HPSG: Background and Basics*

Carl Pollard
Ohio State University

Ivan A. Sag
Stanford University

April 2, 1996

1 Introductory Remarks

We want to emphasize the extent to which HPSG is 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).

The phenomena with which P&S-94 are concerned are among those which 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.

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 upon the proponents of a competing framework to explicate the sense and extent to which the proposed alternative addresses the concerns of that community. For that reason, 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 phrasal types. A number of key GB principles (such as principles A, B, and C

*the first two sections of this paper are adapted from C. Pollard and I. A. Sag. 1994, Chapter 1, with updates to 'HPSG-III'.

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 valence 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 explicated), the attributes of linguistic structure in HPSG are related not by movement but rather by *structure sharing*, i.e. token identity between substructures of a given structure in accordance with lexical specifications or grammatical principles (or complex interactions between the two).[1] In common with a number of linguistic theories, then (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 operations (e.g. move-α). We will argue that, far from being a matter of indifference or mere notational variance, the derivational/nonderivational distinction has important empirical consequences.

A second essential difference between GB and HPSG has to do with the number and nature of structural levels posited. Although both theories posit multiple levels of structure, the inventory is somewhat different. A *sign* (i.e. a word or phrase, the HPSG analog of an expression in GB) is assumed to have (at least) the attributes PHONOLOGY (PHON), SYNTAX-SEMANTICS (SYNSEM), and (in the case of phrases) one or more attributes (e.g. HEAD-DTR and NON-HD-DTR) specifying the number and nature of the immediate constituents. Here PHON on the one hand and the various features for daughters on the other can be regarded as rough analogs of the GB levels PF (phonetic form) and S-structure. But the SYNSEM attribute does not correspond directly to any one level of GB syntactic structure. Rather, it in turn has (at least) three attributes of its own called CATEGORY (CAT), CONTENT (CONT), and CONTEXT. Here CAT plays a role roughly analogous to that of D-structure in GB; CONTENT, on the other hand, is concerned principally with linguistic information that bears directly on semantic interpetation (and is therefore most closely analogous to GB's level of LF (logical form)).[2] It should also be emphasized here that, unlike the situation in GB theory, where only sentences are assumed to have the levels of representation PF, LF, S-structure, and D-structure, in HPSG it is assumed that all signs, be they sentences, subsentential phrases, or words (i.e. lexical signs), have the attributes PHON and SYNSEM, and that all headed phrasal signs have the attributes HEAD-DTR and NON-HD-DTR as well.

Technical detail, of course, is what most work in HPSG consists of. Just a few salient

[1]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. (See especially the formulation in Johnson and Postal 1980 and the references cited therein).

[2]The CONTEXT attribute contains linguistic information that bears on certain context-dependent aspects of semantic interpretation.

respects in which HPSG differs from GB will be mentioned here, to give something of the flavor of the theory; all will be discussed in full in the chapters to come. Perhaps most characteristically, in HPSG tree-configurational notions such as government and c-command are not regarded as linguistically significant; instead, their role is taken over by the relation of *relative obliqueness* that obtains between syntactic dependents of the same head. For example, in HPSG the subject is defined not in terms of a D-structure configurational position, but rather as the least oblique complement of the relevant head, where relative obliqueness is modelled by position on the list which forms the ARGUMENT-STRUCTURE (ARG-ST) indicating relative obliqueness of a head's arguments. The 'valence' features SUBJ, SPR and COMPS take over the role of specifying the particular elements that the lexical head actually must combine with. Another example: in HPSG, principle A (which constrains the possible antecedents of anaphors) makes no reference to c-command or government, but merely requires 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;[3] 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.[4] To put it another way, we deny that transformations themselves model anything in the empirical domain (and therefore HPSG shares the property of 'nonderivationality' 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 SPR specification, which corresponds to the complement's unexpressed 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.[5] 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 a word's SUBJ, SPR and COMPS lists) do *not* always have to be satisfied on the surface (i.e. they do not always correspond to an actual phrase).

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.[6] Thus there is no obstacle to a 'raising-to-object' analysis of sentences like *Kim be-*

[3]But we will propose an alternative, traceless analysis in Chapter 9.

[4]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.

[5]Moreover, since passive is handled by lexical rule rather than within the syntax (see below), the necessity for an analog of NP-trace is obviated altogether.

[6]Postal and Pullum (1988) argue persuasively that this assumption, though conventional, is justified by neither empirical nor GB-internal theoretical considerations.

lieves Sandy to be happy. In HPSG this amounts to structure sharing between the matrix object and the *SUBJ* specification of the CP, VP or AP complement. 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 with an ARG-ST element that is not realized as a constituent at all (this will be true of both raising and long-distance dependencies). 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).

The other core case of head movement in GB, viz. 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 which have object agreement) occur within the corresponding ARG-ST element of the verb; and the role of the tense element of INFL is taken over by the head feature VERB-INFLECTIONAL-FORM (VFORM). Thus whether or not 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 or not 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).

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 which 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).[7]

2 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

[7] For an analogous critique of the notion of *metarule* employed in GPSG, see Pollard 1985.

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.

In our view, a linguistic theory should bear exactly the same relation to the empirical domain of natural language, viz. the universe of possible linguistic objects, as a mathematical theory of celestial mechanics should bear to the possible motions of n-body systems. Thus we insist on being explicit as to what types 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 actually count as a theory in the technical sense of precisely characterizing those modelling structures which are regarded as admissible or well-formed (i.e. corresponding to those imaginable linguistic objects which 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 can not have any determinate empirical consequences.

We emphatically reject the currently widespread view which holds that linguistic theory need not be formalized. 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 *typed feature structures* (Moshier 1988, Pollard and Moshier 1990, King 1989, 1994), which 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 which correspond to the admissible feature structures constitute the predictions of the theory.

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 posited to account for the observed variety of subatomic particles. But the parsimony principle with respect to nonobservable constructs 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 various daughters attributes, e.g. HEAD-DTR, NON-HD-DTRS) 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 PF, HPSG's PHONOLOGY), logical predicate-argument structure (GB's LF, HPSG's CONTENT), or underlying grammatical relations (GB's D-Structure, HPSG's ARG-ST attribute, LFG's f-structure). But for all that a theory that successfully dispensed with a notion of surface constituent structure is to be preferred (other things being equal, of course), the explanatory power of such a notion is too great for many syntacticians to be willing to relinquish it.

But if phrase structures are current syntactic theory's quarks, move-α – as Koster (1987) has remarked – might well be regarded as its phlogiston. As we hope to have made clear by now, we regard transformational operations between levels as constructs that are not motivated by empirical considerations. What we observe, albeit indirectly, is sharing of certain subparts (e.g. between a filler and a gap, between an anaphor and a binder, between an 'understood' subject and a controller). But such sharing is straightforwardly and neutrally accounted for as simple identity; attributing it to derivational processes at best contributes nothing to the theory, and at worst introduces complications and confusions (e.g. ordering paradoxes) of a completely artifactual nature.[8]

There is a further condition of *decidability* that we impose upon a linguistic theory. That is, we require that for a substantial fragment of candidate expressions (i.e. expressions and non-expressions) for a given language under study, it must be determinable by algorithm whether each candidate expression is assigned a well-formed structure by the theory, and if so what that structure is. The condition of decidability is the theory's reflection of two fundamental facts about language use: first, the structures of linguistic expressions are capable in principle of being computed by the resource-bounded information-processing organisms which successfully employ them in a communicative function; and 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).

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 are 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

[8] For further arguments in support of the view that grammars should be formulated as declarative systems of constraints rather than derivational processes, see Johnson and Postal 1980 and Langendoen and Postal 1984.

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 which provides linguistic descriptions that can be shown to be incrementally processable should be regarded as superior to one which 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 modification of the prepositional phrase *on the atom* to be determined well before word-by-word processing of a sentence like (1) is complete.[9]

(1) 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 (2), early accessing of morphological 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.[10]

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

In (3), 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 .

(3) 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 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

[9]Example (1) is an adaptation of an example of Graeme Hirst's (see Hirst 1987).

[10]We owe this sort of example to Martin Kay.

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 produceable 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 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 that stray very far from the grammar of their language run 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.[11] 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

[11]A similar point is made by Bresnan and Kaplan (1982). See also Halvorsen 1983, Sag et al. 1985, and Fenstad et al. 1987.

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 constrast, 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 developed within HPSG exhibit no biases toward one mode of processing rather than another. Because each partial linguistic description is to be viewed denotatively, i.e. 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 order a process may dictate – the constructs are all constraints which, by their very nature, are order-independent and which 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, viz. 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.

3 Some Ontology

To get started, consider the set of linguistic types listed in (4).[12]

(4) *sign, word, phrase, category, head (= part-of-speech), list(σ), set(σ), content, case, index, verb-form*

Note that none of these as an atomic type, but rather corresponds to a general classification that has more specific instances. The theory of grammar has to specify exactly what other

[12]Here σ is a variable ranging over those types which may give rise to lists or sets within our theory.

types are classified by these types, i.e. which types are the immediate subtypes of these types, what additional types are subtypes of these immediate subtypes, and so forth. Such an ontology can be specified by a set of partitions like the following:

(5)

Type	Immediate Subtypes
sign	*word*, *phrase*
phrase	*headed-phrase*, *non-headed-phrase*
list(σ)	*nelist*(σ), *elist*
set(σ)	*neset*(σ), *eset*
content	*relation*, *indexed-obj*
relation	*give-rel*, *walk-rel*,...
head	*noun*, *verb*, *adj*, *prep*, ...
case	*nom*, *acc*
index	*ref*, *non-ref*

non-ref(erential) will be the type of index associated with expletive NPs ('dummies') like *there* and *it*; *indexed-obj* is the type used for the semantics of nominals in general; *nelist* stands for *nonempty-list*, *elist* for *empty-list* ($\langle\,\rangle$), *neset* for *nonempty-set*, and *eset* for *empty-set* ($\{\,\}$). Though the sketch in (5) is preliminary and incomplete (for example, we are simplifying the semantics for the moment, ignoring quantification), in a fully developed grammar, we will define an *atomic* type as one that has no subtypes.

And these types will also participate in feature declarations, which specify the attributes that are appropriate for particular kinds of linguistic objects and what type of value those attributes must take. Here are some examples:

(6)

sign	PHONOLOGY *list*(*phonstring*), SYNSEM *synsem*
headed-phrase	HEAD-DTR *sign*, NON-HD-DTRS *list*(*sign*)
list(σ)	FIRST σ, REST *list*(σ)
walk-rel	WALKER *ref*
give-rel	GIVER *ref*, RECIPIENT *ref*, GIFT *ref*
indexed-obj	INDEX *ref*, RESTRICTION *set*(*relation*)
noun	CASE *case*
verb	VFORM *vform*
index	PER *per*, NUM *num*, GEND *gend*

We are almost ready to illustrate some simple signs, but first we need to introduce a few additional types, those for the information complexes we call *synsem* objects and *local* objects. These types represent hypotheses about what kind of information is available for valence selection and for transmission within unbounded dependency (filler-gap) constructions. It is a consequence of the organization of current HPSG theory that a head selects for a specifier, complement, or subject only in terms of its *synsem* object. Likewise the architecture of the constraints on the relation between fillers and gaps allow only *local* information to be referenced. Thus the very organization of feature structures – the geometry of the sign, will embody hypotheses of considerable substance about the nature of specific linguistic phenomena.

4 Words

A word is a particular kind of feature structure which, if presented in all its detail, would be described in the fashion of (7).

(7)
$$\left[\begin{array}{ll} \textit{word} & \\ \text{PHON} & \langle\text{chased}\rangle \\ \text{SYNSEM} & \left[\begin{array}{ll} \textit{synsem} & \\ \text{LOCAL} & \left[\begin{array}{ll} \textit{local} & \\ \text{CATEGORY} & \left[\begin{array}{ll} \text{HEAD} & \left[\begin{array}{l}\textit{verb}\\ \text{VFORM } \textit{fin}\end{array}\right] \\ \text{SUBJ} & \left\langle \boxed{1}\left[\begin{array}{ll}\textit{synsem} & \\ \text{LOC}|\text{CAT}|\text{HEAD} & \left[\begin{array}{l}\textit{noun}\\ \text{CASE } \textit{nom}\end{array}\right]\end{array}\right]\right\rangle \\ \text{COMPS} & \left\langle \boxed{2}\left[\begin{array}{ll}\textit{synsem} & \\ \text{LOC}|\text{CAT}|\text{HEAD} & \left[\begin{array}{l}\textit{noun}\\ \text{CASE } \textit{acc}\end{array}\right]\end{array}\right]\right\rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{array}\right] \\ \text{CONTENT} & \ldots \\ \text{CONTEXT} & \left[\begin{array}{ll} \text{CONX-INDICES} & \left[\begin{array}{l}\text{SPEAKER} \ldots\\ \text{HEARER} \ldots\end{array}\right] \\ \text{BACKGROUND} & \ldots \end{array}\right] \end{array}\right] \\ \text{NONLOCAL} & \left[\begin{array}{l}\text{SLASH } \{\ \}\\ \ldots\end{array}\right] \end{array}\right] \end{array}\right]$$

But since these descriptions can quickly become unwieldy, we will systematically simplify them. A few abbreviated entries are sketched in what follows. Bear in mind, though, that however partial or abbreviated the descriptions might be, given our modelling assumptions, they always designate a family of fully specified feature structures.

(8) walks
$$\left[\begin{array}{ll} \text{CAT} & \left[\begin{array}{ll} \text{HEAD} & \left[\begin{array}{ll}\textit{verb} & \\ \text{VFORM} & \textit{fin}\end{array}\right] \\ \text{SUBJ} & \langle\ \boxed{2}\text{NP}\left[\textit{nom}\right]_{\boxed{1}[\textit{3sing}]}\ \rangle \\ \text{COMPS} & \langle\ \rangle \\ \text{ARG-ST} & \langle\boxed{2}\rangle \end{array}\right] \\ \text{CONTENT} & \left[\begin{array}{ll}\textit{walk-rel} & \\ \text{WALKER} & \boxed{1}\end{array}\right] \end{array}\right]$$

(9) give

$$\begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{verb} & \\ \text{VFORM} & \textit{fin} \end{bmatrix} \\ \text{SUBJ} & \left\langle \text{NP}[\textit{nom}]_{\boxed{1}[non-3sing]} \right\rangle \\ \text{COMPS} & \left\langle \text{NP}[\textit{acc}]_{\boxed{2}}, \text{NP}[\textit{acc}]_{\boxed{3}} \right\rangle \end{bmatrix} \\ \text{CONTENT} & \begin{bmatrix} \textit{give-rel} & \\ \text{GIVER} & \boxed{1} \\ \text{RECIPIENT} & \boxed{2} \\ \text{GIFT} & \boxed{3} \end{bmatrix} \end{bmatrix}$$

(10) a. chases b. picture c. of

$$\begin{bmatrix} \text{HEAD} & \text{verb [fin]} \\ \text{SUBJ} & \langle \boxed{1}\text{NP[nom]}_{3s} \rangle \\ \text{COMPS} & \langle \boxed{2}\text{NP} \rangle \\ \text{ARG-S} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \quad \begin{bmatrix} \text{HEAD} & \text{noun} \\ \text{SPR} & \langle \boxed{1}\text{Det} \rangle \\ \text{COMPS} & \langle (\boxed{2}\text{PP[of]}) \rangle \\ \text{ARG-S} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \quad \begin{bmatrix} \text{HEAD} & \text{prep[of]} \\ \text{SUBJ} & \langle\ \rangle \\ \text{COMPS} & \langle \boxed{1}\text{NP} \rangle \\ \text{ARG-S} & \langle \boxed{1} \rangle \end{bmatrix}$$

Note that here the SUBJ and COMPS lists 'add up' to the list value of the ARGUMENT-STRUCTURE (ARG-ST). ARG-ST values correspond to the hierarchical argument structure of a word (relevant, for example, to binding theory – see Pollard and Sag (1992)), while the valence features specify the word's combinatoric potential.

Lexical entries such as these contain much information that can in fact be consolidated within an explanatory theory of lexical structure and organization. Indeed, considerable research within HPSG has been concerned with the development of just such theories, namely those which allow complex lexical information to be factored in various ways to reflect appropriate linguistic generalizations. Central to this line of inquiry has been the concept of hierarchical classification – essentially an assignment of words to categories, and an assignment of those categories to superordinate categories. With each category (or *type*), certain attributes are specified to be appropriate and certain constraints are stated that hold for all members of that category. Without stipulation, a word inherits all the features and constraints of the type it is assigned to and, via the technique of *hierarchical inheritance*, all such features and constraints declared for supertypes of that type are also associated with the word in question.

Because particular words (lexical feature structures) are multiply classified, i.e. have more than one immediate supertype, it is possible to express cross-cutting generalizations about words in an elegant, deductive fashion. (See, e.g. Flickinger 1987, Flickinger and Nerbonne 1992, Riehemann 1993, Davis 1996).

Consider the three verbs in (11):

(11) a. chases b. continues c. dies

$$\begin{bmatrix} \text{HEAD} & \text{verb [fin]} \\ \text{SUBJ} & \langle \boxed{1}\text{NP[nom]}_{3s} \rangle \\ \text{COMPS} & \langle \boxed{2}\text{NP} \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \quad \begin{bmatrix} \text{HEAD} & \text{verb[fin]} \\ \text{SUBJ} & \langle \boxed{1}\text{NP} \rangle \\ \text{COMPS} & \langle \boxed{2}\text{VP[inf, SUBJ } \langle \boxed{1} \rangle] \rangle \\ \text{ARG-ST} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \quad \begin{bmatrix} \text{HEAD} & \text{verb [fin]} \\ \text{SUBJ} & \langle \boxed{1}\text{NP[nom]}_{3s} \rangle \\ \text{COMPS} & \langle\ \rangle \\ \text{ARG-ST} & \langle \boxed{1} \rangle \end{bmatrix}$$

None of the feature specifications indicated in (11) needs to be stipulated ad hoc. Words are assigned to types that are subordinate to various others. With each type come certain constraints stating general properties that are true of all elements belonging to that type. Thus by establishing hierarchical relations among types, an individual word inherits all properties (constraints) associated with all its types and all supertypes of those types[13]

The lexical descriptions in (11) are thus a logical consequence of the appropriate type classification of English verbs, which might appear as in (12).

(12)

TYPE	CONSTRAINTS	ISA
verb	$\begin{bmatrix} \text{HEAD} & \textit{verb} \\ \text{SUBJ} & \langle \text{NP} \rangle \end{bmatrix}$	*word*
tran-verb	[COMPS ⟨NP,...⟩]	*verb*
subj-raising	$\begin{bmatrix} \text{SUBJ} & \langle \boxed{1} \rangle \\ \text{COMPS} & \langle \text{XP[SUBJ } \langle \boxed{1} \rangle], \ldots \rangle \end{bmatrix}$	*verb*
strict-intran-verb	[COMPS ⟨ ⟩]	*verb*
obj-raising verb	$\begin{bmatrix} \text{COMPS} & \langle \boxed{1}, \text{XP[SUBJ } \langle \boxed{1} \rangle] \rangle \end{bmatrix}$	*tran-verb*
strict-tran-verb	[COMPS ⟨X⟩]	*tran-verb*
finite-verb	$\begin{bmatrix} \text{HEAD} & [\text{VFORM } \textit{fin}] \\ \text{SUBJ} & \langle \text{NP}[\textit{nom}] \rangle \end{bmatrix}$	*verb*
3rd-person-verb	[SUBJ $\langle \text{NP}_{3sg} \rangle$]	*finite-verb*
base-verb	[HEAD [VFORM *base*]]	*verb*
passive-verb	[HEAD [VFORM *pass*]]	*verb*

The type names at the end of each line in (12) specify 'is a' relations among the types, i.e. they indicate each type's immediately superordinate type(s). The resulting inheritance hierarchy thus allows the particular properties of the lexical entries in (11) to be derived, i.e. deduced, from the type assignments in (13):

(13) chases: *strict-tran-verb* & *3rd-person-verb*
continues: *subj-raising* & *3rd-person-verb*
dies: *strict-intran-verb* & *3rd-person-verb*

Thus a 3rd-person-verb form like *chases* is assigned to two distinct atomic types *strict-transitive-verb* and *3rd-person-verb*, each of which specifies a different subset of the information that *chases* inherits, as shown in (13). Multiple inheritance is thus an essential feature

[13] Up to consistency. I will assume here without argument that subordinate conflicting constraints 'override' more general superordinate constraints.

of lexical organization in a theory like HPSG. It is a fundamental mechanism for expressing common properties of lexical items that are divergent in other respects.

Further generalizations about lexical entries are expressed by lexical rules.[14] As in early work in LFG (Bresnan, ed. 1982), these systematically expand the set of basic (or 'canonical') lexical entries, specifying only particular noncanonical properties that hold of the output forms. Among these is the passive lexical rule, sketched in (14).

(14) Passive Lexical Rule (PLR):

$$\begin{bmatrix} \textit{tran-verb} & \\ \text{SUBJ} & \langle \text{NP}_i \rangle \\ \text{COMPS} & \langle \boxed{2}, \ldots \rangle \end{bmatrix} \Rightarrow \begin{bmatrix} \textit{passive-verb} & \\ \text{SUBJ} & \langle \boxed{2} \rangle \\ \text{COMPS} & \langle \ldots\ (,\text{PP}_i) \rangle \end{bmatrix}$$

Within a lexical rule, all properties of the input (e.g. semantic role assignment) that are not explicitly modifed remain unchanged in the corresponding output. Thus, in virtue of (14), the base form of the lexeme *chase* (looks similar to (11)a above (but is *base* instead of *finite*) gives rise to the appropriately specified passive form *chased*:

(15) chased

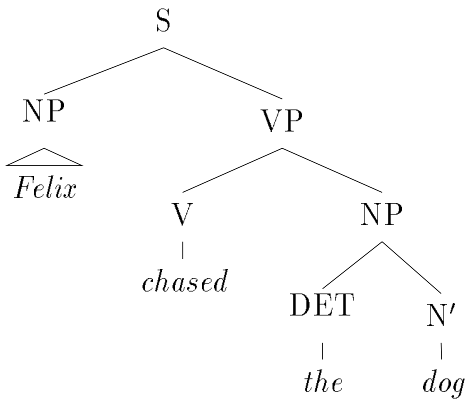
$$\begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{verb} & \\ \text{VFORM} & \textit{pass} \end{bmatrix} \\ \text{SUBJ} & \langle \boxed{2}\text{NP} \rangle \\ \text{COMPS} & \langle (\boxed{3}\text{PP}[\textit{by}]) \rangle \\ \text{ARG-ST} & \langle \boxed{2}, \boxed{3} \rangle \end{bmatrix}$$

This form may then serve as the lexical head of a passive verb phrase, e.g. *chased by the police*. This follows from the interaction of the lexicon, the *head-complement-phrase* type, the HFP and the Valence Principle (see next sections) without the need for any passive-specific machinery.

5 Phrases and Schemata

For expository purposes, HPSG is often presented in terms of the familiar trappings of generative grammar, where syntactic rules or schemata are formal devices that 'generate' word-terminated structures like (16):

[14]For recent attempts to eliminate lexical rules in HPSG in favor of a hierarchically organized theory of morphological structure, see Riehemann 1993, ms.; Kim 1995; Kathol to appear; and Malouf 1994.

(16) 

But this presentation is in fact a distortion of HPSG, where phrases are treated in essentially the same way as words, as feature structures that serve as models of utterance types. The most fundamental type of utterance recognized in HPSG is the *sign*, with its two immediate subtypes: *word* and *phrase*. So, just as lexical entries are descriptions of (or constraints on) feature structures of type *word*, schemata are descriptions of feature structures of type *phrase*. And parochial and universal principles are just further descriptions, i.e. additional constraints that the phrases of the language in question must satisfy.

Word tokens are feature structures like those described below, where certain phonological, morphological, syntactic and semantic features are hypothesized and organized according to a particular feature geometry. Phrases, likewise, have appropriate features of their own, as well as their own feature geometry. The feature geometry of phrases corresponds to phrase structure, as standardly conceived, but the feature-based description of phrase structure looks somewhat different from the familiar presentation. Compare the tree structure in (16) with the feature structure description in (17):

(17)
$$
\begin{bmatrix}
\textit{hd-spr-ph} & \\
\text{PHON} & \langle \text{Felix,chased,the,dog} \rangle \\
\text{SYNSEM} & \text{`S'} \\
\text{NON-HD-DTRS} & \left\langle \begin{bmatrix} \text{PHON} & \langle \text{Felix} \rangle \\ \text{SYNSEM} & \text{`NP'} \end{bmatrix} \right\rangle \\
\text{HEAD-DTR} & \begin{bmatrix}
\textit{hd-comp-ph} & \\
\text{PHON} & \langle \text{chased,the,dog} \rangle \\
\text{SYNSEM} & \text{`VP'} \\
\text{HEAD-DTR} & \begin{bmatrix} \textit{word} \\ \text{PHON } \langle \text{chased} \rangle \end{bmatrix} \\
\text{NON-HD-DTRS} & \left\langle \begin{bmatrix} \text{PHON} & \langle \text{the,dog} \rangle \\ \text{SYNSEM} & \text{`NP'} \end{bmatrix} \right\rangle
\end{bmatrix}
\end{bmatrix}
$$

It may not be obvious whether there is any significant difference between the two conceptions of linguistic structure. However, there are several noteworthy advantages to this 'sign-based' approach that makes phonology and semantics derivative of antecedently generated syntactic structures. First, the tree-based conception of phrase structure is a special case of the sign-based approach – one that uses only concatenation to relate the PHON values of mother and daughters. But generalizing such operations to include wrapping[15] or other operations that permit interleaving (e.g. Reape's *sequence union* operation (Reape 1994, in press)) has proven to be an interesting and promising aproach to the analysis of many problems of word order variation, extraposition, and coordination that have proved challenging for purely concatenative approaches.[16] Second, the explicit role of heads, subjects and complements in the theory provides hybrid structural-functional data structures that allow constraints about linear order, feature 'percolation', etc. to be stated in a uniform way without the introduction of ancillary mechanisms. Third, the bundling of syntactic, semantic and even contextual information into each SYNSEM value makes such information ubiquitous in phrase structure. This flexible access to contextual information is of considerable value, e.g. in the treatment of focus placement and focus inheritance, as demonstrated by Engdahl and Vallduvi (1994), who exploit this crucially in explaining differences between the focus systems of, *inter alia*, English and Catalan. Finally, since the sign-based approach involves hierarchical classification of phrases, it is possible to encode previously unexpressible generalizations about phrasal signs using the same multiple inheritance techniques that have proven useful in the analysis of lexical signs (see below).

6 Universal Grammar

HPSG is thus a constraint-based theory of grammatical competence. All of its representations – lexical entries, rules, and even universal principles – are partial descriptions of (or constraints on) constructs used to model types of linguistic utterances.[17] Hence HPSG linguistic descriptions are declarative, order-independent, and reversible, making them ideally suited for the description of linguistic performance.

Many of the central constructs of HPSG are motivated by its adherence to strict lexicalism, a thesis that entails that syntactic operations cannot operate on or make reference to internal properties of lexical items. Any lexically based theory necessarily employs rich lexical representations and HPSG's UG is a small set of principles that allow the grammar of phrases to be projected from the particular information encoded in lexical heads. One might think of the core of HPSG theory as an attempt to simplify both grammatical structures and their grammar, deriving the effects equivalent to those of head movement, functional categories and the projection principle all from the interaction of X'-theory and strict lexicalism.

All X'-theories embody some variant of the following principle, whose specific formulation presumes that HEAD is a feature taking a feature structure complex as its value:

(18) The Head Feature Principle (HFP)

The HEAD value of a headed phrase is identified with that of its head-daughter.

[15]Various kinds of wrapping have been investigated. See Bach 1979 and Pollard 1984, inter alia.

[16]See Reape 1994, Kathol and Levine 1992, and Kathol 1995, for example.

[17]The idea that such a uniform characterization of linguistic theory is possible is due to Martin Kay.

This familiar principle guarantees that certain grammatical properties, e.g. part-of-speech, case, and form class, are systematically projected onto X'-phrases from lexical items, and from X'-phrases onto maximal phrases. The HEAD value of a word thus contains only information that phrasal projections inherit in virtue of the HFP.

As in Categorial Grammar, phrase maximality is described not in terms of bar level, but rather via combinatoric saturation. That is, a lexical entry bears certain specifications that determine what elements it combines with syntactically. Such specifications are stated in terms of the valence features SUBJ (SUBJECT), COMPS (COMPLEMENTS), and SPR (SPECIFIER).[18] A headed phrase is well-formed only if it satisfies the following principle:

(19) The Valence Principle (VALP)

The F value of a Head-F-Phrase (F = Subj, Spr, or Comps) is the head-daughter's F value minus (the *synsems* of) the realized non-head-daughters. For all other valence features G, the phrase's G value is the head daughter's G value.

The Valence Principle thus plays a role within HPSG much like that of the category cancellation associated with function application in Categorial Grammar. Although such principles are often described informally in terms of a bottom-up phrase generation procedure, notice that (19) is a static constraint on headed phrases.

Universal grammar makes available a small set of phrase types which specify partial information about universally available kinds of phrases. As in GPSG, these types not only abstract away from the principles of X' theory just enumerated, but also from the order of daughter elements, leaving such matters to more general constituent ordering principles. Three types of relevance are illustrated in (20).

(20) a. Head-Specifier-Phrase:

$$X \rightarrow \underset{[\textit{phrase}]}{\text{Head-Dtr}}, \quad \text{Specifier-Dtr}$$

b. Head-Complement-Phrase:

$$X \rightarrow \underset{[\textit{word}]}{\text{Head-Dtr}}, \quad \text{Comp-Dtrs}$$

c. Head-Subject-Phrase:

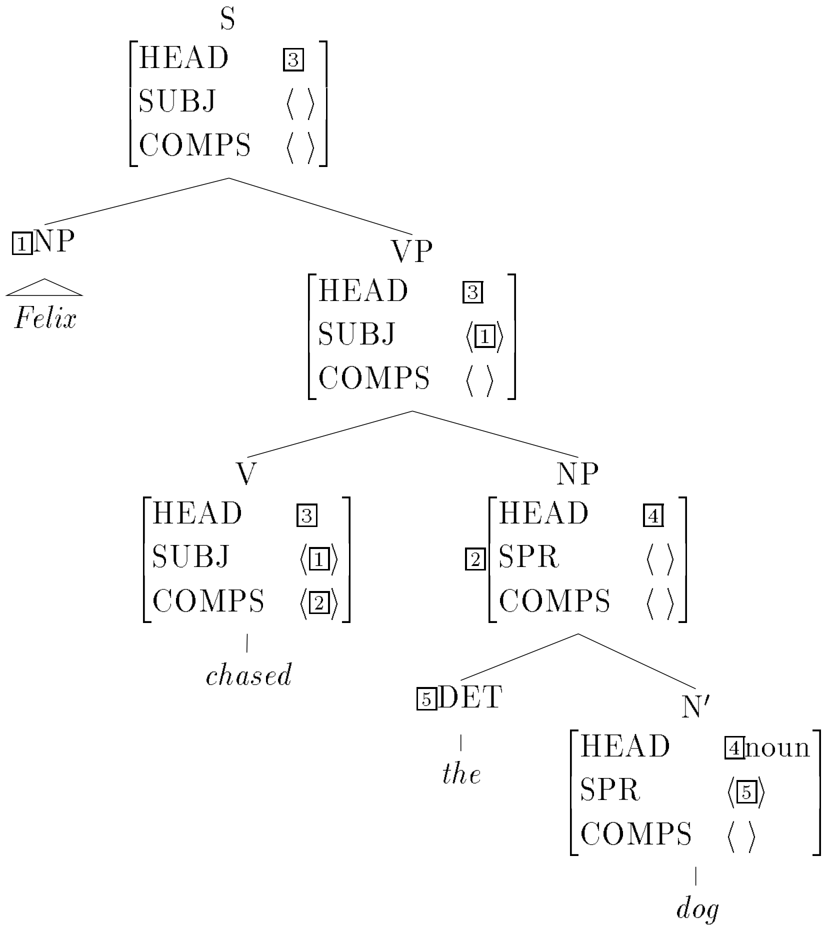
$$X \rightarrow \underset{[\textit{phrase}]}{\text{Head-Dtr}}, \quad \text{Subj-Dtr}$$

Instances of (20a) consist of a phrasal head daughter and a specifier daughter; the phrases that are instances of (20b) consist of a lexical head daughter and any number of complement daughters; instances of (20c) consist of a phrasal head daughter and a subject daughter. Because of X' theory, the head daughter's HEAD information is maximally projected in

[18] This follows innovations in HPSG theory due to Robert Borsley (1989), specifically as adapted by Pollard and Sag (1994: chap. 9).

any given phrase (by the HFP) and the head's valence information determines the elements that the maximal projection contains (in accordance with the VALP). Thus each embedded phrase in the following structure (shown here in familiar subtree format) is an instance of one of the phrasal types and obeys all of the principles of UG:

(21)



The boxed integers in these tree diagrams are variables used to 'tag' certain feature values within the structure as being token identical, as required by the **HFP** or the **VALP**. Thus the part-of-speech information (tagged 4 in (21)) specified in the lexical entry for *dog*, is identified with that of the NP it projects, in accordance with the HFP. (The same would be true for CASE specifications in a language whose nouns were systematically inflected for case.) In like fashion, the lexical entry for *chased* specifies the part-of-speech *verb*, which the HFP ensures is also the part-of-speech of the VP and the S. (The reader should continue to bear in mind that the tree structure shown in (21) is used solely for expository convenience. We're all used to thinking in terms of phrase structure trees, after all. The tree in (21) depicts (albeit in more detail) the very same phrasal sign that we illustrated earlier in feature structure notation.)

The lexical entry for *chased* selects for an NP complement and hence may combine with the phrase *the dog* (whose grammatical information (tagged 2) is identified with the complement selected by *chased*) to form a *head-complement-phrase*. *Chased* similarly selects lexically for an NP subject, and this specification is also part of the VP (in accordance with the Valence Principle). Hence this VP combines with the subject NP to form a *head-subject-phrase*. We refer to such phrases, i.e. to those phrases all of whose valence specifications are empty as *saturated*.

References

Bach, Emmon. 1979. Control in Montague Grammar. *Linguistic Inquiry* 10: 515–531.

Borsley, Robert. 1989. Phrase-Structure Grammar and the *Barriers* Conception of Clause Structure. *Linguistics* 27: 843–863.

Bresnan, Joan, ed. 1982. *The Mental Representation of Grammatical Relations.* Cambridge, MA.: MIT Press.

Bresnan, Joan and Ronald M. Kaplan. 1982. Introduction. In Bresnan, ed. 1982a. Pp. xvii-lii.

Chomsky, Noam. 1957. *Syntactic Structures.* The Hague: Mouton.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax.* Cambridge, MA: MIT Press.

Chomsky, Noam. 1981. *Lectures on Government and Binding.* Dordrecht: Foris.

Chomsky, Noam. 1982. *Some Concepts and Consequences of the Theory of Government and Binding.* Cambridge, MA: MIT Press.

Davis, Anthony. in progress. *Linking and the Hierarchical Lexicon.* Doctoral dissertation, Stanford University.

Engdahl, Elisabet and Enric Vallduvi. 1994. Information Processing in Constraint-Based Grammars. Unpublished paper: University of Edinburgh.

Fenstad, Jens-Erik, Per-Kristian Halvorsen, Tore Langholm and Johan van Benthem. 1987. *Situations, Language, and Logic*. Dordrecht: Reidel.

Flickinger, Daniel. 1987. *Lexical Rules in the Hierarchical Lexicon.* Doctoral dissertation, Stanford University.

Flickinger, Daniel P. and John Nerbonne. 1992. Inheritance and Complementation: A Case Study of *Easy* Adjectives and Related Nouns. *Computational Linguistics* 18: 269–309.

Halvorsen, Per-Kristian: 1983. Semantics for Lexical-Functional Grammar. *Linguistic Inquiry* 14: 567–616.

Hirst, Graeme. 1987. *Semantic Interpretation and the Resolution of Ambiguity*. Cambridge: Cambridge University Press.

Johnson, David and Paul Postal. 1980. *Arc Pair Grammar.* Princeton: Princeton University Press.

Kathol, Andreas. 1995. *Linearization-Based German Syntax*. Doctoral dissertation. Ohio State University.

Kathol, Andreas. to appear. Agreement and the Syntax-Morphology Interface. In Ohio State University Working Papers in Linguistics, Volume 46. Columbus: OSU.

Kathol, Andreas, and Robert Levine. 1993. Inversion as a Linearization Effect. In *Proceedings of NELS 23*. GLSA, Amherst.

Kim, Jong-Bok. 1995. The Grammar of Negation: A Lexicalist, Constraint-Based Perspective. Doctoral dissertation, Stanford University.

King, Paul. 1989. *A Logical Formalism for Head-Driven Phrase Structure Grammar*. Doctoral dissertation. University of Manchester.

King, Paul. 1994. An expanded logical formalism for Head-Driven Phrase Structure Grammar. Arbeitspapiere des sfb 340. University of Tübingen.

Koster, Jan 1987. *Domains and Dynasties*. Dordrecht: Foris.

Langendoen, D. Terrence, and Paul Postal. 1984. *The Vastness of Natural Language*. Oxford: Basil Blackwell.

Malouf, Robert. 1994. Some Theoretical Consequences of Potawatomi Morphology. Paper presented at the First International Conference on Head-Driven Phrase Structure Grammar. Copenhagen.

Moshier, Drew. 1988. *Extensions to Unification Grammars for the Description of Programming Languages*. Doctoral dissertation, University of Michigan, Ann Arbor.

Pollard, Carl. 1984. *Generalized Context-Free Grammars, Head Grammars and Natural Language*. Doctoral dissertation, Stanford University.

Pollard, Carl and Drew Moshier. 1990. Unifying Partial Descriptions of Sets. *Information, Language and Cognition*, Volume 1 of *Vancouver Studies in Cognitive Science*. University of British Columbia Press, Vancouver. PAGES??

Pollard, Carl and Ivan A. Sag. 1992. Anaphors in English and the Scope of Binding Theory. *Linguistic Inquiry* 23: 261–303.

Pollard, Carl and Ivan A. Sag. 1987. *Information-Based Syntax and Semantics, Vol. 1: Fundamentals.*. CSLI Lecture Notes Series. Stanford: CSLI Publications.

Pollard, Carl and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press and Stanford: CSLI Publications.

Pollard, Carl. 1985. Phrase Structure Grammar without Metarules. *Proceedings of the Fourth West Coast Conference on Formal Linguistics*, eds. Jeffrey Goldberg, Susannah MacKaye and Michael Wescoat. Stanford: Stanford Linguistics Association. Pp. 246–261.

Postal, Paul and Geoffrey K. Pullum. 1988. Expletive Noun Phrases in Subcategorized Positions. *Linguistic Inquiry* 19: 635–670.

Reape, Mike. 1990. A Theory of Word Order and Discontinuous Constituency in West Continental Germanic. In Elisabeth Engdahl and Mike Reape, eds., *Parametric Variation in Germanic and Romance: Preliminary Investigations.* ESPRIT Basic Research Action 3175 DYANA, Deliverable R1.1.A.. Edinburgh: Center for Cognitive Science and Department of Artificial Intelligence, 25–39.

Reape, Mike. In press. Getting Things in Order. In A. Horck and Harry Bunt, eds., *Discontinuous Constituency*. Berlin: Mouton de Gruyter.

Riehemann, Susanne. 1993. *Word Formation in Lexical Type Hierarchies: A Case Study of* **bar**-*Adjectives in German*. M.A. thesis, University of Tübingen.

Riehemann, Susanne. ms.. Type-Based Morphology. Unpublished manuscript, Stanford University.

Sag, Ivan A., Ronald Kaplan, Lauri Karttunen, Martin Kay, Carl Pollard, Stuart Shieber and Annie Zaenen. 1985. Unification and Grammatical Theory. *Proceedings of the Fifth West Coast Conference on Formal Linguistics*, eds. Mary Dalrymple, Jeffrey Goldberg, Kristin Hanson, Michael Inman, Chris Piñon, and Stephen Wechsler. Stanford: Stanford Linguistics Association. Pp. 238-254.