Chapter 4

The nature and role of the lexicon in HPSG

Anthony Davis

Southern Oregon University

Jean-Pierre Koenig

University at Buffalo

This chapter discusses the critical role the lexicon plays in HPSG and the approach to lexical knowledge that is specific to HPSG. We describe the tenets of lexicalism in general, and discuss the nature and content of lexical entries in HPSG. As a lexicalist theory, HPSG treat lexical entries as informationally rich, representing the combinatorial properties of words as well as their part of speech, phonology, and semantics. Thus many phenomena receive a lexically-based account, including some that go beyond what is typically regarded as lexical. We then turn to the global structure of the HPSG lexicon, the hierarchical lexicon and inheritance. We show how the extensive type hierarchy employed in HPSG accounts for lexical generalizations at various levels and discuss some of the advantages of default (nonmonotonic) inheritance over simple monotonic inheritance. We then describe lexical rules and their various proposed uses in HPSG, comparing them to alternative approaches to relate lexemes and words based on the same root or stem.

1 Introduction

The nature, structure, and role of the lexicon in the grammar of natural languages has been a subject of debate for at least the last 50 years. For some, the lexicon is a prison that "contains only the lawless" to borrow a memorable phrase from Di Sciullo & Williams (1987), and not much of interest resides there. In some recent views, the lexicon records merely phonological information and some world

knowledge about each lexical entry (see Marantz (1997)). All of the action is in the syntax, save the expression of complex syntactic objects as inflected words. In contrast, lexicalist theories of grammar, and HPSG in particular, posit a rich and complex lexicon embodying much of grammatical knowledge.

This chapter has two principal goals. One is to review the arguments for and against a lexicalist view of grammar within the generative tradition. The other is to survey the HPSG implementation of lexicalism. In regard to the first goal, we begin with the reaction to Generative Semantics, and note developments that led to lexicalist theories of grammar such as LFG and then HPSG. Central to these developments was the argument that lexical processes, rather than transformational ones, provided more perspicuous accounts of derivational morphological processes. The same kinds of arguments then naturally extended to phenomena like passivization, which had previously been treated as syntactic. Once on this path, lexical treatments of other prototypically syntactic phenomena — long distance extraction, *wh*-movement, word order, and anaphoric binding — were advanced as well, with HPSG playing a leading role.

But this does not mean that opposition to lexicalism melted away. Both Minimalism (cross-ref here), in particular Distributed Morphology, and Construction Grammar (cross-ref here) claim that lexicalist accounts fail in various ways. We discuss some of these current issues, including phrasal processes in the lexicon, word-internal ellipsis, and endoclitics, each of which poses challenges for those who advocate a strict separation between lexical and syntactic systems. While we maintain that the anti-lexicalist arguments are not especially strong, and the phenomena they are based somewhat marginal, we acknowledge that these questions are not yet settled. We then turn to the specifics of the lexicon as modeled within HPSG. Lexicalism demands, of course, that lexical entries be informationally rich, encoding not merely idiosyncratic properties of a single lexical item like its phonology and semantics, but also more general characteristics like its combinatorial possibilities. We outline what HPSG lexical entries must contain, and how that information is represented. This leads naturally to the next topic: with so much information in a lexical entry, and so much of that repeated in similar ones, how is massive redundancy avoided? The hierarchical lexicon, in which individual lexical entries are the leaves of a multiple inheritance hierarchy, is a core component of HPSG. Types throughout the hierarchy capture information common to classes of lexical entries, thereby allowing researchers to express generalizations at various levels. Just as all verbs share certain properties, all transitive verbs, all verbs of caused motion, and all transitive verbs of caused motion share additional properties, represented as constraints on types

within the hierarchy. We draw on examples from linking, gerunds, and passives as illustrations, but many others could be added.

Constraints specified on types in the hierarchy are deemed to be inherited by their subtypes, but monotonic inheritance of this kind runs into vexing issues. Most obviously, there are irregular morphological forms; any attempt to represent, say, the phonology of English plurals, as a constraint on a plural noun class in the hierarchical lexicon must then explain why the plural of *child* is *children* and not **childs*. Beyond this simple example, there are ubiquitous cases of lexical generalizations that are true by default, but not always. Various mechanisms for modeling default inheritance have therefore been one focus within HPSG, and we furnish an example of their use in modeling the properties of gerunds cross-linguistically.

Finally, we discuss lexical rules and their alternatives. Along with the "vertical" relationships between classes of lexical entries modeled by types and their subtypes in the hierarchical lexicon, there is a perceived need for "horizontal" relationships between lexical entries that are based on a single root or stem, such as forms of inflectional paradigms. Yet formalizing lexical rules adequately within HPSG has proven tricky; specifying just what information is preserved and what is changed by a lexical rule is one prominent issue. We conclude this chapter by describing alternatives to lexical rules. One is implicational statements on partially specified feature structures; this might be thought of as a kind of "online lexical rule". The second augments the type hierarchy via online type construction, extending the predefined lexical types specified in the hierarchy to include "virtual types" that combine the information from multiple predefined types.

2 Lexicalism

2.1 Lexicalism and the origins of HPSG

Lexicalism began as a reaction to Generative Semantics, which treated any regularity in the structure of words (derivational patterns, broadly speaking) as only ephiphenomenally a matter of word structure and underlyingly as a matter of syntactic structure (see Lakoff (1970), among others). In the Generative Semantics view, all grammatical regularities are a matter of syntax (much of it, in fact, logical syntax). Chomsky (1970) presented many arguments that lexical knowledge differs qualitatively from syntactic knowledge and should be modeled differently. Jackendoff (1975) is an explicit model of lexical knowledge that follows Chomsky's insights, although it focuses exclusively on derivational morpholog-

ical processes. The main insight that Jackendoff formalizes is that relations between stems and words (say, between *destruct* and *destruction*) are to be modeled not via a generative device but through a redundancy mechanism that measures the relative complexity of a lexicon where these relations are present or not present (the idea is that a lexicon where *construct* and *construction* are related is simpler than one where they are not). Bochner (1993) is the most formalized and detailed version of this approach to lexical relations. Lexicalist approaches, including LFG and HPSG, took their lead from Jackendoff's work. LFG relied heavily on treating relations between stems and between words as lexical rules, rather than the kind of generative devices that one finds in syntax. But, as accounts of linguistic phenomena in LFG focused increasingly on the lexicon, the question of whether lexical rules retained the character of redundancy rules or turned into yet another kind of generative device arose. Consequently, the necessity of lexical rules has been questioned as well (see Koenig & Jurafsky (1994) and Koenig (1999) for potential issues that arise once lexical rules are assumed to be involved in the creation of new lexical entries).

Lexicalism, at least within HPSG, embodies two distinct ideas. First is the idea that parts of words are invisible to syntactic operations (*lexical integrity*, see Bresnan & Mchombo 1995), so that relations between stems and between word forms cannot be the result of or follow syntactic operations, as in distributed morphology, or other linguistic models that assign no special status to the notion of word. Relations between words are therefore not modeled via syntactic operations (hence the appeal to Jackendoff's lexical rules). Second is the idea that the occurrence of a lexical head in distinct syntactic contexts arises from distinct variants of words. For instance, the fact that the verb *expect* can occur both with a finite clause and an NP+VP sequence (see (1) vs. (2)) means that there are two variants of the verb *expect*, one that subcategorizes for a finite clause and one where it subcategorizes for an NP+VP sequence.[1] Not all lexicalist theories, though, cash out these two distinct ideas the same way. The net effect of lexicalism within HPSG is that words and phrases are put together via distinct sets of constructions and that words are syntactic atoms. These two assumptions justify positing two kinds of signs, *phrasal-sign* and *lexical-sign* and go hand in hand with the surface-oriented character of HPSG and what one might call a principle of surface combinatorics: If expression A consists of B⊕C, then all grammatical

---

[1]As this chapter is an overview of the approach to lexical knowledge HPSG embodies rather than a description of particular HPSG analyses of phenomena, we will sample liberally from various illustrative examples and simplify whenever possible the analyses so that readers can see the forest and not get lost in the trees.

Draft of 29th May 2019, 13:58

constraints that make reference to B and C are circumscribed to A.

(1) I expected to leave yesterday

(2) I expected that I would leave yesterday.

An evident concern regarding this view of the lexicon is the potential proliferation of lexical entries, replete with redundant information. Will it be necessary to specify all the information in these two entries for *expect* without regard for the large amount of duplication between them? Will the same duplication be needed for the verb *hope*, which patterns similarly? How will somewhat similar verbs, such as *imagine* and *suppose*, which allow finite complements but not infinitive ones, be represented? We will describe HPSG's solutions to these questions below, in our discussion of the hierarchical lexicon. First, however, we turn to recent arguments against lexicalism, and then discuss in more detail just what kinds of information should be in HPSG lexical entries.

2.2 Recent challenges to lexicalism

As there are been several challenges to lexicalism (see Bruening (2018) and Haspelmath (2011) among others for some recent challenges), we now explore lexicalism and lexical integrity in HPSG in more detail. We first note that lexicalism does not imply that word and phrase formation are necessarily different "components" as is often claimed (see Marantz 1997, Bruening (2018)). Some lexicalist approaches *do* assume that word formation and phrase building belong to two different components of a language's grammar (this is certainly true of Jackendoff 1975), but they need not. Within HPSG, there are approaches that treat every sign-formation (be it word-internal or word-external) as resulting from typed mother-daughter configurations (this is the hypothesis pursued in Koenig (1999), and is also the approach frequently taken in implementations of large-scale grammars where lexical rules are modeled as unary-branching trees, see the English Resource Grammar at http://www.delph-in.net/erg/). Furthermore, recent approaches to inflectional morphology model realizational rules through the very same tools the rest of a language's grammar uses (see Crysmann & Bonami (2016) and the chapter on morphology in this volume). There are also approaches to phrases where the same analytical tools developed to model lexical knowledge (see Section 4) are employed to model phrase-structural constructions (see Sag's 1997 analysis of relative clauses, for example). So, both in terms of the formal devices and in terms of analytical tools used to model datasets, words and phrases can be treated the same way in HPSG (alhtough they need not be). Somewhat

ironically, and despite claims to the contrary, word formation in the syntactico-centric approach Marantz or Bruening advocate *does* make use of distinct formal machinery to model word formation, namely realizational rules to model inflectional morphology (see Halle & Marantz 1993).

With this red herring out of the way, we concentrate on the two most important challenges Bruening (2018) and Haspelmath (2011) present to lexicalist views. The first challenge are cases of phrasal syntax feeding the lexicon, purportedly exemplified by sentences such as (3).

(3)   I gave her a don't-you-dare! look. (example (1a) in Bruening 2018)

We can provisionally accept for the sake of argument Bruening's contention that *don't-you-dare!* is a word in (3), despite its reliance on the (unjustified) assumption that the secondary object in (3) involves N-N compounding rather than an AP N structure (we refer readers to Bresnan & Mchombo (1995) or Müller (2010) for counter-arguments to Bruening's claim). Crucially, though, examples such as (3) have no bearing on HPSG's model of lexical knowledge, as HPSG-style lexicalism does not preclude constructions that form words from phrases. Nothing, as far as we know, rules out constructions of the form *phrase → stem/word* in HPSG. The two assumptions underlying HPSG brand of lexicalism we mentioned above do not preclude a *lexical-sign* having a *phrasal-sign* as sole daughter (although we do not know of any HPSG work that exploits this possibility) and examples such as (3) are simply irrelevant to whether HPSG's lexicalist stance is empirically correct.

The second challenge to lexicalism presented in Bruening (2018) bears more directly on HPSG's assumption that words are syntactic atoms. Word-internal conjunction/ellipsis examples, illustrated in (4) (adapted from Bruening's (31a)), seem to violate the assumption that syntactic constraints cannot "see" the internal structure of words, as ellipsis in these kinds of examples seems to have access to the internal part of the word *over-application*. In fact, though, such examples do not violate lexical integrity if one enriches the representation of composite words (to borrow a term from Anderson 1992) to include a representation of their internal phonological parts as proposed in Chaves (2008; 2014).

(4)   Over- and under-application of stress rules plagues Jim's analysis.

Chaves' analysis assumes that the phonology of compound words and words that contain affixoids (to borrom a term from Booij 2005) is structured. The MorphoPhonology or MP attribute of words (and phrases) is a list of phonological

forms and morphs information. The MP of compound words and words that contain affixoids includes a separate member for each member of the compound, or for the affixoid and stem. Thus in (4), the MPs of *overapplication* and *underapplication* each contain two elements: one for *over/under*, and one for *application*. Given this enriched representation of the morphophonology of words like *under/overapplication*, a single ellipsis rule can apply both to phrases and to composite words, eliding the second member of the word *overapplication*'s MP. As Chaves makes clear (p.304) such an analysis is fully compatible with lexical integrity, as there is no access to the internal structure of composite words, only to the (enriched) morphophonology of the entire word.

Haspelmath (2011) similarly challenges the view that syntactic processes may not access the internal structure of words, although Haspelmath's point is merely that what is a word is cross-linguistically unclear. So-called suspended affixation in Turkish (see (5)) also shows that word parts can be elided. We cannot discuss here whether Chaves' analysis can be extended to cases like (5) where suffixes are seemingly elided or whether lexical sharing (where a single word can be the daughter of two c-structure nodes à la McCawley 1982), as proposed in Broadwell (2008) is needed. What is important for current purposes is that these putative challenges to lexical integrity such as (4) or (5) do not necessarily render a substantive version of it implausible. The same is true of another potential challenge to lexical integrity which neither Bruening nor Haspelmath discuss, endoclitics, which we discuss next.

(5) kedi ve köpek-ler-im-e
     cat and dog-PL-1SG-DAT
     'to my cat(s) and dogs'

Endoclitics are clitics that at least appear to be situated within a word, rather than immediately preceding or following it, as clitics often do. (cross-reference to Abeillé & Penn chapter on clitics) In many cases, endoclitics appear at morphological boundaries, as in the well-studied pronominal clitics of European Portuguese (Crysmann 2001a). An approach similar to what we have referenced above for composite words and elided morphology may well extend to these as well. But some trickier cases have also come to light, in which the clitic appears within a morpheme, not at a boundary. Two of the best documented cases from the Northeast Caucasian language Udi (Harris 2000) and from Pashto (Tegey 1977; Roberts 2000; Dost 2007). Here are examples from Udi (6) and Pashto (7), where the clitics appear in the middle of verbs.

(6) q'ačaɣ-ɣ-on bez tänginax   baš=**q'un**-q'-e
     thief-PL-ERG my money.DAT steal₁-3PL-steal₂-AORII
     'Thieves stole my money.' (root *bašq'*, 'steal')

(7) a. ṭəlwɑhə́=**me**
         push.IMPF.PST.3SG-cl.1SG
         'I was pushing it.' (from Tegey 1977; Dost 2007)

     b. ṭə́l=**me**-wɑhə
         push₁-cl.1SG-push₂.PF.PST.3SG
         'I pushed it.' (from Tegey 1977; Dost 2007)

In these cases. as with clitics in general, there is a clash between the phonological criteria for wordhood, under which the clitics would be regarded as incorporated within words, and the syntactic constituency and semantic compositionality. But what makes these particularly odd is that these clitics are situated word-internally, even morpheme-internally. Udi subject agreement clitics such as *q'un* in (6) typically attach to a focused constituent, which can be a noun, a questioned constituent, or a negation particle as well as a verb (Harris 2000). Under certain conditions, as in (6), none of these options is available or permitted, and the clitic is inserted before the final consonant of the verb root, dividing it in two pieces, neither of which has any independent morphological status. Its position in this instance is apparently phonologically determined; it cannot appear word-finally or word initially, and as there is no morphological boundary within the word it must therefore appear within the monomorphemic root. Pashto clitics seek "second position," whether at the phrasal, morphological, or phonological level; *me* in (7) appears to be situated after the first stressed syllable (or metrical foot), which, in the case of (7b), also divides the verb into two parts that lack any independent morphological status.

If clitics are viewed as a syntactic phenomenon ("phrasal affixes", as Anderson (2005) puts it), these endoclitics must have "visibility" into the internal structure of words (be it morphological, prosodic, or something else), thereby seemingly violating lexical integrity. Anderson's brief account invokes a reranking of optimality theoretic constraints from their typical ordering, whereby the clitic's positional requirements outrank lexical integrity requirements. Crysmann (2001b) proposes an analysis, paralleling in many respects his account of European Portuguese clitics in Crysmann (2001a), using Kathol's topological fields (Kathol 1999). The "morphosyntactic paradox in Udi" is effectively "resolved on the basis of discontinuous lexical items"; this account then "parallels HPSG's represen-

tation of syntactic discontinuity." (cross reference to Müller's chapter on word order here)

For Pashto, researchers generally agree that the notion of second position is crucial, but that it can be defined at various levels— phrasal, lexical, and phonological. In this last case clitics can appear within a word following the first metrical unit, as illustrated above. (cross reference to Tseng's chapter on phonology, and Crysmann's on morphology here). Dost (2007) invokes word order domains (Reape 1994) and topological fields (Kathol 1999) at these various levels to account for this distribution of clitics. In this analysis, some words contain more than one order domain at the prosodic level. Lexical integrity is preserved to the extent that, while domains at the prosodic level are "visible" to clitics in Pashto, syntatic processes do not reference the internal makeup of words.

Still, these accounts of endoclitics in Udi and Pashto appear to breach the wall of the strictest kind lexical integrity, requiring that they have access to some of the internal structure of lexical entries through a partial decomposition of their morphophonology into distinct order domains. Yet we would not wish to advocate models that permit unconstrained violations of lexical integrity, either. The troublesome cases we have noted here are relatively marginal or cross-linguistically rare, and limited in scope to prosodic or morphophonological information and seem to only pertain to phonological interactions (ellipsis, insertion). As Broadwell (2008) points out when comparing possible analyses of Turkish suspended affixation, rejecting lexicalism altogether may lead to an unconstrained theory of the interaction between words/stems and phrases and incorrect predictions (e.g., that all affixes in Turkish can be suspended). Likewise, we would not expect to find a language in which endoclitics positioning is utterly unconstrained, and thus we would not wish to see grammatical theories abandon lexical integrity altogether.

3 Lexical entries in HPSG

3.1 What are lexical entries?

A consequence of HPSG's lexicalist stance is that there will be many lexical entries where one might at first glance expect a single entry. We will see below how HPSG handles multiple entries and classes of entries while avoiding redundancy, but it is important at the outset to clarify what a lexical entry is in HPSG. One of the misunderstandings about lexical knowledge is that it confuses descriptions and entities being described, or the distinction between constructions

and constructs (lexical entry vs. fully instantiated lexeme). As the chapter on the formal foundations of HPSG discusses, grammars in HPSG consist of *descriptions* of structures, and the lexicon thus consists of descriptions of what are fully specified lexemes. What is stored in the lexicon is descriptions of fully instantiated lexemes. To see the importance of the distinction between descriptions (stored entries) and the fully instantiated entries that are being described, consider HPSG's model of subcategorization and consider the relevant portion of the tree for sentence (2). HPSG's model of the dependency between heads and complements involves identity between the syntactic and semantic information of each complement (the value of the SYNSEM attribute) and a member of the list of complements the head subcategorizes for. Since there are infinitely many SYNSEM values, on the assumption that there are infinitely many clausal meanings (a point Jackendoff (1990) emphasizes), there are, in principle, infinitely many fully instantiated entries for the verb *see* subcategorizing for a clausal complement (as in (2)). But each of these fully instantiated entries for *expect*, one for each clausal sentence that corresponds to the tree in (8) corresponds to a single abstract description, and it is this description that the lexicon contains.

(8)

$$\left[\text{COMPS} \ \langle\boxed{1}\rangle\right] \quad \left[\text{SYNSEM} \ \boxed{1}\right]$$

The formal status of lexical entries has engendered a fair amount of theoretical work and some debate, particularly over the question of whether lexical entries must be fully specified. We will touch on some aspects of this further below, in connection with online type construction. For further discussion of these kinds of issues, see the chapters on Basic Properties and Elements and on Formal Background.

3.2 What information is in lexical entries?

Because lexical items play a critical role in accounting for the syntax of natural languages, lexical entries are informationally rich in HPSG. Aside from the expected phonological and semantic information, specific to each lexeme, they include morphological and combinatorial potential information. Morphological information serves as input to inflectional rules, but is also used to select the appropriate types of phrases (via their projection through the Head-Feature Principle), as shown in (9). Some verbs, for instance, select for a PP headed by a particular preposition; others select for VPs whose verb is a gerund, or a bare

infinitive, and so forth. Lexical entries thus include as much morphological information as both (inflectional) morphology and syntactic selection require.

(9)　a.　John conceived *of/*about* the world's tastiest potato chip..

　　b.　John regretted *going/*(to) go* to the party.

We illustrate the second leading idea behind HPSG or LFG's lexicalism (that there are different variants of lexical heads for different contexts in which heads occur) with the French examples in (10). The verb *aller* 'go' in (10a) combines with a PP headed by *à* that expresses its goal argument and a subject that expresses its theme argument. The same verb in (10b) combines with the so-called non-subject clitic *y* that expresses its goal argument. We follow Miller & Sag (1997) and assume here that French non-subject clitics are prefixes. Since the context of occurrence of the head of the sentence, *aller*, differs across these two sentences (NP____PP[*à*] and NP *y*____ , respectively and informally), there will be two distinct entries for *aller* for both sentences, shown in (11) and (12) (we simplify the entries' feature geometry for expository purposes).

(10)　a.　Muriel va　　　　　à　Lourdes.
　　　　　Muriel go-PRES.3RD.SG at Lourdes.

　　b.　Muriel y　　va.
　　　　　Muriel there go-PRES.3RD.SG

(11)

$$
\begin{bmatrix}
\text{MORPH} & \begin{bmatrix} \text{FORM} & \boxed{5} \\ \text{I-FORM} & \boxed{5}\ \textit{va} \\ \text{STEM} & \textit{v-} \end{bmatrix} \\[4ex]
\text{CAT} & \begin{bmatrix}
\text{HEAD} & \begin{bmatrix} \textit{verb} \\ \text{VFORM} & \begin{bmatrix} \text{MOOD} & \textit{indic} \\ \text{TNS} & \textit{pres} \\ \text{AGR} & \textit{3rdsing} \end{bmatrix} \end{bmatrix} \\[6ex]
\text{VAL} & \begin{bmatrix} \text{SUBJ} & \langle \boxed{1} \rangle \\ \text{COMPS} & \langle \boxed{2} \rangle \end{bmatrix} \\[3ex]
\text{ARG-ST} & \langle \boxed{1}\text{NP}[\textit{3rdsg}]_{\boxed{3}},\ \boxed{2}\text{PP}[\textit{à}]_{\boxed{4}} \rangle
\end{bmatrix} \\[8ex]
\text{CONT} & \begin{bmatrix} \textit{go-rel} \\ \text{THEME} & \boxed{3} \\ \text{GOAL} & \boxed{4} \end{bmatrix}
\end{bmatrix}
$$

(12)
$$
\begin{bmatrix}
\text{MORPH} & \begin{bmatrix}
\text{FORM} & \boxed{y\text{-}va} \\
\text{I-FORM} & va \\
\text{STEM} & v\text{-}
\end{bmatrix} \\[2em]
\text{CAT} & \begin{bmatrix}
\text{HEAD} & \begin{bmatrix}
verb \\
\text{VFORM} \begin{bmatrix}
\text{MOOD} & indic \\
\text{TNS} & pres \\
\text{AGR} & 3rdsing
\end{bmatrix}
\end{bmatrix} \\[2em]
\text{VAL} & \begin{bmatrix}
\text{SUBJ} & \langle \boxed{1} \rangle \\
\text{COMPS} & \boxed{\langle \rangle}
\end{bmatrix} \\[2em]
\text{ARG-ST} & \langle \boxed{1}\text{NP}[3rdsg]_{\boxed{3}}, \boxed{\text{PP}[p\text{-}aff,loc]_{\boxed{4}}} \rangle
\end{bmatrix} \\[2em]
\text{CONT} & \begin{bmatrix}
go\text{-}rel \\
\text{THEME} & \boxed{3} \\
\text{GOAL} & \boxed{4}
\end{bmatrix}
\end{bmatrix}
$$

CATegory information in both entries include part of speech information (including morphologically relevant features of verb forms), ARGument-STructure information and VALence information. MORPH information includes both stem form information, inflected form information (I-FORM) and, in case so-called clitics are present, the combination of the clitic and inflected form information. Both entries illustrate how informationally rich lexical entries are in HPSG. But, postulating informationally rich entries does not mean stipulating all of the information within every entry. In fact, only the stem form and the relation denoted by the semantic content of the verb *aller* need be stipulated within either entry. All the other information can be inferred once it is known which classes of verbs these entries belong to. In other words, most of the information included in the entries in (11) and (12) is not specific to these individual entries, an issue we take up in Section 4. The entry-specific information in (11) and (12) is in black font while the shared information is in gray font; the informational difference between the two entries for *va* and *y va* is included in shadowed boxes in the respective entries. The first difference between the two variants of *va* 'goes' is in the list of complements: the entry for *y va* does not subcategorizes for a locative PP since the affix *y* satisfies the relevant argument structure requirement. This difference in the realization of syntactic arguments (via phrases and pronominal affixes) is recorded in the type of the PP members of ARG-ST, *p-aff* in (12) but not in (11). Finally, the two entries differ in the FORM of the verb, which is the same

as the inflected form of the verb in (11) (as indicated by the identically numbered ⑤), but not in (12) whose FORM includes the prefix *y*.

One other question arises with regard to the information in lexical entries. Are there attributes or values that occur solely within lexical signs, and not in phrasal ones? If so, they would provide a diagnostic for distinguishing lexical signs from others. Specific phonological information, for instance, is something we would expect to be introduced by lexical entries, and not elsewhere. Some information is claimed to be specific to lexical signs, such as phonological information (cross reference to Tseng here) and the ARG-ST list, on the premise that lexical items alone specify combinatorial requirements (but see Przepiórkowski (2001) for a contrary view, and see the chapter on Construction Grammar for other views questioning this assumption). But HPSG researchers have generally not typically explored this question in depth, and we will leave this issue here.

3.3 The role of the lexicon in HPSG

As we hope is evident by now, the lexicon plays a critical role in HPSG's explanatory mechanisms, as words encode their distributional potential, as well as their idiosyncratic phonological and semantic charateristics. Much of the information contained in lexical entries is geared to modeling the combinatorial potential of words. As detailed in the chapter on Argument Structure, their combinatorial potential is recorded using two kinds of information, a list of syntactic arguments or syntactic requirements to be satisfied, and distinct lists that indicate how these requirements are to be satisfied (as local dependents, as non-local dependents, as clitics/affixes). Not only are syntactic arguments recorded; so is their relative obliqueness (in terms of grammatical function), as per the partial hierarchy in (13) from Pollard and Sag 1992.

(13)   SUBJECT < PRIMARY OBJ < SECOND OBJ < OTHER COMPLEMENTS

We illustrate this explanatory role by alluding to the role of the lexicon in HPSG's approach to binding, as described in Pollard & Sag (1992) (see the chapter on Binding for details). As lexical entries of heads record both syntactic and semantic properties of their dependents, constraints between properties of heads and properties of dependents, e.g. subject-verb agreement, or between dependents, e.g. binding constraints illustrated in (15), can be stated as constraints on classes of lexical entries. The principle in (15) is such a constraint.

(14)   a. Mathilda$_i$ saw herself$_i$ in the mirror.
       b. *Mathilda$_i$ saw her$_i$ in the mirror.

(15)   An anaphor must be coindexed with a less oblique co-argument, if there
       is one.

   Principle (15) is, formally, a constraint on lexical entries that makes use of
the fact that an entry's argument structure records the syntactic and semantic
properties of a word's dependents. The three argument structures in (16) illus-
trate permissible and ungrammatical entries. (16a) illustrates exempt anaphors
as there is no less oblique syntactic argument than the anaphoric NP; (16b) il-
lustrates a non-exempt anaphor properly bound by a less oblique, co-indexed
non-anaphor; (16c) illustrates an ungrammatical lexical entry that selects for an
anaphoric syntactic argument that is not co-indexed by a less oblique syntactic
argument, despite not being an exempt anaphor (i.e., not being the least oblique
syntactic argument).

(16)   a. $\left[\text{ARG-ST} \ \langle \text{NP}_{i,+ana}, ... \rangle \right]$

       b. $\left[\text{ARG-ST} \ \langle \text{NP}_{i,-ana}, ..., \text{NP}_{i,+ana}, ... \rangle \right]$

       c. *$\left[\text{ARG-ST} \ \langle \text{XP}_{j}, ..., \text{NP}_{i,ana}, ... \rangle \right]$

   Our purpose here is not to argue in favor of the specific approach to binding
just outlined. Rather, we wish to illustrate that in a theory like HPSG where
much of syntactic distribution is accounted for by properties of lexical entries,
co-occurrence restrictions treated traditionally as constraints on trees (via some
notion of command) are modeled as constraints on the argument structure of
lexical entries. It is tempting to think of such a lexicalization of binding princi-
ples as a notational variant of tree-centric approaches. Interestingly, this is not
the case, as argued in Wechsler (1999). Wechsler argues that the difference be-
tween argument structure and valence is critical to a proper model of binding in
Balinese. Summarizing briefly, voice alternations in Balinese (e.g., objective or
agentive voices) do not alter a verb's argument structure but do alter its valence,
which is the subject and object it subcategorizes for. As binding is sensitive to
relative obliqueness within ARG-ST, binding possibilities are not affected by voice
alternations within the same clause, which are represented with different valence
values. In the case of raising, on the other hand, the argument structure of the
raising verb and the valence of the complement verb interact, as the subject of
the complement verb is part of the argument structure of the raising verb. An
HPSG approach to binding therefore predicts that voice alternations within the
embedded clause will not affect binding of co-arguments of the embedded verb,
but will affect binding of the raised NP and an argument of the embedded verb.
This prediction seems to be borne out, as the examples in (17) show.

(17) a. Ia$_i$ nawang awakne$_i$/Ia$_{*i}$ lakar tangkep polisi.
       3rd AV.know self/3rd   FUT OV.arrest police
       'He$_i$ knew that the police would arrest self$_i$./him$_{*i}$.'

    b. Cang ngaden ia$_i$ suba   ningalin awakne$_i$/ia$_{*i}$
       1sg AV.think 3rd already AV.see   self/3rd
       'I believe him$_i$ to have seen himself$_i$/ him$_{*i}$.'

    c. Cang ngaden awakne$_i$ suba   tingalin=a$_i$.
       1sg AV.think self$_i$ already OV.see=3
       'I believe him to have seen himself.'

Sentence (17a) shows that the proto-agent (the first element of ARG-ST) of the subject-to-object raising verb *nawang* 'know' can bind the raised subject (which in this case corresponds to the proto-patient of the complement verb *tangkep* 'arrest' since that verb is in the objective voice). Sentence (17b) shows that the raised (proto-agent) subject of the complement verb can bind its proto-patient argument. Critically, sentence (17c) shows that the raised proto-patient (second) argument of the complement verb can be bound by the complement verb's proto-agent. The contrast between sentences (17b) and (17c) illustrates that while binding is insensitive to valence alternations (the same proto-agent binds the same proto-patient argument in both sentences), raising is not (the proto-agent argument is raised in (17b) and the proto-patient argument in (17c)). As Wechsler argues, this dissociation between valence subjects and less oblique argumentson the ARG-ST list is hard to model in a configurational approach to binding that equates the two notions in terms of c-command or the like. What is important for our purposes is that a 'lexicalization' of argument structure, valence, and binding has explanatory power beyond tree configurations, illustrating some of analytical possibilities informationally rich lexical entries create.

3.4 Lexical vs. constructional explanations

As we have noted above, HPSG posits that much of the combinatorics of natural language syntax is lexically determined; lexical entries contain information about their combinatorial potential and, if a word occurs in two distinct syntactic contexts, it must have two distinct combinatorial potentials. Under this view, phrase-structure rules are boring and few in number. They are just the various ways for words to realize their combinatorial potential. In the version of HPSG presented in Pollard & Sag (1994), for example, there are only a handful of general phrase-structural schemata, one for a head and its complements, one for a head

and its specifier, one for a head and a filler in an unbounded dependency and so forth and the structure of clauses is relatively flat in that relations between contexts of occurrence of words is done "at the lexical level" rather through operations on trees.

In a transformational approach, on the other hand, relations between contexts of occurrence of words are seen as relations between trees, and the information included in words can be thus rather meager. In fact, in some recent approaches, lexical entries contain nothing more than some semantic and phonological information, so that even part of speech information is something provided by the syntactic context (see Borer 2003; Marantz 1997). In some constructional approaches (Goldberg (1995), for example), part of the distinct contexts of occurrence of words comes from phrase structural templates that words fit into. So again, there can be a single entry for several contexts of occurrence.

HPSG's approach to lexical knowledge is quite similar to that of Categorial Grammar (to some degree this is due to HPSG's borrowing from Categorial Grammar important aspects of its view on subcategorization). As in HPSG, the combinatorial potential of words is recorded in lexical entries so that two distinct contexts of occurrence correspond to two distinct entries. The difference from HPSG lies in how lexical entries relate to each other. In Categorial Grammar (be it Combinatorial or Lambek-calculus style), relations between entries are the result of a few general rules (e.g., type raising, function composition, hypothetical reasoning …) and the assumption is that those rules are universally available (although those rules could be organized in a type hierarchy, as in Baldridge (2002)). Relations between entries in HPSG can be much more idiosyncratic and language-specific. We note, however, that nothing prevents lexical rules constituting a part of a Categorial Grammar (see Carpenter 1992a), so that this difference is not necessarily qualitative, but concerns how much of researchers' efforts are typically spent on extracting lexical regularities; HPSG has focused much more, it seems, on such efforts.

4 The hierarchical lexicon

We have seen that lexicalism demands that lexical entries be information rich, in order to encode what might otherwise be represented as syntactic rules. To avoid massive and redundant stipulation within each lexical entry, we need mechanisms to represent regularities within the lexicon. Two main mechanisms have been used in HPSG to represent these regularities. The first mechanism is the organization of information shared by lexical entries or parts of entries into a hi-

erarchy of types in a way quite similar to semantic networks within Knowledge Representation systems (see among others Brachman & Schmolze 1985). This hierarchy of types (present in HPSG since the beginning, Pollard & Sag (1987) and the seminal work of Flickinger (1987)) ensures that individual lexical entries only specify information that is unique to them. The second mechanism is lexical rules, which relates variants of entries, and more generally, members of a lexeme's morphological family (which consists of a root or stem as well as all stems derived from that root or stem). In this section, we discuss the hierarchical organization of the lexicon into cross-cutting classes of lexical entries at various levels of generality.

4.1 Inheritance

All grammatical frameworks classify lexical entries to some extent, of course. Basic part of speech information is one obvious case. This high-level classification is present in HPSG, too, as part of the hierarchy of types of heads. That information is recorded in the value of the HEAD feature. A simple hierarchy of types of heads is depicted in (18).

(18)

$$
\begin{array}{c}
\textit{head} \\
\diagup \quad | \quad \diagdown \\
\textit{noun} \quad \textit{verb} \quad \textit{...}
\end{array}
$$
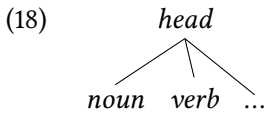
Each of these types is a partial specification of a lexical entry's head properties. Typing of HEAD information allows the ascription of appropriate properties to different classes of lexical entries. For example, case information is only relevant to nouns, and whether a verb is an auxiliary or not is only relevant to verbs. Each type in (18) includes in its definition a specification of which features are appropriate for it, as shown in (19). (19) specifies what it means to be a noun or a verb in a particular language (of course, there will strong similarities in these properties across languages).

(19) a. *noun* $\Rightarrow$ $\begin{bmatrix} \text{CASE} & \textit{case} \end{bmatrix}$

b. *verb* $\Rightarrow$ $\begin{bmatrix} \text{AUX} & \textit{boolean} \\ \text{TENSE} & \textit{tns} \\ \text{ASPECT} & \textit{asp} \end{bmatrix}$

More technically, each type of head imposes some constraint on lexical entries of that type. Thus, (19a) requires all noun lexemes to be eligible for case information. Here, the constraint on each type specifies the value of SYNSEM|LOC|CAT|HEAD.

as the atomic value *noun*, *verb*, etc. Since these atomic values are disjoint, and since the HEAD value is unique for each lexical entry, the types in (18) are also disjoint. If there's a type corresponding to each possible HEAD value, then they constitute a partition of lexical entries as well. Lexical entries for particular lexemes make use of the definitions of types like (19) to abstract information that is shared across classes of entries. Thus, the pronoun *him* need only include the fact that its HEAD is of type *noun*; the fact that it might bear case can be inferred. Similarly, the entry for the verb *can* need only include information that its head information include the specification [AUX +] for us to be able to infer that it is a *verb*.

So far, this is merely an HPSG implementation of a part of speech taxonomy, but once we consider subtypes with additional constraints the utility of the hierarchical lexicon within a lexicalist framework becomes apparent. There are interesting generalizations to be made about more specific classes, such as transitive verbs, or past participles, or predicators denoting caused motion (regardless of their part of speech). In the hierarchical lexicon, we can represent these "interesting" classes as types. Which classes are worth instantiating in the grammar of a given language depends on its grammar; thus we expect lexical classes to specify a mix of cross-linguistically common (maybe, in some cases, universal) and language-particular constraints. Consider some of the subtypes of verbs shown in (20) adapted from Bouma et al. (2000):

(20)
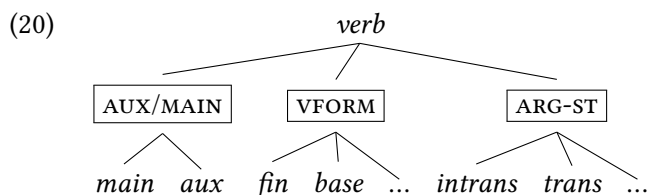
Again, each subtype specifies additional information constraining the lexical entries belonging to it. The boxed labels have no independent formal status (although they play a role in the framework of online type construction, discussed below), indicating simply that the parent type, here *verb*, is partitioned by the subtypes under each box. Typically, this means that each subtype specifies an atomic value for a particular attribute, out of a set of mutually disjoint values, as in the part of speech types above. Thus *main-verb* and *aux-verb* are disjoint subtypes of *verb*, with the values + and – for the attribute SYNSEM|LOC|CAT|HEAD|AUX.

More specific verb subtypes can combine the constraints of the types depicted in (20), through multiple inheritance. Infinitive forms of transitive verbs, for example, inherit the constraints of both *infinitive-verb* and *transitive-verb*. Provided

that the constraints do not conflict (i.e., the descriptions of the two types unify), such a type can exist and have members. Whether it is useful to reify such a type is another matter; not all possible combinations of constraints yield linguistically interesting classes of lexical entries (this is another issue we address in the discussion of online type construction). It seems desirable, for instance, to avoid a proliferation of types for each form of a verb in a paradigm in an inflectionally complex language, as the number of forms, and thus types, would be extremely large (there are, for example, 2,494 combinations of inflectional prefixes in Oneida, a Northern Iroquoian language, Karin Michelson, p.c.). While an economical type-based description of regular morphological paradigms may prove descriptively adequate, it is implausible in assuming that each form of every fully regular verb is reified as a lexical entry. We will discuss mechanisms (lexical rules and online type construction) that offer better accounts of morphologically regular and productive word formation below.

In general, types are posited in the hierarchy when there is some additional constraint to state about them. We now briefly examine some of the lower levels of the lexical hierarchy; that is, some more specific lexical types that illustrate how types is one way to reduce the amount of information that needs to be stipulated in individual lexical entries and are one of the tools HPSG employs to represent lexical generalizations. We begin with the *transitive-verb* type (*trans-vb* for short). Apart from requiring its ARG-ST list to contain two NPs, *trans-vb* is further constrained, at least in English, to be a main verb rather than an auxiliary verb (see the value of HEAD *main* in (21) and the hierarchy of verbal head information in (20)): there are no transitive auxiliaries in English. So, *trans-vb* includes information constraining the feature values of transitive verbs that goes beyond simply specifying the nature of the ARG-ST list.

The partial representation of the type *transitive-verb* in (21).

(21) $\begin{bmatrix} \textit{trans-vb} \\ \text{HEAD} \quad \textit{main} \\ \text{ARG-ST} \ \langle \text{NP, NP, ...} \rangle \end{bmatrix}$

A more specific subtype of *trans-vb* is *caused-motion-transitive-verb*, which states information about the semantics of verbs in the class as well as their subcategorization, as in (22). (↑ indicates that the type that follows it is a supertype of the type indicated in the feature structure; in this case that *caused-mot-trans-vb* is a subtype of *trans-vb*.)

(22)
$$
\begin{bmatrix}
\textit{caused-mot-trans-vb}(\uparrow \textit{trans-vb}) \\
\text{CONTENT} \begin{bmatrix} \textit{caused-motion-rel} \\ \text{CAUSER} \quad \boxed{1} \\ \text{MOVED} \quad \boxed{2} \end{bmatrix} \\
\text{ARG-ST} \quad \left\langle \text{NP}_{\boxed{1}}, \text{NP}_{\boxed{2}} \right\rangle
\end{bmatrix}
$$

The information in each type constitutes constraints on objects of that type. With the types situated in a hierarchy, each type inherits all the constraints of its supertypes. Thus constraints will be inherited from supertypes. But additional constraints can be added at the level of that type as well; this is the principal fashion in which generalizations about classes of lexical entries can be stated. For example, following Davis & Koenig (2000) and Koenig & Davis (2003) we might state a contraint on argument realization on the type *caused-motion-transitive-verb*, to ensure that the causer is linked to the subject and the entity that is caused to move to the direct object, as in (23). (cross-reference to chapter on argument structure and linking here) Here, we make use of Richter's logic (Richter 1999) to encode constraints on information that is included in lexical entries. The constraint in (23) says that if a verb's semantic content is a cause relation, the causer arguments corresponds to the index of the first NP on the ARG-ST list and that if a verb's semantic content is a motion relation, the moved entity is realized as an NP. Implicational constraints such as (23) relieve some of the burden of encoding generalizations over lexical entries exclusively through the lexical type hierarchy, and can lead to a simpler model of lexical generalizations in some cases, as Koenig & Davis (2003) point out. When it is preferable to use a hierarchy of lexical types or conditional constraints on the information included in lexical types remains an open issue.

(23) $\begin{bmatrix} \text{CONTENT} & \textit{cause-rel} \end{bmatrix} \Longrightarrow \begin{bmatrix} \text{CONT} & \begin{bmatrix} \text{CAUSER} & \boxed{1} \end{bmatrix} \\ \text{ARG-ST} & \left\langle \text{NP:}\boxed{1}, ... \right\rangle \end{bmatrix}$

$\begin{bmatrix} \text{CONTENT} & \textit{move-rel} \end{bmatrix} \Longrightarrow \begin{bmatrix} \text{CONT} & \begin{bmatrix} \text{MOVED} & \boxed{1} \end{bmatrix} \\ \text{ARG-ST} & \left\langle ...\text{NP:}\boxed{1}, ... \right\rangle \end{bmatrix}$

More specific classes of transitive caused-motion verbs, such as the *spray* verbs in English that exhibit locative alternations, inherit the additional constraints in (22) and further specify additional semantic constraints that characterize these alternating verbs. The hierarchical organization of lexical types allows us to state these additional restrictions, which are often language-particular, in the appropriate place without additional formal mechanisms. The range of ditransitive constructions, to take one such case, varies across languages, with some lacking

them entirely and others freely allowing them in, e.g., morphologicallly productive causatives of any transitive verb. For those languages, like English, in between these extremes, semantic (and possibly other) constraints can be placed on the type *ditransitive-verb*, limiting such verbs to those involving, e.g., transfer of possession.

We now illustrate how the organization of the lexicon in a hierarchy of lexical types minimizes the information that needs to be specified within individual entries, such as those for the forms of the French verb *va* we provided earlier (see (11)). We start with semantics and how it links to the argument structure. We can infer that the use of *va* illustrated in (10a) includes two arguments, a theme and a goal, from the hierarchy of semantic relations, which ensures that all types of directed motion events, of which *go-rel* is a subtype, includes these two arguments (see Davis (2001) for such an approach to semantic relations). The linking of these arguments to an NP and PP follows either from linking types, as in Davis & Koenig (2000) or Davis (2001), or from constraints similar to those we show above in (23) for English caused-motion verbs. The relation between the argument structure of *va* and its subcategorization requirements for a subject and PP complement follows from general constraints on words and a general type for intransitive verbs, analogous to (21) for transitive verbs. The inflectional features of this form are instantiations of possible combinations of values of mood, tense, and agreement information within French verbs. Finally, the expression of these inflectional features is the result of either general lexical rules (see Miller & Sag (1997) for some examples) or, as in more recent work in HPSG, a network of associations between morphosyntactic features and forms at various positions in the word (see Crysmann & Bonami 2016). In the end, nothing but the meaning of this use of *va* and the fact that the stem form is *v-* need be stipulated in the entry.

4.2 The lexicon as repository of generalizations at various levels

The hierarchical lexicon makes it possible to specify constraints on classes of lexical entries at any level, not just, e.g., all nouns, or a single word. An illustrative example, drawn from Ackerman & Webelhuth (1998), involves German passives, which come in several varieties, each with its own constraints. Each passive construction uses a different auxiliary: (*werden*, *sein*, or *bekommen*) and two of these constructions require a participial form of the verb, while the *sein* passive requires *zu* followed by an infinitive VP. Additionally, passives appear attributively, as NP modifiers, as well as predicatively. Here are two examples of the *zu* + infinitive passive, the first attributive, the second predicative:

(24) a. de dem Mann von Johann zu schenkenden Blumen
          the the man   by Johann to give   flowers
          'the flowers that must be given to the man by Johann'

     b. weil   die Blumen den Mann von Johann zu schenken sind
          because the flowers the man   by Johann to give   are
          'because the flowers must be given to the man by Johann'

Ackerman & Webelhuth's account of German passives posits a multiple inheritance hierarchy of lexical types in German, a portion of which is shown in Figure 1.
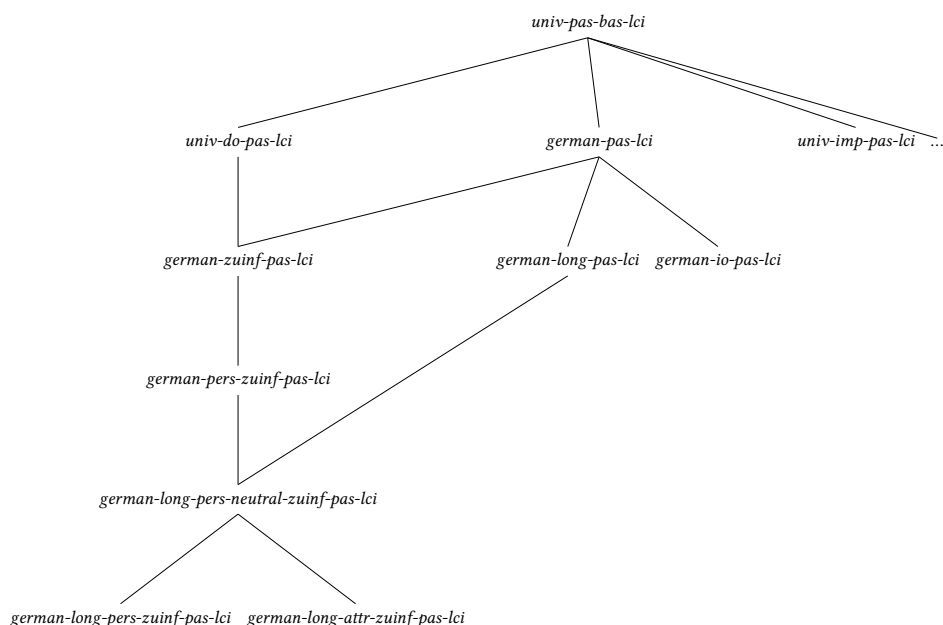
Figure 1: A portion of the hierarchy of passive lexical types according to Ackermann and Webelhuth, p.244

While all passives share the constraint that a logical subject is demoted, as stipulated on a general *univ-pas-bas-lci* passive type, the other requirements for each kind of passive are stated on various subtypes. The *zu*+infinitive passive, for instance, requires not only that *sein* is the auxiliary and that the main verb is infinitive, but that the semantics involves necessity or obligation. This differs from the other passives, which simply maintain the semantics of their active counterparts. However, the types of the passive verb *schenken(den)* in (24) both

inherit from several passive verb supertypes. As mentioned, at a general level, there is information common to all German passives, or indeed to passives universally, namely that the "logical subject" (first element of the basic verb's ARG-ST list) is realized as an oblique complement of the passive verb, or not at all. A very common subtype, which Ackerman & Webelhuth also regard as universal, rather than specific to German, specifies that the base verb's direct object is realized as the subject of its passive counterpart; this defines personal passives. Once in the German-specific realm, an additional subtype specifies that the logical subject, if realized, is the object of a *von*-PP; this holds true of all three types of German personal passives. Among its subtypes is one that requires *zu* and the infinitive form of the verb; moreover, although Ackerman & Webelhuth do not spell this out in detail, this subtype specifies the modal force characteristic of this passive construction but not of the others. Finally, both the predicative and attributive forms are subtypes of all the preceding, but these inherit also from distinct supertypes for predicative and attributive passives of all kinds. The supertype for predicative passives constrains them to occur with an auxiliary; its subtype for *zu* + infinitive passives further specifies that the auxiliary is *sein*. The attributive passive type, on the other hand, inherits from modifier types generally, which do not allow auxiliaries, but do require agreement in person, number, and case with the modified noun. In summary, the hierarchical lexicon is deployed here to factor out the differing properties of the various German passive constructions, each of which obtains its particular combination of properties via multiple inheritance.

The most specific types of the lexical hierarchy, where individual lexical entries reside, is where constraints pertaining solely to a given word or root – its phonological form, inflectional class, specific semantics, register, and so forth – are stated. Specific information about a word needs to be spelled out somewhere in any grammatical framework. In a hierarchically organized lexicon we can view this as just the narrowest, most particular case of specifying information about a class of linguistic entities. But where information is shared across a broader set of lexical entries, it need not be stated separately for each one. Thus, the phonology of the word *spray* and the precise manner of motion of the particles or liquid caused to move in a spraying event are unique to this lexical entry. However, much of its syntactic and semantic behavior— it is a regular verb, participating in a locative alternation, involving caused ballistic motion of a liquid or other dispersable material— is shared with other English verbs such as *splash*, *splatter*, *inject*, *squirt*, and *drizzle*. To the extent that these "narrow conflation classes," as Pinker (1989) terms them, are founded on clear semantic criteria, we can readily

state syntactic and semantic constraints at the appropriate level in the hierarchical lexicon (some, however, such as Briscoe & Copestake (1999), cast doubt on the feasibility of formulating such constraints for dative and other alternations in English, suggesting that lexical rules might be a better alternative). Given this semantic similarity, it may be that much of the semantics of a verb like *spray* need not be specified at the level of that individual lexical entry. Apart from the broad semantics of caused motion, shared by numerous verbs, the verbs in the narrow conflation class containing *spray* share the selectional restriction, noted above, that their objects are set in motion by an initial impulse and that they are liquid or particulate material. We might therefore posit a subtype of the type *caused-motion-rel* to represent this shared semantics triggering the locative alternation, with further subtypes of that for the semantics of the individual verbs. Note that not all these constraints apply to precisely the same class (there are other verbs with somewhat different semantics, like *load* and *wrap*, exhibiting the locative alternation, for example), so a multitude of types in the hierarchy is crucial.

4.3 Default inheritance in the lexicon

So far, we have assumed rigid, monotonic inheritance of all information in supertypes to their subtypes; none of the inherited information can be overridden. This runs into difficulties when dealing with lexical entries that appear to be exceptional in some way, the obvious examples being morphological irregularities. How can productive regular forms such as **childs* be blocked, and only *children* allowed as a lexical entry?

While several approaches to exceptions have been proposed, we will focus here on *default unification*; that is, weakening monotonic inheritance in some circumstances. Then, although the plural of *child* might inherit the information from the pertinent lexical entry and from the *plural-noun* type, which would entail the phonology for **childs*, this regular plural form is overridden. Various complex issues arise in attempting to formulate a workable system of default unification and inheritance. See, e.g., Briscoe & Copestake (1999) for a brief overview of various ways that default unification might be defined. Lascarides & Copestake (1999) list several desirable criteria, including:

- Non-default information is always preserved

- Default unification behaves like monotonic unification whenever possible

- Default unification is order-independent

They explore the properties of their system, called YADU, in considerable detail. The intent is to preserve the behavior of non-default unification in cases where no default information is present, and for defeasible information at more specific level in the type hierarchy to override defeasible information at a more general level.

As another example of the use of default, nonmonotonic inheritance, outside of morphology, consider the account of the syntax of gerunds in various languages developed by Malouf (2000). Gerunds exhibit both verbal and nominal characteristics, and furnish a well-known example of seemingly graded category membership, which does not accord well with the categorical assumptions of mainstream syntactic frameworks. Roughly speaking, English gerunds, and their counterparts in other languages, act much like verbs in their "internal" syntax, allowing direct objects and adverbial modifiers, but function distributionally ("externally") as NPs. To take but a couple of pieces of evidence (see Malouf, op.cit. p.27 et seq. for more details), gerunds can be the complement of prepositions when finite clauses cannot (see (25)); conversely, adverbs, but not adjectives can modify gerunds, but adjectives must modify deverbal nouns (see (26)).

(25) a. Pat is concerned about Sandy('s) getting arrested.

   b. * Pat is concerned about (that) Sandy got arrested.

(26) a. Pat disapproved of (me/my) *quiet/quietly leaving before anyone noticed.

   b. Pat disapproved of my quiet/*quietly departure.

In contrast to accounts that attempt to model this dichotomy directly, via syntactic rules that allow an NP to be expanded as a constituent internally headed by a verb, Malouf posits a lexical rule, which converts the lexical category of a verb to *noun*, but otherwise preserves its verbal properties, such as subcategorization. This would pose problems with strictly monotonic inheritance, however, as it would force us to abandon generalizations about nouns other than gerunds (e.g., they do not take direct object complements, as many verbs and their gerunds do). Default inheritance provides one way to model the observed phenomena, without weakening the constraints on parts of speech to the point where no meaningful constraints distinguish them.

Malouf notes that some possible combinations of noun-like and verb-like attributes are frequently attested cross-linguistically in gerunds and their equivalents, while others are rare or unattested. Cross-linguistically, gerunds vary in their subcategorization possibilities: some allow subjects and complements,

while some allow only complements and no subjects. But there appear to be no cases of gerund-like lexical items that can take a subject but cannot take complements. Malouf invokes default inheritance (Lascarides & Copestake 1999) as a mechanism to represent these generalizations. In his account, there are both "hard" constraints – a verb lexical entry, for example, must have a HEAD value of type *relational* (encompassing verbs, adjectives, and adpositions) – and "soft", overridable constraints – a verb lexical entry by default has a HEAD value of type *verb*. In addition, following Bouma et al. (2001), he posits the types *ext-subj* and *ext-spr*. The former constrains the HEAD value to *relational* and the first element of the ARG-ST list to be the SUBJ (only adjective, adpositions, and verbs have subjects), while the latter constrains the HEAD value to *noun* and the first element of the ARG-ST list to be the SPR (only nouns have specifiers), as shown in (27).

(27) a.
$$\begin{bmatrix} \textit{ext-subj} \\ \text{HEAD} \quad \textit{relational} \\ \text{VAL} \quad [\text{SUBJ} \; \boxed{1}] \\ \text{ARG-ST} \; \langle \boxed{1}, ... \rangle \end{bmatrix}$$
b.
$$\begin{bmatrix} \textit{ext-spr} \\ \text{HEAD} \quad \textit{noun} \\ \text{VAL} \quad [\text{SPR} \; \boxed{1}] \\ \text{ARG-ST} \; \langle \boxed{1}, ... \rangle \end{bmatrix}$$

Malouf then specifies default HEAD values for the lexical classes *n* and *v* (see (29) for the latter's definition). As gerunds have both properties of nominal and relational heads, they are subtypes of both, as shown in the multiple inheritance hierarchy in (28). The *v* type, which concerns us here, has a default HEAD value *verb*, as shown in (29) in addition to the non-default, more general type *relational* it also includes (default information follows /).

(28)
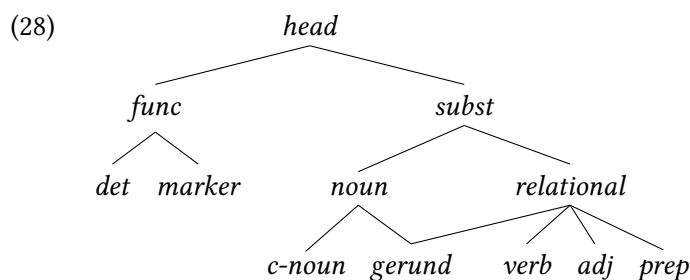
(29)
$$\begin{bmatrix} \textit{v} \\ \text{HEAD} \quad \textit{relational /verb} \\ \text{CONT} \quad \textit{psoa} \end{bmatrix}$$

However, the default value *verb* is overridden in the subtype *vger*, in which the HEAD value is *gerund*, which is a subtype of both *noun* and *relational*, but not of *verb*. The type *vger* is shown in (30); where *f-ing* is a function the produces the *-ing* form of an English verb from its root.

(30) $\begin{bmatrix} \textit{vger} \\ \text{MORPH} \begin{bmatrix} \text{ROOT} & \boxed{1} \\ \text{I-FORM} & \textit{f-ing}(\boxed{1}) \end{bmatrix} \\ \text{HEAD} & \textit{gerund} \end{bmatrix}$

The type *vger* is thus compatible with "verb-like" characteristics; in particular, it has an ARG-ST list. But, as its HEAD is also a subtype of *noun*, it lacks a SUBJ attribute and instead has a SPR attribute. Gerunds therefore allow complements (unlike ordinary nouns), but not subjects (unlike ordinary verbs). Malouf's hierarchy of types makes this prediction, in effect, because the *ext-spr* type requires that the "external argument" (the first on the ARG-ST list) is realized as the value of SPR.

While it would be possible to construct type hierarchies of lexical types, HEAD types, and so on that would allow for this kind of "reverse gerunds" – those that would act externally as nouns, allow subjects, but not permit complements – this would require reorganizing these type hierarchies to a considerable extent. Given that many nouns besides gerunds – nominalizations, for example – are relational (that is, have a CONTENT value of type *psoa*), it could be difficult to model a hypothetical language that permits only the reverse gerunds rather than the normal ones.

Malouf further notes a key difference between gerunds and exceptions like **childs/children*: English gerunds are productive (and completely regular morphologically). If the same mechanisms of default unification are involved in both, what accounts for this difference? His answer is that productive and predictable processes involve online type construction (see Section 5.2 for details). The irregular form *children* must of course be learned and stored, not generated online. The default mechanisms described above, however, are employed at higher levels of the lexical hierarchy, and the individual gerunds forms *are* productively generated online. Note that, in contrast to the morphological and syntactic consistency among gerunds, English nominalizations display some idiosyncrasies that suggest at least some of them must be stored as distinct lexical items. Thus, as Malouf emphasizes, modeling prototypicality in the lexicon within HPSG can draw on both default inheritance and online type construction; together, they make "the connection between prototypicality, and productivity."

5 Lexical rules

In this section we describe the role lexical rules play in HPSG as well as their formal nature, i.e., how they model "horizontal" relations among elements of the

lexicon. These are relations between variants of a single entry (be they subcat-egorizational or inflectional variants) or between members of a morphological family, as opposed to the "vertical" relations modeled through inheritance. Thus they provide a means to represent the intuitive notion of "derivation" of one lexeme from another.

While lexical rules or similar devices have been invoked within HPSG since its inception, formalizing their nature and behavior was deferred until somewhat later. The intent, however, has always been, as Lahm (2016) stresses, to treat lexical rules (typically written $A \mapsto B$) to mean that for every lexeme or word described by A there is one described by B that has as much in common with A as possible.

Copestake & Briscoe (1991), Briscoe & Copestake (1999), Meurers (2001), and many others formalize the notion of lexical rule within HPSG by introducing a type, say *lex-rule*, with the attributes IN and OUT, whose values are respectively the rule's input and output lexical entries. As Briscoe & Copestake (1999) note, lexical rules of this form also bear a close relationship to default unification. The information in the input is intended to carry over to the output by default, except where the rule specifies otherwise and overrides this information. But, as Lahm (2016) points out, a sound basis for the formal details of how lexical rules work is not easy. Meurers' careful analysis of how to apply lexical rules to map a descrip-tion of an entry A into the description B does not always work as intended in that what would expect to be licit inputs are actually not and no output description results as a consequence. Fortunately, it is not clear that this is a severe problem in practice, and Lahm notes that he has not found an example of practical im-port where Meurers's lexical rule formulation would encounter the problems he raises.

In a slight variant of the representation of lexical rules proposed by Copestake & Briscoe and Meurers, the OUT attribute can be dispensed with; the information in the lexical rule type not within the IN value then constitutes the output of the rule. In this variant, lexical rules could alternatively be viewed as be subtypes of a *derived-word* type, which could combine with other types in the lexical hi-erarchy, merely adding the derivational source via the IN value. Formulated in either fashion, lexical rules are essentially equivalent to unary syntactic rules, with the IN attribute corresponding to the daugther and the OUT attribute (or the rest of the information in the rule, if the OUT attribute is done away with) to the mother. This is the way lexical rules are implemented in the English Re-source Grammar (see http://www.delph-in.net/erg/ for demos and details about this large-scale implemented grammar of English). (cross-reference to Bender &

Emerson's chapter on computational linguistics and language engineering here)

5.1 Phenomena accounted for by lexical rules

Lexical rules have been put to many uses, derivational and inflectional morphology (Copestake & Briscoe 1995; Emerson & Copestake 2015), complex predicate formation (Müller 2010), and diathesis alternations (Davis 2001). Moreover, proposals for lexical rules in HPSG have extended beyond what are traditionally or evidently viewed as lexical phenomena, to include treatments of extraction, unbounded dependencies, and adjuncts. In this section, we describe the use of lexical rules to model the realization of arguments as extracted dependents or affixes, rather than complements. We concentrate on these two cases, which we will contrast with alternative analyses not involving lexical rules presented by the same authors (see the next section). They thus provide a good illustration of some of the analytical choices available to model relations between variant lexical entries that are based on a single stem.

We begin with the Complement Extraction Lexical Rule (hereafter, CELR) proposed in Pollard & Sag (1994) shown in (31). The input to the rule is any lexeme that selects for a syntactic argument ([3]) that the lexeme requires be expressed as a complement (as indicated, this syntactic argument is also a member of the COMPS list). The ouput stipulates that this same syntactic argument is no longer a member of the COMPS list; however, the SLASH set now includes a new element, which is the local information of this syntactic argument ([1]). Informally stated, the input entry specifies that a syntactic argument must be realized as a complement, whereas the output entry specifies that the same syntactic argument must be realized by a non-local dependent (see Pollard & Sag (1994) for why only LOCAL information is shared between syntactic arguments and fillers that realize them).

$$
(31)\quad
\begin{bmatrix}
\text{ARG-ST} & \langle ..., \boxed{3}, ... \rangle \\
\text{COMPS} & \langle ... \boxed{3}[\text{LOC } \boxed{1}], ... \rangle \\
\text{SLASH} & \boxed{2}
\end{bmatrix}
\mapsto
\begin{bmatrix}
\text{ARG-ST} & \left\langle ..., \boxed{4}\begin{bmatrix}\text{LOC } \boxed{1} \\ \text{SLASH } \boxed{1}\end{bmatrix}, ... \right\rangle \\
\text{COMPS} & \langle ... \rangle \\
\text{SLASH} & \{\boxed{1}\} \cup \boxed{2}
\end{bmatrix}
$$

A similar use of lexical rules to model alternative realizations of arguments can be found in Monachesi (1993), who analyzes alternations between complements and so-called object clitics in Italian in a way that parallels the French examples in (10). In the output of her lexical rule, in (32), a subset of the list of complements in the input ([2]) corresponds to a list of clitic SYNSEMs, realized as prefixes through inflectional rules not shown here.

$$(32) \quad \begin{bmatrix} word \\ \text{HEAD} & verb \\ \text{VAL|COMPS} & \boxed{1} \bigcirc \boxed{2} \\ \text{CLTS} & elist \end{bmatrix} \mapsto \begin{bmatrix} word \\ \text{VAL|COMPS} & \boxed{1} \\ \text{CLTS} & \boxed{2} list(cl\text{-}ss) \end{bmatrix}$$

Here as well, a lexical rule is employed in an analysis of what might well be considered a syntactic phenomenon. The possibility of treating phenomena like extraction and clitic placement at a lexical level, however, makes sense when they are considered fundamentally as matters of the combinatorial requirements of predicators, rather than effects of movement.
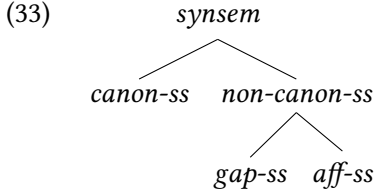
Before turning to the alternatives, we note in passing that lexical rules are inherently "directional", with an input and an output. This seems intuitively correct in the cases we've discussed, but might not always be so. Is there inherent directionality, for example between the causative and inchoative alternants of verbs such as *melt* or *slide*? In constrast, the alternatives to lexical rules described in the following section lack this notion of directionality.

5.2 Alternatives to lexical rules

In this section we briefly examine two alternatives to lexical rules, each involving underspecification. The types of members of the ARG-ST list might be underspecified so that a lexical entry accounts for more than one subcategorization. Or the type of the entry itself may be underspecified, so that it subsumes multiple inflectional or derivations forms. In both cases, the intent is that sufficiently underspecified information covers multiple entries that would otherwise have to be specified and related by lexical rules. We begin with alternatives to the complement extraction and clitic lexical rules in (31) and (32), proposed in Bouma et al. (2001) and Miller & Sag (1997).

In both cases, the idea is to distinguish between "canonical" and "non-canonical" realizations of syntactic arguments, as shown in the hierarchy of *synsem*s in (9). "Canonical" means local realization as a complement or subject/specifier, and "non-canonical" means realization as an affix or filler of an unbounded dependency. Linking constraints between semantic roles (values of argument positions) and syntactic arguments (members of ARG-ST) do not specify whether the realization is canonical or not; thus they retain their original form. Only canonical members of ARG-ST must be structured-shared with members of valence lists. The two constraints that determine the non-canonical realization of fillers are shown in (34). (34a) specifies what it means to be a *gap-ss*), namely that the argument is extracted (its local information is "slashed,") whereras (34b) prohibits any *gap-ss* member from being a member of the COMPS list. As these two constraints

are compatible with either a canonical or extracted object, there is no need for the lexical rule is (31).

(33)

$$
\begin{array}{c}
\textit{synsem} \\
\diagdown \\
\textit{canon-ss} \quad \textit{non-canon-ss} \\
\diagdown \\
\textit{gap-ss} \quad \textit{aff-ss}
\end{array}
$$

(34)   a. $\textit{gap-ss} \Longrightarrow \begin{bmatrix} \text{LOC} & \boxed{1} \\ \text{SLASH} & \boxed{1} \end{bmatrix}$

     b. $\textit{word} \Longrightarrow \begin{bmatrix} \text{SUBJ} & \boxed{1} \\ \text{COMPS} & \boxed{2} \ominus \textit{list(gap-ss)} \\ \text{DEPS} & \boxed{1} \oplus \boxed{2} \end{bmatrix}$

Miller & Sag (1997) make a similar use of non-canonical relations between the ARG-ST list and the valence lists, eschewing lexical rules to model French clitics and proposing instead the constraint in (35), where a subset of ARG-ST members, those that are realized as affixes (of type *aff*) are not also subcategorized for as complements.

(35)

$$
\begin{bmatrix} \text{MORPH} & \begin{bmatrix} \text{FORM} & F_{PRAF}(\boxed{1}, \dots) \\ \text{I-FORM} & \boxed{1} \end{bmatrix} \\[2ex] \text{SYNSEM} & \begin{bmatrix} \text{LOC} | \text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{verb} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \boxed{2} \\ \text{COMPS} & \boxed{3} \textit{list(non-aff)} \end{bmatrix} \\ \text{ARG-ST} & (\boxed{2} \oplus \boxed{3}) \bigcirc \textit{nelist(aff)} \end{bmatrix} \end{bmatrix} \end{bmatrix}
$$

The second alternative to lexical rules based on underspecification was presented in Koenig & Jurafsky (1994) and Koenig (1999). Typically in HPSG, all possible combinations of types are reified in the type hierarchy (in fact, they must be present, per the requirement that the hierarchy be sort-resolved, Carpenter 1992b,Pollard & Sag 1994). Thus, if one partitions verb lexemes into transitive and intransitive and, orthogonally, into, say, finite verbs and gerunds (limiting ourselves to two dimensions here for simplicity), the type hierarchy must also contain the combinations transitive+finite, transitive+gerund, intransitive+finite, and intransitive+gerund. Naturally, this kind of fully enumerated type system is unsatisfying. For one thing, there is no additional information that the combination subtype transitive+finite carries that is not present in its the two supertypes transitive and finite, and similarly for the other combinations. In contrast to

the "ordinary" types, posited to represent information shared by classes of lexemes, these combinations seem to have no other function than to satisfy a formal requirement of the mathematical structure of a type hierarchy (namely, that it forms a lattice under meet and join). Second, and related to the first point, this completely elaborated type hierarchy is redundant. Once you know that all verbs fall into two valence classes, transitive and intransitive, and simultaneously into two inflectional classes, finite and gerund, and that valence and inflection are two orthogonal dimensions of classification of verbs, you know all you need to know; the type of any verb can be completely predicted from these two orthogonal dimensions of classification and standard propositional calculus inferences.

In (36) is a simplified version of the hierarchy in (20), where the boxed labels in small caps VFORM and ARG-ST are mnemonic names of orthogonal dimensions of classification of subcategories of verbs (and are not themselves labels of subcategories). Inheritance links to the predictable subtypes are dashed and their names grayed out; this indicates that these types can be inferred, and need not be declared explicitly as part of the grammar. A grammar of English would include statements to the effect that head information about verbs includes a classification of verbs into finite or base forms (of course, there would be more types of verb forms in a realistic grammar of English) as well a classification into intransitive and transitive verbs (again, a realistic grammar would include many more types).

(36)
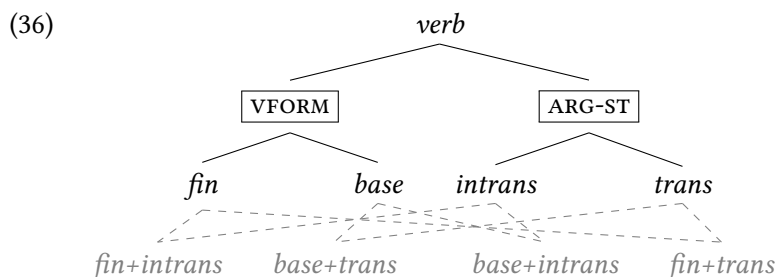
Crysmann & Bonami (2016) have shown how this *online type construction*, where predictable combinations of types of orthogonal dimensions of classification are not reified in the grammar, is useful when modeling productive inflectional morphology. Consider, for example, exponents of morphosyntactic features whose shape remains constant, but whose position within a word's template (to speak informally here) varies. One case like this is the subject and object markers of Swahili, which can occur in multiple slots in the Swahili verb template. For reasons of space we illustrate the usefulness of this dynamic ap-

proach to type creation, the Type Underspecified Hierarchical Lexicon (TUHL) with an example from Koenig (1999), the cross-cutting classification of syntactic/semantic information and stem form in the entry for the French verb *aller* (see Bonami & Boyé (2001) for a much more thorough discussion of French stem allomorphy along similar lines; Crysmann and Bonami's much more developed approach to stem allomorphy would model the same phenomena differently and we use Koenig's simplified presentation for expository purposes only). The forms of *aller* are based on four different suppletive stems: *all-* (1st and 2nd person plural of the indicative and imperative present, infinitive, past participle, and imperfective past), *i-* (future and conditional), *v-* (1st-3rd person singular and 3rd person plural of the indicative present), and *aill-* (subjunctive present). These four suppletive stems are shared by all entries (i.e., senses) of the lexeme *aller*: the one which means 'to fit' as well as the one which means 'to leave', as shown in (37) (see Koenig, op.cit, p.40-41). The cross-cutting generalizations over lexemes and stems are represented in Figure 2. Any *aller* stem combines one entry and one stem form. In a traditional HPSG type hierarchy, each combination of types (grayed out in Figure 2), would have to be stipulated. In a TUHL, these combinations can be dynamically created when an instance of *aller* needs to be produced or comprehended.

(37) a. Marc est   allé   à Paris.
          Marc be-PR.3RD.SG go-PPT to Paris
          'Marc went to Paris.'

  b. Marc s'en   ira.   'Marc will leave.'
      Marc 3.REFL-of.it go-FUT.3RD.SG

  c. Ce   costume te   va   bien.
      This suit   you go-PR.3.SG well
      'This suit becomes you.' (lit. goes well to you)

  d. Il faut   que j'y   aille.
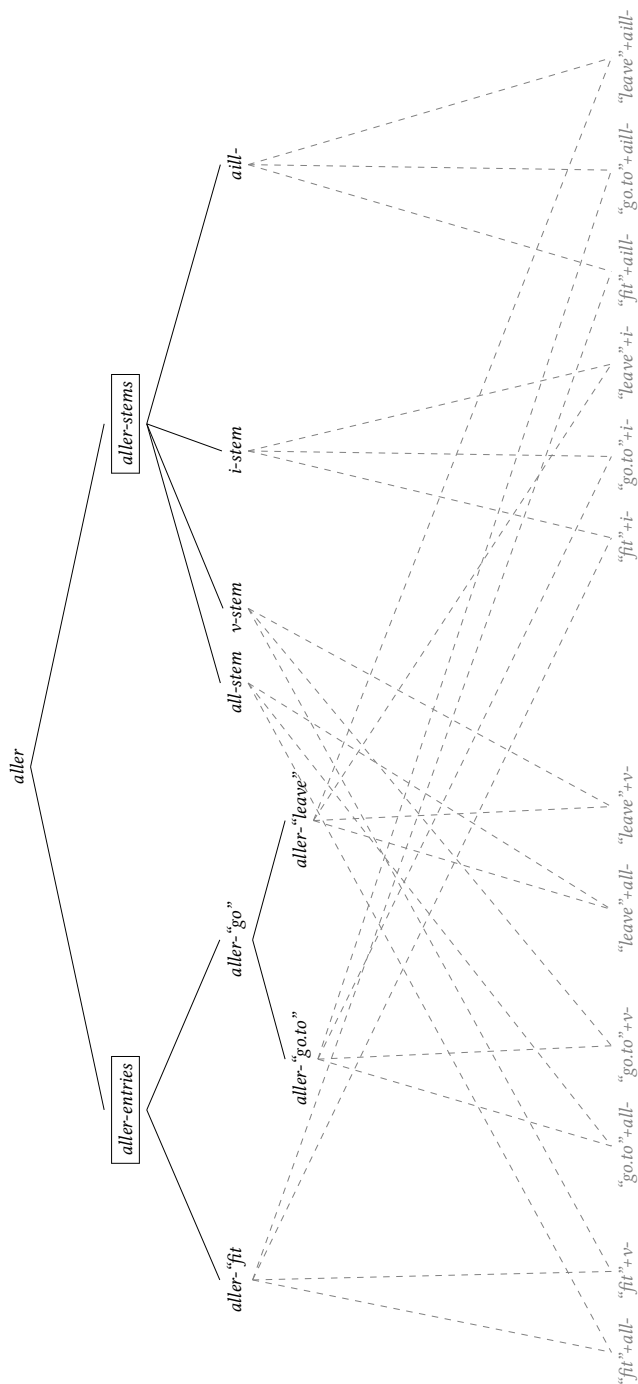   It must that I.to.there go-SUBJ.PR.1.SG
      'I must go there.'

Figure 2: A hierarchy of lexical entries and stem-forms for the French verb *aller*, from Koenig (1999)

Both *synsem* and type underspecification avoid conflict between the information specified in the variants of words based on a single lexeme (e.g., conflicts on how syntactic arguments are realized); they abstract over the relevant pieces of conflicting information. Underspecifying information included in lexical entries or lexical types allows a single entry or type to stand for the two distinct entries or types that would be related as input and output by lexical rules. The third alternative to lexical rules eschews informational conflict by adding internal structure to stems and words. This is the approach to derivational morphology taken by Riehemann (1998). Example (38) (Riehemann's (1)) illustrates *-bar* suffixation in German, a process by which an adjective that includes a modal component can be derived from verb stems (similar to English *-able* suffixation). A lexical rule approach would posit a verb stem input and derive an adjective output. As Riehemann stresses, though, there are many different subtypes of *-bar* suffixation, some productive, some unproductive, all sharing some information. This combination of productive and unproductive variants of a lexical process is exactly what the type hierarchy is meant to capture and what Riehemann's *Type-Based Derivational Morphology* capitalizes on. (39) presents the relevant information of Riehemann's type for regular *-bar* adjectives (see p.68 for more details). Critically, *-bar* adjectives include a singleton-list base (the value of MORPH-B) that records the information of the adjective's verbal base (what would be the lexical rule's input). Because of this extra layer, the local information in the base (*local*1) and the *-bar* adjective (*local*2) can differ without being in conflict.

(38) Sie   bemerken die Veränderung. Die Veränderung ist bemerkbar.
      They notice   the change.   The change   is   noticeable.

(39)
$$
\begin{bmatrix}
\textit{reg-bar-adj} \\
\text{PHONOLOGY} \quad \boxed{1}\textit{+bar} \\
\text{MORPH-B} \quad \left\langle \begin{bmatrix} \textit{trans-verb} \\ \text{PHON} \quad \boxed{1} \\ \text{LOCAL} \quad \textit{local}_1 \end{bmatrix} \right\rangle \\
\text{SYNSEM|LOCAL} \quad \textit{local}_2
\end{bmatrix}
$$

Lexical rules played a critical role in the rise of lexicalist approaches to syntax. But the three alternative analytical tools we discussed in this section (which, of course, can be combined in an analysis) have chipped away at their use in HPSG. Inflectional morphology is now dealt with through lexical types associating morphosyntactic features with forms/positions and constraints on words (ensuring that all morphosyntactic features are realized). (cross-reference the chapter on Morphology) Derivational morphology is handled via lexical types

too, but ones that add an extra internal layer (the ᴍᴏʀᴘʜ-ʙᴀsᴇ in Riehemann's analysis and (39)). Non-canonical realization of syntactic arguments as affixes or fillers in unbounded dependencies is now modeled by distinguishing kinds of members of the ᴀʀɢ-sᴛ list and constraints on words that relate valence, argument structure, and dependents lists. So, what remains of the case for lexical rules now? Müller (2006; 2010) argues that diathesis phenomena, broadly speaking, favor a lexical rules approach over a phrase-structural constructional approach à la Goldberg (1995) or an online type construction approach suggested in Kay (2002). The arguments are convincing, but it should be noted that some of the data involves derivational morphology (e.g., causatives) or passive morphemes, which, arguably, could be handled via a Type-Based Derivational Morphology of the kind Riehemann argues for (such an approach was suggested in Koenig (1999: Chapter 4)). It is unclear to us whether there are always motivated types for morphologically derived stems to dispense entirely with lexical rules of the kind Müller argues for. On the other hand, if one adopts the version of lexical rules in which the ᴏᴜᴛ attribute is eliminated and lexical rules are subtypes of *derived-lexeme*, little will be at stake formally, as lexical rules and derivational processes "look the same."

Abbreviations

Acknowledgements

References

Ackerman, Farrell & Gert Webelhuth. 1998. *A theory of predicates.* Stanford, CA: CSLI Publications.

Anderson, Stephen. 1992. *A-morphous morphology.* Cambridge: Cambridge University Press.

Anderson, Stephen. 2005. *Aspects of the theory of clitics.* Oxford: Oxford University Press.

Baldridge, Jason. 2002. *Lexically specified derivational control in combinatory categorial grammar.* Edinburgh: University of Edinburgh dissertation.

Bochner, Harry. 1993. *Simplicity in generative morphology.* Berlin: Mouton de Gruyter.

Bonami, Olivier & Gilles Boyé. 2001. Suppletion and dependency in inflectional morphology. In Frank Van Eyde, Lars Hellan & Dorothee Beermann (eds.), *Proceedings of the 2001 HPSG conference*, 51–70. Stanford: CSLI Publications.

Booij, Gert. 2005. Compounding and derivation: Evidence for Construction Morphology. In Wolfgang Dressler, Dieter Kastovsky, Oskar Pfeiffer & Franx Rainer (eds.), *Morphology and its demarcations*, 109–132. Amsterdam: John Benjamins Publishing Co.

Borer, Hagit. 2003. Exo-skeletal vs. endo-skeletal explanations: syntactic projections and the lexicon. In John Moore & Maria Polinsky (eds.), *The nature of explanation in linguistic theory*, 31–67. Stanford: CSLI Publications.

Bouma, Gosse, Frank Van Eynde & Dan Flickinger. 2000. Constraint-based lexicons. In Frank Van Eynde & Dafydd Gibbon (eds.), *Lexicon development for speech and language processing* (Text, Speech and Language Technology). Dordrecht: Kluwer Academic Publishers.

Bouma, Gosse, Robert Malouf & Ivan Sag. 2001. Satisfying constraints on extraction and adjunction. *Natural Language and Linguistic Theory* 19. 1–65.

Brachman, Ronald & James Schmolze. 1985. An overview of the KL-ONE knowledge representation system. *Cognitive Science*. 2nd ser. 9. 171–216.

Bresnan, Joan & Sam Mchombo. 1995. The lexical integrity principle: evidence from Bantu. *Natural Language and Linguistic Theory* 13. 181–254.

Briscoe, Ted & Ann Copestake. 1999. Lexical rules in constraint-based grammars. *Computational Linguistics* 25(4). 487–526.

Broadwell, George Aaron. 2008. Turkish suspended affixation is lexical sharing. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG 2008 conference*, 198–213. Stanford, CA: CSLI Publications. http://csli-publications.stanford.edu/LFG/13/, accessed 2018-2-25.

Bruening, Benjamin. 2018. The lexicalist hypothesis: both wrong and superfluous. *Language* 94. 1–42.

Carpenter, Bob. 1992a. Categorial grammars, lexical rules, and the English predicative. In Bob Levine (ed.), *Formal grammar: theory and implementation*, 168–242. New York, NY: Oxford University Press.

Carpenter, Bob. 1992b. *The logic of typed feature structures* (Tracts in Theoretical Computer Science). Cambridge: Cambridge University Press.

Chaves, Rui. 2008. Linearization-based word-part ellipsis. *Linguistics and Philosophy* 31. 261–307.

Chaves, Rui. 2014. On the disunity of right-node raising phenomena: extraposition, ellipsis, and deletion. *Language* 90. 834–886.

Chomsky, Noam. 1970. Remarks on nominalization. In Roderick A. Jacobs & Peter S. Rosenbaum (eds.), *Readings in english transformational grammar*, 184–221. Boston: Ginn.

Copestake, Ann & Ted Briscoe. 1991. Lexical operations in a unification-based framework. In James Pustejovsky & Sabine Bergler (eds.), *Lexical semantics and knowledge representation*, 101–119. Berlin: Springer Verlag.

Copestake, Ann & Ted Briscoe. 1995. Semi-productive polysemy and sense extension. *Journal of Semantics* 12. 15–67.

Crysmann, Berthold. 2001a. Clitics and coordination in linear structure. In Birgit Gerlach & Janet Grijzenhout (eds.), *Clitics in phonology, morphology and syntax* (Linguistik Aktuell/Linguistics Today 36), 121–159. Amsterdam: John Benjamins Publishing Co.

Crysmann, Berthold. 2001b. On the placement and morphology of Udi subject agreement. In Frank Van Eynde, Lars Hellan & Dorothee Beermann (eds.), *Proceedings of the HPSG-2001 Conference, Norwegian University of Science and Technology, Trondheim.* Stanford, CA: CSLI Publications. http://csli-publications.stanford.edu/HPSG/2001/, accessed 2015-12-23.

Crysmann, Berthold & Olivier Bonami. 2016. Variable morphotactics and Information-based Morphology. *Journal of Linguistics* 52. 311–374.

Davis, Anthony. 2001. *Linking by types in the hierarchical lexicon.* Stanford: CSLI Publications.

Davis, Anthony & Jean-Pierre Koenig. 2000. Linking as constraints on word classes in a hierarchical lexicon. *Language* 76. 56–91.

Di Sciullo, Anna-Maria & Edwin Williams. 1987. *On the definition of word.* Cambridge, MA: MIT Press.

Dost, Ascander. 2007. A domain-based approach to second position clitics in Pashto. In Frederick Hoyt, Nikki Seifert, Alexandra Teodorescu & Jessica White (eds.), *Texas linguistic society ix: the morphosyntax of underrepresented languages.* Austin, TX.

Emerson, Guy & Ann Copestake. 2015. Lacking integrity: HPSG as a morphosyntactic theory. In Stefan Müller (ed.), *Proceedings of the 22nd International Conference on Head-Driven Phrase Structure Grammar, Nanyang Technological University (NTU), Singapore*, 75–95. Stanford, CA: CSLI Publications. http://csli-publications.stanford.edu/HPSG/2015/, accessed 2018-2-25.

Flickinger, Daniel. 1987. *Lexical rules in the hierarchical lexicon.* Stanford University dissertation.

Goldberg, Adele E. 1995. *Constructions: A Construction Grammar approach to argument structure* (Cognitive Theory of Language and Culture). Chicago/London: The University of Chicago Press.

Halle, Morris & Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In Kenneth Hale & Samuel Jay Keyser (eds.), *The view from building 20*, 111–176. Cambridge, MA: MIT Press.

Harris, Alice. 2000. Where in the word is the Udi clitic? *Language* 76. 593–616.

Haspelmath, Martin. 2011. The indeterminacy of word segmentation1 and the nature of morphology and syntax. *Folia Linguistica* 45. 31–80.

Jackendoff, Ray. 1990. *Semantic structures*. Cambridge, MA: MIT Press.

Jackendoff, Ray S. 1975. Morphological and semantic regularities in the lexikon. *Language* 51(3). 639–671.

Kathol, Andreas. 1999. *Linearization-based syntax*. Oxford: Oxford University Press.

Kay, Paul. 2002. An informal sketch of a formal architecture for construction grammar. *Grammars* 5. 1–19.

Koenig, Jean-Pierre. 1999. *Lexical relations* (Stanford Monographs in Linguistics). Stanford, CA: CSLI Publications.

Koenig, Jean-Pierre & Anthony Davis. 2003. Semantically transparent linking in HPSG. In Stefan Müller (ed.), *Proceedings of the HPSG 2003 conference*, 222–235. Stanford: CSLI Publications.

Koenig, Jean-Pierre & Daniel Jurafsky. 1994. Type underspecification and on-line type construction in the lexicon. In *Proceedings of WCCFL XIII*, 270–285. Stanford: CSLI Publications.

Lahm, David. 2016. Refining the semantics of lexical rules in HPSG. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway-King & Stefan Müller (eds.), *Proceedings of the joint 2016 conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar*, 360–379. Stanford, CA: CSLI Publications. http://cslipublications.stanford.edu/HPSG/2016/headlex2016-lahm.pdf.

Lakoff, George. 1970. *Irregularity in syntax*. New York: Holt, Rinehart, & Winston.

Lascarides, Alex & Ann Copestake. 1999. Default representation in constraint-based frameworks. *Computational Linguistics* 25(1). 55–105.

Malouf, Robert. 2000. *Mixed categories in the hierarchical lexicon*. Stanford, CA: CSLI Publications.

Marantz, Alec. 1997. No escape from syntax: don't try morphological analysis in the privacy of your own lexicon. In A. AkmajianDimitriadis, L. Siegel, C. Surek-Clark & A. Williams (eds.), *Upenn working papers in linguistics*, vol. 4, 201–225. Philadelphia: University of Pennsylvania Linguistics Department.

McCawley, James D. 1982. Parentheticals and discontinuous constituent structure. *Linguistic Inquiry* 13. 91–106.

Meurers, Detmar. 2001. On expressing lexical generalizations in HPSG. *Nordic Journal of Linguistics* 24. 161–217.

Miller, Philip & Ivan Sag. 1997. French clitic movement without clitics or movement. *Natural Language and Linguistic Theory* 15. 573–639.

Monachesi, Paola. 1993. Object clitics and clitic climbing in Italian HPSG grammar. In Steven Krauwer, Michael Moortgat & Louis des Tombe (eds.), *Sixth Conference of the European Chapter of the Association for Computational Linguistics. Proceedings of the conference*, 437–442. Uetrecht: Association for Computational Linguistics.

Müller, Stefan. 2006. Phrasal or lexical constructions? *Language* 82(4). 850–883.

Müller, Stefan. 2010. Persian complex predicates and the limits of inheritance-based analyses. *Journal of Linguistics* 46(3). 601–655. DOI:10.1017/S0022226709990284

Pollard, Carl J. & Ivan A. Sag. 1987. *Information-based syntax and semantics* (CSLI Lecture Notes 13). Stanford, CA: CSLI Publications.

Pollard, Carl & Ivan Sag. 1992. Anaphors in English and the scope of binding theory. *Linguistic Inquiry*. 2nd ser. 23. 261–303.

Pollard, Carl & Ivan Sag. 1994. *Head-driven phrase-structure grammar*. Chicago, IL: Chicago University Press.

Przepiórkowski, Adam. 2001. ARG-ST on phrases: evidence from Polish. In Dan Flickinger & Andreas Kathol (eds.), *Proceedings of the 7th international conference on head-driven phrase structure grammar*, 267–284. CSLI Publications.

Reape, Mike. 1994. Domain union and word order variation in german. In John Nerbonne, Klaus Netter & Carl Pollard (eds.), *German in head-driven phrase structure grammar*, 151–197. Stanford University: CSLI Publications.

Richter, Frank. 1999. RSRL for HPSG. In Valia Kordoni (ed.), *Tübingen studies in head-driven phrase structure grammar* (Arbeitsberichte des SFB 340 132), 74–115. Tübingen: Universität Tübingen.

Riehemann, Susanne Z. 1998. Type-based derivational morphology. *The Journal of Comparative Germanic Linguistics* 2. 49–77.

Roberts, Taylor. 2000. *Clitics and agreement*. Cambridge, MA: MIT dissertation.

Tegey, Habibullah. 1977. *The grammar of clitics: evidence from Pashto and other languages*. Urbana-Champaign: University of Illinois at Urbana-Champaign dissertation.

Wechsler, Stephen. 1999. HPSG, GB, and the Balinese bind. In Gert Webelhuth, Jean-Pierre Koenig & Andreas Kathol (eds.), *Lexical and Constructional aspects of linguistic explanation* (Studies in Constraint-Based Lexicalism 1), 179–195. Stanford, CA: CSLI Publications.

Draft of 29th May 2019, 13:58