

# 7

## Phrase structure grammar

James P. Blevins and Ivan A. Sag

### 7.1 Origins of phrase structure analysis

To understand the properties of modern phrase structure grammars, it is useful to place their development in a wider formal and historical context. Phrase structure grammars and associated notions of phrase structure analysis have their proximate origins in models of Immediate Constituent (IC) analysis. Although inspired by the programmatic syntactic remarks in Bloomfield (1933), these models were principally developed by Bloomfield's successors, most actively in the decade between the publication of Wells (1947) and the advent of transformational analyses in Harris (1957) and Chomsky (1957). The central intuition underlying models of IC analysis was that the structure of an expression could be exhibited by dividing the expression into parts (its immediate constituents), further subdividing these parts, and continuing until syntactically indivisible units were obtained. This style of analysis was motivated in part by a belief in the LOCALITY of syntactic relations, in particular the view that the most important relations held between immediate constituents.

The process of analyzing syntax is largely one of finding successive layers of ICs and of immediate constructions, the description of relationships which exist between ICs, and the description of those relationships which are not efficiently described in terms of ICs. The last is generally of subsidiary importance; most of the relationships of any great significance are between ICs. (Gleason 1955:133)

Within the Bloomfieldian tradition, there was a fair degree of consensus regarding the application of syntactic methods as well as about the analyses associated with different classes of constructions. Some of the general features of IC analyses find an obvious reflex in subsequent models of analysis. Foremost among these is the idea that structure involves a part-whole relation between elements and a larger superordinate unit,

rather than an asymmetrical dependency relation between elements at the same level. The Bloomfieldians' preference for binary branching analyses likewise reemerges in later models of phrase structure, and their practice of extending syntactic analysis below the word level, to include stems and inflectional formatives, survives largely intact in the transformational tradition. Some other features of IC analyses are less faithfully preserved. These include general properties such as the recognition of discontinuous and overlapping constituents or the representation of intonation.<sup>1</sup> More specific proposals, such as the classification of elements (notably coordinating conjunctions) as 'markers' (Hockett 1958:153) were not rehabilitated until nearly thirty years later (Gazdar *et al.* 1985, Sag *et al.* 1985, Pollard and Sag 1994: Chapter 1). The encoding of dependency relations within a part-whole analysis (Nida 1966) was also suppressed until the development of feature-based models such as Lexical-Functional Grammar (LFG) (Kaplan and Bresnan 1982 and Chapter 6 of this volume) and Head-driven Phrase Structure Grammar (HPSG) (Pollard and Sag 1987 and Section 7.3.1 below) that could explicitly express valence dependencies within syntactic representations.

### 7.1.1 Procedures of IC analysis

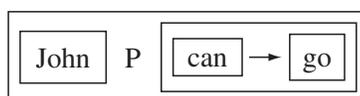
The development of constituent structure analysis within the Bloomfieldian tradition was held back by, among other things, the lack of a perspicuous format for representing syntactic analyses. The formats explored by the Bloomfieldians were cumbersome, ranging from annotated circuit diagrams in Nida (1966) through the chart representation in Table 7.1 or the 'Chinese box' arrangements in Table 7.2.<sup>2</sup>

The shortcomings of these representational formats were particularly evident in the analysis of the discontinuous and overlapping constituents recognized by the Bloomfieldians. While generally preferring continuous (and binary) analyses, they also admitted a range of constructions that violated these preferences.

**Table 7.1.** Chart-based IC Analysis (Hockett 1958: Chapter 17)

John	is	here

**Table 7.2.** Chinese box-based IC Analysis (Gleason 1965:157)





successive segmentation of an expression into sub-expressions, each of which was annotated with a word class label and, usually, other types of information. It was not until the early transformational accounts that IC analyses were incorporated into explicit grammar formalisms rather than treated as procedures of classification, and the fact that these procedures were first formalized by the Bloomfieldians' successors had the effect of simplifying them, much as the Bloomfieldians had themselves simplified Bloomfield's more intricate constructional perspective (Manaster-Ramer and Kac 1990). In Chomsky (1956), phrase structure grammars are proposed as "the form of grammar [that] corresponds to [the] conception of linguistic structure" expressed by IC analysis (p. 111). Chomsky's insight consisted in recognizing how informal procedures for segmenting and classifying expressions could be expressed by means of rules of the form  $A \rightarrow \omega$  that would 'rewrite' a single word class label  $A$  by a string  $\omega$  (which could consist of labels along with words and formatives). Thus a rule such as  $S \rightarrow NP VP$  would rewrite a sentence  $S$  by a subject  $NP$  and a  $VP$  predicate, and the rule  $V \rightarrow took$  would classify *took* as a verb.

By starting with the sentence label 'S' and applying a sequence of phrase-structure rules, one could define a 'derivation' that terminated in the expression that would be the starting point for procedures of IC analysis. The syntactic analysis assigned to an expression by a phrase structure grammar was conventionally represented by a phrase structure tree, though in Chomsky's initial formulations, analyses are represented by stringsets that he termed PHRASE MARKERS.<sup>3</sup> These sets contain strings from equivalence classes of derivations differing from one another solely in that they apply the same rules in a different order (e.g., a derivation where the subject  $NP$  is rewritten before rewriting the  $VP$  and a second derivation where the  $VP$  is rewritten first).

## 7.2 Extended phrase structure systems

As clarified particularly in Scholz and Pullum (2007) and Pullum (2011), phrase structure (and transformational) grammars represent linguistic applications of the general string rewriting systems developed in Post (1943, 1947). Despite the evident success attained by grammatical models based on rewriting systems, it was soon apparent that standard systems were not always ideally suited to the description of natural languages.

### 7.2.1 'The difficult question of discontinuity'

In particular, initial formulations of phrase structure grammars were incapable of representing the classes of discontinuous constituents

recognized by the Bloomfieldians, a point that Chomsky (1975a) was initially freely prepared to concede.

This [the treatment of ‘long components’ in the sense of Harris 1951] is an important question, deserving a much fuller treatment, but it will quickly lead into areas where the present formal apparatus may be inadequate. The difficult question of discontinuity is one such problem. Discontinuities are handled in the present treatment by construction of permutational mappings from **P** [the level of phrase structure, JPB/IAS] to **W** [the level of word structure, JPB/IAS], but it may turn out that they must ultimately be incorporated somehow into **P** itself. (Chomsky 1975a:190)

The transformational tradition never did reconsider whether discontinuities could be handled better within a phrase structure analysis and no general approach to this issue was explored within constituency-based grammars until the development of Head Grammars (Pollard 1984) and linearization-based models of HPSG (Reape 1996, Stefan Müller 1999, 2004, Kathol 2000). These models rehabilitated many of the same intuitions about syntactic and semantic units that had been explored in ‘wrapping’ analyses in the Montague grammar tradition, particularly in the accounts of Bach (1979) and Dowty (1982). However, Chomsky sought to reinforce the case for ‘permutational mappings’ (i.e., transformations) by disputing the feasibility of applying procedures of IC analysis to ‘derived’ constructions such as polar and information questions.

The case for indirect representation, not based on the relation of membership, becomes even stronger when we consider such sentences as “did they see John” or “whom did they see”. THESE ARE SENTENCES THAT NO LINGUIST WOULD EVER CONSIDER AS THE STARTING POINT FOR APPLICATION OF TECHNIQUES OF IC ANALYSIS – i.e., no one would ask how they can be subdivided into two or three parts, each of which has several constituents, going on to use this subdivision as the basis for analysis of other sentences, and so on. Yet there is nothing in the formulation of principles of procedure for IC analysis that justifies excluding these sentences, or treating them somehow in terms of sentences already analyzed.

(Chomsky 1958/1962:131f.; emphasis added JPB/IAS)

In the emphasized passage, as elsewhere in Chomsky’s writings about the Bloomfieldians, a position possibly consistent with the practice of Zellig Harris is incorrectly attributed to the Bloomfieldians as a group. Virtually all leading American linguists of the time, including Hockett, Gleason, Nida, Pike, and Wells, among others, not only considered applying – but in fact DID apply – procedures of IC analysis to questions in English. In particular, the analysis of polar questions was regarded as a solved problem and presented as such in the introductory textbooks of the day. In the passage below, Gleason gives what he takes to be an uncontroversial IC

analysis of polar questions to exemplify the notion of discontinuous constituents.

In English, discontinuous constituents occur. One common instance occurs in many questions: *Did the man come?* This is clearly to be cut *did ... come | the man.* (Gleason 1961:142)

This discrepancy between procedures of IC analysis and phrase structure grammars is of more than purely historical interest. One of the criticisms levelled by Chomsky against phrase structure grammars turned on their inability to represent discontinuous dependencies, particularly within auxiliary verb phrases.

To put the same thing differently, in the auxiliary verb phrase we really have discontinuous elements ... But discontinuities cannot be handled within  $[\Sigma, F]$  grammars [i.e., phrase structure grammars, JPB/IAS]. (Chomsky 1957:41)

### 7.2.2 Generalized Phrase Structure Grammar (GPSG)

For the most part, modern phrase structure systems preserve Chomsky's preference for describing discontinuous dependencies indirectly, usually in terms of relations between different parts of a single structure or correspondences between different types of structures. However other restrictions on phrase structure systems have been more comprehensively revised. The most severe of these was the assumption that the 'non-terminal' vocabulary of a phrase structure grammar should consist solely of atomic labels such as 'S,' 'NP,' 'V,' etc. The case for relaxing this restriction is made initially by Harman (1963), who objects that "it is irrational to restrict the amount of information expressed by the grammar to statements about grammatical category" (p. 604). The response in Chomsky (1965:210f.) dismisses Harman's proposal out of hand as a "terminological equivocation" and appears to construe any refinement of phrase structure grammars as a case of a patent infringement rather than as a genuine attempt to understand the scope and limits of constituent structure grammars. Partly as a consequence, Harman's 'defense of phrase structure' had little direct influence on the field at the time. Hence, the descriptive potential of feature 'decompositions' of atomic symbols was not fully realized until the later work on unbounded dependencies and coordination (Gazdar 1981).

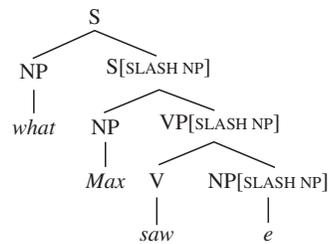
By this time, a limited amount of feature decomposition had been incorporated into transformational models that adopted some version of the X-bar conventions. However, features were assigned a tightly circumscribed role in Chomsky (1970a), and these restrictions were preserved in subsequent accounts. Two constraints were particularly decisive. The first of these restricted propagation through an endocentric X-bar projection to

the word class features  $\pm N$  and  $\pm V$  (Chomsky 1970a:52f.), excluding other types of lexical and inflectional properties. The second constraint limited feature ‘percolation,’ as it came to be known, more generally by “tak[ing] feature complexes to be associated only with lexical categories, and permitting] complex symbols to dominate a sequence of elements only within the word” (Chomsky 1970a:48).<sup>4</sup> These restrictions precluded the use of constituent structure links as a conduit for the propagation of complex feature bundles. Likewise, although the ‘non-distinctness’ condition on complex symbols in Chomsky (1965:84) anticipated the unification operations of later constraint-based formalisms, this condition could play no role in regulating the distribution of features within a projection.

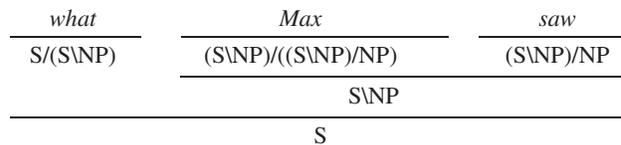
### 7.2.2.1 Non-local dependencies

As with the representational constraints that barred discontinuities, restrictions on the ‘flow’ of feature information prevented feature-based mechanisms from encroaching on the role reserved for structure-changing rules and derivational operations in transformational models. By relaxing these restrictions, extended phrase structure models could exploit the descriptive value of feature information for describing local and non-local grammatical dependencies. Unbounded dependencies had long been taken to require the power of a transformational grammar, or at any rate to defy analysis in terms of phrase structure grammars, as suggested in the quotation from Chomsky (1958/1962) above. Hence the rehabilitation of phrase structure analysis began, somewhat counterintuitively perhaps, with an analysis of unbounded dependencies that was developed in the late 1970s but first published in Gazdar (1981). The simple intuition developed in this work was that the constituent structure links of a phrase structure tree provided a suitable conduit for the flow of information about displaced elements. The components of the analysis were equally straightforward: feature attributes that could take categories as values, the insight that information about ‘missing’ elements could be treated in terms of a feature (Bear 1982), and feature ‘passing’ conditions that could match features between the ‘mother’ and ‘daughter’ nodes in a phrase structure tree. By passing the value of a category-valued attribute along a chain of local mother–daughter nodes, a phrase structure analysis could match the properties of a ‘missing’ element at an ‘extraction site’ with those of the ‘dislocated’ element that typically occurred at the periphery of a construction.

The components of what came to be known as the ‘slash category’ analysis of unbounded dependencies are exhibited in the analysis of the English indirect question in Figure 7.1 below. The lowest occurrence of the category-valued `SLASH` feature encodes the properties of the missing object NP that is governed by the transitive verb *saw*. These properties



**Figure 7.1.** 'Slash category' analysis of indirect question

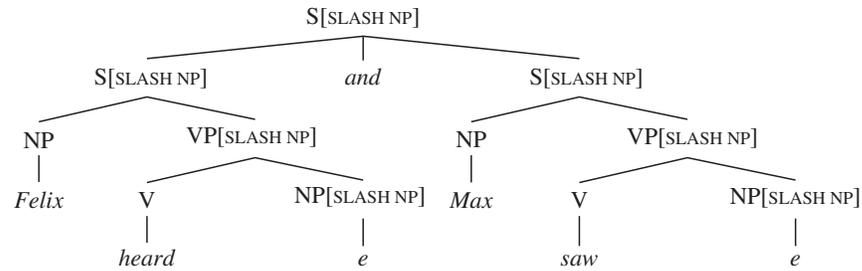


**Figure 7.2.** Gap-free categorial analysis (cf. Steedman 2000b)

are passed up successively to the superordinate VP and S nodes until they can be matched against the properties of the 'filler' *what*.

To a large degree, the early phrase structure analyses carried over prevailing assumptions about the structure of unbounded dependency constructions from transformational accounts. In contrast to the IC analyses adumbrated in the descriptivist tradition, the structure in Figure 7.1 assumes that the dislocated element *what* is higher in the tree as well as to the left of the extraction site. This assumption is retained in most subsequent analyses of unbounded dependencies. In addition, the structure in Figure 7.1 preserves the assumption that the extraction site is occupied by an empty placeholder 'gap.' Since this assumption had no internal motivation within phrase structure models, the analysis developed in Sag and Fodor (1994) and Bouma *et al.* (2001) dispensed with null terminals. These analyses nevertheless retain the strategy of using dedicated attributes to represent information about extracted elements. In this respect, they are unlike categorial analyses, such as Steedman (2000b), which use the slash notation both to indicate the argument of a functor and to encode information about extracted elements. In the categorial analysis in Figure 7.2, the category '(S\NP)/NP' marks the transitive verb *saw* as a functor that looks rightward for an NP to form a functor that in turn looks leftward for an NP to form an S. The overloading of this notation is reflected in the fact that the category 'S\NP' encodes the 'missing' object NP in the expression *Max saw*.<sup>5</sup>

As recognized by those working to extend the empirical coverage of phrase structure models, category-valued features offered a novel perspective on a range of phenomena that interacted with unbounded dependencies. In particular, the assumption that information about missing constituents formed part of the syntactic information associated with a node interacted with the independent assumption that coordination was restricted to syntactically like elements. One immediate consequence was



**Figure 7.3.** ‘Across-the-board’ extraction from coordinate indirect question

an account of the parallelism that Ross (1967a) had termed ‘across-the-board’ extraction. The central observation was that in a coordinate structure, if one conjunct contained an extraction site, then all of the conjuncts must.<sup>6</sup> In transformational models, this condition had been attributed to dedicated devices, such as the Coordinate Structure Constraint (Ross 1967a) or the Across-the-Board convention of Williams (1978), which, as Gazdar *et al.* (1982b) noted, incorporated a construction-specific and somewhat imprecise extension to the notion of phrase marker. In contrast, the parallelism requirement on extraction from coordinate structures followed on a phrase structure analysis. Two conjuncts of category  $X[\text{SLASH } YP]$  were syntactically alike, whereas a conjunct of category  $X[\text{SLASH } YP]$  and one of category  $X$  were not. In the analysis in Figure 7.3, the two conjuncts of category  $S[\text{SLASH } NP]$  are syntactically alike and can be conjoined, but neither could be conjoined with a full  $S$  to yield unacceptable examples such as *\*what Felix heard and Max saw the intruder* or *\*what Felix heard the intruder and Max saw*.<sup>7</sup>

Gazdar (1981) also clarified how constraints on extraction, which had typically been described in terms of conditions on rule application, could be recast in terms of restrictions on the ‘paths’ of category-valued features that connected extraction sites to dislocated fillers. In classical transformational accounts, there had been no reason why information about missing constituents should trace a path along the constituent structure links of a tree.<sup>8</sup> But once extraction was characterized in terms of the sharing of category-valued features along a sequence of mother-daughter links, it became clear that any restrictions on the extraction of elements out of specified ‘island’ domains (Ross 1967a) would correspond to paths in which those domains occurred somewhere along the path between extraction sites and fillers.

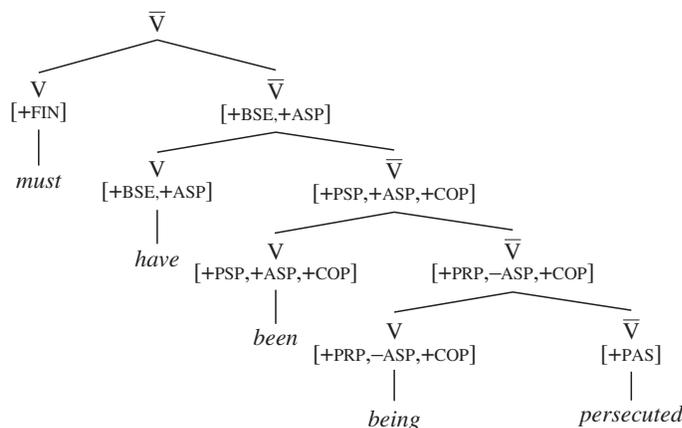
### 7.2.2.2 Local dependencies

The demonstration that complex-valued features could provide an analysis of unbounded dependencies inspired surface-based analyses of more local syntactic phenomena within the nascent community that had begun to explore the potential of monostratal models.<sup>9</sup> The English auxiliary system had long been an obvious candidate for reanalysis. The system consisted of a finite inventory of modal and auxiliary elements, which

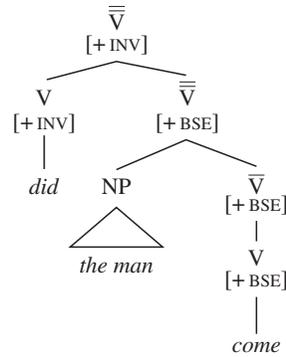
were subject to ordering constraints that determined a (short) maximum expansion. The expansions were indeed almost as restricted as pronominal clitic sequences in Romance languages, and, like these sequences, exhibited some of the ordering rigidity characteristic of morphological formations. Even the selectional dependencies tended to relate pairs of adjacent elements. So there was nothing that presented any intrinsic difficulties for a phrase structure analysis.

The ‘affix hopping’ analysis of Chomsky (1957) had long been held to be one of the crowning achievements of transformational approaches. However, Gazdar *et al.* (1982a) showed that the strategy of ‘hopping’ affixes from one point in a terminal string to another was a solution to a self-inflicted problem and hence dispensable in a model with complex-valued features. If one auxiliary element could select the verb form of the head of a phrasal complement, there was no need to assemble inflected forms in the course of a syntactic derivation. Instead, the admissible expansions could be determined by the subcategorization demands of individual elements. The first component of this analysis is a feature classification of verbal elements that distinguishes tense, aspect, and voice properties, along with form variants, such as participles, infinitives, etc. The second is a generalization of the X-bar feature conventions that allows these ‘head’ features to be shared between a mother and head daughter node. The final ingredient is, again, category-valued features that permit a verbal element to select a complement headed by a particular form variant.

These components are set out in detail in Gazdar *et al.* (1982a) and in much subsequent work within Generalized Phrase Structure models. One type of analysis that they define is illustrated in Figure 7.4. The advantages of this analysis are summarized in Gazdar *et al.* (1982a:613ff.), though one immediate benefit was the avoidance of the formal problems that had plagued the ‘affix-hopping’ analysis since its initial formulation (see, e.g., Akmajian and Wasow 1975, Sampson 1979).



**Figure 7.4.** Passive auxiliary expansion (cf. Gazdar *et al.* 1982a:601)



**Figure 7.5.** Subject–auxiliary ‘inversion’

The analyses in Gazdar *et al.* (1982a) thus established that the same basic feature-passing strategy used in the treatment of unbounded dependencies could provide an account of local dependencies. Patterns of subject–auxiliary inversion were amenable to a similar analysis using grammar rules systematically related to the basic rules via metarules, a device whose utility in the grammar of programming languages had previously been established. Figure 7.5 exhibits the analysis of the polar question cited by Gleason (1955) above. The invertibility of modals and auxiliaries is encoded here via compatibility with the [+INV] specification that is required of the verbal head in a phrase structure rule licensing the ‘inverted’ structure. Independent motivation for this feature comes from lexical restrictions on the distribution and interpretation of auxiliary elements. Some elements, such as *lsg aren’t*, are obligatorily inverted, while others, such as *better*, are obligatorily uninverted, and yet others, such as *may*, have a different range of meanings depending on whether or not they are inverted.

### 7.2.3 Node admissibility and constraint satisfaction

More generally, it turned out that all of the alternations and dependencies that had been described by transformational models had simple – and, in at least some cases, arguably superior – phrase structure analyses.<sup>10</sup> One might have expected that this result would have produced a general rapprochement between transformational and phrase structure approaches and an attempt to arrive at broadly accepted criteria for evaluating the different strategies for describing these patterns. In fact, just the opposite occurred. Transformational models abandoned their flirtation with a ‘representational’ interpretation, a perspective that had been developed particularly in Koster (1978a, 1987), and adopted a more resolutely derivational orientation.

While transformational accounts were following the developmental pathway that led to current Minimalist models (see Chapter 4), extended phrase structure models began to incorporate insights and perspectives from other monostratal approaches. Following McCawley (1968a), models

of Generalized Phrase Structure Grammar (Gazdar *et al.* 1985) had already adopted – and, indeed, refined – a ‘node admissibility’ interpretation of phrase structure rules. On this interpretation, a rule such as  $S \rightarrow NP VP$  is interpreted as directly ‘licensing’ a local subtree in which *S* immediately and exhaustively dominates *NP* and *VP* daughters, and the *NP* daughter immediately precedes the *VP* daughter. A node admissibility interpretation immediately eliminated the need for string-rewrite derivations and string-based representations of phrase structure (‘phrase markers’). Instead, rules could be regarded as partial DESCRIPTIONS of the subtrees that they sanctioned and the admissibility of a tree could be defined in terms of the admissibility of the subtrees that it contained.

In large part, this reinterpretation of phrase structure productions supplied graph-theoretic modeling assumptions that were a better fit for the classes of IC analyses initially proposed by the Bloomfieldians. The schematization adopted within models of X-bar Theory similarly deprecated phrase structure rules within transformational models, though without substantially revising the string-based model of phrase structure represented by phrase markers (as discussed in note 3).

Furthermore, a node admissibility interpretation clarified the fact that conventional phrase structure rules bundle information about structure (mother–daughter links) together with information about order (linear arrangement of daughters). GPSG accounts showed how these two types of information could be expressed separately, by means of a set of Immediate Dominance (ID) rules that just constrained mother–daughter relations and a set of Linear Precedence (LP) statements that applied to sisters in a local tree. For example, the information represented by the phrase structure rule  $S \rightarrow NP VP$  would be expressed by an ID rule  $S \rightarrow NP, VP$  and the general LP statement  $NP \prec VP$ . The absence of an applicable LP rule would not sanction unordered trees, but rather trees in which the *NP* and *VP* occurred in either order.

An overriding consideration in the development of GPSG was the goal of keeping analyses as explicit as possible and the underlying grammatical formalism as formally restrictive as possible. The central role of context-free phrase structure grammars largely reflected the fact that their properties were well understood and provided a formal basis for transparent analyses. In some cases, analyses were constrained so that they did not take GPSG models outside the class of phrase structure grammars. For example, requiring that sets of ID rules and LP statements must operate over the same local domains, ensured that they could in principle be ‘reconstituted’ as phrase structure grammars. LP statements were thus restricted to apply to sister nodes. As a consequence, LP statements could allow free or partial ordering of *VP*-internal elements, but they could not impose any ordering of subjects and *VP*-internal elements other than those that followed from the ordering of a subject and full *VP* expansion. Yet there was no direct empirical support for this restriction.

Hence the tight association between the domains of ID rules and LP statements undermined the fundamental separation of structure and order in the ID/LP format since certain types of ordering variation dictated a flat structure. This was perhaps acceptable as long as there was some independent motivation for remaining within the class of context-free phrase structure grammars. But by 1985, the demonstration of non-context-free patterns in Swiss German subordinate clauses (Shieber 1985) and Bambara compounds (Culy 1985) had weakened the empirical grounds for this restriction and the non-transformational community shifted their focus to identifying restricted classes of weakly context-sensitive grammars that were descriptively adequate. This was a natural development within the family of phrase structure approaches, given that the interest in context-free grammars had been driven by an interest in explicit formalisms with clearly defined and well-understood properties. Hence the move from the limited word order freedom defined by the ID/LP format in GPSG to 'domain union' in HPSG (Reape 1996) extended the dissociation of structure and order in ways that allowed for the interleaving of non-sisters in an explicit but non-context-free formalism.<sup>11</sup>

### 7.3 Model-theoretic grammar

In the subsequent development of phrase structure grammars, the interpretation of rules as partial descriptions of trees provided the model for a more comprehensive constraint-based or model-theoretic perspective. As in models of Lexical-Functional Grammar (Kaplan and Bresnan 1982; Chapter 6 of this volume), rules and grammatical principles came to be construed as `CONSTRAINTS` that described or were satisfied by corresponding types of `STRUCTURES`. This move to a uniform model-theoretic orientation permitted much of the complexity that had been associated with representations to be confined to the constraint language that described structures. In addition, a general model of constraint satisfaction provided a conception under which the diverse feature distribution principles of GPSG could be subsumed. The gradual accretion of constraints and conditions in GPSG had led over time to a theory in which the components that regulated feature information included feature co-occurrence restrictions and feature specification defaults, in addition to the immediate dominance rules, linear precedence statements, and other devices, such as metarules. As detailed in Gazdar *et al.* (1985: Chapter 5), the constraints in these components exhibited fairly intricate interactions. On a description-theoretic interpretation, these constraints and interactions could be modeled in a more uniform and transparent way.

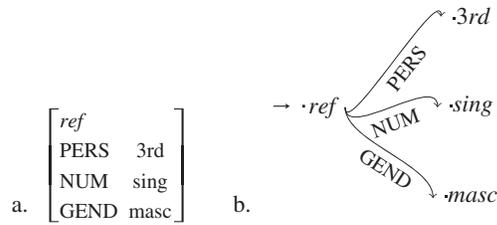
The emergence of a constraint-based perspective was accompanied by the adoption of richer sets of structures and more expressive constraint languages. These developments provided clearer conceptions of the lexicon, valence, and valence alternations than had been possible in GPSG.

The phrase structure systems proposed in Chomsky (1956) had offered only very rudimentary treatments of the lexicon and valence demands, and incorporated no notion of a lexical valence-changing process. The closest counterparts of ‘lexical entries’ in these simple systems were rules of the form  $V \rightarrow \textit{sleep}$ , which rewrote a non-terminal symbol as a terminal element. The valence of a predicate was likewise represented implicitly by the other elements that were introduced in the same rule expansions. GPSG enriched this spartan conception by locating terminal elements within lexical entries that specified distinctive grammatical features of an element other than word class. Corresponding to the pre-terminal rules of a simple phrase structure grammar was a class of ‘lexical ID rules’ which introduced lexical heads indexed by a subcategorization index. This index (technically the value of a `SUBCAT` feature) was then cross-referenced with a class of lexical entries. For example, the rule  $VP \rightarrow H[l]$  would license a local VP subtree that dominated a unary tree whose mother was  $V[l]$  and whose daughter was an intransitive verb, such as *sleep*, whose entry contained the index *l*.<sup>12</sup>

In effect, the use of subcategorization indices achieved a limited type of context sensitivity within a context-free formalism. Yet, as Jacobson (1987:394ff.) pointed out, the fact that lexical items did not directly represent valence information created numerous complications in GPSG. The most acute arose in connection with the treatment of valence alternations. There was no way to formulate a passive rule that mapped the transitive entry for *devour* onto a (syntactically) detransitivized entry *devoured*, because entries themselves contained no direct representation of transitivity. This led to an analysis of passivization in terms of metarules that mapped a ‘transitive expansion’ such as  $VP \rightarrow W, NP$  to a ‘detransitivized expansion’ such as  $VP[\textit{PAS}] \rightarrow W$  (where *W* is any string). However, it then became necessary to constrain metarules so that they only applied to lexical ID rules. But lexical ID rules were serving as proxies for underinformative entries, so the obvious solution lay in associating valence information directly with lexical items and introducing a class of lexical rules to map between entries, as suggested by Pollard (1984).

### 7.3.1 Head-Driven Phrase Structure Grammar (HPSG)

The models of Head-driven Phrase Structure Grammar outlined in Pollard and Sag (1987, 1994) develop a number of these revisions in the context of a broad constraint-based conception of grammar. A central component of these models is the set of assumptions that have come to be known as the ‘formal foundations’ of HPSG. As in LFG, grammatical constraints and lexical entries are interpreted as partial descriptions of structures, though the representational conventions and model theories of the two theories differ significantly. One representational difference concerns the interpretation of attribute-value matrices (AVMs). Whereas in LFG, AVMs of the



**Figure 7.6.** Descriptions and structures in HPSG

sort illustrated in Figure 7.6a are used to represent f(unctional)-structures, in HPSG they represent descriptions, i.e., sets of constraints. The structures DESCRIBED BY (or which SATISFY) a set of constraints are represented as graphs like the one in Figure 7.6b.

A distinctive aspect of the HPSG model theory is the role assigned to the TYPE system. The core idea is that each kind of structure is associated with certain sorts of attributes, and that each attribute is associated with a type of value. For example, a referential index (object of type *ref* in Figure 7.6) is associated with the attributes PERS(ON), NUM(BER), and GEND(ER). Each attribute takes values from a partitioned value space, which in the present case just represents traditional person, number, and gender contrasts. The empirical effects of this type system derive from two additional assumptions. The first is that structures must be TOTALLY WELL-TYPED (Carpenter 1992: Chapter 6) in the sense that they must be assigned a value for each appropriate attribute. This constraint precludes, for example, the assignment of a number-neutral structure as the analysis of English *sheep*, given that number is distinctive for English nouns (each occurrence of *sheep* is unambiguously singular or plural). A separate requirement that structures must be SORT-RESOLVED (Pollard and Sag 1994:18) permits only ‘fully specific’ feature values and thus bars disjunctive case values from occurring in a well-formed structure. Hence *sheep* could not be treated as neutral by assigning the attribute NUM a maximally general value such as *number*, which subsumes the resolved values *sing* and *plur*. Given that entries are interpreted as descriptions of lexical structures, the English lexicon can still contain a single underspecified ENTRY for *sheep*, one that specifies either no NUM attribute or a NUM attribute with a non-sort-resolved value. But the lexical structures described by the entry must be totally well-typed and sort-resolved.

These general assumptions have the effect of ensuring that structures are maximally specific and that all underspecification is confined to descriptions. A neutral description is not satisfied by a correspondingly underspecified structure but by a set of structures, each of which supplies different, fully resolved values for underspecified attributes. This technical point has a number of consequences. On the positive side, the assumption that structures must be totally well-typed and sort-resolved does some of the work of the completeness and coherence conditions in LFG, and facilitates type-based inferencing within HPSG. However, these

assumptions also lead to apparent difficulties in accounting for the types of patterns described in Ingria (1990), in which the neutrality of an item seems to permit it to satisfy incompatible demands simultaneously, most prominently in coordinate structures.<sup>13</sup>

Note further that in a model theory that only contains fully specified structures, it is somewhat anachronistic to describe the processes that determine feature compatibility in terms of feature structure UNIFICATION, as had been the practice in GPSG and PATR-based formalisms (Shieber 1986). A more accurate characterization of a model-theoretic linguistic framework would be as CONSTRAINT-based, a term that has garnered a certain acceptance in the non-transformational community. Within HPSG, configurations in which a single object occurs as the value of multiple attributes are described in terms of STRUCTURE-SHARING, a term that refers to reentrance in the graph-theoretic models typically assumed in HPSG.

### 7.3.1.1 Valence, raising, and control

Raising constructions illustrate how structure sharing interacts with complex-valued attributes to provide an insightful analysis of grammatical dependencies. The term ‘raising’ derives from transformational analyses in which the subject of a complement is taken to be ‘raised’ to become an argument of the raising verb. However, complex-valued features permit an analysis in which raising involves the identification of arguments within the argument structure of a raising predicate. Patterns involving the sharing of purely morphological properties offer the clearest support for this analysis. As discussed by Andrews (1982), among others, modern Icelandic contains verbs that may govern ‘quirky’ non-nominative subjects. One such verb is *vanta* ‘to want,’ which occurs with the accusative subject *hana* ‘her’ in (1a). These quirky case demands are preserved by raising verbs such as *virðist* ‘seems.’ As example (1b) shows, *virðist* is, in effect, ‘transparent’ to the accusative case demands of *vanta*, which are imposed on its own syntactic subject.

#### (1) Quirky case in Icelandic raising constructions (Andrews 1982)

- a. *hana vantar peninga*  
her.ACC lack.3SG money.ACC  
‘she lacks money’
- b. *hana virðist vanta peninga*  
her.ACC seem.3SG lack money.ACC  
‘she seems to lack money’

The constraints in Figure 7.7 first identify *hana* as a 3sg feminine accusative NP, and indicate that the verb *vanta* selects an accusative subject and complement (though category is suppressed here). In place of the integer-valued SUBCAT feature of GPSG, HPSG represents the core valence demands of a verb by means of list-valued SUBJ and COMPS features.



$$vir\dot{d}ist: \left[ \begin{array}{l} \text{SUBJ} \langle \boxed{1} \rangle \\ \text{COMPS} \langle \boxed{2} [\text{SUBJ} \langle \boxed{1} \rangle] \rangle \\ \text{ARG-ST} \langle \boxed{1} \boxed{2} \rangle \end{array} \right]$$

Figure 7.9. Lexical entry of subject raising verb

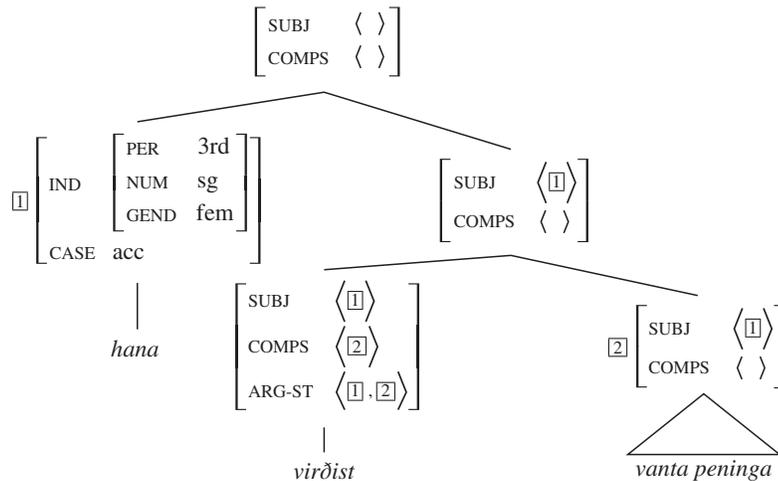


Figure 7.10. Subject 'raising' as structure sharing

two SUBJ values (via the boxed integer  $\boxed{1}$ ) indicates that the SUBJ attribute of *vir\dot{d}ist* literally shares its value with the SUBJ value of its complement. Identifying the values of the two SUBJ attributes ensures that any constraints that apply to the SUBJ of the complement of *vir\dot{d}ist* will apply to its own syntactic SUBJ. Hence when *vanta* occurs as the head of the complement, as in Figure 7.10, its accusative SUBJ demands will be identified with the SUBJ demands of *vir\dot{d}ist*. Only an accusative subject such as *hana* can satisfy these demands. So this analysis forges a direct association between *hana* and the complement *vanta peninga*, but the association is established by means of structure sharing, rather than through constituent structure displacements.

This analysis shows how the complex-valued features that provide an account of basic valence demands in Figure 7.8 interact with structure-sharing to allow the subject demands of a raising verb to be identified with those of its complement.<sup>14</sup> Furthermore, precisely the same elements offer an analysis of 'control' constructions, in which the higher controller merely identifies the reference of the subject of the complement. The properties of control constructions are discussed in detail in Sag and Pollard (1991), but they can be broadly subsumed under the generalization that control verbs are not transparent to the syntactic demands of the head of their complement. The contrast with raising verbs is reflected in the fact that the subject of the control verb *vONA* 'hope' in (2b) follows the default nominative pattern and does not inherit



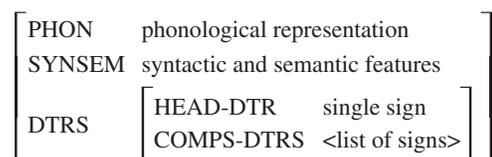
### 7.3.1.2 Lexical rules

The treatment of argument structure in terms of a list-valued ARG-ST feature also provides the formal basis for a lexical analysis of valence alternations in HPSG. Lexical rules can apply to an entry and modify the ARG-ST list in various ways, by adding, deleting, permuting, or reassociating elements. The new entries that are defined by these types of rules will then have different combinatory and interpretive properties, due to the cross-referencing of ARG-ST elements with valence features and with semantic representations. For example, different versions of passive lexical rules are proposed in Pollard and Sag (1987:215) and Manning and Sag (1999), and a number of other valence-changing lexical rules are proposed in the HPSG literature (see, e.g., Wechsler and Noh (2001) and Müller (2002)). However, the study of valence alternations has been less a primary focus of research within HPSG than in, say, LFG (see the discussion of Lexical Mapping Theory in Chapter 6).

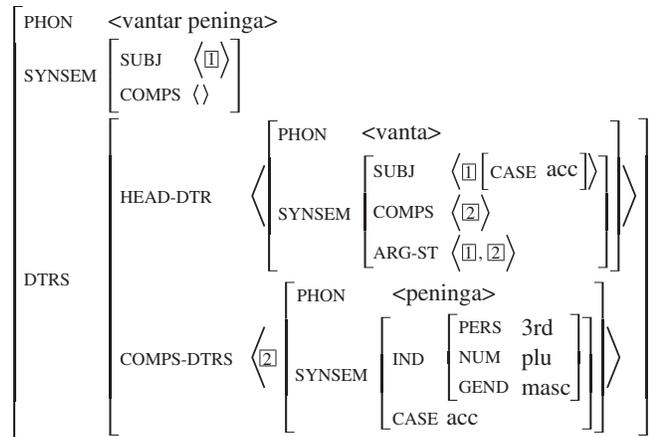
### 7.3.1.3 The architecture of signs

Figures 7.10 and 7.12 illustrate the tree-based diagrams that are often used to exhibit HPSG analyses. These representations show the usefulness of tree-based diagrams for isolating particular aspects of an analysis, in the present case the role of valence features and structure sharing. However, the familiarity of this representational format comes at a cost, as it slightly misrepresents the SIGN-BASED nature of HPSG.<sup>16</sup> In GPSG, feature structures are labels that annotate the nodes of a phrase structure tree. But HPSG inverts this conception and incorporates constituent structure links within general data structures termed SIGNS. Within the version of HPSG expounded in Pollard and Sag (1994), a head-complement sign has the general structure in Figure 7.13. There are two innovative aspects of this analysis. The first is that syntactic and semantic features are consolidated into a single type of data structure, termed a *synsem*. The second is that constituent structure is represented by DTRS ('daughters') attributes that take signs or lists of signs as values. Hence the VP from Figure 7.8 above is represented, albeit somewhat less perspicuously, by the sign in Figure 7.14.

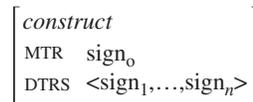
Even the fairly rich analysis in Figure 7.14 suppresses syntactic detail (not to mention all of the semantic properties incorporated within *synsem* objects). Although the highly explicit nature of the HPSG formalism may seem somewhat imposing, the formal character of the formalism is designed with the dual goals of broad-coverage theoretical description and large-scale practical implementation in mind.<sup>17</sup> For students (and



**Figure 7.13.** Structure of head-complement signs



**Figure 7.14.** Partial analysis of *vantar peninga*



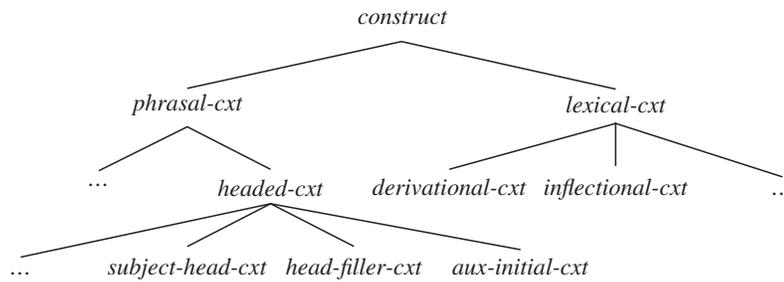
**Figure 7.15.** General structure of a *construct* (Sag 2010a: 497)

general linguists) who mainly want to understand the basic intuitions and desiderata that underlie HPSG models, a more streamlined version of the formalism is presented in Sag *et al.* (2003).

### 7.3.2 Sign-Based Construction Grammar (SBCG)

In much the same way that initial models of HPSG drew on ideas from categorial grammar and adapted techniques from AI and theoretical computer science, current models of Sign-Based Construction Grammar integrate key empirical insights from the Berkeley Construction Grammar tradition (Goldberg 1995, Kay and Filmore 1999). The conceptual unification of these traditions rests on the insight that the rich construction inventories investigated in Construction Grammar can be modeled by organizing individual constructions into inheritance networks. The formal architecture required by this analysis is already fully present in standard models of HPSG, in the form of the type hierarchies that cross-classify individual signs representing words, phrases, and clauses. The main prerequisite for a construction-based extension of HPSG is then a type of feature structure that represents constructions.

As noted in Sag (2010b, 2012), feature structure counterparts of the local trees from GPSG provide suitable candidates. Individual constructions can be represented by feature structures exhibiting the organization in Figure 7.15, where *MTR* represents the mother sign and *DTRS* a list of daughter signs. Many of the construction-specific properties investigated in the modern Construction Grammar literature (typified by Kay and Filmore (1999)) can be integrated into these unified data structures.



**Figure 7.16.** Partial construction type hierarchy (Sag 2010a: 499)

As in HPSG, the inheritance of properties within a construction inventory can be modeled by type hierarchies. The partial hierarchy in Figure 7.16 represents natural classes of constructions relevant to the analysis of extraction in English.

The detailed treatment of English relative and filler-gap clauses in Sag (1997, 2010a) presents a sustained argument for extending HPSG models to include a notion of construction. At the same time, these studies make a case for reconceptualizing grammatical constructions in the context of a constraint-based architecture, rather than in the exemplar-based terms assumed in traditional grammars.

These studies also illustrate the ways that phrase structure models continue to evolve, driven in part by the logic of their basic organizing principles, and in part by their ability to incorporate and extend insights from other traditions. From their origins in the string rewriting systems in Chomsky (1956), extended phrase structure models have assumed their modern form by successively integrating traditional perspectives on grammatical features and units with more formal notions such as inheritance hierarchies and constraint satisfaction. In addition to providing analyses of a wide range of syntactic constructions, these models have clarified how explicit mechanisms for regulating the distribution of grammatical information within a single syntactic representation can achieve any of the benefits that had, beginning with the work of Harris (1957), been claimed to accrue to derivational analyses.

## Notes

We thank Stefan Müller and an anonymous reviewer for comments on an earlier version of this chapter.

1. A partial list of constructions that were analyzed as discontinuous by the Bloomfieldians would include parentheticals (Bloomfield 1933:186, Nida 1966:21), coordination (Bloch 1946:229), ditransitives (Pike 1943:77), complex predicates (Nida 1966:46), phrasal verbs (Wells 1947:105–106), polar questions (Pike 1943:77, Hockett 1958:158, Gleason 1955:142, Nida 1966:28), non-subject relatives (Nida 1966:27), non-subject constituent questions (Nida 1966:46, Gleason 1955:155).

2. The analysis in Table 7.2 also represents functional and even dependency information, as it illustrates the convention that “the arrow points towards the head” in a modifier-head construction and that “the **P** always faces the predicate” in a subject–predicate construction (Gleason 1965:157).
3. Chomsky appears to maintain the representational assumptions in Chomsky (1975a:chapter VII) when he later suggests that “We take these objects [i.e., levels of linguistic representation, JPB/IAS] to be phrase markers in the familiar sense represented conventionally by trees or labelled bracketings” (Chomsky 1995c:34).
4. In the continuation of this passage, Chomsky notes that the second constraint has merely been carried over from Chomsky (1965:188) and appears willing to countenance the idea “that certain features should also be associated with nonlexical phrase categories.” Yet the accompanying footnote immediately characterizes the arguments supporting previous proposals as “very weak,” and Chomsky does not in fact propose a general relaxation of the constraints on complex feature bundles that would allow the inflectional features associated with a lexical category to be propagated to or shared with a phrasal projection.
5. See Steedman and Baldridge (2011) for recent synopsis of combinatory categorial approaches.
6. Though subsequent work has called into question whether this parallelism is restricted to coordinate structures (Postal 1998) and whether the constraint ultimately reflects more general semantic or discourse factors (Goldsmith 1985, Lakoff 1986, Kehler 2002).
7. The analysis in Figure 7.3 introduces the coordinating conjunction as a sister of the conjuncts, rather than associating it with the second conjunct, though nothing here hinges on this difference.
8. Indeed, there was a considerable delay before the tradition even addressed the challenge of assigning a derived constituent structure to transforms, an issue that had been raised as early as Stockwell (1962).
9. See Ladusaw (1988) for discussion of the contrast between linguistic ‘levels’ and ‘strata.’
10. The analyses of unbounded dependencies and auxiliary selection/inversion outlined above were followed by phrase structure treatments of, among others, an expanded range of extraction constructions (Pollard and Sag 1994: Chapters 4–5, Levine and Hukari 2006), passives (Pollard and Sag 1987:215, Manning and Sag 1999), control constructions (Sag and Pollard 1991), anaphoric binding (Pollard and Sag 1992). Contemporary work in Lexical–Functional Grammar (Dalrymple *et al.* 1995) and Tree Adjoining Grammar (Joshi and Schabes 1997) explored a similar range of

- empirical extensions. See also Johnson and Lappin (1999) for a comprehensive comparison of constraint-based and derivation perspectives.
11. Reape (1996) was widely circulated in draft form, as of 1990.
  12. By virtue of the Head Feature Principle (aka the Head Feature Convention), the metavariable ‘H[1]’ would license a preterminal V[] that shared the word class features of the VP mother.
  13. See Blevins (2011) for a recent review and discussion of these types of cases.
  14. A similar analysis is proposed within LFG in terms of ‘functional control’ (Bresnan 1982c).
  15. There is more to say about Icelandic raising constructions and the mechanisms that allow quirky ‘lexical’ case to take priority over default ‘structural’ case. See Andrews (1982, 1990) for some discussion.
  16. See also Orgun (1996) for a sign-based model of morphology compatible with HPSG assumptions.
  17. Repositories of information, publications, and materials related to current HPSG implementations can be found at <http://lingo.Stanford.edu/>, <http://hpsg.fu-berlin.de/Projects/core.html>, and <http://hpsg.fu-berlin.de/~stefan/Babel/>. See also Müller (2010:187f.)
- .