Remarks on Locality*

Ivan A. Sag
Stanford University

April 20, 2005

1 Introduction

This paper deals with a two issues having to do with locality in natural language. The first – locality of syntactic selection – is the problem of delimiting what syntactic and semantic information lexical items select for. The second, closely related matter – the locality of construction – is the problem of delimiting the syntactic and semantic information accessible to grammar rules. These issues have considerable history in the field, though matters of locality are sometimes left implicit in theoretical discussions.

After providing some necessary background, I will propose a modification of HPSG theory that embodies a particular hypothesis about locality. In the general theory I propose, the feature geometry serves to delimit the grammatical information accessible for lexical selection or constructional constraints. My proposal is consistent with the relevant cross-linguistic facts that I am familiar with, which I will survey.

2 Background

Among the theoretical issues that were hotly debated during the 1960s is the locality of selection. For example, Chomsky (1965, Ch. 2) proposed that the lexical entries of verbs and other lexical 'formatives' include 'strict subcategorization restrictions' like those shown in (1):

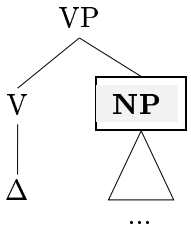
(1) a. prove, V, [+ __ NP]

b. run, V, [+ __ DIR]

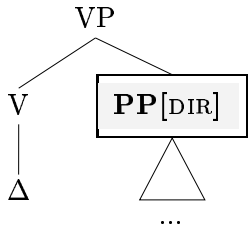
*I would like to thank Emily Bender, Bill Croft, Bruno Estigarribia, Charles Fillmore, Dan Flickinger, Adele Goldberg, Andreas Kathol, Paul Kay, Bob Levine, Detmar Meurers, Laura Michaelis, Carl Pollard, Jan Strunk, and Tom Wasow for valuable discussions about locality. I am particularly grateful to Detmar Meurers and Stefan Müller for detailed comments on an earlier draft of this paper. This work was supported in part by grant BCS-0094638 from the National Science Foundation to Stanford University and in part by the Research Collaboration between NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation and CSLI, Stanford University.

Context-sensitive lexical insertion transformations (which involved the substitution of a lexical formative for a dummy symbol 'Δ') were subject to a 'matching condition' that required the subcategorization restrictions to match the local context in the deep structure phrase marker. Chomsky proposed that the matching condition obeyed a principle of 'strict locality', which stipulated that strict subcategorization restrictions like those illustrated in (1) could only make reference to (could only be matched against) elements that are dominated by the VP directly dominating the V in deep structure subtrees like (2a,b):

(2) a.



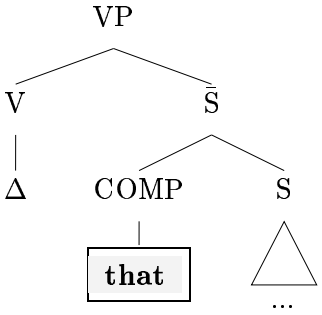
b.



Strict locality imposed an upper bound on the domain of subcategorization, but not a lower bound. That is, an element referred to by a subcategorization restriction did not have to be a sister of the V; it could be an element embedded within a sister of the V, as in (3):

(3) a. believe, V, [+ __ *that* S]

b.



But strict locality sharply distinguished subcategorization restrictions from selectional restrictions, the similar device introduced by Chomsky to analyze semantic cooccurrence restrictions. The selectional restrictions of a verb, for example, were permitted to access properties of the subject NP, but the strict subcategorization restrictions were not.

This matter was taken up anew by Kajita (1967), who argued that Chomsky's notion of strict locality was both too strong and too weak. In particular, Kajita (p. 96) argued, on the basis of contrasts like (4a,b), that subcategorizational domains should be extended to include a verb's subject:[1]

[1]This argument, of course, turns on the assumption that English has hierarchical clause structure, and not the flat structure assumed, for example, in a number of proposals for German, Japanese and other language with considerable word order freedom. Assuming the flat structure for clauses, the subject is accessible to a verb without modifying Chomsky's theory of strict subcategorization.

(4) a. That Kim was right bothered me.

b.*That Kim was right loved me.

Although contrasts like this might be explained away as semantic (selectional) in nature, there are other contrasts that perhaps make Kajita's point more convincingly:

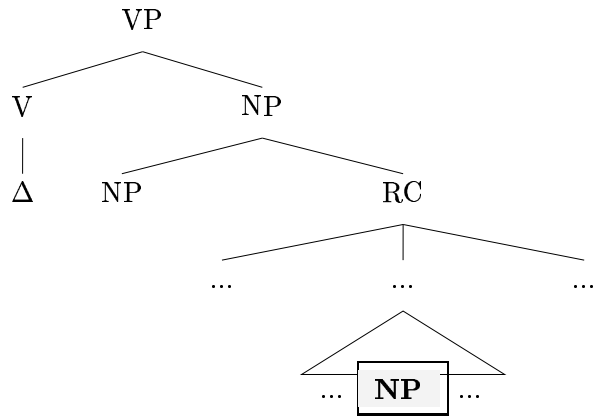
(5) a. The question of whether Kim was right perplexed me.

b. *?Whether Kim was right perplexed me.

In any case, it is now well established that many languages have verbs that select for a subject with idiosyncratic case properties (e.g. Icelandic verbs with 'quirky' dative, accusative or genitive subjects; Thrainsson (REF)). Hence, given that case information is (at least partly) syntactic in nature, permitting the syntactic selection of subjects, as Kajita suggested, provides the most straightforward account of quirky subject case and related phenomena.

Chomsky's strict locality was too weak, Kajita argued, because it allowed subcategorization restrictions to access elements deeply embedded within a verb's complement. For example, under Chomsky's definition of strict locality, an object within a relative clause would be locally accessible to a verb whose object was modified by that relative clause, as in (6):

(6)



Yet verbs, as far as we know, never select for such deeply embedded constituents, in any embedded clause. If there were such a language, it would allow lexical subcategorization restrictions for a hypothetical verb *fu* that would predict contrasts like the following:

(7) a. Lee **fued** someone who proved a theorem.

b.*Lee **fued** someone who ran into the room.

(8) a. Lee **fued** that someone proved a theorem.

b.*Lee **fued** that someone ran into the room.

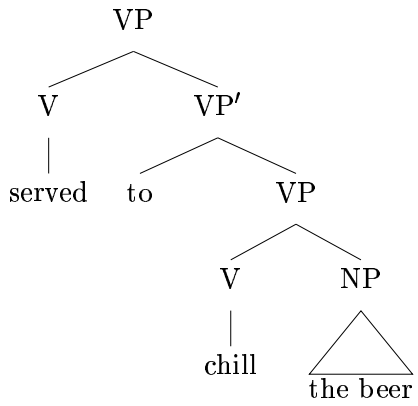
Kajita may well have been the first to point out that there is a serious theoretical issue involving the lower bound on subcategorization restrictions.

Kajita considered examples like (9), arguing that the verb *serve* requires an infinitival VP complement (an S, in his system) that contains a direct object NP:

(9) a. The ice served to chill the beer.

b.*The ice served to melt

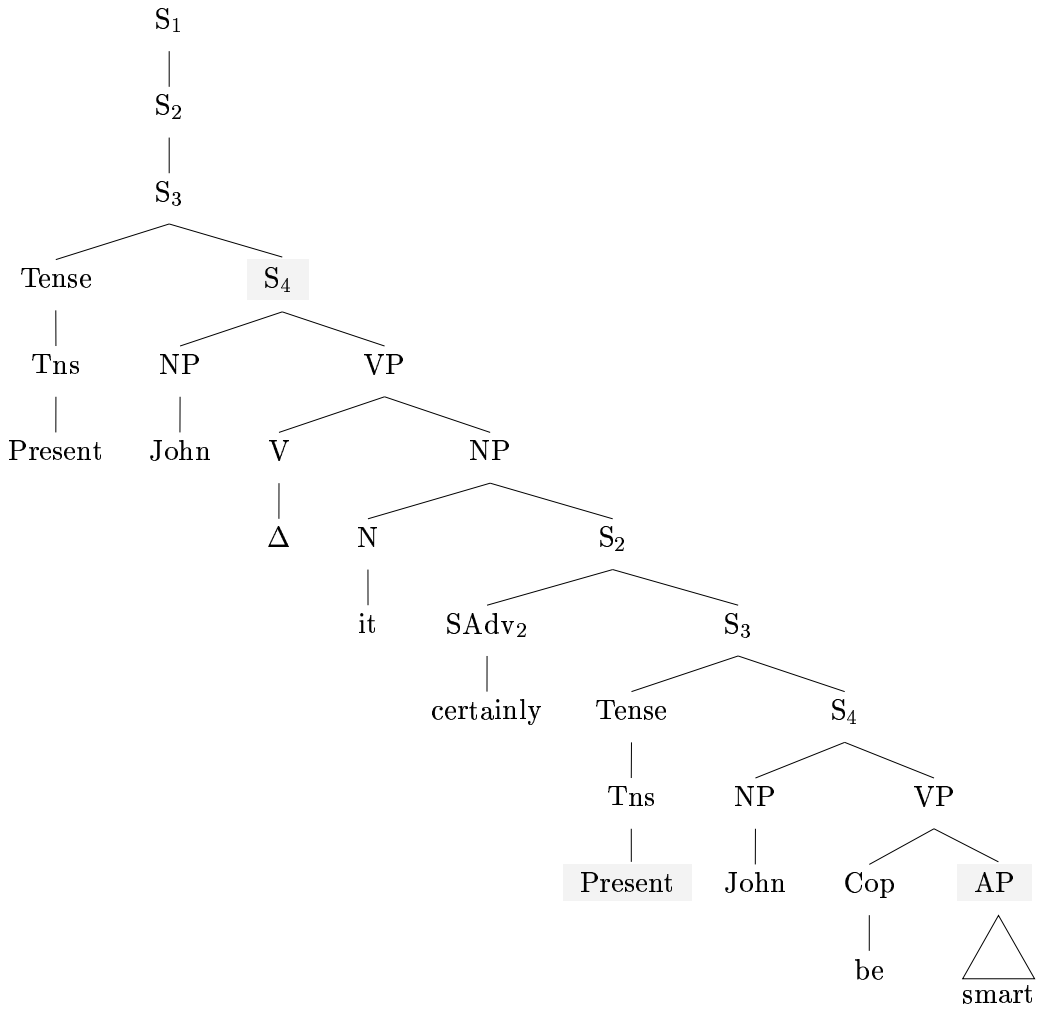
(10)



To accommodate this contrast and other data he considers, Kajita (1967: 105) proposed that the upper bound of a verb's subcategorization domain be the minimal S node that dominates if and that the lower bound be determined by a constraint requiring that the path from the upper bound to the selected constituent contain at most one S node.[2] Kajita's theory must be understood in terms of the particular theory of deep structure that he was assuming, which countenanced deep phrase markers like (11):

[2]I am loosely paraphrasing Kajita's theory, replacing his distinction 'width' and 'depth' of the subcategorization by 'upper' and 'lower bound'.

(11)

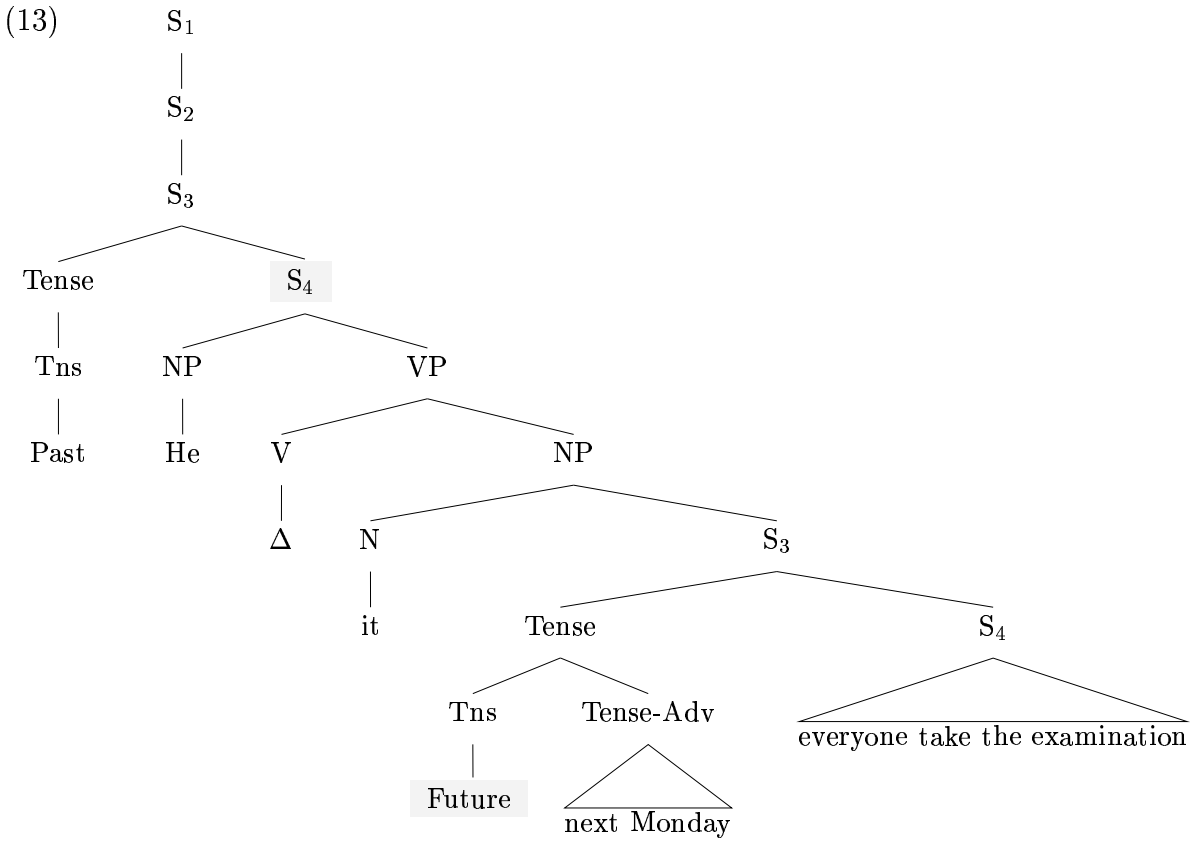


The intent here was to rule out the possibility of verb substituted for Δ whose lexical entry contained a subcategorization restriction that referenced, say, the Present Tense of the embedded clause or the AP within that clause's VP.

However, Kajita treated subjunctive selection in examples like (12) in terms of selection for future tense in structures like (13):

(12) He demanded that everyone take the examination next Monday.

(13)



Thus under his assumptions about clausal structure (which were justified in admirable detail), it was crucial that subcategorizatonal domains be allowed to cross exactly one sentential node.

However, Kajita's conclusions about the verb *serve* were reassessed by Higgins (1973: 173, nt. 5), who argued that the correct generalization is a semantic requirement: the unexpressed subject of the VP complement of the verb *serve* must be interpretable as an instrument. Shieber supported this conclusion by observing (p.c. reported in Pollard and Sag (1987, p. 145) that examples like the following follow Higgins' constraint, but not Kajita's:

(14) *Kim served to break the window with a hammer.

The verb *serve* thus imposes semantic constraints on the unexpressed subject of its VP complement, but makes no direct reference to the internal syntactic properties (e.g. the presence of an object NP) within that VP.

I am not aware of transformational studies that have sought to refine or update Kajita's conclusions. Indeed, the question of locality of subcategorization seems to have fallen by the wayside within mainstream generative grammar. It is important to realize, however, that '$\bar{\text{X}}$ Theory', as first developed in Chomsky 1970 (but cf. Harris 1946), bears on this question. A verb that is subcategorized by an NP complement (a transitive verb) is really selecting for a phrase with a nominal head. And $\bar{\text{X}}$ Theory, which relies on the reformulation of syntactic categories as feature structures, provides a way of projecting

the category information of the lexical head 'up' to its maximal projection (e.g. the maximal NP headed by a given noun, the maximal AP headed by a given adjective, etc.). $\bar{\text{X}}$ Theory thus plays a crucial role in considerations of locality – a verb can refer to the category features of the phrases it combines with, i.e. the phrases (NP, AP, etc.) that are sisters of the verb, and $\bar{\text{X}}$ Theory will ensure that those phrases will be headed by a word of the appropriate category.

These ramifications of $\bar{\text{X}}$ Theory played an important analytic role in Generalized Phrase Structure Grammar (GPSG). Gazdar (1981, 1982) argued that $\bar{\text{X}}$ Theory, with a slightly enriched inventory of syntactic features, provides the basis for a wholesale revision of linguistic theory, one that eliminates transformational operations altogether. GPSG researchers proposed that the 'HEAD' features, those whose specifications were passed up from head daughter to mother in a headed structure, included not only N and V, which (following Chomsky) were used to (coarsely) distinguish grammatical categories, but also the following:

(15) a. CASE (values in $\{nom, acc\}$, specified for all NPs, but distinguishing among pronouns, e.g. she vs. her),

b. VFORM (values in $\{fin, base, inf, prp, psp, ger\}$ distinguishing the various inflected forms of the V that heads a VP or S),

c. NFORM (values in $\{norm, it, there\}$, distinguishing referential nominals from dummies),

d. PFORM (values in $\{to, of, \text{loc}, \text{dir}\}$, distinguishing the various kinds of prepositions (and PPs) that can be subcategorized),

e. PRED (values in $\{+, -\}$, distinguishing the predicative Xs (and XPs) from their nonpredicative counterparts),

f. AUX (values in $\{+, -\}$, distinguishing the auxiliary verbs (and VPs) from their nonauxiliary counterparts), and

g. SLASH (values in sets or lists of categories, distinguishing 'complete' phrases from those that contain one or more unbound gaps of a particular kind).

With this feature inventory, the explanatory domain of $\bar{\text{X}}$ Theory is expanded to include not only the locality of category selection, but also the locality of case assignment, verb form government, selection of expletives, preposition selection, auxiliary selection, and the selection of phrases containing gaps of a particular kind (e.g. by *tough*-adjectives in English). Assuming that the values for these features are 'percolated up' from lexical heads to the phrases they project (by the Head Feature Principle (HFP), an uncontroversial principle of $\bar{\text{X}}$ Theory), the information relevant to all these phenomena becomes locally accessible to the lexical items that combine with those phrasal projections.

In other words, once $\bar{X}$ Theory and an expanded inventory of HEAD features was assumed, proponents of GPSG were able to reformulate grammar rules as shown in (16), where verbs are subcategorized only by properties of their sister constituents:[3]

(16) a. VP → H[1], where V[1] is in $\{$*walk*, *die*, *laugh*, ...$\}$

b. VP → H[2] , NP[*acc*], where V[2] is in $\{$*prove*, *clarify*, *reveal*, ...$\}$

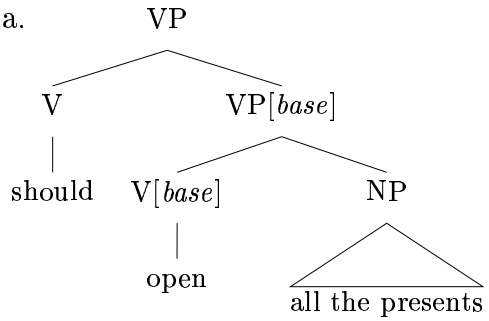
c. VP → H[3] , VP[*base*], where V[3] is in $\{$*can*, *should*, *may*, *must*, ...$\}$

d. AP → A[27] , VP[*inf*]/NP[*acc*], where A[27] is in $\{$*tough*, *easy*, ...$\}$

...

This 'context-free' theory of subcategorization relies on the HFP and other general principles (e.g. the Foot Feature Principle) to define the domain in which subcategorization restrictions hold, e.g. in structures like the following:

(17) a.

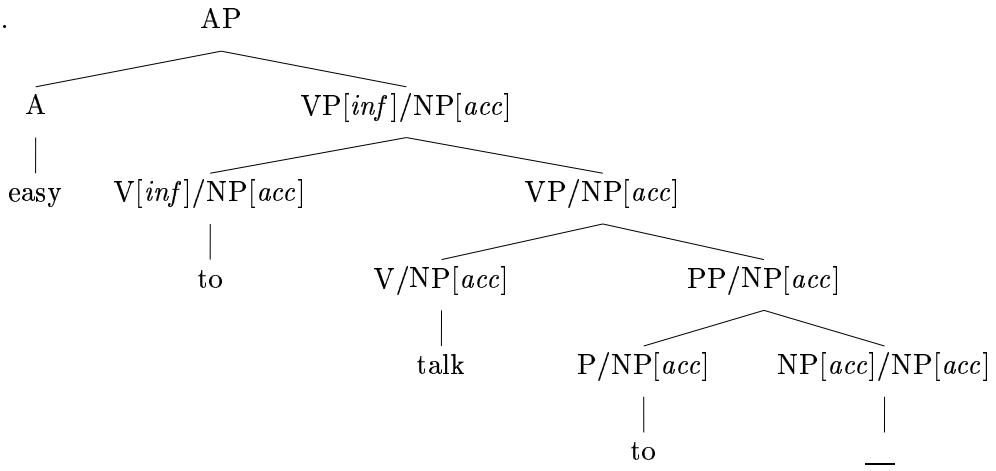


[3]Note that here the following abbreviations are used:

$$\text{V}[i] = [\textsc{subcat}\ \ i\ ] \qquad \text{NP}[acc] = \begin{bmatrix} \text{N} & + \\ \text{V} & - \\ \textsc{case} & \text{acc} \\ \textsc{bar} & 2 \end{bmatrix}$$

$$\text{VP}[base] = \begin{bmatrix} \text{N} & - \\ \text{V} & + \\ \textsc{vform} & base \\ \textsc{subj} & - \end{bmatrix} \qquad \text{VP}[inf]/\text{NP} = \begin{bmatrix} \text{N} & - \\ \text{V} & + \\ \textsc{vform} & inf \\ \textsc{subj} & - \\ \textsc{slash} & \text{NP}[acc] \end{bmatrix}$$

b.



In fact, given the possibility of modification and unbounded expansion of 'slashed' constituents, the domain over which subcategorization takes place is now in principle unbounded, as it should be, given examples like the following:

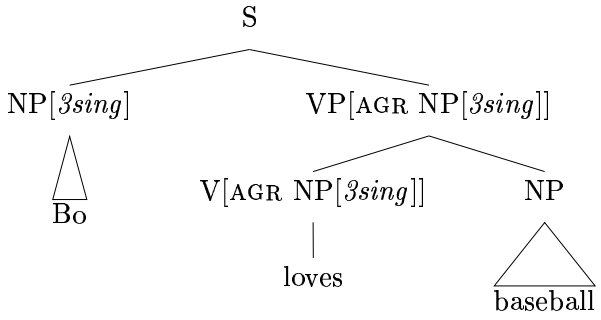
(18) a. Kim should [carefully [not [open all the presents]]].

b. People like Pat are hard [to [imagine [Sandy [being [willing [to [talk [to __]]]]]]]].

To put it somewhat differently, GPSG did not deny that there were long-distance dependency phenomena of the sort just illustrated. Rather, the claim made by GPSG (and also by the HPSG approach developed in subsequent sections) is that non-local dependency phenomena are a consequence of strictly local constraints (e.g. lexical specifications involving the category, meaning, case, etc. of a word's selected dependents) and their interaction with independent principles of grammar, such as the HFP.

GPSG accommodated subcategorization by subjects in terms of another HEAD feature AGR, which allowed a verb to 'pass up' information (again, via the HFP) to its VP projection, whose AGR value had to be identified with the subject NP, by a separate principle (the Control Agreement Principle):

(19)

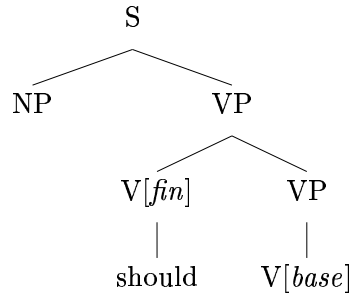


Note also that the difference between subjunctive verbs ([VFORM *base*]) and indicative verbs ([VFORM *fin*]) is projected by the HFP, thus providing an account of Kajita's example in (12) above in terms of context-free subcategorization: *believe* is subcategorized by an S[*fin*] complement; *demand* is subcategorized by an S[*base*]. GPSG was thus

able to achieve a theory of subcategorization that embodied a notion of locality quite similar to the one proposed by Kajita. The GPSG theory is not about deep structure phrase markers, of course; GPSG embraced the ambitious goal of generating surface syntactic structures directly. And in the GPSG theory, Kajita's domain stipulation, as well as the exceptions to it that must be countenanced in a surface-based subcategorization theory, are actually derived from the nature of the subcategorization mechanism and its interaction with independently motivated grammatical principles – a welcome result.

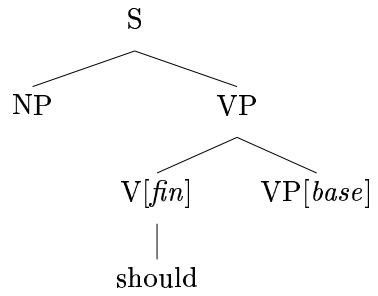
GPSG's approach to subcategorization is based on local trees and the decomposition of categories via syntactic features. The best known tree-based approach to subcategorization, however is probably Tree-Adjoining Grammar (TAG) (REF), which differs from GPSG in grounding sentence generation not in local trees, but rather in elementary trees that can be viewed as approximating Kajita's local domains:

(20)



In TAGs, lexically anchored elementary trees like this can undergo two kinds of operations: a tree structure can be substituted for any of the unexpanded nodes in (20) or else an auxiliary tree can be grafted into the middle of (20) by the adjunction operation. In this set-up, the question of locality is in essence the question of how deep elementary trees can be. One might attempt to retain $\bar{X}$ Theory within TAG, for example, and replace (20) with a shallower tree like (21), possibly providing a tighter theory of locality:

(21)



But one theory's tree is another's lexical stucture. In Categorial Grammar (CG), for instance, a notion of locality is built into the structure of lexical categories (assuming function application as the only mode of combination). The GPSG grammar rules in (16) above correspond to lexical entries like the following, where NP, S, and NP are abbreviations for feature structure categories similar to those considered above:

(22) a. S\NP: {*walk*, *die*, *laugh*, ... }

b. (S\NP)/NP[*acc*]: {*prove*, *clarify*, *reveal*, ... }

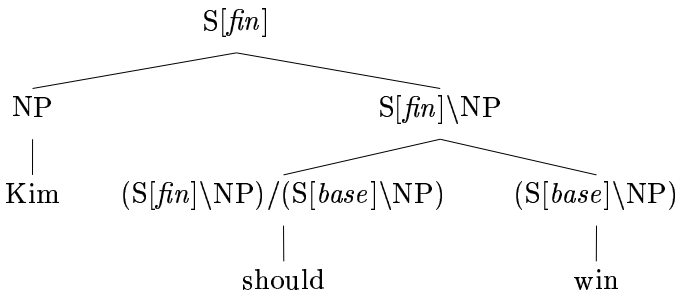
c. (S\NP)/(S[*base*]\NP): {*can*, *should*, *may*, *must*, ... }

d. (AP)/((S\NP)/NP[*acc*]): {*tough*, *easy*, ... }

...

Hence, assuming a simple regime of function application for the construction of basic sentences, as illustrated in (23), these lexical representations provide an extended locality domain for subcategorization that is, again, quite like Kajita's:

(23)



Here too, because of modification and the composition employed by, e.g., Steedman's (REF) analysis of filler-gap phenomena, subcategorization dependencies are extended over an unbounded domain in predictable ways.

In sum, grammatical theory must embody some hypothesis about the domain in which subcategorization dependencies hold. Such hypotheses involve basic lexical subcategorization restrictions which function within narrowly specified domains and which interact with other grammatical principles to account for the fact that local subcategorization domains are extended in fixed, precisely characterizable ways. The mechanisms for handling basic subcategorization dependencies vary from theory to theory, ranging from the pristine lexical categories of CG to the intricacies of Chomsky's (1965) proposal, which is formulated in terms of pre-terminal phrase markers, lexical substitution transformations, and a matching condition. The mechanisms for extending local subcategorization domains also varies from theory to theory: the work is variously done by transformations (Chomsky/Kajita's model), adjunction (TAGs), composition and modification (CG) and general principles of feature inheritance (GPSG). HPSG analyses, e.g. those discussed below, have attempted to integrate the basic lexical subcategorization mechanism of CG (reformulated slightly in terms of valence lists) with the general principles of feature inheritance that were pioneered within GPSG.

As we will see, the locality of basic subcategorization restrictions, and with it the locality of agreement, case assignment, and government, raises various issues in HPSG. In essence, however, the basic locality of these phenomena follows from the nature of the arguments on valence lists, as in CG. The issue of how to restrict the amount of structure that is 'visible' to grammar rules (which the issue of locality hinges upon in TAG) remains an interesting independent problem in HPSG that I will attempt to solve in section 4 below.

3 Locality in HPSG

3.1 The SYNSEM Locality Hypothesis

The background provided in the preceding section sets the stage for discussions of locality within HPSG. The feature geometry proposed by Pollard and Sag (1987) [henceforth P&S-87] (sketched in (24)), taken together with their Subcategorization Principle (sketched in (25)), failed to place sufficient constraints on which elements could be selected by a given word.[4]

(24) Feature Geometry of Pollard and Sag 1987:

$$\begin{bmatrix} \textit{phrase} \\ \text{PHONOLOGY} \\ \text{SYNTAX} \begin{bmatrix} \text{LOCAL} \begin{bmatrix} \text{HEAD} \\ \text{SUBCAT} \quad \textit{list}(\textit{sign}) \end{bmatrix} \\ \text{NONLOCAL} \quad \ldots \end{bmatrix} \\ \text{SEMANTICS} \\ \text{DTRS} \begin{bmatrix} \text{HD-DTR} \quad \textit{sign} \\ \text{COMP-DTRS} \quad \textit{list}(\textit{sign}) \\ \quad \ldots \end{bmatrix} \end{bmatrix}$$

(25) Subcategorization Principle (P&S-87: 71):

$$\left[\text{DTRS} \quad \textit{head-struc}\right] \Rightarrow \begin{bmatrix} \text{SYN}|\text{LOC}|\text{SUBCAT} \quad \boxed{A} \\ \text{DTRS} \begin{bmatrix} \text{HD-DTR} \quad \left[\text{SYN}|\text{LOC}|\text{SUBCAT} \quad \boxed{A} \oplus \boxed{B}\right] \\ \text{COMP-DTRS} \quad \boxed{B} \end{bmatrix} \end{bmatrix}$$

In this set-up, since phrasal signs have daughters, the elements on a verb's SUBCAT list do too. Hence a lexical entry could easily be written for a verb that is subcategorized for a VP complement that must contain a direct object NP, as Kajita would have wanted, or (even more permissively) for an S whose VP contained an S whose VP contained an object specified as, say, [CASE dative]. Early HPSG thus embodied little in the way of a theory of subcategorization locality.

The proposals made by Pollard and Sag (1994) [henceforth P&S-94] embodied an attempt to remedy this defect of their earlier system. By introducing the feature SYNSEM and the syntactico-semantic complexes ('synsem objects') that served as values of SYNSEM, P&S-94 were able to limit the information that was accessible under lexical selection:

[4] For uniformity of presentation, I here reverse the order of elements on SUBCAT lists from that assumed in P&S-87. The symbol '$\oplus$' denotes list concatenation (also referred to as the 'addition' or the 'appending' of two lists.

(26) Feature Geometry of Pollard and Sag 1994:

$$\begin{bmatrix} \textit{phrase} \\ \text{PHONOLOGY} \\ \text{SYNSEM} \begin{bmatrix} \textit{synsem} \\ \text{LOCAL} \begin{bmatrix} \text{CATEGORY} \begin{bmatrix} \text{HEAD} \\ \text{SUBCAT} \quad \textit{list}(\textit{synsem}) \end{bmatrix} \\ \text{CONTENT} \end{bmatrix} \\ \text{NONLOCAL} \ldots \end{bmatrix} \\ \text{DTRS} \begin{bmatrix} \text{HD-DTR} \quad \textit{sign} \\ \text{COMP-DTRS} \quad \textit{list}(\textit{sign}) \\ \ldots \end{bmatrix} \end{bmatrix}$$

This feature geometry worked together with a revised Subcategorization Principle, formulated in (27):[5]

(27) Subcategorization Principle (a formalization of P&S-94: 34):

$$\left[\text{DTRS} \quad \textit{head-struc}\right] \Rightarrow \begin{bmatrix} \text{SS}|\text{LOC}|\text{CAT}|\text{SUBCAT} \quad \boxed{A} \\ \text{DTRS} \begin{bmatrix} \text{HD-DTR}|\text{SS}|\text{LOC}|\text{CAT}|\text{SUBCAT} \quad \boxed{A} \oplus \texttt{s2s}(\boxed{B}) \\ \text{COMP-DTRS} \quad \boxed{B} \end{bmatrix} \end{bmatrix}$$

We may refer to the feature geometry in (26), taken together with the Subcategorization Principle in (27), as the SYNSEM Locality Hypothesis (SSLH).

The SSLH ensures that if a lexical entry includes a constraint on a member of the SUBCAT list, that constraint will apply to the SYNSEM value of the corresponding valent (subject, complement, or specifier) that that word cooccurs with. There is no direct access to information about any element that appears within those valents, e.g. a direct object within a VP complement, or an object within a sentential complement of a sentential complement. There is only indirect access to such elements whenever certain SYNSEM properties of a given valent are determined by or correlated with those of some element it contains, as discussed below.

The SSLH embodies a quite particular claim: taken together with a theory of what SYNSEM values are, it ensures that the grammatical constraints that concern the following phenomena all function within the same locality domain: **category selection** (strict subcategorization in Chomsky's sense), **case assignment**, **government** (of the form of a complement's head), and **(non-anaphoric) agreement**. In many clear cases, these predictions are correct, though there remain certain issues of controversy, some of which I discuss below.

[5]The function `s2s` (`signs-to-synsems`) maps a list of signs onto the corresponding list of synsem objects.

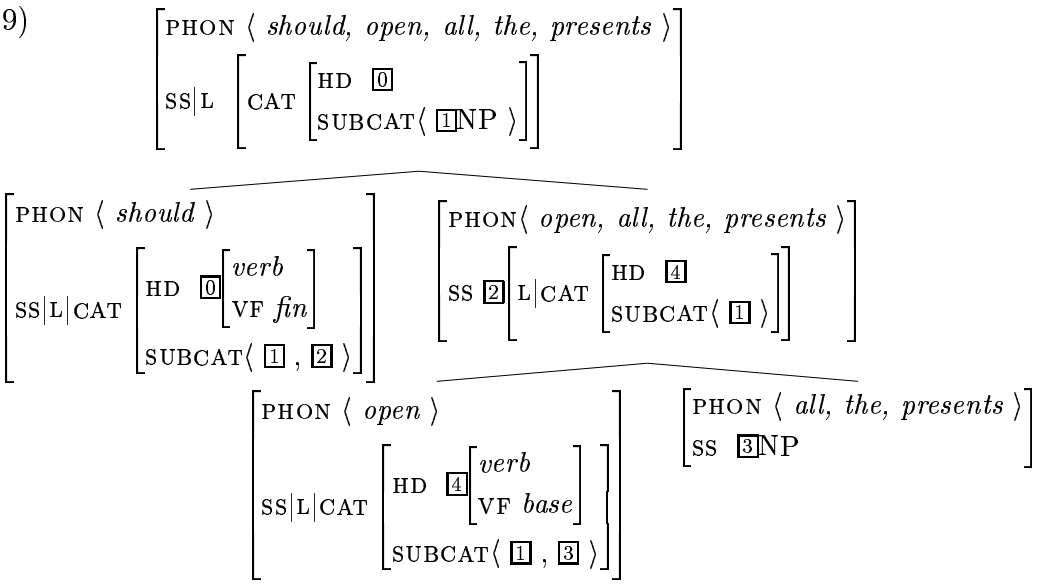
3.1.1 Category Selection

Since the complex categories of HPSG are specified within SYNSEM, the SSLP interacts with the HFP to make locality predictions similar to those made by earlier work in GPSG with regard to locality of form government and case assignment. That is, among HPSG's HEAD features are CASE, VFORM, PFORM, PRED, and AUX, and a phrase's value for HEAD must be the same as that of its head daughter in order for the HFP to be satisfied. For example, verbs like *depend* or *rely* require that the prepositional head within their PP complement be *on* or *upon* and this is ensured by a lexical specification like (28):

(28) $\left[\text{S|L|CAT|SUBCAT} \ \langle \ \text{NP} \ , \ \text{PP[PF } \textit{on}] \rangle\right]$

Similarly, the modal verbs select for a VP complement whose verbal head is specified as [VFORM *base*]. This will have the intended effect on the VP's head daughter, as sketched in (29):[6]

(29)



Note in addition that since a verb's SUBCAT list includes reference to its subject (the first valent), the domain of locality is automatically extended to include subjects without the introduction of the AGR feature discussed earlier in connection with GPSG.

[6]For ease of exposition, I will eliminate the feature NONLOCAL from feature structure descriptions. In fact, I believe the features LOCAL and NONLOCAL are unnecessary, a matter I will return to briefly in section 4 below. Here I abbreviate as follows: PHON = PHONOLOGY, L = LOCAL = LOC, SS = SYNSEM, CAT = CATEGORY, HD = HEAD , VF = VFORM, PF = PFORM. Roman caps (N, VP, etc.) are used to abbreviate feature structures of type *synsem*; italic caps (*NP*, *PP*, etc.) abbreviate feature structures of type *local*.

3.1.2 Agreement

To see how this set-up also imposes locality on agreement, consider the following well attested agreement patterns:

(30) a. Verb-subject agreement

b. Verb-object agreement

c. Noun-possessor agreement

d. Determiner-noun agreement

e. Modifier-modified agreement

Once verb-subject, verb-object, noun-possessor, and noun-determiner selection are all treated (as they have been in HPSG) in terms of selection features, the agreement types in (30a-d) are all naturally accommodated as constraints imposed by a lexical head. That is, a third-singular verb (e.g. *runs*) is simply a form that selects (via SUBCAT) for a third-singular subject; a feminine plural form of an adjective (in a Romance language, for example) is just a form that is lexically required (via MOD) to modify a feminine plural nominal, as sketched in (31):

(31) a. *runs*: [SS|L|CAT|SUBCAT ⟨NP[*3rd,sing*]⟩]

b. *grandes* (French feminine plural adjective): [SS|L|CAT|HEAD|MOD N[*fem,pl*]]

Assuming that agreement information is specified in terms of HEAD features like PERS, NUM and GEND, all such agreement is localized, by the SSLH.[7] However, recall from the discussion in section 2 that this means that agreement phenomena, like government, are local in their basic case, but will be extended within complex structures to induce indirect, long distance agreement, as guaranteed by other, independently motivated grammatical principles. Notable among such examples is the long-distance agreement of reflexive pronouns that is mediated by the theories of binding, control and SUBCAT selection in examples like (32):

(32) $\left\{\begin{array}{l}\text{They}\\ *\text{She}\\ *\text{He}\end{array}\right\}$[may [want [to [consider [trying [to [get **themselves** on the ballot]]]]]]] .

[7] If, however, agreement information is associated with semantic indices (as proposed by Pollard and Sag (1994)), then something more needs to be said to ensure locality of agreement. See below.

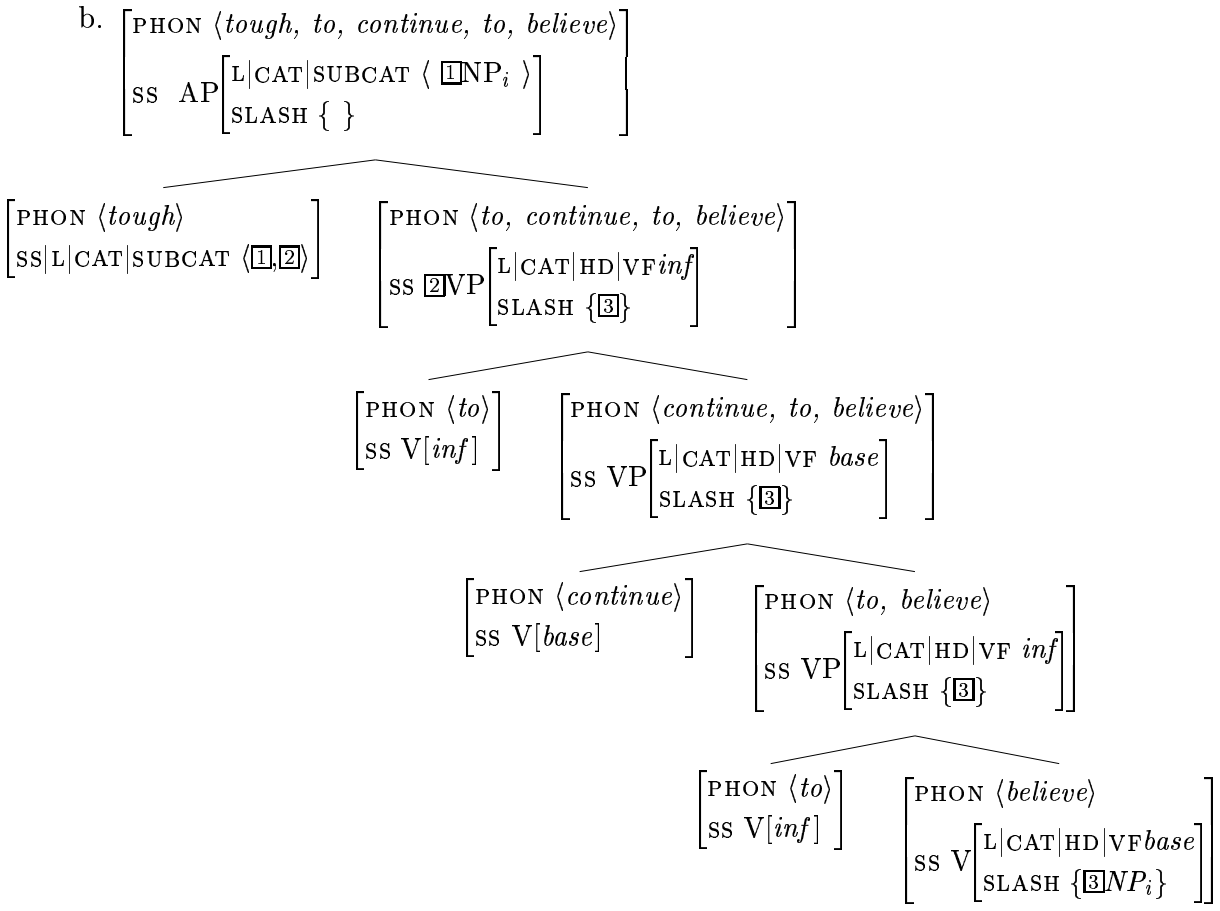
3.1.3 Lexical Gap Binding

This same mode of analysis extends to *tough*-adjectives, which select for an infinitival complement which must contain a gap. Because SLASH appears within SYNSEM, SLASH values are accessible under subcategorization, and so it is possible to formulate a lexical entry for *tough* that requires it to combine with a complement that contains an NP gap, moreover one that is coindexed with the subject of *tough*:

(33)
$$\textit{tough}\text{:}\ \left[\begin{array}{l}\text{SS|L|CAT|SUBCAT}\ \left\langle \text{NP}_i\ ,\ \left[\begin{array}{l}\text{L|CAT|HD|VF}\ \textit{inf}\\ \text{SLASH}\ \ \{NP_i\}\end{array}\right]^{\text{VP}}\right\rangle\end{array}\right]$$

General principles of HPSG theory require that the nonempty SLASH specification required by (33) will be 'passed down' into the VP complement of *tough* until a 'gap' position is encountered. The lexical specification in (33) thus entails only that a gap will be present somewhere within the VP complement, as illustrated in (34):

(34) a. (The claims being made are) tough to continue to believe __ .

b.


Note that under these assumptions it is not possible to write a lexical entry that selects for a gap appearing at some fixed level of embedding. That is, the 'localist' analysis of filler-gap dependencies that has emerged from the GPSG/HPSG tradition comes close to predicting (correctly, to the best of my knowledge) that no language will ever have an arbitrary depth filler-gap dependency. The positions in which the gap can appear are always determined by general constraints on the 'inheritance' of SLASH specifications.[8]

Finally, note further that by requiring that the information about filler-gap dependencies be locally encoded at each stage along the path connecting the filler and its gap(s) leads to the expectation that elements appeaing along this extraction path might be sensitive to the presence or absence of a filler-gap dependency. This prediction has in fact been confirmed now by a considerable number of languages where words or constructions exhibit local sensitivity to the presence or absence of a global filler-gap dependency. The relevant extraction-sensitive elements include complementizers in Irish, verbs bearing exhibiting a particular morphological pattern in Chamorro, Palauan, and Thompson Salish, suppression of tonal downstep in Ewe, and certain inversions in Spanish, French, and Yiddish. All of these phenomena point to the conclusion that information about a filler-gap dependency is systematically registered along the entire extraction path, and hence is accessible to verbs and complementizers that appear in these intermediate configurations. For an overview and further discussion, see Zaenen 1983, Hukari and Levine 1995, Bouma et al. 2001, and other refs?.

3.1.4 Semantic Selection

The SSLH circumscribes the information about a linguistic expression that is accessible for purposes of selection. Without further sources of constraint, however, an analysis based on SSLH does not yet predict the full range of information that is apparently unavailable to lexical selection. This is particularly apparent with respect to matters of semantic selection and semantic role assignment. Since synsem objects also include content, it follows that the semantic type of a head's complement can be selected for. For example, Ginzburg and Sag (2000) classify verbal complements in terms of the following semantic distinctions:

(35) a. *outcome* (**prefer**)

b. *proposition* (**claim**)

c. *question* (**wonder**)

d. *fact* (**regret**)

[8]This should be compared with a different approach that could also be incorporated within HPSG, namely the use of regular expressions to characterize the relation between fillers and gaps. Under this alternative (cf. its deployment within LFG under the rubric of 'functional uncertainty'), one could write a lexical entry that forced that gap to appear at some fixed depth within the infinitival complement of *tough*, an expressible, but cross-linguistically non-occurring possibility.

e. *prop-constr*, i.e. *question* or *fact* (**amazing**)

Since the relevant constructional schemata guarantee that the semantic objects (values of SS|CONT) that are assigned to declarative, interrogative and factive clauses are of the appropriate semantic type, verbs like those indicated in (35) can locally select for the semantically appropriate complement. Similarly, since semantic role assignment in HPSG is treated simply as the identification of a complement's INDEX value with one of the role arguments within the head's semantics, the index of a subcategorized element is always available for role assignment.

However, neither the feature geometry of P&S-94 nor that of Ginzburg and Sag (2000) circumscribes semantic information in such a way as to predict the fact that semantic selection and role assignment is universally local. For example, there is no verb (in any language, as far as we know) that requires its second argument to be a proposition built up from a relation whose second argument is itself a proposition. That is, no human language includes a verb *thlink* that would give rise to semantic contrasts like the following:

(36) a. Kim thlinks that Sandy believes that the earth is flat.

b.#Kim thlinks that Sandy died.

c.#Kim thlinks that Sandy loves Pat.

Similarly, there is no verb *fask* that requires its second argument to be an animate *wh*-question, determining semantic contrasts like (37):

(37) a. Bo fasked who left.

b.#Bo fasked whether Carrie had left.

c.#Bo fasked what Carrie had left.

The presumably universal impossibility of such verbs is not predicted by HPSG feature geometries, because the value of CONTENT is analyzed as a feature structure with complex internal structure. This means that the SSLH does not preclude formulating lexical entries for verbs that require nonlocal components of meaning to be selected or assigned a semantic role. Such lexical entries would be more complex than those that contain only local constraints. Hence, one might appeal to a 'simplicity measure' to treat such nonlocal semantic selection as more 'marked', i.e. less likely, in the world's languages. However, without a demonstration that the phenomenon of nonlocal semantic selection occasionally manifests itself in some human language, this approach to the problem seems too weak. Finally, note that the possibility of non-local semantic selection persists within the framework of Minimal Recursion Semantics (see Copestake et al. (to appear)), where all the semantic predications of embedded phrases are present on a phrase's RELATIONS list, and hence are 'locally visible', just as they are in earlier approaches to semantics within HPSG. Particular levels of embedding would be quite difficult to identify within MRS, where the order of elements on the RELATIONS list has no semantic significance. Hence one might argue that MRS provides sufficient prediction in this domain.

3.2 Some Analytic Issues

Research in HPSG has provided detailed analyses of various languages that have assumed an inventory of grammatical features that includes the HEAD features CASE, VFORM, PFORM, PRED, AUX, AGR, as well as others such as SUBCAT (or SUBJ and COMPS), CONTENT, and SLASH. As illustrated in the previous section, these features, along with the principles that govern their distribution, have led to analyses of category selection, form government, agreement, lexical gap binding, semantic selection, semantic role assignment and case assignment, that are stated exclusively in terms of local constraints. In particular, general grammatical principles like the HFP or the Nonlocal Feature Principle entail that the SS value of a given phrase includes feature specifications that can be required or disallowed by a governing head, according to the SUBCAT constraints specified in its lexical entry.

This general picture, of a head imposing restrictions directly on the SYNSEM values of its syntactic arguments (or of a modifier imposing such restrictions on the head it modifies), is perhaps now a hallmark of HPSG analyses. It is interesting, however, that a number of researchers have observed phenomena in a variety of languages that appear to defy analysis in terms of such local constraints. In this section, I will review a number of these, indicating ways that they can be reconciled with localist analysis.

3.2.1 Nonlocal Case Assignment in English

English *for/to* clauses present an interesting analytic challenge for the locality of case assignment. In order to analyze contrasts like the one in (38), it is necessary that an accusative case constraint be imposed somehow:

(38) a. I prefer [for [*they to be happy]]

b. I prefer [for [them to be happy]].

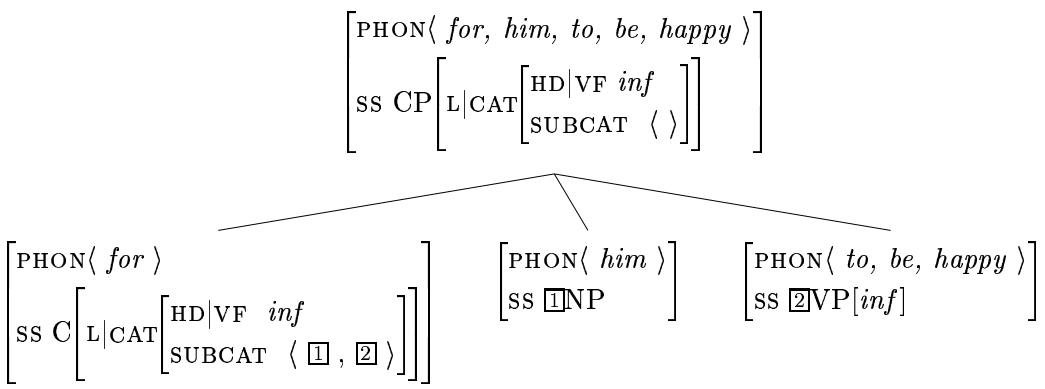
But given the standardly assumed structure in (38), the subject NP of the infinitive is not locally accessible to the complementizer *for*, which selects for the infinitival S either as a head (via SUBCAT) or as a marker (via SPEC). Nor can the infinitive marker *to* assign accusative case to its subject, as in examples like (39), that subject must be compatible with nominative case:

(39) [He/*Him seems [to be happy]].

A solution to this problem is proposed by Sag (1997), who argues that the standard structure for *for/to*-clauses should be replaced by the flat head-complement structure in (40):[9]

[9]Here and throughout this section, I have regularized valence features and the attendant feature geometry to conform with the preceding discussion.

(40)



Assuming this structure, rather than the one in (39), the lexical entry for the complementizer *for* can simply require that its first SUBCAT element be an accusative NP. The problematic NP is now locally accessible.

Moreover, the structure in (40) is independently motivated, for it provides an immediate account of contrasts like the following, first noted by Emonds (1976):

(41) a. Mary asked me [if, in St. Louis, [John could rent a house cheap]].

b. He doesn't intend [that, in these circumstances, [we be rehired]].

c.*Mary arranged for, *in St. Louis*, John to rent a house cheap.

d.*He doesn't intend for, *in these circumstances*, us to be rehired.

Assuming that only finite CPs have the traditional structure indicated in (41a-b), there is no constituent for the italicized modifiers to modify in (41c-d). The deviance of these examples follows from the same constraints that disallow the indicated modifiers in (42a-b), whose structure is analogous to the new *for/to*-clausal structure:

(42) a.*Kim persuaded *in St. Louis* Sandy to rent a house cheap.

b.*Lee believed *in these circumstances* Sandy to be in the right.

3.2.2 Case Stacking Languages

Perhaps the best-known examples of apparent nonlocal case assignment come from languages that allow case 'stacking', as in the following examples from Martuthunira, a Pama-Nyungan language:

(43) Ngayu nhuwa-lalha tharnta-a kupuyu-marta-a thara-ngka-marta-a.
1SG.NOM spear-PAST euro-ACC little-PROP-ACC pouch-LOC-PROP-ACC
'I speared a euro with a little one in its pouch.'
(Dench and Evans (1988))

(44) Ngunhu wartirra puni-lha ngurnu-ngara-mulyarra kanyara-ngara-mulyarra
the woman go-PAST that-PL-ALL man-PL-ALL
kapunmarnu-marta-ngara-mulyarra jirli-wirra-marta-ngara-mulyarra.
shirt-PROP-PL-ALL arm-PRIV-PROP-PL-ALL
'That woman went towards those men with shirts without sleeves.'
(Andrews 1996)

The operant generalization about these examples is that nominals within NPs are inflected not only in accordance with their local grammatical function, but also so as to reflect the function of the NPs that contain them. The unbounded case dependency phenomenon illustrated in (43)–(44) seems to pose a serious challenge for any locality hypothesis, and certainly for SSLH.

However, an elegant analysis of this phenomenon in terms of purely local constraints has been developed by Malouf (2000). Malouf proposes that in case stacking languages the value of the feature CASE is not an atomic case, but rather a list of such atoms. Assuming that nouns select for their NP dependents, the lexical entry for the noun *tharnt* 'euro'[10] looks like (45):

(45)
$$\begin{bmatrix} \textsc{phon} & \langle\ \text{tharnt-}\ \rangle \\ \textsc{ss|l|cat} & \begin{bmatrix} \textsc{head} & \begin{bmatrix} \textit{noun} \\ \textsc{case} & \boxed{B} \end{bmatrix} \\ \textsc{subcat} & \left\langle \text{NP}[\textsc{case}\ \langle \textit{prop} \rangle \oplus \boxed{B}\ ] \right\rangle \end{bmatrix} \end{bmatrix}$$

The key thing to see here is that every word formed from this stem will bear a particular case specification that is then passed on to the NP on that word's SUBCAT list.

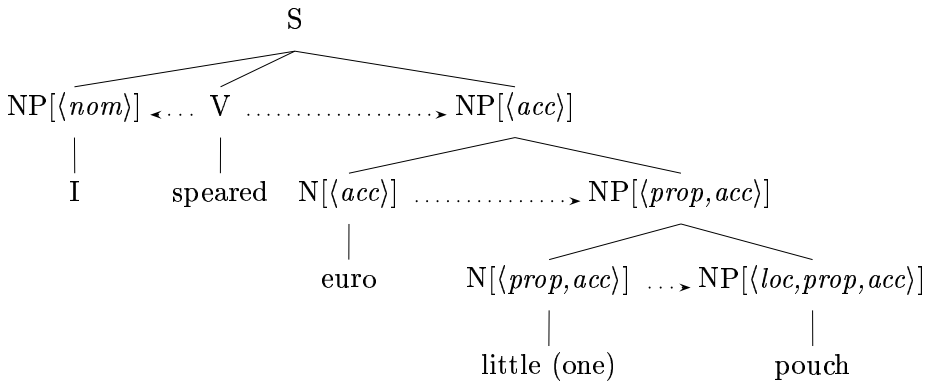
Malouf's treatment of nouns interacts with the analysis of verbs, which is sketched in (46):

(46)
$$\begin{bmatrix} \textsc{phon} & \langle\ \text{nhuwalalha}\ \rangle \\ \textsc{ss|l|cat} & \begin{bmatrix} \textsc{head} & \begin{bmatrix} \textit{verb} \\ \textsc{case} & \boxed{B}\langle\ \rangle \end{bmatrix} \\ \textsc{subcat} & \langle\ \text{NP}[\langle \textit{nom} \rangle \oplus \boxed{B}],\ \text{NP}[\langle \textit{acc} \rangle \oplus \boxed{B}]\ \rangle \end{bmatrix} \end{bmatrix}$$

Finite verbs bear an empty CASE specification. However, (46) is formulated so as to illustrate the general principle that lexical heads add their own CASE value to that of their dependents. As a result of this case addition, CASE values become longer with embedding, as shown in (47):

[10] A euro is a kind of marsupial distinct from kangaroos, wallabies, pademelons, and potoroos.

(47)



Long-distance case stacking is thus a consequence of CASE specifications that pass the case properties of a superordinate context down into a subordinate one, adding only the case information that reflects the local grammatical function of a given head-dependent combination. The morphological case inflections are thus based on local CASE specifications, just as they are in languages that lack case stacking. But when multiple case affixes are present (e.g. on *pouch* in (47), it follows that the CASE specification of the noun is non-singleton. This in turn entails that the immediately embedding syntactic context (e.g. little (one)) must introduce an appropriate case specification. Otherwise, the maximal NP in (47) would fail to meet the SUBCAT requirements of the verb *speared*. The local constraints of lexical items and general grammatical principles thus interact to guarantee a long-distance case dependency that is bounded only by the complexity of the embedding environment.

3.2.3 Selecting Words

Various researchers have discussed data indicating that selection via SUBCAT or other features must be able to distinguish words from phrases. P&S-87's discussion of English prenominal adjectives, for example, introduced the feature LEX: adjectival prenominal modifiers were required to be [LEX +]. However, this proposal also assigns the [LEX +] specification to adjectival phrases that contain, for example, internal premodifiers like *very*, allowing such phrases (e.g. *very tall*) to appear prenominally. Hence it is not really a proposal to allow selection for words.

Similarly, certain complementation patterns in French (e.g. the flat auxiliary structures first defended by Abeillé and Godard (1994) and the flat structures for the complements of causatives argued for by Abeillé et al. (1998)) seem to require an analysis in which a verb selects for a complement that is a single word (a past participle in the case of tense auxiliaries; an infinitive in the case of causatives). But here too it seems that the requisite distinction is not word vs. phrase. Rather, as shown by Abeillé and Godard (2000), the relevant phenomena are governed by a notion of grammatical weight that is only indirectly correlated with lexicality. Thus, there are 'light' phrases (coordinate structures or head-modifier strcutures, for example) that can appear in the same environments as the words selected by tense auxiliaries or causatives.

Much the same is true in the case of German verbal complexes, which Gunkel (2003: 159) suggests should be analyzed in terms of selection for objects of type *word*. However,

as pointed out to me by Stefan Müller (personal communication), coordinations of words appear in verbal complexes like (48):

(48) ...weil Peter Maria [lieben oder hassen] wird.
...because Peter Maria [love or hate] will
...because Peter will either love or hate Maria

The proper generalization here also seems to involve a set of entities that could be specified as [LEX +] (See Müller 2002 for a precisely worked out analysis along these lines). This would include words, but also certain phrases constructed via coordination or modification.

Given the current state of play, it might seem that there are no cases showing that a governing element must be able to select for a word, rather than a phrase. However, one such case is the emotive expressions that modify interrogative *wh*-words in various Germanic and Romance languages. In English, these include *the hell*, *the heck*, *in tarnation*, *in the world*, *the blazes*, *the devil*, *in god's name*, and so forth. Ginzburg and Sag (2000) discuss the following contrasts:

(49) a. [[Who] the hell] did you assign to that task?

b. [[[Who the hell]'s] book] did you assign?

c.*[[Whose book] the hell] did you assign?

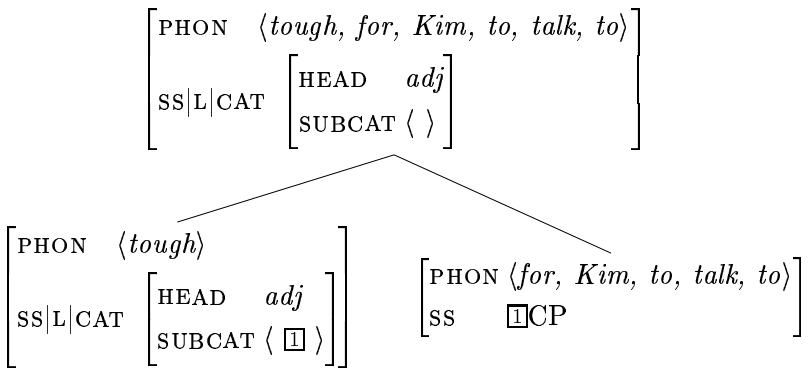
In light of these data, Ginzburg and Sag argue that the correct generalization is that the emotive modifiers are lexically specified (via the feature MOD) to modify an object of type *word*; hence the ungrammaticality of examples like (49c). If this analysis is correct, then values of the feature MOD make reference to the *word/phrase* distinction, which is impossible in the set-up of P&S-94.

3.2.4 The Role of Subjects

Earlier I mentioned the presumed locality of semantic role assignment. However, as a number of researchers have recently argued, there are phenomena in a variety of languages whose analysis requires, for example, that a verb selecting a sentential complement must be able to place constraints on the subject realized within that complement. Indeed, the majority of papers in the present volume are concerned with this and related issues.

The locality of role assignment within *tough*-constructions has been challenged by Levine (2000), who argues that the constituency of English APs headed by *tough* is as indicated in (50):

(50)



However, the subject within the complement clause is nonetheless a semantic argument of *tough*, a fact supported by the ungrammaticality of examples like (51b), where the subject of the *for/to*-clause is a nonreferential NP:

(51) a. Pat is tough [for Kim to talk to].

b. **tough**(**kim**,**talk-to**(**kim**,**pat**))

c.*Pat is [*tough* [for there to be a conversation with]].

Levine argues that the semantics of (51a) is as sketched in (51b), i.e. he argues that Kim is a semantic argument of the *tough* predication, even though it is not a syntactic argument of *tough*.

Similarly, Meurers (1999) discusses German examples (due originally to Haider; see Haider 1990) where fronted constituents like those in (52) and (53) are licensed by the case assigning properties of the finite verb:

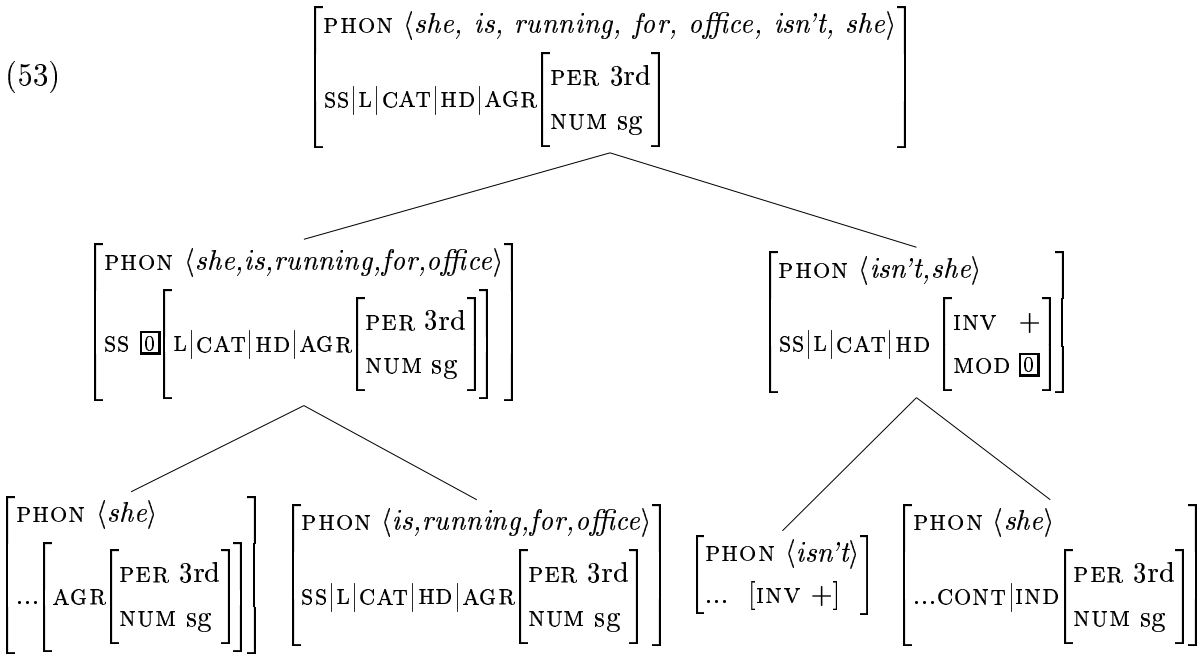
(52) a. [Ein Außenseiter gewonnen] **hat** hier noch nie __
An[nom] outsider won has here still never __
'An outsider has never won here yet.'

b. [Einen Außenseiter gewinnen] **läßt** Gott hier nie __
An[acc] outsider win lets god here never __
'God never lets an outsider win here.'

As Meurers argues at length, the category of the fronted phrase in these examples (which is identified with the category of the gap) must include information about the subject NP within the fronted phrase, so that the verb governing the gap can assign the appropriate case to it. For a related analysis, see Kathol 2002, and for a third proposal that converges with the approach developed below, see Meurers this volume.

In a similar vein, Bender and Flickinger (1999) analyze English tag questions by allowing agreement information about the subject NP to percolate to the clausal level, where it is accessible to the tag constituent, which they treat as an S-modifier:

(53)



Their analysis uses a lexical rule to make modifiers out of [INV +] auxiliary verbs. These modifiers select for a single complement whose *index* value is identified with the AGR value of the S to be modified. As Bender and Flickinger show, this treatment affords an analysis of the agreement patterns illustrated in (54), where the 'androgynous' pronoun *they* is referentially linked to the subject *everyone*, but engenders its own (AGR-based) plural agreement within the tag constituent:

(54) Everyone wins, don't they?

However, it remains to be seen whether the separation of AGR and INDEX will suffice to deal with the full range of agreement 'mismatches' discussed in the literature, for example the following:[11]

(55) a. Sears is having a sale, aren't they?

b. At least one of us is sure to win, aren't we?

c. The crowd is getting agitated, arent' they?

Despite this uncertainty, it seems clear that there is some (semantic) agreement relation that must be established between the two subjects here. Hence, as long as the structure for tags is along the lines sketched in (53), the agreement relation between the two subjects is non-local, i.e. it involves agreement between two elements that are neither sisters nor in a dependency relation.

[11] For discussion, see Oehrle 1987, Culicover 1992, and Kay 2002.

In sum, there appear to be numerous phenomena in a variety of languages that motivate modifying grammatical theory to allow external access to subjects realized within a clause. These include English 'copy raising' (Rogers 1974, Potsdam and Runner 2001, Asudeh 2002, Snider to appear), controlled pronominal subjects in Serbo-Croatian (Zec 1987), Halkomelem Salish (ref?) and other languages, raising across Polish prepositions (Przepiórkowski 1999), and complementizer agreement in Eastern Dutch dialects (Höhle 1997), to name just a few.

A principled solution to all of these problems, suggested independently by a number of these researchers, is the introduction of a feature (distinct from SUBCAT) that passes up to a given phrase information about one of the daughters used to construct that phrase. Kiss (1995) proposed such a feature for the subject of nonfinite verbal clauses in German, calling it SUBJECT, and this is the feature used by Levine and by Meurers. However, Fillmore et al. (to appear) motivate the use of a feature to make genitive pronouns realized within a given NP available for selection by elements outside that NP. In addition, the Polish preposition raising phenomenon discussed by Przepiórkowski (1999) and Dickinson (this volume?) appears to motivate an analysis where the object of certain prepositions is available for selection by elements external to the PP that the preposition projects.

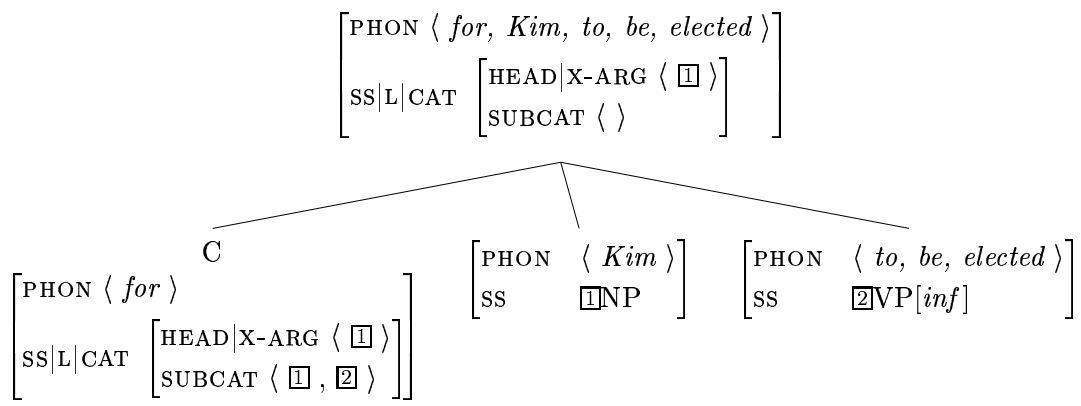
In sum, there is some variation as to which element within a phrase is externally accessible. Since 'subject' is too narrow a notion, empirically, I submit that SUBJECT is an inappropriate name for the feature in question. I propose instead that we borrow the name X-ARG from earlier HPSG analyses[12] and generalize its value as follows to provide a unified account of the phenomena just mentioned:

(56) The HEAD feature XARG is used to specify a distinguished element (e.g. subject, possessor, or object) within a given phrase. The value of XARG is either the empty list or a list containing exactly one sign.

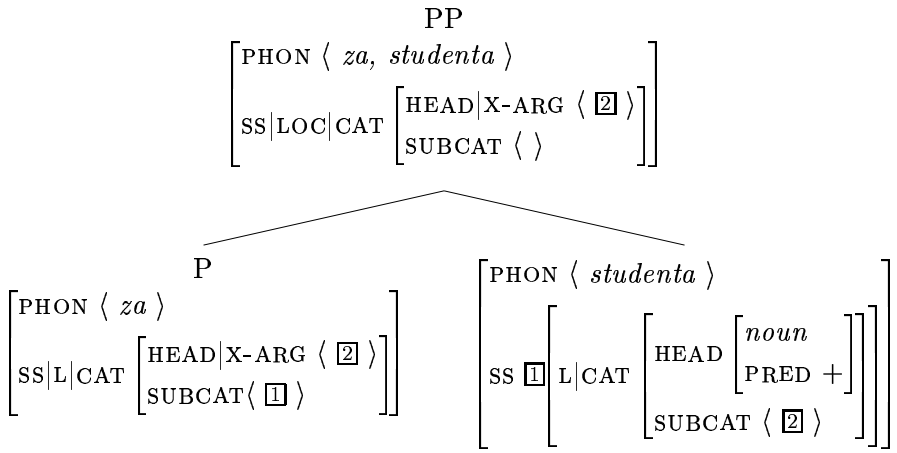
Because XARG is a HEAD feature, it percolates information about a designated phrasal constituent, as illustrated in (57):

[12]Sag and Pollard (1991) proposed a semantic feature EXTERNAL-ARGUMENT (XARG), which makes only the index of the subject argument available at the clausal level. This analysis has been incorporated into Minimal Recursion Semantics (and the English Resource Grammar) by Bender and Flickinger (2003).

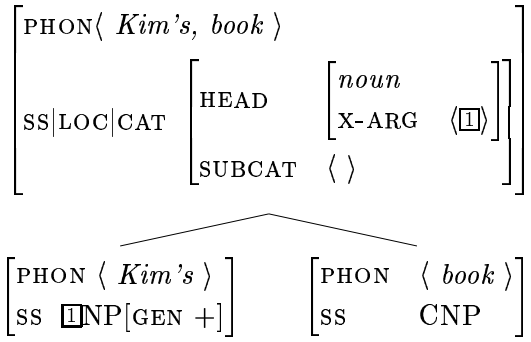
(57) a.



b.



c.



This in turn allows a strictly local treatment of role assignment into *for*/*to* clauses by *tough*-adjectives (along the lines suggested by Levine (2000)), of Polish prepositional

raising (see Dickinson (this volume)), and of English idioms with obligatory control of possessor pronouns, which has resisted analysis within the approach of P&S-94:

(58) a. He$_i$ lost his$_i$ way. (*her way/his$_j$ way...)

b. He$_i$ lost his$_i$ mind. (*her mind/his$_j$ mind...)

Once the category of NPs like *his way* includes information about the internal possessor, it is possible to treat the obligatory coreference in examples like (58) via local selection, as illustrated in (59):

(59)
$$\begin{bmatrix} \textsc{phon} & \langle\ \textit{lose}\ \rangle \\ \textsc{ss|loc|cat} & \begin{bmatrix} \textsc{head} & \textit{verb} \\ \textsc{subcat} & \left\langle \mathrm{NP}_i\ , \begin{bmatrix} \textsc{xarg} & \langle\ \mathrm{NP}[\textit{p-pro}]_i\ \rangle \end{bmatrix}^{\mathrm{NP}} \right\rangle \end{bmatrix} \end{bmatrix}$$

Finally, note that the external argument analysis also provides a new approach to the problem of how to unify the treatment of head-complement constructions. Recall that P&S-94's Schema 2 (the head-complement schema) required that the head daughter's SUBCAT list include a subject synsem that becomes the lone member of the mother's SUBCAT list. In other words, complements could be realized only if the head governing the complements also selected a subject. This problem was resolved in Chapter 9 of P&S-94 by separating SUBCAT into two distinct features SUBJ and COMPS (following suggestions made originally by Borsley (1987) and restricting Schema 2 to the realization of complements, not subjects.

But XARG provides another way to make the relevant distinction, without introducing separate features for subject and complement selection. Let us assume that the XARG value of verbal lexemes is a singleton list whose member is also the first member of the verb's SUBCAT list, as shown in the following lexical entry for the lexeme *love*:

(60)
$$\begin{bmatrix} \textit{lexeme} & & \\ \textsc{phon} & \langle\ \text{love}\ \rangle & \\ \textsc{ss|loc} & \begin{bmatrix} \textsc{cat} & \begin{bmatrix} \textsc{hd|xarg} & \langle\ \boxed{1}\ \rangle \\ \textsc{subcat} & \langle\ \boxed{1}\mathrm{NP}_i\ ,\ \mathrm{NP}_j\ \rangle \end{bmatrix} \\ \textsc{cont} & \textbf{love}(i,j) \end{bmatrix} \end{bmatrix}$$

By contrast, so-called 'case-marking' prepositions like *to* or *of* will have an empty XARG list:

(61)
$$\begin{bmatrix} \textit{lexeme} & & \\ \textsc{phon} & \langle\ \text{of}\ \rangle & \\ \textsc{ss|loc} & \begin{bmatrix} \textsc{cat} & \begin{bmatrix} \textsc{hd|xarg} & \langle\ \rangle \\ \textsc{subcat} & \langle\ \mathrm{NP}[\textsc{cont}\ \boxed{1}]\ \rangle \end{bmatrix} \\ \textsc{cont} & \boxed{1} \end{bmatrix} \end{bmatrix}$$

With these lexical contrasts in place, we can now analyze the different properties of VP and case-marking PPs without positing distinct head-complement constructions. We constrain the general head-complement schema as shown in (62):

(62) Schema 2: A phrase whose SUBCAT value is equal to its lexical head-daughter's XARG value.

What (62) says is that a head-complement phrase can be constructed from a lexical head daughter together with all the complements that it subcategorizes for, except the external argument, if there is one. The mother's SUBCAT value is then the head daughter's XARG list, which will be singleton in the case of a verb, but empty when the head daughter is a case-marking preposition (in English). This allows phrases like the following to be constructed:

(63) a.

$$\begin{bmatrix} \text{PHON} & \langle \text{ loves , Pat } \rangle \\ \text{SS|LOC} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HD|XARG} & \langle \boxed{1} \rangle \\ \text{SUBCAT} & \langle \boxed{1}\text{NP}_i \rangle \end{bmatrix} \\ \text{CONT} & \textbf{love}(i,j) \end{bmatrix} \end{bmatrix}$$

$$\begin{bmatrix} \text{PHON} & \langle \text{ loves } \rangle \\ \text{SS|LOC} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HD|XARG} & \langle \boxed{1} \rangle \\ \text{SUBCAT} & \langle \boxed{1}\text{NP}_i \text{ , } \boxed{2} \rangle \end{bmatrix} \\ \text{CONT} & \textbf{love}(i,j) \end{bmatrix} \end{bmatrix} \quad \begin{bmatrix} \text{PHON} & \langle \text{ Pat } \rangle \\ \text{SS} & \boxed{2}\text{NP}_j \end{bmatrix}$$

b.

$$\begin{bmatrix} \text{FORM} & \langle \text{ of , Pat } \rangle \\ \text{SS|LOC} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HD|XARG} & \langle\ \rangle \\ \text{SUBCAT} & \langle\ \rangle \end{bmatrix} \\ \text{CONT} & \boxed{2} \end{bmatrix} \end{bmatrix}$$

$$\begin{bmatrix} \text{FORM} & \langle \text{ of } \rangle \\ \text{SS|LOC} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{HD|XARG} & \langle\ \rangle \\ \text{SUBCAT} & \langle \boxed{1} \rangle \end{bmatrix} \\ \text{CONT} & \boxed{2} \end{bmatrix} \end{bmatrix} \quad \begin{bmatrix} \text{PHON} & \langle \text{ Pat } \rangle \\ \text{SS} & \boxed{1}\text{NP}[\text{LOC|CONT } \boxed{2}] \end{bmatrix}$$

Note that in both cases, the Subcategorization Principle of P&S-94 is satisfied.

3.2.5 Idiomatic Expressions

Idioms have also been thought to pose a locality issue. It is well known that certain idiomatic interpretations arise only when the particular pieces of the idiom are in construction with one another. The proper characterization of the notion of 'in construction

with', however, remains controversial. Since Nunberg et al. 1994, it has generally been agreed that syntactic flexibility is related to semantic decomposability. Thus a particularly decomposable idiom like *pull strings*, occurs flexibly in a variety of configurations, as illustrated in (64):

(64) a. Sandy *pulled strings* to get Kim the job.

b. *Strings* were *pulled* to get Kim the job.

c. The *strings* that seem likely to have been *pulled* to get Kim the job were an offense to man and nature.

d. We objected to the *strings* that Sandy had to *pull* to get Kim the job.

e. Sandy *pulled* the *strings* that got Kim the job.

f. The *strings* that Sandy *pulled*, nobody else could have *pulled*.

Idioms vary considerably in terms of their syntactic flexibility and it is perhaps unclear where to draw the line between an idiomatic sentence that should be allowed by the grammar and an extension of the grammar (or 'language play'). However, it is reasonably clear that copredication is a necessary condition for idiomaticity. That is, in order for *pull strings* to receive its idiomatic interpretation, the second semantic argument of *pull* must also have *strings* predicated of it, however the grammar allows for that to happen.

In recent work by Sailer (2003), the proposal is made to augment HPSG with lexical constraints (called 'conditions on lexical licensing' (COLL)) that can access arbitrarily distant elements within a given phrasal structure. As the data in (64) indicate, the COLL constraints associated with *pull* and *strings* will have to be highly schematic, making reference to pieces of a sentence's semantic structure that are quite remote. Sailer argues that the domain of COLL constraints should be the entire sentence (a sentential sign) in which the idiomatic word occurs. This is necessary, he claims, in order to describe what he takes to be purely syntactic restrictions on particular idiom 'chunks'.

That is, Sailer assumes that contrasts like the following should be grammatically stipulated:

(65) a. The strings that got Kim the job, nobody else could have pulled.

b.*The beans, John spilled.

To this end, he introduces elaborate machinery which includes a complex definition of the notion 'minimal clause', and stipulates of the idiomatic *spill* that it 'find' the idiomatic *beans* within its minimal clause, which excludes the topicalized position in examples like (65b).

Sailer's theory has clearly thrown to the winds all considerations of locality. His COLL conditions, given the way his theory is set up, have access to any aspect of the

global syntactic and semantic context. A theory of this sort allows lexical entries to be formulated with COLL conditions that do everything that I have been arguing in the previous sections that natural languages do not do. Is this theoretically backward step really warranted? I would like to suggest that it is not.

First of all, even accepting Sailer's claim about the data, his minimal clause condition emodies the wrong generalization, for it would wrongly exclude *wh*-interrogative instances of the idiom, e.g. (66), which are not hard to find on the internet:

(66) How many beans do you spill?[13]

(67) Uh....I don't know how many beans I can spill for you without having to kill you guys![14]

(68) I figure Lockhart and Cleland are coming forth to put their version on the record, because they're afraid of what beans Burkett will spill.[15]

(69) What beans did Deepthroat spill?[16]

(70) ... regret that trip! And what other beans did Mother spill?[17]

I think the proper account of whatever deviance inheres in (65b) has to do with semantic and pragmatic factors that make the idiomatic NP *the beans* ill-suited as the filler of a topicalizaed clause. We can find topicalized occurrences of *the beans* in poems without too much difficulty, as the following examples indicate:

(71) ...
Just misleading
Pleading purrs,
All cuddly kitten,
'I'll have tea first,'

And then I'll tell,
The beans I'll spill,
Of daring do...
But he never will![18]

[13][http://m1.mny.co.za/MBTax.nsf/0/C2256A2A0059A83942256AA2002B4801?OpenDocument] (March 20, 2005)
[14][www.cinecon.com/interviews/mattcedeno-part2.html](March 20, 2005)
[15][www.indcjournal.com/cgi-bin/mt/mt-comments.cgi?entry_id=974](March 20, 2005)
[16][www.talkaboutgovernment.com/group/ alt.politics.clinton/messages/1313165.html](March 20, 2005)
[17][www.halhkmusic.com/toogood/c11.html](March 20, 2005)
[18][http://www.moggies.co.uk/html/george.html](March 21, 2005)

(72) ...
Bloggers unite!
Defend what is right.
Keep the pressure on the CBS chiefs.
Annika want their collective head,
Rather's career dead, dead, dead:
Go thee hence and read her briefs.

The forgery stain
Will surely remain
Unless CBS shows its goodwill.
Give us the name!
Free speech will remain!
Divulge! Reveal! The beans you must spill![19]

But even if such examples all turn out to be instances of 'poetic license', it is still true that the syntactic COLL constraint offered by Sailer fails to draw the appropriate distinction. A semantic/pragmatic approach, as urged by Nunberg et al. (1994) seems to be preferable. My hope would be that whatever nonlocal licensing must be recognized for idiomatic expressions can be confined to constraints on the semantics of such expressions, which, as noted in section 3.1.4 above, is already a possible locus of nonlocal selection.

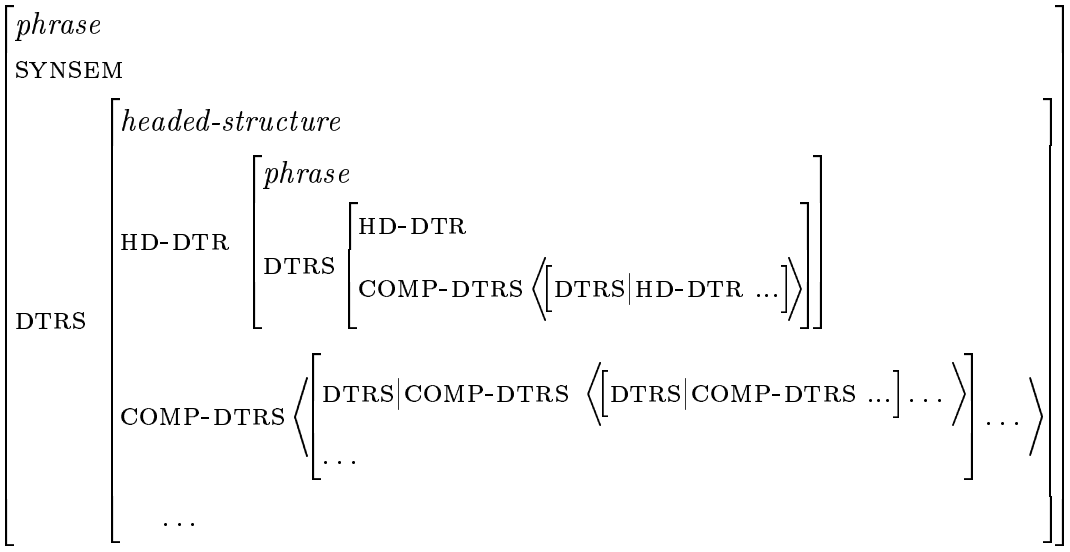
4 Locality of Construction

4.1 Background

Since the inception of work in HPSG, it has been assumed that there are two kinds of signs – words and phrases, with the feature DAUGHTERS (DTRS) declared appropriate for the type *phrase*. Grammar schemata were introduced in PS-94 as the HPSG analog of grammar rules. These schemata specified an inventory of phrase types, where phrases had the geometry shown in (73):

[19][nakedvillainy.com/archive_dir/2004_09_01_nakedvillainy_archive.html](March 21, 2005)

(73) Pollard and Sag 1994 (Chap. 9):



There was nothing in these schemata that imposed any notion of locality. Nothing but an unspoken 'gentleman's agreement' prevented the HPSG grammarian from writing schemata that directly reference the mother and a daughter's daughters, or in fact elements that appear at any arbitrary depth of embedding. HPSG had evolved far from its GPSG (CFG) roots, an evolutionary path that did not go unnoticed. For example, Copestake (1993) observes that:

> [...] it is unclear that the HPSG account of phrasal signs as feature structures which incorporate their daughters is the best one to adopt. Constraint resolution can be used to perform operations which cannot be straightforwardly mimicked by more conventional grammar rules. [...]. However, it is not clear to me whether HPSG currently takes advantage of this possibility in any very significant way. There have to be good reasons to adopt an approach which makes most known parsing technology inapplicable.

Copestake's observation is probably still correct today, leaving aside work on linearization theory,[20] which uses a DOMAIN feature to allow 'liberation' of embedded elements to higher levels of phenogrammatical structure. Apart from this line of research, which is in any case argued to be empirically deficient by Müller (2005, to appear), there are to my knowledge no HPSG analyses that propose a grammatical schema making direct reference to embedded structure. The practice of the HPSG community seems to adhere to something like the notion of locality that is inherent in CFGs.

[20] See, for example, Reape 1994, 1996 and Kathol 2000.

4.2 Signs, Constructions, and Constructs

I would like to propose that HPSG theory be modified so as to incorporate the strong constraints of the actual practice of the HPSG community. To this end, I suggest first that phrases not be endowed with the feature DTRS. Phrases, like words, specify values for the features PHONOLOGY, SYNTAX, and SEMANTICS. Second, I propose to distinguish signs from the constructions that license them. (What I mean by this will become clear in a moment.)

A construction, like a schema in PS-94, is intuitively a constraint defining a local pattern of sign combination. That is, a construction places restrictions on what properties signs must have if they are to directly combine with one another and in addition puts constraints on the sign that results from such a combination. On this conception, a construction is essentially a fancy context-free grammar rule. The objects defined by constructions are thus configurations of signs: a set of daughter signs and one more sign that is the mother of those daughters. Let us call each such configuration a 'construct'. Constructs may be thought of as local trees licensed by some construction:[21]

(74)

$$\begin{bmatrix} \text{PHON} & \langle\ \text{Kim}\ ,\ \text{walks}\ \rangle \\ \text{SYN} & \text{S} \\ \text{SEM} & \mathbf{walk}(\mathbf{k}) \end{bmatrix}$$

$$\begin{bmatrix} \text{PHON} & \langle\ \text{Kim}\ \rangle \\ \text{SYN} & \text{NP} \\ \text{SEM} & \mathbf{k} \end{bmatrix} \qquad \begin{bmatrix} \text{PHON} & \langle\ \text{walks}\ \rangle \\ \text{SYN} & \text{V} \\ \text{SEM} & \mathbf{walk} \end{bmatrix}$$

Notice that we may now return to a simpler feature geometry like the one in PS-87, eliminating the feature SYNSEM. In addition, with no distortion of the grammar's intended effect, we may reformulate constructs as feature structures:

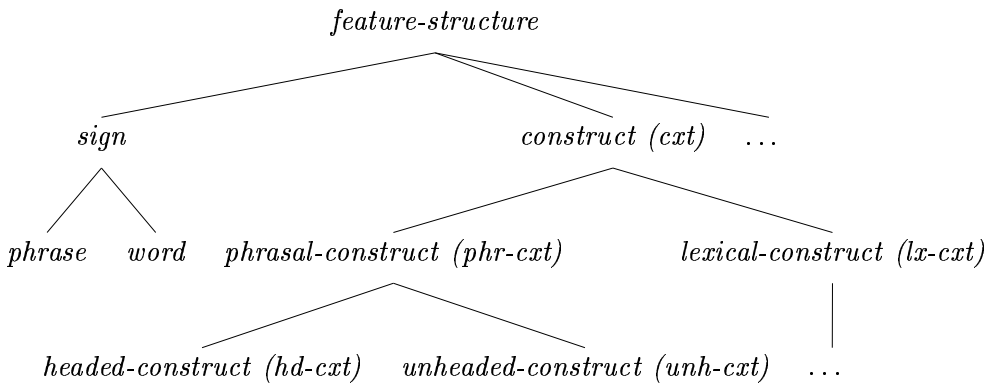
(75)

$$\begin{bmatrix} \textit{phr-cxt} & & \\ \text{MTR} & \begin{bmatrix} \textit{phrase} & \\ \text{PHON} & \langle\ \text{Kim}\ ,\ \text{walks}\ \rangle \\ \text{SYN} & \text{S} \\ \text{SEM} & \mathbf{walk}(\mathbf{k}) \end{bmatrix} & \\ \text{DTRS} & \left\langle \begin{bmatrix} \text{PHON} & \langle\ \text{Kim}\ \rangle \\ \text{SYN} & \text{NP} \\ \text{SEM} & \mathbf{k} \end{bmatrix}, \begin{bmatrix} \text{PHON} & \langle\ \text{walks}\ \rangle \\ \text{SYN} & \text{V} \\ \text{SEM} & \mathbf{walk} \end{bmatrix} \right\rangle \end{bmatrix}$$

[21] The SYNTAX and SEMANTICS values here are simplified for expository purposes only.

This last move is in fact easily achieved by the grammar signature sketched in (76) and (77):

(76)



(77) a. $cxt : \begin{bmatrix} \textsc{mother} & sign \\ \textsc{dtrs} & list(sign) \end{bmatrix}$

b. $ph\text{-}cxt : \begin{bmatrix} \textsc{mother} & phrase \end{bmatrix}$

c. $hd\text{-}cxt : \begin{bmatrix} \textsc{hd-dtr} & sign \end{bmatrix}$

d. $sign : \begin{bmatrix} \textsc{phon} & list(phon\text{-}structure) \\ \textsc{form} & list(morph\text{-}form) \\ \textsc{syntax} & syn\text{-}obj \\ \textsc{semantics} & sem\text{-}obj \end{bmatrix}$

Of course, this system of grammar doesn't define complex expressions until we include a principle like the the following, which allows recursive application of constructions:

(78) The Principle of Construction (PoC):

Given a grammar that includes a lexicon Lex (a set of 'basic' lexical entries) and a set of Constructions Cxx (a set of construct descriptions), Φ is a sign only if:

a. there is some lexical entry L in Lex such that Φ satisfies L, or

b. there is a construction in C in Cxx and a construct c satisfying C, such that c's MOTHER value is Φ.

It follows from this theory, as a matter of principle, that a construction cannot have direct access to properties of a mother and its granddaughters. If we observe that there is some such dependency, then we must provide an analysis in terms of some property of

the granddaughter that is systematically encoded on the daughter, and hence rendered locally accessible at the higher level. This has the virtue of making explicit exactly where nonlocality resides in linguistic descriptions.

The fundamental principles of P&S-94 are now recast as constraints on constructions, as shown in (79):[22]

(79) Head Feature Principle

$$\textit{hd-cxt} \Rightarrow \begin{bmatrix} \text{MTR} & \left[\text{SYN}|\text{LOC}|\text{CAT}|\text{HEAD} \quad \boxed{1}\right] \\ \text{HD-DTR} & \left[\text{SYN}|\text{LOC}|\text{HEAD} \quad \boxed{1}\right] \end{bmatrix}$$

Subcategorization Principle

$$\textit{hd-cxt} \Rightarrow \begin{bmatrix} \text{MTR} & \left[\text{SYN}|\text{LOC}|\text{CAT}|\text{SUBCAT} \quad \boxed{A}\right] \\ \text{DTRS} & \boxed{B} \bigcirc \langle \boxed{1} \rangle \\ \text{HD-DTR} & \boxed{1}\left[\text{SYN}|\text{LOC}|\text{SUBCAT} \quad \boxed{A} \oplus \boxed{B}\right] \end{bmatrix}$$

Note that the Subcategorization Principle is stated here without appeal to the *signs-to-synsems* relation.

Finally, this proposal also provides a new way of making sense of lexical rules, i.e. by treating them as varieties of lexical construction. We may posit three subtypes of lexical construct: *inflectional-construct*, *derivational-construct*, and *post-inflectional-construct*, each with its own properties. Following in the main Sag et al. 2003 (see especially chapter 16), we may assume that lexical entries describe feature structures of type *lexeme*, not *word*. Hence derivational constructions involve constructs (of type *deriv-cxt*) whose mother is of type *lexeme*; inflectional constructions involve unary constructs (of type *infl-cxt*) whose mother is of type *word* and whose daughter is of type *lexeme*; and post-inflectional constructions involve unary constructs (of type *post-infl-cxt*) where both mother and daughter are of type *word*. Instances of these constructions (i.e. constructs that satisfy these constructions) are shown in (80a-c):[23]

[22] '$\bigcirc$' is Reape's domain union operator: '$\boxed{A} \bigcirc \boxed{B}$' is satisfied by any list containing exactly the elements of $\boxed{A}$ and $\boxed{B}$, as long as any α which precedes some β in $\boxed{A}$ or in $\boxed{B}$ also precedes β in $\boxed{A} \bigcirc \boxed{B}$. '$\bigcirc$' is thus a 'shuffle' operator.

[23] The construction illustrated in (80c) is the constructional analogue of the Adverb Addition lexical rule first proposed by Van Noord and Bouma (1994). For further discussion of this construction, see Levine 2002 and Sag 2005.

(80) a.
$$\begin{bmatrix} \textit{deriv-cxt} \\ \textsc{mtr} & \begin{bmatrix} \textit{lexeme} \\ \textsc{form} & \langle \text{ finance , committee } \rangle \\ \textsc{syn} & \text{N} \end{bmatrix} \\ \textsc{dtrs} & \left\langle \begin{bmatrix} \textit{lexeme} \\ \textsc{form} & \langle \text{ finance } \rangle \\ \textsc{syn} & \text{N} \end{bmatrix}, \begin{bmatrix} \textit{lexeme} \\ \textsc{form} & \langle \text{ committee } \rangle \\ \textsc{syn} & \text{N} \end{bmatrix} \right\rangle \end{bmatrix}$$

b.
$$\begin{bmatrix} \textit{infl-cxt} \\ \textsc{mtr} & \begin{bmatrix} \textit{word} \\ \textsc{form} & \langle \text{ walks } \rangle \\ \textsc{syn} & \text{V} \end{bmatrix} \\ \textsc{dtrs} & \left\langle \begin{bmatrix} \textit{lexeme} \\ \textsc{form} & \langle \text{ walk } \rangle \\ \textsc{syn} & \text{V} \end{bmatrix} \right\rangle \end{bmatrix}$$

c.
$$\begin{bmatrix} \textit{post-infl-cxt} \\ \textsc{mtr} & \begin{bmatrix} \textit{word} \\ \textsc{form} & \langle \text{ reads } \rangle \\ \textsc{syn|l|cat|subcat} & \langle \text{ NP , NP , ADV } \rangle \end{bmatrix} \\ \textsc{dtrs} & \left\langle \begin{bmatrix} \textit{word} \\ \textsc{form} & \langle \text{ reads } \rangle \\ \textsc{syn|l|cat|subcat} & \langle \text{ NP , NP } \rangle \end{bmatrix} \right\rangle \end{bmatrix}$$

This proposal thus provides a unified approach to the construction of words and phrases, allowing for hierarchical generalizations of varying grain, without the need for ancillary devices.

4.3 Further Issues

There are a number of points to be made about the revised version of HPSG theory presented in the preceding section. First, recall that in order to derive locality of selection, previous versions of HPSG had to make use of constraints stated in term of the relation *signs-to-synsems*, e.g. as in (27) above. But now that we have removed DTRS from feature structures of type *phrase*, signs contain no information about the daughters that were used to construct them. The elements on SUBCAT lists (or ARG-ST lists, or those of whatever selection features we posit) can now simply be signs. This permits a simple, previously unavailable analysis of the *a/an* alternation in English in terms

of phonological selection. Further (as noted above), locality now follows from subcategorization constraints that are stated without using *signs-to-synsems*, as illustrated in (81):

(81) a.

$$\textit{hd-comp-cxt} \Rightarrow \begin{bmatrix} \text{MTR} & \left[\text{SYN|L|CAT|SUBCAT } \boxed{A}\right] \\ \text{DTRS} & \left[\text{FIRST } \boxed{1}\right] \\ \text{HD-DTR} & \boxed{1}\begin{bmatrix} \textit{word} \\ \text{SYN|L|CAT|HD|XARG } \boxed{A} \end{bmatrix} \end{bmatrix}$$

b.

$$\textit{hd-subj-cxt} \Rightarrow \begin{bmatrix} \text{DTRS} & \left[\text{REST } \begin{bmatrix} \text{FIRST } \boxed{0}\left[\text{SYN|L|CAT|HEAD|INV } -\right] \\ \text{REST } \langle\ \rangle \end{bmatrix}\right] \\ \text{HD-DTR} & \boxed{0} \end{bmatrix}$$

Note that the constraint in (81a) requires the mother's SUBCAT value to be the head (first) daughter's XARG value, allowing a case-marking preposition to project a saturated phrase just in case its XARG value is the empty list, as described above. And the constraint in (81b) stipulates only that the head (second) daughter be [INV −] and that there is no third daughter. These constraints interact with the HFP and the Subcategorization Principle to derive constructs like those illustrated in (63) above and also standard constructs like (82):

(82)

$$\begin{bmatrix} \text{PHON } \langle\ \textit{Kim, left}\ \rangle & \\ \text{SYN|L|CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{verb} \\ \text{X-ARG } \langle\boxed{1}\rangle \end{bmatrix} \\ \text{SUBCAT} & \langle\ \rangle \end{bmatrix} \end{bmatrix}$$

$$\boxed{1}\begin{bmatrix} \text{PHON} & \langle\ \textit{Kim}\ \rangle \\ \text{SYN} & \text{NP} \end{bmatrix} \qquad \begin{bmatrix} \text{PHON} & \langle\ \textit{left}\ \rangle & \\ \text{SYN|L|CAT} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{verb} \\ \text{X-ARG } \langle\boxed{1}\rangle \end{bmatrix} \\ \text{SUBCAT} & \langle\boxed{1}\rangle \end{bmatrix} \end{bmatrix}$$

The revisions I have proposed here, in addition to guaranteeing locality of selection and locality of construction, also fit well with a number of simplifications that are naturally envisaged for HPSG theory, including the elimination of the features SUBJ, COMPS, SPEC, HEAD, NONLOCAL, and LOCAL, leaving a feature geometry for signs looking like (83), or even (84):

(83)
$$\begin{bmatrix}
\textsc{phon} & \textit{list}(\textit{phon-structure}) & \\
\textsc{form} & \textit{list}(\textit{morph-form}) & \\
\textsc{syn} & \begin{bmatrix}
\textsc{cat} & \begin{bmatrix} \textsc{xarg} & \textit{list}(\textit{sign}) \\ \textsc{subcat} & \textit{list}(\textit{sign}) \\ \ldots & \end{bmatrix} \\
\textsc{slash} & \textit{list}(\textit{sign}) & \\
\textsc{wh} & \textit{list}(\textit{sign}) & \\
\ldots & &
\end{bmatrix} \\
\textsc{sem} & \ldots & \\
(\textsc{arg-st} & \textit{list}(\textit{sign})) & \\
\ldots & &
\end{bmatrix}$$

(84)
$$\begin{bmatrix}
\textsc{phon} & \textit{list}(\textit{phon-structure}) & \\
\textsc{form} & \textit{list}(\textit{morph-form}) & \\
\textsc{syn} & \begin{bmatrix} \textsc{xarg} & \textit{list}(\textit{sign}) \\ \textsc{subcat} & \textit{list}(\textit{sign}) \\ \ldots & \end{bmatrix} \\
\textsc{sem} & \ldots & \\
\textsc{slash} & \textit{list}(\textit{sign}) & \\
\textsc{wh} & \textit{list}(\textit{sign}) & \\
(\textsc{arg-st} & \textit{list}(\textit{sign})) & \\
\ldots & &
\end{bmatrix}$$

As far as I have been able to ascertain, the revisions I have suggested here are compatible with all previous work done in HPSG. For example, I believe the detailed proposals made by Ginzburg and Sag (2000) are all consistent with the present framework: their phrase hierarchies simply become construction hierarchies. But it is important to understand that the separation of signs from constructions proposed here in no way commits one to an approach based on fine-grained distinctions of construction type. For example, the grammar of PS-94 can readily be ported to this framework.

Let us now suppose that there are some lexical items that select for words, as opposed to phrases, as argued by Ginzburg and Sag. This phenomenon can now be described without appeal to a feature LEX, since the objects selected by heads are not *synsem*-objects, but rather signs, which are naturally differentiated into words and phrases. This recurring (though still controversial) issue is rendered unproblematic by the proposals I've made here.

5 Conclusion

In this paper, I have surveyed numerous problems concerning the matter of locality in the theory of grammar. I have proposed a version of HPSG theory that is based on a distinction between signs and constructions. Drawing the distinctions in the way I have outlined provides numerous benefits, including the following:

- Lexical selection is localized.
- Previous results in HPSG are maintained.
- Valence specifications are simplified, eliminating relational constraints.
- Lexical selection may have access the *word/phrase* distinction.
- Phrasal schemata (constructions) are localized, i.e. they are fundamentally like CFG grammar rules.

References

Abeillé, Anne, and Danièle Godard. 1994. The Complementation of Tense Auxiliaries in French. In *Proceedings of the Thirteenth Annual Meeting of the West Coast Conference on Formal Linguistics*. Stanford: SLA, CSLI Publications.

Abeillé, Anne and Danièle Godard. 2000. French Word Order and Lexical Weight. In Robert Borsley, ed., *Syntactic Categories* Syntax and Semantics series, volume 32. San Diego: Academic Press.

Abeillé, Anne, Danièle Godard, and Ivan A. Sag. 1998. Two Kinds of Composition in French Complex Predicates. In Erhard Hinrichs, Andreas Kathol, and Tsuneko Nakazawa, eds., *Complex Predicates in Nonderivational Syntax*. New York: Academic Press. Pp. 1–41.

Andrews, Avery D. 1996. Semantic case-stacking and inside-out unification. *Australian Journal of Linguistics* 16:1–55.

Asudeh, Ash. 2002. Richard III. In Mary Andronis, Erin Debenport, Anne Pycha and Keiko Yoshimura (eds.), *CLS 38: The main session. Papers from the 38th Meeting of the Chicago Linguistic Society*, vol. 1. Chicago, IL: Chicago Linguistic Society.

Bender, Emily M. and Dan Flickinger. 1999. Peripheral constructions and core phenomena: Agreement in tag questions. In Webelhuth, Gert, Jean-Pierre Koenig and Andreas Kathol, (Eds.) *Lexical and Constructional Aspects of Linguistic Explanation*. Stanford: CSLI. Pp. 199–214.

Borsley, Robert. 1987. Subjects and Complements in HPSG. Report no. CSLI-87-107, Center for the Study of Language and Information, Stanford University.

Bouma, Gosse, Robert Malouf, and Ivan A. Sag. 2001. Satisfying Constraints on Extraction and Adjunction. *Natural Language and Linguistic Theory.* 19.1: 1–65.

Culicover, Peter. 1992. English Tag Questions in Universal Grammar. *Lingua* 88.193–226.

Dickinson, Markus. (this volume??)

Flickinger, Dan and Emily M. Bender. 2003. Compositional Semantics in a Multilingual Grammar Resource. Presented at the Workshop on Ideas and Stratgies for Multilingual Grammar Development, ESSLLI 2003, Vienna. Pp. 33–42.

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

Copestake, Ann. 1993. *The Compleat LKB.* Technical report 316, University of Cambridge Computer Laboratory.

Dench, Alan, and Nick Evans. 1988. Multiple case-marking in Australian languages. *Australian Journal of Linguistics* 8:1–47.

Emonds, Joseph E. 1976. *A Transformational Approach to English Syntax.* New York: Academic Press.

Gazdar, Gerald. 1981. Unbounded Dependencies and Coordinate Structure. *Linguistic Inquiry* 12: 155–184. Reprinted in Walter J. Savitch, Emmon Bach, William Marsh and Gila Safran-Naveh, eds. *The Formal Complexity of Natural Language* Dordrecht: Reidel. Pp. 183-226 (1987).

Gazdar, Gerald. 1982. Phrase Structure Grammar. In Pauline Jacobson and Geoffrey K. Pullum, eds., *The Nature of Syntactic Representation.* Dordrecht: Reidel, 131-186. Reprinted in J. Kulas, J.H. Fetzer, and T.L. Rankin (eds.) *Philosophy, Language, and Artificial Intelligence.* Dordrecht: Kluwer, 163-218, 1988.

Ginzburg, Jonathan and Ivan A. Sag. 2000. *Interrogative Investigations: the form, meaning, and use of English Interrogatives.* Stanford: CSLI Publications. [Distributed by University of Chicago Press]

Haider, Hubert. 1990. Topicalization and Other Puzzles of German Syntax. In Grewendorf, Günther, and Wolfgang Sternefeld, eds., *Scrambling and Barriers.* Pp. 93–112. Amsterdam: John Benjamins Publishing Co.

Harris, Zellig S. 1946. From Morpheme to Utterance. *Language* 22: 161–183. Reprinted in Martin Joos, ed., (1957) *Readings in linguistics I.* Pp. 142–153. Chicago: University of Chicago Press.

Higgins, Francis Roger. 1973. *The Pseudo-cleft Construction in English.* Doctoral dissertation, MIT. Published by Garland Publishers, 1979.

Höhle, Tilman N. 1994. Featuring Creatures of Darkness. Handout for a talk given at the International HPSG Workshop 94, 7 Sept. 1994, Institute for Logic and Linguistics, IBM Germany, Heidelberg. Downloadable at `http://www.uni-tuebingen.de/Deutsches-Seminar/hoehle/`.

Höhle, Tilman. 1997. Vorangestellte Verben und Komplementierer sind eine natürliche Klasse. In C. Dürscheid, K. H. Ramers, and M. Schwarz (Eds.), Sprache im Fokus. Festschrift für Heinz Vater zum 65. Geburtstag, pp. 107-120. Tübingen: Max Niemeyer Verlag. A discussion in English of selected excerpts can be found in Höhle 1994.

Hukari, Thomas E., and Robert D. Levine. 1995. Adjunct Extraction. *Journal of Linguistics* 31: 195–226.

Kajita, Masaru. 1968. *A Generative-Transformational Study of Semi-Auxiliaries in Present-Day American English.* Tokyo: Sanseido.

Kathol, Andreas. 1995. *Linearization-based German syntax.* Unpublished doctoral dissertation, Ohio State University.

Kathol, Andreas. 1999. Agreement and the Syntax-Morphology Interface in HPSG. In Robert Levine and Georgia Green (eds.) *Studies in Contemporary Phrase Structure Grammar.* Cambridge: Cambridge University Press. Pp. 223–274.

Kathol, Andreas. 2000. *Linear Syntax.* Oxford: OUP.

Kathol 2002. Subjects in Fronted German VPs and the Problem of Case and Agreement: Shared Argument Structures for Discontinuous Predicates. In Jongbok Kim and Stephen Wechsler, eds., *Proceedings of the 9th International Conference on Head-Driven Phrase Structure Grammar, Kyung-Hee University, Seoul.* Stanford: CSLI Publications. http://cslipublications.stanford.edu/HPSG/3/. Pp. 91–108.

Kay, Paul. 2002. English Subjectless Tagged Sentences. *Language* 78.3: 453–481.

Kiss, Tibor. 1995. *Infinitive Komplementation.* Number 333 in Linguistische Arbeiten. Tübingen: Max Niemeyer Verlag. Published version of 1992 doctoral dissertation, Bergische Universität–Gesamthochschule Wuppertal.

Levine, Robert. 2000. 'Tough' complementation and the extraclausal propagation of argument descriptions. (On-line) Proceedings of the 7th International Conference on Head-Driven Phrase Structure Grammar, UC Berkeley. (`http://csli-publications.stanford.edu/HPSG/1/hpsg00-toc.html`)

Levine, Robert D. 2002. Adjunct Valents: Cumulative scoping adverbial constructions and impossible descriptions. In Jongbok Kim and Stephen Wechsler, eds., *Proceedings of the 9th International Conference on Head-Driven Phrase Structure Grammar, Kyung-Hee University, Seoul.* Stanford: CSLI Publications. http://cslipublications.stanford.edu/HPSG/3/. Pp. 209–232.

Malouf, Robert. A Head-Driven Account of Long-Distance Case Assignment. In Cann, Ronnie, Claire Grover, and Philip Miller, eds., *Grammatical Interfaces in HPSG* Stanford: CSLI Publications. Pp. 201–214.

Meurers, Detmar. 1999. Raising Spirits (and Assigning Them Case). In Werner Abraham, ed., *Groninger Arbeiten zur Germanistischen Linguistik*, 43.

Meurers, Detmar. 2000. *Lexical Generalizations in the Syntax of German Non-Finite Constructions.* Arbeitspapiere des SFB 340, Nr. 145. [PhD Dissertation, Universität Tübingen, 1999].

Meurers, Detmar. 2001. On Expressing Lexical Generalizations in HPSG. *Nordic Journal of Linguistics* 24:2.

Meurers, Detmar. this volume...

Müller, Stefan. 1995. Scrambling in German – Extraction into the *Mittelfeld.* In *Proceedings of the tenth Pacific Asia Conference on Language, Information and Computation*, ed. by Benjamin K. T'sou and Tom Bong Yeung Lai. City University of Hong Kong. Pp. 79–83.
[Available on-line at: http://www.cl.uni-bremen.de/~stefan/Pub/scrambling.html]

Müller, Stefan. 1999. *Deutsche Syntax Deklarativ. Head-Driven Phrase Structure Grammar für das Deutsche.* Tübingen: Max Niemeyer Verlag. Linguistische Arbeiten, No. 39.
[Available on-line at: http://www.cl.uni-bremen.de/~stefan/Pub/hpsg.html]

Müller, Stefan. 2002 *Complex Predicates: Verbal Complexes, Resultative Constructions, and Particle Verbs in German*, Studies in Constraint-Based Lexicalism No. 13. Stanford: CSLI Publications.

Müller, Stefan. 2004. Continuous or Discontinuous Constituents? A Comparison between Syntactic Analyses for Constituent Order and Their Processing Systems. *Research on Language and Computation 2.2, Special Issue on Linguistic Theory and Grammar Implementation*, ed. by Walt Detmar Meurers and Shuly Winter and Erhard W. Hinrichs.
[Available online at: http://www.cl.uni-bremen.de/~stefan/Pub/discont.html.]

Müller, Stefan. 2005. Zur Analyse der deutschen Satzstruktur. *Linguistische Berichte*, 201: 3–39.

Müller, Stefan. to appear. Continuous or Discontinuous Constituents? Elliptical Constructions, Multiple Frontings, and Surface-Based Syntax. To appear in *Proceedings of Formal Grammar 2004*, ed. by Gerhard Jäger, Paola Monachesi, Gerald Penn und Shuly Winter. Stanford: CSLI Publications. [Draft of February 26, 2005 available at http://www.cl.uni-bremen.de/~stefan/Pub/surface.html]

Oehrle, Richard T. 1987. Multi-Dimensional Categorial Grammars and Linguistic Analysis. In *Categories, Polymorphism and Unification*, ed. by Ewan Klein and Johan van Benthem. Edinburgh: Centre for Cognitive Science, University of Edinburgh.

Pollard, Carl, and Ivan A. Sag. 1987. *Information-Based Syntax and Semantics; Volume One - Fundamentals.* CSLI Lecture Notes Series No. 13. Stanford: CSLI Publications.

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

Potsdam, Eric and Jeff Runner. 2001. Richard Returns: copy raising and its implications. In *Papers from the 37th Regional Meeting of the Chicago Linguistics Society.* Chicago: CLS. Pp. 453–468.

Przepiórkowski, Adam. 1999. *Case Assignment and the Complement-Adjunct Dichotomy: A Non-Configurational Constraint-Based Approach.* Doctoral dissertation, University of Tübingen.

Reape, Mike. 1994. Domain Union and Word Order Variation in German. In John Nerbonne and Klaus Netter and Carl J. Pollard, eds., *German in Head-Driven Phrase Structure Grammar.* Pp. 151–197. CSLI Lecture Notes, number 46. Stanford University: CSLI Publications.

Reape, Mike. 1996. Getting Things in Order. In Harry Bunt and Arthur van Horck, editor(s), *Discontinuous Constituency.* Pp. 2090-253. Natural language processing, number 6. Berlin, New York: Mouton de Gruyter.

Rogers, Andy. 1974 A Transderivational Constraint on Richard? In *Papers from the 10th Regional Meeting of the Chicago Linguistics Society.* Chicago: CLS. Pp. 551–558.

Sag, Ivan A. 1997. English Relative Clause Constructions. *Journal of Linguistics* 33.2: 431–484.

Sag, Ivan A. 2005. Adverbial Extraction: A defense of tracelessness.

Sag, Ivan A., and Carl Pollard. 1991. An Integrated Theory of Complement Control. *Language* 67.1: 63–113.

Sag, Ivan A., Thomas Wasow, and Emily Bender. 2003. *Syntactic Theory: A formal introduction: second edition.* Stanford: CSLI Publications.

Sailer, Manfred. 2003. *Combinatorial Semantics and Idiomatic Expressions in Head-Driven Phrase Structure Grammar.* Arbeitspapiere des SFB 340. Nr. 161. Eberhard-Karls-Universität Tübingen. [Revision of 2000 PhD Dissertation]

Snider, Neal. to appear?

Zaenen, Annie. 1983. On Syntactic Binding. *Linguistic Inquiry* 14(3): 469–504.

Zec, Draga. 1987. On Obligatory Control in Clausal Complements. In Masayo Iida, Stephen Wechsler, and Draga Zec (eds.) *Working Papers in Grammatical Theory and Discourse Structure.* Lecture Notes Series, No. 11. Stanford: CSLI Publications. Pp. 139–168.