTRICTION psoas, however, BACKGROUND psoas are not part of the CONTENT value but should rather be considered as felicity conditions on the utterance context. To put it another way, CONTENT values represent contributions to literal (truth-conditional) meaning, while BACKGROUND values represent conditions on anchors that correspond to presuppositions or conventional implicatures. For example, in the present case the single background psoa corresponds to the presupposition that the referent of the English feminine pronoun she must be female. This is how we capture the fact that English is a 'natural gender' language (as opposed to, say, French, which is a standard example of a 'grammatical gender' language). By contrast, the French feminine pronoun elle does not introduce this BACKGROUND psoa: the referent of elle need not be a female. (For further discussion of these and related issues, see Chapter 2).

To take another example, the BACKGROUND value of the proper NP John would contain the psoa shown in (6):

\[
\text{RELATION} \begin{bmatrix}
\text{BEARER} & \text{1} \\
\text{NAME} & \text{John}
\end{bmatrix}
\]

Here the atomic value John refers to the name John, not to an individual named John. The psoa (6) corresponds to the presupposition that the referent of a use of the name John be named John, or, to be somewhat more precise, that the referent be identifiable in the utterance context by means of the name John. A somewhat different kind of example will be provided in Chapter 2, where we analyze Korean honorific verbs as introducing a background psoa corresponding to the condition that the speaker be honoring the referent of the verb's subject. Such conditions have often been regarded as a kind of conventional implicature.

Let us now examine lexical entries for some verbs. To ease the notation, we introduce in (7) some abbreviations for certain AVM diagrams of sort synsem that will be employed repeatedly; as always, sort symbols will be omitted when recoverable from context:

21. The BACKGROUND attribute fulfills much the same function as the INDICES attribute in P&S-87. Later, we will add a second context attribute called CONTEXTUAL-INDICES (C-INDS), which represents information about various indexical coordinates such as SPEAKER, ADDRESSEE, indices of spatiotemporal location, etc.
Like all tensed verbs in English, this bears the VFORM value finite (fin) and
subcategorizes for a nominative subject.23 And like all third-singular verbs, it
requires that the index on the subject be [PER 3rd, NUM sing]; in other words,
for a verb, being third-singular is simply a matter of selecting a third-singular
subject. Finally, the CONTENT value expresses the fact that the verb walks
makes reference to the walk relation, and that the WALKER role of the walk
relation is filled by the referential index of the subject.

The lexical entry for see—(9)—is similar, but now in addition to a third-
singular nominative subject (whose index fills the SEE role), an accusative
primary object is also selected whose index is assigned to the SEEN role. And
gives (in one of its valence alternants) further selects an accusative secondary
object, with the assignment to semantic roles as shown in (10). Note that the
order of the SUBCAT elements in (10), in virtue of English-specific principles
of linear precedence, will correspond to the (focus-neutral) order in which the
corresponding complements are realized (see P&S-87, Chapter 7):

(9) sees

(10) gives

There are a number of general points to be noted with respect to lexical
entries like those in (8)-(10). First, unlike most work in GB and GPSG (but
like LFG), we assume that subjects are selected by verbs just as nonsubject complements are. (For a defense of this position, see P&S-87, sec. 5.4.)

Second, assignment of case to complements, including nominative case assignment to the subject of finite verbs, is simply treated as part of subcategorization, no different in principle from selection of a VP complement bearing a particular VFORM value or of a PP complement bearing a particular PFORM value. There is no separate theory of case (or of Case). Nominative case assignment takes place directly within the lexical entry of the finite verb, without the intercession of an abstract INFL element. Loosely speaking, though, we might say that nominative case in English is assigned by the VFORM value finite, which is HPSG's closest analog to GB theory's AGR-bearing INFL. As in GB, we assume that the subject SUBCAT element of a nonfinite verb (as well as a predicative PP, AP, or NP) does not have a CASE value specified.

Third, the assignment of semantic roles is uniform, with all roles, including that of the subject, being assigned directly within the verb's lexical entry; in particular, there is no notion of the subject's role being assigned 'externally' or by the intercession of the verb's phrasal projection. In all cases the role assignment comes about by structure sharing of a SUBCAT element's index with the value of some attributes (i.e. semantic role) of the verb's CONTENT value; thus it is not the complements themselves (subject, object, etc.) that are assigned roles, but rather their indices. Thus the CONTENT value of a verb directly embodies the verb's underlying predicate-argument structure; however, unlike the GB account of role-assignment under subcategorization, roles are explicitly modelled as attributes of the verb's CONTENT. To summarize, in HPSG the selection of complements (including subjects), as well as the assignment of their cases and semantic roles, all takes place within the lexicon.

24. This assumption bears comparison to certain proposals within GB (e.g. Kitagawa 1986) to the effect that subjects originate under the maximal projection of V. In Chapter 9, we will consider an alternative position, wherein subjects and (nonsubject) complements are selected by distinct SUBJ and COMPS features.

25. We do acknowledge, however, that for languages with more complex case systems, some sort of distinction analogous to the one characterized in GB work as 'inherent' vs. 'structural' is required. For example, the analysis of Icelandic case assignment proposed in Sag et al. 1992 also distinguishes between structural (default) case and lexically assigned case specifications. For some discussion in the context of German, see Pollard, to appear.

26. This feature is shared with GB proposals wherein subjects originate under VP. See n. 24 above.

27. A similar intuition seems to underlie the suggestion by Williams (1989) that the binding theory applies to 0-roles.

28. As mentioned above, another difference is that we do not assume that every nonsubject complement is assigned a role. We return to this question in our account of 'subject-to-object raising' (Chapter 3).

1.4 Phrasal Signs and Principles of Universal Grammar

The principal type of object with which our theory is concerned, of course, is the sign (corresponding to the feature structure sort sign); and we assume that signs fall into two disjoint subtypes, phrasal signs (sort phrase) and lexical signs (sort word). We have seen in the previous section some representative words, though these need not be given simply as a list; considerable redundancy can be eliminated by organizing the lexicon as a multiple-inheritance hierarchy along the lines proposed in Chapter 6 of P&S-87 (see below).

But how do we specify the well-formed phrases of a given language? The answer we give is similar to the one given in GB theory: a candidate phrase will be well-formed provided it satisfies all the principles of grammar, including both universal principles and language-specific principles. We do not rule out as a possible long-term goal the elimination of language-specific principles (e.g. principles of constituent ordering) in favor of 'parametrized' universal principles; and we will from time to time propose certain variants of universal principles in response to the empirical demands of some language or other, which might be regarded as parametrized forms of some more general principle. But in our opinion the notion of a parameter as it is currently employed in much syntactic research is not sufficiently constrained and well-defined to bear much real empirical weight. In the absence of a list, however tentative, of posited parameters and their range of settings, together with a substantial, worked-out fragment for at least one language, a specification of the settings for that language, and a reasonably detailed account of how those settings account for the array of facts covered in the fragment, we are inclined to view parameter-based accounts of cross-linguistic variation as highly speculative.

In this section, we consider the structure of some simple phrases and explain two of the universal principles proposed in HPSG, the Head Feature Principle (HFP) and the Subcategorization Principle, which function as rough analogs of GB's Projection Principle.

Unlike words, phrases have the attribute DAUGHTERS (DTRS) (in addition to PHON and SYNSEM), whose value is a feature structure of sort constituent-structure (con-struc) representing the immediate constituent structure of the phrase. The sort con-struc has various subsorts characterized by the kinds of daughters that appear in them. For the most part we will be concerned in this book with the constituent structure subsort headed-structure (head-struc).

29. Formally, this assumption is expressed by statements to the effect that, in the partial ordering of the sorts, sign is the meet or greatest lower bound of word and phrase, but word and phrase have no join or least upper bound (i.e. they are mutually inconsistent).

30. For further discussion, see P&S-87, pp. 55–59. Our view of hierarchically classified sorts of structures, or constructions, bears a striking resemblance to independent work on 'construction grammar.' See, e.g., Fillmore and Kay to appear.
which is employed in all headed constructions. Appropriate attributes for the sort head-struc include HEAD-DAUGHTER (HEAD-DTR) and COMPLEMENT-DAUGHTERS (COMP-DTRS); other attributes such as ADJUNCT-DAUGHTER (ADJ-DTR), FILLER-DAUGHTER (FILLER-DTR), and MARKER-DAUGHTER (MARKER-DTR) appear on more specific subsets of head-struc to be discussed later. Thus we assume the minimal overall structure of a headed structure to be as shown in (11):

(11) \[
\text{head-struc} \\
\begin{array}{l}
\text{HEAD-DTR (a sign)} \\
\text{COMP-DTRS (a list of signs)}
\end{array}
\]

Note that every headed structure has a unique head daughter, but there may be many (or no) complement daughters; as with SUBCAT lists, order on the COMP-DTRS list is by increasing obliqueness.\(^{31}\)

We now consider certain subsets of head-struc, beginning with the sort head-complement-structure (head-comp-struc). In this sort, the only daughters are the head daughter and complement daughters (if any).\(^{32}\) The structure of a very simple phrase whose DTRS value is a head-complement structure is sketched in (12); numerous inessential details are omitted:

(12) \[
\text{phrase} \\
\begin{array}{l}
\text{PHON} (\text{Kim, walks}) \\
\text{SYNSEM } s_{[fin]}
\end{array}
\]

\[
\text{DTRS} \\
\begin{array}{l}
\text{HEAD-DTR (walks)} \\
\text{COMP-DTRS (Kim)}
\end{array}
\]

For the sake of familiarity—and typographical convenience—we adopt certain abbreviatory conventions for displaying phrasal structures as illustrated in (13):

(13) \[
\text{LOC} | \text{CAT} \\
\begin{array}{l}
\text{HEAD-} \\
\text{SUBCAT (sandy, gives)}
\end{array}
\]

Let us consider (14) in bottom-up fashion, beginning with the lexical head gives. Note that this is the same as the lexical entry (10).\(^{33}\) The next node up, representing the VP gives Sandy Fido, has three daughters: the lexical head already considered (H), the primary object Sandy (C), and the secondary object Fido (C). There are two important points to note here. First, each of the complement daughters of the VP has its SYNSEM value (indicated by the tags '2' and '3') token-identical to one of the elements on the SUBCAT list of the head daughter, while the SUBCAT value of the VP itself consists of the head daughter's SUBCAT list minus those requirements that were satisfied by one of the complement daughters. This state of affairs exemplifies the Subcategorization Principle, which is stated in (15):

(14) \[
\text{LOC} | \text{CAT} \\
\begin{array}{l}
\text{HEAD-} \\
\text{SUBCAT (sandy, gives)}
\end{array}
\]

31. Again, this reverses the convention of P&S-87.

32. We make the standard assumption that there are head-complement structures where the list of complement daughters is empty. Such structures are employed to form phrases whose only daughter is a noun or verb lexical head, so that there is a distinction, say, between the word walk and the VP walk with the word walk as its head (and only) daughter.

33. Note that here the numerical tags within the SUBCAT value refer to the synsem objects on the SUBCAT list, not to their indices as in (10).
(15) **Subject Categorization Principle:**

In a headed phrase (i.e., a phrasal sign whose DTRS value is of sort *head-struc*), the SUBCAT value of the head daughter is the concatenation of the phrase's SUBCAT list with the list (in order of increasing obliqueness) of SYNESEM values of the complement daughters.

The effect of this principle is to 'check off' the subcategorization requirements of the lexical head as they become satisfied by the complement daughters of its phrasal projections; at the same time, the SUBCAT elements themselves are token-identical to the SYNESEM values of the corresponding complements. Thus the Subject Categorization Principle works much the same way as cancellation in categorial grammar.

The second point to note with respect to the VP node is that its HEAD value (indicated by the tag '4') is token-identical to that of gives. This state of affairs exemplifies the **Head Feature Principle**, which is stated in (16):

(16) **Head Feature Principle (HFP):**

The HEAD value of any headed phrase is structure-shared with the HEAD value of the head daughter.

The effect of the HFP is to guarantee that headed phrases really are 'projections' of their head daughters. Notice that when the token identities in (14)—indicated by '1' and '3'—are taken into consideration, the CAT value of the VP node is as shown in (17), namely, the category of VP[fin]:

(17) \[
\begin{align*}
\text{HEAD} & : \text{verb[fin]} \\
\text{SUBCAT} & : \langle \text{NP[nom][3rd, sing]} \rangle
\end{align*}
\]

Finally, we consider the top node, representing the whole sentence, which has two daughters. One, the head daughter (H), is the VP just considered, while the complement daughter (C) is the subject *Kim*. In accordance with the Subcategorization Principle, the SYNESEM value of the subject is token-identical to the remaining element (indicated by '1' on the SUBCAT list of the VP); and in accordance with the HFP, the HEAD value is token-identical to that of the VP (and therefore to that of the head verb). Taking into consideration the token identity indicated by '4', the CAT value of the S node is as shown in (18), namely, the category of S[fin]:

(18) \[
\begin{align*}
\text{HEAD} & : \text{verb[fin]} \\
\text{SUBCAT} & : (\_)
\end{align*}
\]

It should be clear from the consideration of this example that in any saturated (SUBCAT (\_)) phrase where every phrasal constituent is headed, it will always be the case that (1) every SUBCAT requirement of the lexical head is satisfied by (i.e., token-identical to the SYNESEM value of) precisely one complement daughter of one of the lexical head's phrasal projections; and (2) the HEAD value of the entire phrase will be token-identical to that of the lexical head. Thus the combined effect of the Subcategorization Principle and the HFP is somewhat analogous to that of GB's Projection Principle, inasmuch as it guarantees that the information in the CAT of the lexical head (which should be seen as analogous to the D-structure of the whole sentence) is in some sense respected in the sentence itself. However, unlike GB, HPSG does not adopt the hypothesis that every lexical sign have a saturated projection; instead, we will see some examples of lexical signs that do not in the following section.

It should be noted in passing that the example in (14) also illustrates the handling of subject-verb agreement in HPSG: it arises from the token identity of the subject's SYNESEM value (and therefore the INDEX value) with that of the least oblique SUBCAT element of the lexical head. This should be compared with the GB analysis sketched in (19):

(19) **D-structure:**

    [D = S(19)]

    \[
    \begin{array}{c}
    \text{NP} \\
    \text{[3rd, sing]} \\
    \text{Kim} \\
    \text{INFL} \\
    \text{[3rd, sing]} \\
    \text{V} \\
    \text{NP} \\
    \text{NP} \\
    \text{S} \\
    \end{array}
    \]

    **S-structure (or LF):**

    \[
    \begin{array}{c}
    \text{NP} \\
    \text{[3rd, sing]} \\
    \text{Kim} \\
    \text{INFL} \\
    \text{[3rd, sing]} \\
    \text{V} \\
    \text{NP} \\
    \text{NP} \\
    \text{S} \\
    \end{array}
    \]

    \[
    \text{move-C} \Rightarrow \quad \text{give-Sandy-Fido} \quad \text{V} \\
    \text{NP} \\
    \text{NP} \\
    \text{S} \\
    \end{array}
    \]

Unlike the GB account (Chomsky 1986b, 4–6, 68–69; Chomsky and Lasnik, to appear), there is no movement from V to INFL (or from INFL to V—see, e.g., Chomsky and Lasnik, to appear). Indeed, verb inflection, which we view as a lexical matter, is not accounted for within the syntax at all. There is no INFL node, which we consider an unmotivated conflation of two pieces of information (tense and agreement) that are best kept apart. 34 Neither is there any need to explain how INFL becomes coindexed with the subject NP, or how it is that coindexing causes the agreement features to be copied (if indeed they

34. But we will posit a weakened form of this hypothesis in Chapter 9.

35. Proposals by Pollock (1989) and others to eliminate the INFL node in favor of distinct nodes for tense, agreement, and modality are perhaps motivated by analogous considerations. Our analysis differs in the respect that we do not propose distinct phrase-structural nodes for these different features.
are copied). Instead, our account has the general form shown in (20) (which abbreviates the relevant aspects of (14)): subject-verb agreement arises directly from the lexical specifications of the finite verb in conjunction with the structure sharing imposed by the Subcategorization Principle.

(20)

As is evident from this discussion, the theory of grammar presented here relies crucially on complex lexical information, which determines, in accordance with general principles such as the HFP and the Subcategorization Principle, the essential grammatical properties of phrasal expressions. This does not mean, however, that HPSG relies on complex lexical stipulations, or that the presence of distinct lexical entries with shared properties leads to massive redundancy within the lexicon. As described in P&S-87 and Flickinger 1987 (see also Flickinger et al. 1985; Flickinger and Nerbonne 1992; and Fraser and Hudson 1992), properties of lexical entries and relationships among them are expressed in a concise and principled fashion in terms of classification by a multiple inheritance hierarchy (P&S-87, sec. 8.1) and lexical (redundancy) rules (P&S-87, sec. 8.2), respectively.

The multiple-inheritance architecture for the lexicon employs standard knowledge-representation techniques to cross-classify words according to properties shared across word-classes (which are represented by 'generic' lexical entries, similar to defined types in knowledge representation systems). Thus, each generic lexical entry specifies certain constraints (i.e. values of certain features or relationships among values of different features) that must hold of all actual lexical entries that instantiate the generic entry. The hierarchical organization of lexical entries (both generic and actual) has the effect of amalgamating the information associated with any one actual entry with the information associated with all of the generic entries that it instantiates, thus requiring that a word inherit the properties of all word classes to which it (directly or indirectly) belongs. For example, one lexical entry for the word *gives*

might specify that it inherits all constraints imposed by the two generic entries *ditransitive* and *third-singular-finite* (each of which in turn will inherit other properties from more general generic entries higher in the hierarchy); the only specific information about *gives* that has to be stipulated is the phonology of the base form, the semantic relation of its CONTENT value, and the assignment of semantic roles to grammatical relations. In consequence of this architecture, the lexical entry in question acquires its seemingly complex syntactic information in a maximally general and highly deductive fashion.

Lexical rules, similar to those employed in early lexical-functional grammar (Bresnan ed. 1982), may be interpreted as rules of inference that derive lexical entries of inflected, derived, or compound words from those of simpler words. Passive (see Chapter 3) is one such lexical rule in our treatment of English; others include rules to produce inflected verb forms, valence alternants of 'the same verb' (e.g. the inchoative/causeative alternation), nominalizations, etc. Thus lexical rules too have a highly deductive character.

For detailed discussion of both lexical hierarchy and lexical rules, the reader is referred to the relevant sections of P&S-87 cited above. Although numerous technical aspects of the syntactic theory have been revised since that account was presented, we do not anticipate any particular difficulties in adapting this overall lexical architecture to the current theory.

1.5 Immediate Dominance Schemata

Among the principles of universal grammar, there is a set of principles that have occupied a distinguished status within syntactic theory. These are the principles, variously known as 'grammar rules,' 'immediate dominance (ID) rules,' 'phrase structure rules,' or 'X-schemata,' that in effect serve as templates for permissible local phrase structure trees or configurations of immediate constituency. In the past decade, there has been a culmination of the general trend in syntactic theory toward the 'lexicalization of grammar,' and concomitantly a tendency to collapse or eliminate construction-specific rules in favor of highly schematic immediate dominance templates. For example, most GB work assumes the existence of schematic X-rules such as those shown in (21): 37

\[
(21)
\begin{align*}
\text{a.} & \quad X'' \rightarrow \ y'' \ x' \\
& \text{ specifier) }
\text{b.} & \quad X' \rightarrow \ x \ y'' \\
& \text{ (complement) }
\end{align*}
\]

37. It is frequently assumed that such X-schemata should be derivable from deeper principles. We are sympathetic with such a view, but are not aware of any proposal in this regard that is worked out with sufficient precision to qualify as a genuine theory of possible ID configurations.

36. For an initial attempt at formulating a 'linking theory' for HPSG, i.e. a theory of the relation between semantic roles and grammatical relations, see Wechsler 1991.
The set of such schemata should be regarded as a disjunctively specified principle of universal grammar. That is, one of the universal well-formedness conditions on a phrase is that its set of immediate constituents satisfy one of the schemata.\footnote{\textsuperscript{38}}

Immediate dominance (ID) schemata in HPSG occupy a position in the theory analogous to that of X-schemata in GB theory: they are a small, universally available set of disjunctive constraints on the immediate constituency of phrases, from among which each language makes a selection. Thus the disjunction of the ID schemata itself constitutes a universal principle, which we call the Immediate Dominance Principle (IDP). A preliminary version, containing only two disjuncts (schemata), is stated informally in (22); successive versions will include more—but not many more—schemata.\footnote{\textsuperscript{39}}

\begin{equation}
\text{(22) IMMEDIATE DOMINANCE PRINCIPLE (IDP, preliminary version):}
\end{equation}

The universally available options for a well-formed phrase are:

\begin{enumerate}
\item[(Schema 1)] a saturated ([\textsc{subcat} ( ])) phrase with DTRS value of sort \textit{head-comp-struct} in which the HEAD-DTR value is a phrasal sign and the COMP-DTRS value is a list of length one; or
\item[(Schema 2)] an almost-saturated (\textsc{subcat} list of length one) phrase with DTRS value of sort \textit{head-comp-struct} in which the HEAD-DTR value is a lexical sign; or
\end{enumerate}

\begin{footnotesize}
\begin{itemize}
\item \ldots
\end{itemize}
\end{footnotesize}

Let us consider these schemata one by one. Schema 1 (called ‘Rule 1’ in P&S-87, p. 149) is analogous to the X-schema (21a); it licenses saturated phrases with a phrasal head daughter and one other daughter that is a complement. It is an immediate consequence of the Subcategorization Principle that this complement daughter must have its \textsc{synsem} value token-identical to the single element on the \textsc{subcat} list of the head daughter, and that moreover that element must be the subject (least oblique \textsc{subcat} element) of the lexical head. Thus any phrase licensed by Schema 1 will have the general form shown in (23) (disregarding the surface order of the daughters); note the structure sharings indicated by the tags ‘\textsc{1}’ and ‘\textsc{2},’ which arise from the HFP and the Subcategorization Principle, respectively.\footnote{\textsuperscript{40}}

\begin{equation}
\text{(23)}
\end{equation}

\begin{footnotesize}
\begin{itemize}
\item \textsc{head}
\item \textsc{subcat (2)}
\end{itemize}
\end{footnotesize}

Thus Schema 2 subsumes all conventional phrase structure rules that expand a phrase as a lexical head together with its (nonsubject) complements; an example is the VP[\textit{fin}] node in (14).

For all that the schemata in (22) are analogous to X-schemata, two important differences should be noted.\footnote{\textsuperscript{41}} First, the notion of ‘bar level’ plays no role, being superseded in part by the lexical/phrasal distinction and in part by the relative degree of saturation. Second and more important is a difference in the relationship between the schemata and notions of grammatical function such as subject/specifier and object/(nonsubject) complement. In X-theory, such no-

\begin{equation}
\text{(24)}
\end{equation}

\begin{footnotesize}
\begin{itemize}
\item \textsc{head}
\item \textsc{subcat (2)}
\end{itemize}
\end{footnotesize}

38. Thus X-theory has the same status as any other disjunctively specified principle of UG, e.g. versions of the ECP (Empty Category Principle) that require that any non-pronominal empty category be either lexically governed or antecedent governed.

39. As with all disjunctive principles, it is a goal of the theory to have the precise inventory of possible disjuncts follow from deeper principles.

40. To enhance readability, nodes in constituent structure tree diagrams will often be labelled with only the \textsc{loc} | \textsc{cat} value at each node, rather than the entire \textsc{synsem} value.
tions are defined by the schemata; for example, the subject/specifier in a phrase is the subject/specifier by virtue of occupying a particular position in (some instantiations of) a particular schema, namely, the 1” position in (21a). Thus in GB grammatical functions are wholly configurational notions, where ‘configurational’ in turn is understood in terms of immediate dominance trees. In HPSG, though, such notions as subject and (nonsubject) complement are lexically defined in terms of position on the SUBCAT list of lexical heads. ID schemata, then, do not define grammatical functions, but only constrain how they can be realized configurationally.

This last point can be made clearer by introducing our third ID schema, given in (25):

(25)  (Schema 3) A saturated ([SUBCAT ()]) phrase with DTR value of sort head-comp-struc in which the HEAD-DTR value is a lexical sign.

By the HFP and the Subcategorization Principle, any phrase licensed by Schema 3 will have the general form shown in (26), that is, a phrase in which all complements (including the subject) are realized as sisters of the lexical head:

(26)  

It is this schema, we assume, that licenses clausal structures, standardly analyzed by GB in terms of ‘scrambling,’ in free constituent order languages such as Warmpi; languages like Japanese and German where the complements (including the subject) can be more or less freely ordered are analyzed in a similar fashion, except that parochial conditions require the verb to be clause-final in Japanese and either clause-initial or clause-final in German according as the clause is root or subordinate.43 We also consider VSO structures such as Welsh finite clauses or English ‘subject-auxiliary inversion’ clauses, sometimes analyzed by GB in terms of head movement of the auxiliary from INFL into complementizer position (Sproat 1985), to be instances of Schema 3.44

43. We follow much recent work in analyzing V2 word order in German root clauses in terms of fronting of a constituent from a verb-initial finite clause.

44. Thus ‘Rule 3’ of P&S-87 (p. 156) should be regarded as a special case of Schema 3. In Chapter 9, we will consider an alternative view of this matter.

By way of illustration, we consider briefly an example of subject-auxiliary inversion. The heart of the analysis lies in the lexical entry of the auxiliary verb, exemplified by (27):

(27)  CAT value of the auxiliary verb can:

Note that auxiliaries are simply treated as verbs bearing the specification [+AUX]; there is another head feature INV (INVERTED) appropriate for verbs, for which most finite auxiliaries are unspecified.45 Like all finite verbs, can subcategorizes for a nominative subject; and in common with other modals, it subcategorizes for a base-form VP complement.46 Now in languages like English, where only auxiliaries can invert, we assume that Schemata 1 and 2 are employed in uninveted structures, while Schema 3 is employed in inverted structures. Technically, this means that, in availing itself of Schemata 1 and 2, English imposes on them the further parochial condition that the HEAD value must bear the specification [−INV], while in availing itself of Schema 3, English imposes on it the parochial condition that the HEAD value must bear the specification [+INV], which arises only in cases the lexical head is a finite auxiliary.47

Thus an uninveted sentence like Kim can go is analyzed as in (28):

45. But auxiliaries like first-singular aren’t that must invert are lexically specified as [+INV], while those like better that must not, as well as all verbs that are nonfinite or nonauxiliary, are [−INV]. It is these subtle distributional facts, to the best of our knowledge not treated in other accounts, that motivate the INV feature. For further discussion, see P&S-87, pp. 63–64.

46. In fact the VP complement will be further specified as [SUBCAT {}], i.e. the subject of the auxiliary is structure-shared with the complement subject. This kind of sharing is characteristic of raising-to-subject verbs (discussed in Chapter 3), of which we consider auxiliaries to be a subspecies.

47. Such parochial conditions on ID schemata can be regarded as candidates for parameters of cross-linguistic variation; but as a point of methodology, we are cautious about proposing parameters in the absence of a precise specification of the range of possible values for the parameter. The important point is that the same schemata might well be employed in other languages, with gross typological differences corresponding to different parochial conditions on the structures that realize those schemata. For example, in languages such as Welsh, where finite clauses are VSO but nonfinite VPs and clauses (S)V(NO), we might propose that Schema 3 carry the further condition (roughly) that the HEAD value be specified as fn while Schemata 1 and 2 are nonfinite.
Here the S and VP nodes are licensed by Schema 1 and Schema 2, respectively. Except for the fact that the auxiliary is taken to be a verb rather than INFL, this is grossly similar to the standard GB analysis shown in (29):

In the case of an inverted sentence like *can Kim go*, however, the situation is quite different. On the HPSG analysis, the structure is licensed by Schema 3, as in (30):

Presumably, in most general terms, the point of such an analysis is to capture the relationship between the two sentences. But on our view, the relationship between *Kim can go* and *can Kim go* is captured perfectly in the lexical heads of (30) and (28), which are identical except that the values of the INV feature are instantiated as − and +, respectively. In both cases the NP[nom] *Kim* is the subject and the VP[bse] *go* the oblique complement (whose own subject is controlled by *Kim*). But these facts arise out of the subcategorization of the lexical head, not from the way that the different grammatical functions are variously realized configurationally in (29) and (32). (For a more complete HPSG analysis of the English auxiliary system, see Warner (1992, to appear).)
1.6 Complementizers

In Chomsky 1986b and much subsequent GB work it is assumed that complementizers are heads. Thus, for example, the gross structure of the complementized clause that John left is taken to be that shown in (33):

(33)

Among the arguments adduced for this structure are the following: (1) it extends the X system to complementizers, thereby contributing to greater uniformity of the overall grammar; (2) if we assume, as much current GB research does, that many instances of VSO word order (including subject-auxiliary inversion) arise from movement of the finite verb from INFL to COMP (see (32)), then we predict that inverted sentences do not occur with complementizers; and (3) if we further adopt the standard GB assumption that unbounded dependency constructions (such as wh-questions and relativization) arise from movement of a constituent into the specifier of CP position, then we also predict that ‘dislocated’ constituents appear to the left of the complementizer (e.g. in Scandinavian languages).

Suggestive though they may be, we do not consider such arguments persuasive, since as much as all three arguments depend on GB-internal assumptions that we have not adopted here (e.g. X-theory, the analysis of inversion as movement to COMP, and the analysis of unbounded dependencies as movement to specifier of CP). In addition, as is well known, some verbs select a complementized sentential complement headed by a base-form (rather than finite) verb, as in (34):

(34) I demand that he leave/*leaves immediately.

If the complementizer heads the complementized clause, then it seems difficult to explain how demand selects the complement’s head verb inflection. On the other hand, if S is the head of that-S, then we need only assume that demand subcategorizes for S[base] rather than S[fin].

We are not claiming that the analysis of complementizers as heads is untenable, only that the fundamental intuition underlying such proposals raises as many questions as it answers; enough questions, in fact, to lead one to consider alternative approaches. But if complementizers are not heads, then what are they? We will take the position that they are a subspecies of marker. On our account, a marker is a word that is ‘functional’ or ‘grammatical’ as opposed to substantive, in the sense that its semantic content is purely logical in nature (perhaps even vacuous). A marker, so-called because it formally marks the constituent in which it occurs, combines with another element that heads that constituent. In addition to the complementizers that and for, other examples of markers include the comparative words than and as, the case-marking postclitics of Japanese and Korean, and perhaps nonpredicative adpositions in (the vast majority of) languages where adposition stranding does not occur.

Technically, as noted above, we posit a new part of speech marker (mark). Markers are distinguished from each other, and from nonmarkers, by a new attribute of categories (in addition to HEAD and SUBCAT) called MARKING, with values of sort marking. The sort marking in turn has the subsets marked and unmarked. Here unmarked is the default value, in the sense that it is the value borne by words other than markers. As we will see presently, constituents with a marker daughter inherit the MARKING value from that daughter. Different classes of markers are distinguished by different subsets of marked. For present purposes, we need consider only the subset complementizer (comp), with subsets of its own that and for.

Although markers are not heads, we assume that they resemble heads inasmuch as they select the phrases that they mark; for example, that selects S[fin \lor base] while for selects S[inf]. To this end, we propose that markers bear the head feature SPECIFIED (spec) whose value is of sort synsem. This value will be structure-shared with the SYNSEM value of the (head) sign that the marker combines with to form a phrase. And such combinations are effected by the new ID schema given in (35):

48. It is unclear whether interrogative whether is best treated as a marker or as an adjunct.

49. For arguments that Japanese case clitics are markers rather than heads (‘postpositions’), see Tomobechi 1989.

50. In fact, finite and ‘subjunctive’ forms of English verbs are probably best analyzed in terms of two subsets of a tensed sort (rather than the two unrelated sorts finite and base that we assume here, borrowing from work in GPSG). Under this alternative, the disjunction S[fin \lor base] is replaced by the supersort tensed in the description of the type of phrase that that selects.

51. We must also ensure that phrases without marker daughters are specified [MARKING unmarked]. One way to do this is to assume that each ID schema except the head-marker schema introduces this specification on the mother. Alternatively, we could let MARKING values be determined by a principle along the lines of (i):

(i) MARKING PRINCIPLE:
   In a headed structure, the MARKING value coincides with that of the marker daughter if there is one, and with that of the head daughter otherwise.

We do not know of any empirical basis for choosing between these alternatives.
Here the sort head-marker-structure is a subset of headed-structure that has no complement daughters (i.e., it is specified as \( \text{COMP-DTR} (\ ) \)) and that bears an additional attribute MARKER-DTR. Thus Schema 4 licenses phrases of the form shown in (36):

\[
\text{HEAD} \quad \begin{array}{c}
\text{MARKING} \\
3
\end{array} \\
\text{M} \\
\text{H}
\]

\[
\text{SUBCAT} (\ ) \\
2
\]

\[
\text{MARKING} (\ !\text{marked})
\]

Given complementizer lexical entries like the one in (37), the analysis of that John left will be as in (38):\(^{52}\)

\[
\begin{array}{c}
\text{HEAD} \quad \begin{array}{c}
\text{MARKING} \\
2
\end{array} \\
\text{M} \\
\text{H}
\end{array}
\]

\[
\text{SUBCAT} (\ ) \\
\]

\[
\text{MARKING} (\ !\text{marked})
\]

\[
\text{that} (\text{fin}, \text{that})
\]

\[
\text{John left}
\]

\(^{52}\) The SPEC value of that is specified as unmarked, thus blocking *that that John left.
be the assumption that all quantifiers 'start out in storage'; the final scope that a quantifier receives will depend on which node it is retrieved at and on the order of its retrieval relative to other quantifiers retrieved at the same node. Roughly speaking, stored quantifiers can be thought of as being passed from constituents to their mothers according to the principle stated in (42). \(^{56}\)

(42) **Quantifier Inheritance Principle (QIP, informal version):**

The Quantifier-Store (QSTORE) value of a phrasal node is the union of the QSTORE values of the daughters less those quantifiers that are retrieved at that node.

The retrieved quantifiers themselves will then appear, with their scopes determined, in the CONTENT value of that node.

In anticipation of this analysis, we now revise the treatment of quantificational NPs like *every book* as follows. Instead of appearing as the CONTENT value, the quantifier (39) will appear in the QSTORE value. Only the nominal-object (an object of sort npo) of that quantifier will be 'left behind' as the CONTENT value. Thus the phrase *every book* will have QSTORE and CONTENT as shown in (43):

(43)

\[
\begin{align*}
\text{SYNSEM} & \mid \text{LOCAL} \mid \text{CONTENT} 2 \\
\text{QSTORE} & \{ \text{RESTIND} 2 \mid \text{INDEX} 1 \mid \text{RESTR} \{ \text{RELN book} 1 \} \} \}
\end{align*}
\]

Notice that on this analysis a quantificational NP's CONTENT value is of the same sort as that of a nonquantificational NP (e.g. *she* in (5)), namely nominal-object (i.e. an index together with a restriction on that index). In fact, the CONTENT value of *every book* will be precisely the same as that of the head N* book*. This is an instance of a general constraint that we state provisionally as (44): \(^{57}\)

(44) **Semantics Principle (preliminary version):**

The CONTENT value of a phrase is token-identical to that of the head daughter.

56. In spite of the procedural ring of this informal description, in Chapter 8 quantifier retrieval will actually be formulated declaratively as a relation among the CONTENT and QSTORE values of a sign and those of the sign's daughters.

57. A careful formulation of the Semantics Principle will be given in Chapter 8. Two important cases not covered by the version given here are (1) the case where a quantifier is retrieved, and (2) the case where there is an adjunct daughter. In the former case, any retrieved quantifiers will have to be scoped (in some order) over the head daughter's content. In the latter case, as we will see in the next section, the mother's content must come not from the head daughter, but rather from the adjunct daughter.

Now according to the Quantifier Inheritance Principle, the stored quantifier in (43) must have been present on either the head noun or the determiner. There is good reason to assume that the quantifier does not originate from the head noun. For one thing, we would like to leave open the possibility that a nonquantificational analysis is available for indefinite NPs like *a book* (perhaps along the lines suggested by Kamp (1981)); moreover, some uses of common nouns (e.g. as prenominal modifiers) never become quantified. For these reasons, we assume that the stored quantifier in (43) arises from the determiner.

Let us now consider the phrase structure (i.e. the DTRS value) of *every book*. We assume that common nouns subcategorize for their determiners (perhaps optionally in the case of plural common nouns). \(^{58}\) Thus the lexical sign for *book* is as shown in (45):

(45) \[
\begin{align*}
\text{PHON} \mid \text{book} \\
\text{SYNSEM} \mid \text{LOC} \\
\text{CONT} \mid \text{CAT} \mid \text{HEAD noun} \\
\text{SUBCAT} \{ \text{DetP} \} \\
\text{INDEX} 1 \mid \text{RESTR} \{ \text{RELN book} 1 \}
\end{align*}
\]

Here the symbol 'DetP' abbreviates the synsem structure given in (46):

(46) \[
\begin{align*}
\text{LOC} \mid \text{CAT} \mid \text{HEAD det} \\
\end{align*}
\]

Schema 1 will now license the phrase *every book* with complement daughter *every* and head daughter *book* as shown in (47):

(47) \[
\begin{align*}
\text{CAT} \mid \text{HEAD N} \\
\text{SUBCAT} \{ \} \\
\text{CONT} 2
\end{align*}
\]

\[
\begin{align*}
\text{QSTORE:} & \{ \text{5} \} \\
\text{C} & \text{H} \\
\text{4} & \text{5} \mid \text{DET forall} \\
\text{QSTORE:} & \{ \text{5} \} \\
\end{align*}
\]

\[
\begin{align*}
\text{CAT} \mid \text{HEAD det} \\
\text{SUBCAT} \{ \} \\
\text{CONT} 5 \mid \text{DET forall} \\
\text{QSTORE:} & \{ \text{5} \} \\
\end{align*}
\]

\[
\begin{align*}
\text{CAT} \mid \text{HEAD noun} \\
\text{SUBCAT} \{ \} \\
\text{CONT} 2 \mid \text{INDEX} 1 \mid \text{RESTR} \{ \text{RELN book} 1 \}
\end{align*}
\]

58. We leave open here the question of whether bare plural NPs should be analyzed as having phonetically null determiners.
Note that (47) incorporates a specific hypothesis about the CONTENT value of quantificational determiners (that it is token-identical to the stored quantifier). This analysis satisfies all four of the universal principles posited thus far: the HFP, the Subcategorization Principle, the QIP, and the Semantics Principle.

However, one important question about the analysis remains. How does it come about that the npo element (the RESTIND value) within the content of the determiner is token-identical to the content of the noun? At first blush it might be thought that this identity could be lexically specified within the SUBCAT value of the noun. This will not work, however, since if the head noun were modified (say, by relative clauses or attributive adjectives), the semantic contribution of the modifiers would not be taken into consideration in the content of the determiner (or in the RESTIND value of the stored quantifier). Under present assumptions, it is clear that the determiner must in some sense be able to select its N’ sister, in order to ‘have its hands on’ the N’ sister’s content. This situation bears comparison with the widely adopted hypothesis that determiners head NPs and therefore select their N’s, rather than the other way around.\(^9\)

We will make a more conservative move, however. While continuing to assume that N’s are the heads of NPs and subcategorize for their determiners, we will also assume that determiners reciprocally select their N’ sisters. We effect this selection by means of a mechanism introduced in the preceding section, namely, the SPEC feature. Recall that this is a head feature, appropriate for markers, by which the marker selects (the SYNSEM value of) the phrase that it marks. We now assume that determiners also bear the SPEC feature. That is, both kinds of functional categories, markers and determiners, will be assumed to share the property that they select their head sisters. Thus far, our only argument that determiners select their N’ sisters is a semantic one. In Chapter 9, we will present evidence from German that this selection is (at least sometimes) syntactic as well. Under this assumption, the lexical sign for every is as in (48):

\[
\begin{align*}
\text{PHON (every)} \\
\text{SYNSEM} & \underset{\text{LOC}}{\text{CAT}} \quad \text{HEAD det\{SPEC N:\{2\}\}} \\
\text{CONT} & \underset{\text{RESTIND}}{\text{DET forail}} \\
\text{QSTORE:} & \{5\}
\end{align*}
\]

Here ‘N:\{2\}' abbreviates the synsem structure in (49):

\[
\begin{align*}
\text{CAT} & \text{HEAD noun} \\
\text{SUBCAT} & \{\text{DetNP}\} \\
\text{CONT} & \{2\}
\end{align*}
\]

Finally, we adopt the following general principle to ensure that SPEC values of nonheads get the right values:

\[
\text{SPEC PRINCIPLE:}
\]

If a nonhead daughter in a headed structure bears a SPEC value, it is token-identical to the SYNSEM value of the head daughter.

This concludes our analysis of quantificational determiners. Incidentally, (50) applies to markers as well as to determiners. As a consequence, Schema 4 in (35) can now be simplified to the form (51):

\[
\text{(Schema 4) a phrase with DTRS value of sort head-marker-structure whose marker daughter is a marker with MARKING value token-identical to that of the mother.}
\]

1.8 Possessives

We turn next to possessives, such as my and Mary’s.\(^{60}\) The analysis of these items might be thought to pose something of a challenge, inasmuch as they behave like determiners in some respects but like pronouns in others. More specifically, on the one hand forms like my form NPs in combination with Ns and introduce a stored quantifier; on the other hand they pattern like personal pronoun subjects with respect to the binding theory (Chapter 6). In HPSG, however, the coexistence of these properties in a single expression does not pose an insurmountable problem. To see why, consider the lexical entry for my in (52):

\[
\begin{align*}
\text{PHON (my)} \\
\text{SYNSEM} & \underset{\text{LOC}}{\text{CAT}} \quad \text{HEAD prn\{SPEC N:\{2\}\}} \\
\text{CONT} & \underset{\text{RESTIND}}{\text{DET forail}} \\
\text{QSTORE:} & \{5\}
\end{align*}
\]

60. In P&S-87 we assumed that such signs had a part of speech distinct from det, provisionally called prn. This assumption entailed that all lexical common nouns have two lexical entries, one subcategorizing for a determiner and the other for a possessive phrase. With possessives assimilated to determiners, this is no longer necessary, although this may still be the analysis of choice for languages like Hungarian where nouns exhibit different forms according as they appear with nonpossessive or possessive determiners. See also Chapter 9.
There are a number of points to note here. First, observe that the CAT value is exactly as for the determiner every, except that we have identified the index and restriction of the selected N' head with tags (’[1]’ and ’[2]’, respectively) so that they can be referred to elsewhere in the structure. The reason for this will be explained presently. Second, the CONTENT is exactly as for the first singular personal pronoun I: the agreement features of the referential index (labelled ’[3’]) identify it as first-singular. Together, the CAT and CONTENT values explain how it is that my behaves like a determiner in some respects, and like a personal pronoun in others. The CONTEXT value, which shows that the index introduced by my is structure-shared with the SPEAKER value, provides a typical example of how indexicality is handled in HPSG. Finally, we consider the stored quantifier, with which the NP mother of my will eventually make its contribution to semantic interpretation. Of course the index bound by this quantifier is [1], the index introduced by the determiner’s N’ sister. Particular attention should be paid to the restriction on this index (the set value of the RESTR feature within the stored quantifier): it is the union of the restrictions contributed by the N’ itself ([2]) with the singleton set containing the psoa (53):

\[ \text{RELN} \quad \text{poss} \]
\[ \text{POSSESSOR} \quad 3 \]
\[ \text{POSSESSED} \quad 1 \]

The RELATION value poss should be understood here as a ‘wild card’ that will be instantiated, in a given context, as some contextually salient binary relation R (see Downing 1977); likewise, POSSESSOR and POSSESSED will be instantiated as the roles of R. And [3] indicates the index introduced by my, which as we noted is coindexed with the speaker. If my combines with (say) book (see (45)) to form the NP my book, the stored quantifier (which will be passed up to the NP by the QIP) will be as in (54):

\[ \text{DET} \quad \text{the} \]
\[ \text{INDEX} \quad 1 \]
\[ \text{RESTR} \quad \text{RELN} \quad \text{book} \quad \text{pos} \quad \text{poss} \quad 3 \]
\[ \text{POSSESSED} \quad 1 \]

We conclude this section with an analysis of possessive ‘s. The basic idea is that ‘s combines with an NP sister to form a sign that is essentially like a possessive determiner, except that (CONTENT or CONTEXT) restrictions on the possessor will now come from the NP sister. More precisely, we analyze ‘s as an unsaturated determiner that subcategorizes for a nonpronominal NP sister, as shown in (55):

\[ \text{PHON} \quad (s) \]
\[ \text{CAT} \quad \text{SPEC N} \]
\[ \text{INDEX} \quad 1 \]
\[ \text{RESTR} \quad \text{RELN} \quad \text{inst} \quad \text{book} \quad \text{poss} \quad \text{poss} \quad 3 \]
\[ \text{POSSESSED} \quad 1 \]

Now suppose ‘s takes as complement the NP Mary, whose SYNSEM[LOC] value is shown in (56):

61. By the way, (52) provides a good illustration of the fact that lexical entries, like ID schemata ((22), (23)) and principles of UG ((15), (16)), must be represented theoretically as constraints on (i.e. descriptions of) feature structures, not simply as feature structures. That is, (52) constitutes a constraint that (the feature structure models of) any token of my must satisfy; but there is no one feature structure that satisfies (52), since for different tokens of my, the set value indicated by [2] will be instantiated by different sets of psoas. By the same token, it is the description (4), not the feature structure (2), that should be thought of as the lexical entry for she; the difference is that in this case, the description is complete, so that (up to structural isomorphism) there is only one (totally well-typed, sort-resolved) feature structure that satisfies it.
In P&S-87 we considered the question of whether adjuncts select the heads that they modify or whether heads select their adjuncts. The tentative solution we adopted there assumed the latter, but as it has turned out, that solution has resisted extension to a satisfactory account of how adjuncts contribute their content to the content of the phrases they occur in. In this section we explore an alternative approach, wherein adjuncts select their heads. This approach has much in common with the analysis of adjuncts in categorial grammar, where it is assumed that adjuncts are functions that take heads as arguments. The basic idea of our analysis is that, in a head-adjunct structure, the content of the mother is token-identical to that of the adjunct; the content of the head is incorporated not by functional application but rather by structure sharing with a substructure of the head’s content.

In order to explicate this analysis, we take as our point of departure the SYNSEM value of the lexical entry for the attributive adjective red given in (58):

To enable an adjunct to select its head, we introduce the head feature MODIFIED (MOD), which is analogous to the SPEC feature employed by markers and determiners for the same purpose. The specification [PRD −] is used to distinguish attributive adjectives from predicative adjectives. For semantically restrictive adjectives like red, the content has a particularly simple form: it is just a nominal object whose index (I) coincides with that of the head noun, and whose restriction is the set of psosas obtained by adding to the restrictions imposed by the head noun one further restriction imposed by the adjective itself, in this case the psos in (59):

63. There remains the question as to which ID schema licenses signs like (57). The obvious candidate is Schema 1, but it will not apply in its present form because’s is not phrasal. Two obvious alternatives are (1) to extend Schema 1 to allow clitic as well as phrasal heads, or (2) to introduce new schemata for clitics (presumably different schemata would be required according as a given clitic is analyzed as a head, a complement, or a marker). We leave this question unresolved here.

64. For discussion, see P&S-87, pp. 64–67.
An adjunct and a head that it selects can combine via the ID schema in (60): 65  

\[(60) \quad \text{(SCHEMA 5) a phrase with DTRANS value of sort head-adjunct-structure (head-adj-struc), such that the MOD value of the adjunct daughter is token-identical to the SYNESEM value of the head daughter.} \]

Here head-adj-struc is a subsort of head-struc whose COMP-DTRANS value is specified as \( ( ) \) and that bears an additional attribute ADJ-DTRANS. For example, Schema 5 will license the N' red book with adjunct daughter red and head daughter book; the SYNESEM phrase will then be as in (61):

\[(61) \quad \begin{array}{c}
\text{CAT} \\
\text{HEAD} \\
\text{SUBCAT (DTRANS)} \\
\text{INDEX} \\
\text{CONT} \\
\text{RESTR} \\
\text{RELM \{book, [INST \{1\}, [RELM \{red, [ARG \{1\} \}]]\}}
\end{array}\]

One point that remains to be accounted for is the structure sharing of the mother's content with that of the adjunct daughter. This is a consequence of the Semantics Principle, which we now reformulate as in (62): 66

\[(62) \quad \text{SEMANTICS PRINCIPLE (second version):} \]

In a headed phrase, the CONTENT value is token-identical to that of the adjunct daughter if the DTRANS value is of sort head-adjunct-struc, and with that of the head daughter otherwise.

Various problems remain in extending this analysis of attributive adjectives to a general theory of adjuncts. One of these is the problem of ensuring that phrasal adjuncts bear the appropriate MOD value. In certain cases this can be handled by introducing the MOD feature on the adjunct's lexical head, letting the HPF carry it to the top of the adjunct. For instance, with-predicative adjunct clauses like the one in (63) can be treated this way:

\[(63) \quad \text{With Kim gone, the project fell apart.} \]

\[(64) \quad \text{* With, the project fell apart.} \]

Roughly, the idea is to treat with as the head of the adjunct, letting it select the small clause Kim gone via the SUBCAT feature and the finite clause by the MOD feature; we can block the nonsense (64) by adding to Schema 5 the requirement that the adjunct daughter be saturated ([SUBCAT ( )]). 67 Further refinements are required, however, in light of the fact that some adjuncts are not saturated. For example, with-less predicative adjuncts can occur either with or without subjects, as illustrated in (65):

\[(65) \quad \begin{array}{c}
a. \text{His hands trembling violently, Sandy loomed in the doorway.} \\
b. \text{Trembling violently, Sandy loomed in the doorway.} \\
\end{array}\]

Similar problems arise in the grammar of relative clauses, where the value on the whole relative clause for the feature MOD cannot be determined lexically, though languages where the verbal head of a relative clause bears identifying morphology (e.g. Korean) may well be analyzed in terms of lexically specified MOD values. The MOD value of a relative clause in English is intimately related to the values of the binding features SLASH and RELATIVE (REL), in ways that are discussed further in Chapters 4 and 5.

Another difficulty is presented by 'polymorphic' adjuncts that can adjoin to different kinds of heads with varying semantic effects. Consider, for example, the uses of the locative PP in Chicago in (66):

\[(66) \quad \begin{array}{c}
a. \text{A man in Chicago claims that the Axiom of Infinity is inconsistent.} \\
b. \text{Kim slept in Chicago.} \\
c. \text{In Chicago (at last), Kim slept soundly.} \\
d. \text{Kim is in Chicago.} \\
\end{array}\]

In (66a), the adjunct is a postnominal predicative modifier ('reduced relative') that restricts the index introduced by man. In (66b), it is a VP adjunct that might be analyzed as restricting an index corresponding to the location of the sleeping. In (66c), it is a sentence adjunct that might be analyzed as contributing a background psa to the effect that Kim is in Chicago. And in (66d), it is not an adjunct at all but a predicative complement to the copula. Such polymorphism might be approached in terms of multiple lexical entries, perhaps related by lexical rules, but we will not speculate further about such matters here.

1.10 Conclusion

In this chapter, we have sketched the outlines of a linguistic theory, one whose descriptive power and empirical adequacy we will examine in detail in the chapters to come. Modelling types of conceivable linguistic entities as rooted labelled graphs of a special kind—totally well-typed, sort-resolved feature

65. As discussed in Chapter 9, this can be seen as a special case of the more general principle that nonhead daughters must be saturated with respect to (nonsubject) complements.
structures—we formulate universal grammar and grammars of particular languages as a system of constraints on those feature structures. Only those feature structures that satisfy the constraints are taken to model (types of) grammatically well-formed linguistic entities. The distinction between the system of constraints and the collection of linguistic entities that satisfies it can be viewed as corresponding both to Chomsky's (1965a) distinction between I-language and E-language and to Saussure's (1916) distinction between langue and parole. Though only the latter is directly observable, only the former can be embodied as a mental computational system shared by members of a linguistic community. The constraints themselves can be classified roughly along the following lines:

(67) **Universal Grammar:**

Linguistic ontology: the inventory of universally available sorts of linguistic entities, together with a specification of their appropriate attributes and their value sorts.

Schema: a small, fixed inventory of universally available phrase types (schematic immediate dominance rules), for example, head-complement, head-adjunct, head-marker, etc.

Universal constraints on well-formed phrases: Head Feature Principle, Subcategorization Principle, etc.

**Particular Grammar:**

Lexicon: a system of lexical entries (possibly interrelated by lexical rules).

Linguistic ontology: selection from and further articulation of the universal linguistic ontology.

Schema: selection from and further specification (e.g. for the constituent order) of the universally available schemata.

Thus, what expressions (words and phrases) are well-formed in a particular language depends on the interactions among a complex system of universal and parochial constraints, which themselves must ultimately be realized in computable form in the minds of the speakers of that language.

Broadly speaking, this conception of language and of linguistic theory has much in common with the GB approach, for all that HPSG and GB differ with respect to matters of technical detail (e.g. feature structures vs. labelled trees, structure sharing vs. transformation, the formulation of specific constraints). However, the approach we take here is also properly viewed as a further development of the one pioneered in Gazdar et al. 1985, not only because of obvious influences at the technical level (e.g. numerous aspects of the feature system, or constraints like the Head Feature Principle), but more importantly because...