Chapter 12

At First Blush on Tenterhooks About Selectional Restrictions Imposed by Nonheads[1]

JAN-PHILIPP SOEHN AND MANFRED SAILER

University of Jena and University of Tübingen
E-mail: jp@soehn.net and mf@sfs.uni-tuebingen.de

ABSTRACT. We will discuss German nouns which can only occur as the complements to a particular preposition. We will show that their distribution can best be accounted by generalizing HPSG's selectional mechanisms so that any nonhead can in principle select the head with which it combines. Then, we will explore the ramifications of this proposal: we will account for data that have been discribed as collocational and argue that our account makes empirically correct predictions on possible collocational patterns.

12.1 Introduction

Selection is one of the major mechanisms in formal grammar such as HPSG (Pollard and Sag, 1994), and is used for two purposes which are usually applied in the same contexts: First, syntactic and semantic arguments are specified and their relation to the selector is indicated. This is usually expressed in terms of thematic role assignment. Second, the properties of elements that may combine are restricted. When a verb selects its complements, it imposes both kinds of restrictions at the same time. On the other hand, if the selector is a nonhead, as in the case of modifiers, the question of thematic role assignment is less clear. When it comes to complementizers which select a clause, only the second function of selection appears to be relevant. In this paper, we will present evidence that all nonheads, including complements, can impose that kind of selectional restrictions on the heads they combine with.

[1]This paper originated within the Special Research Center 441 at the University of Tübingen. The full data and a more comprehensive discussion of the approach is given in Soehn (2003). We are grateful to Koenraad Kuiper, Stefan Müller, Frank Richter, Beata Trawiński and to the FGVienna reviewers for helpful comments and to Carmella Payne for help with the challenges of English.

Proceedings of Formal Grammar 2003
G. Jäger, P. Monachesi, G. Penn & S. Wintner (editors).

As a starting point, we will look at PPs that contain a *unique nominal complement* (UNC), i.e., a noun that does not occur outside combinations with a particular preposition. English examples are *on tenterhooks* (*nervous, worried*), or *in a trice* (*in an instant*). Note that the underlined nouns cannot be used in any other environment. For German a large number of PPs with UNCs have been attested. We will show that the data can be adequately described if we allow nonheads to select heads. Thus, we will generalize the HPSG treatment of adjuncts (which select by MOD) and of functional words like determiners and complementizers (which select by SPEC) to all nonheads. This analysis requires only minor changes to the HPSG architecture, whereas alternative attempts are either unable to account for the data or would require a far stronger mechanism. In this context, we will discuss a constructional analysis (Riehemann, 2001) and a collocational analysis (Sailer and Richter, 2002). While the need for constructions is indisputable for other phenomena (Fillmore et al., 1988), we will show that not only can our new architecture account for the data that Sailer and Richter have used to motivate their collocational approach, but this architecture also implies a number of restrictions on the kinds of collocational effects that may arise. These predictions are confirmed, with respect to a data base of approximately 300 highly collocationally restricted lexical items in German.

12.2 Data

The data we use as motivation for our analysis are *unique nominal complements* (UNC) of prepositions. They can occur only within a PP, as the example *auf Anhieb* (*at first go*) illustrates. To use Tseng (2000)'s terminology: *Anhieb* is an *internal trigger* of the preposition *auf*. In the following examples, the unique element is underlined and the preposition is printed in bold face.

(1) Das "PC-Kummerbuch" ist **auf** Anhieb auf großes Interesse gestoßen.
the PC-sorrow-book is at first-go on great interest hit

'The "PC Troubleshooting Companion" triggered interest right away.'

It is important to say that combining another preposition with *Anhieb* is ungrammatical, see (2) for comparison to a non-unique noun.

(2) **auf/*bei** Anhieb (at first go) vs.
auf den/beim ersten Versuch (at first attempt)

Similar English expressions are *by rote* (*mechanical*) or *in a trice* (*as quickly as possible*). For German, we have compiled about 300 highly collocationally restricted lexical elements from the phraseological literature (Dobrovol'skij, 1988; Dobrovol'skij and Piirainen, 1994; Fleischer, 1989, 1997). This list contains about 90 PPs with UNCs. These comprise:

(3) **im** Brustton (der Überzeugung) (with utter conviction), **zum** Steinerweichen (bitterly), **von/durch** Geisterhand (by magic), **um** Haaresbreite (by a hair's breadth), **im** Nachhinein (with hindsight), **aus/nach** Herzenslust (ad libitum), **ohne** Unterlass (without intermission)

These expressions have a highly restricted environment, namely one or only a few prepositions which are possible.

We are now faced with the question how these expressions can be analyzed. One possibility would be to put the entire P+UNC sequence in the lexicon as a multiword lexeme (like *of course*). That would not be adequate for three reasons: Firstly, the whole expression is syntactically formed like a regular PP and can receive a compositional interpretation (see 4). Following Fillmore et al. (1988) we want to assume phrasal lexical entries only for phrases with idiosyncratic properties. Secondly, the complements of the preposition can be modified (see 5). However, this is not a property of all UNCs and the modifiability has to be examined case by case. Thirdly, a UNC can be coordinated with another UNC or even with a free noun (see 6 and 7). From here it follows that the relevant PPs are not unstructured entities. Rather they are formed like regular PPs and even their semantics is decomposable. Consequently, regarding them as multiword lexemes would postulate too much irregularity of the language.

(4) Sie erhalten dieses Modell **auf** Anfrage.
you get this model on request

'This model is available upon request.'

(5) Er nahm an dem Gewinnspiel **aus** reiner Daffke teil.
He took at the lottery ex pure fun part

'He took part in the lottery just for fun.'

(6) Das Mädchen schmetterte ein Lied **aus** [Herzenslust und Leibeskräften].
the girl belted out a song from heart.delight and body.power

'The girl belted out a song to the top of her bent and with all her heart.'

(7) Damit brachte er sie **in** [Teufels Küche und jede Menge Schwierigkeiten].
with-it brougth he her in devil's kitchen and every amount trouble

'With that he got her in hot water and big trouble.'

The next step is now to make sure that the UNC combines with (and only with) the right preposition. Moreover, where a UNC cannot be modified, the possibility of modification has to be excluded. In the following section we will have a closer look at some approaches to analyze the data.

12.3 Alternative Analyses

12.3.1 Preposition selecting the NP

In all grammar frameworks, selection or subcategorization is a central issue, which is realized within HPSG by valence lists in the CATEGORY value of a sign. Subcategorization is rather general: a typical domain would be e. g. verb subcategorizing a subject and an object, both noun phrases being capable of receiving a particular theta-role. To account for the vast variety of grammatical expressions, values of subcategorized signs are specified only to a certain extent. Underspecification is the medium that allows selection of many different signs and keeps subcategorization as general as required.

If we wanted a preposition to select a UNC as its complement, we would encounter some problems. First of all, even if we introduced something like selection of particular lexemes, we could not prevent the UNC to occur elsewhere, because of underspecification. Additionally, if we defined a preposition that selected a particular UNC, we would have to postulate a certain amount of homophonous prepositions, where each one subcategorized for a different UNC. This postulation would be rather counterintuitive. Even worse, the appearance of a UNC in coordination as in (6) and (7) would be rendered ungrammatical.

Thus, the present mechanism of selection of HPSG is not sufficient, as selection of the argument by the preposition would not be able to correctly restrict the occurrence of a UNC to a particular preposition and it would wronlgy exclude the coordination data above.

12.3.2 Constructional Approach

Riehemann (2001) offers a concrete analysis of idioms within HPSG using ideas of Construction Grammar (Fillmore et al., 1988; Sag, 1997). She describes idioms as units of several words, i. e. constructions. One of her central ideas is to relate each word in an idiom with its literal counterpart. For our data this approach is not adequate, because the lack of a literal lexical counterpart is a defining feature of UNCs like *Anhieb* or *Daffke*. If we introduced such lexical entries, we could not prevent the words from occurring freely in other contexts. Alternatively, we can adopt a constructional analysis in which the UNCs are *idiomatic words* not related to any literal counterpart.

Riehemann (2001) introduces a complex machinery to restrict the distribution of idiomatic words to only the right contexts, defining new sorts and attributes which globally keep track of all parts of idiomatic expressions that occur in an utterance. The components of this machinery are not independently motivated and, as Riehemann acknowledges, the occurrence of pronominalized parts of idioms cannot be handled.

We conclude that the UNC data do not fit Riehemann's basic assumptions on idioms and, thus, do not justify the application of her complicated apparatus.

12.3.3 Collocation

A collocation module has been proposed by Sailer (2000) and Sailer and Richter (2002), introducing the COLL feature. The value of that feature is a list. This list is empty for non-lexical signs, and contains a single sign for each lexical sign. The authors define a principle of grammar making the following constraint: the overall utterance must be identical with the sign in the COLL lists of all the lexical signs of this utterance. Thus, the entire utterance is available at the COLL value of every lexical sign so that occurrence restrictions can be specified. This means that any property of the utterance can be mentioned in the lexical entry of such a sign. For example, the word *Anhieb* specifies in its COLL list a sign dominating a PP with the head *auf*, *Anhieb* being the complement of *auf*.

This module is explicitly designed for elements with limited distribution. However, it is extremely powerful, because the element of the COLL list makes the structure of the whole expression available. Therefore the mechanism does not impose any principled constraints on the kinds of occurrence restrictions that may be found. For our data, i.e. very local dependencies, the mechanism seems far too powerful and requires quite a few changes within the grammar. In the next section we will argue for a different approach, using an existing mechanism in HPSG and making only minor changes.

12.4 External Selection

As indicated already in the introduction, there is a mechanism of external selection provided within the HPSG framework. External selection means that nonheads select their heads. This has been assumed for adjuncts (which select by MOD) and for functional words like determiners and complementizers (selecting by SPEC). We now generalize this mechanism and apply it to all nonheads by specifying a feature XSEL (for *external selection*) for the sort *head*. By XSEL the features MOD and SPEC are subsumed and become dispensable. Pollard and Sag (1994, p. 55) themselves state that these two features are analogous in their use. The function of MOD and SPEC is now incorporated into XSEL and everything that they have accomplished can now be achieved by the new attribute. The value of XSEL is *synsem*. This results in the analysis in (8) whose details will be examined below.

(8)

	PP	
HEAD		COMP
P		NP
$\boxed{1}\left[\text{LOC CAT}\begin{bmatrix}\text{VAL COMPS} & \langle\boxed{2}\rangle \\ \text{ARG-ST} & \langle\boxed{2}\rangle\end{bmatrix}\right]$		$\boxed{2}\left[\text{LOC CAT HEAD XSEL } \boxed{1}\right]$
auf		*Anhieb*

The changes to be made affect only two points of our grammar. First, as already noted, we define an attribute XSEL for the sort *head* and remove MOD and

SPEC from the signature. As we have introduced XSEL for the sort *head*, the value percolates via the HEAD FEATURE PRINCIPLE (HFP). Second, we redefine the SPEC-PRINCIPLE as PRINCIPLE OF EXTERNAL SELECTION, stated in (9).

(9) PRINCIPLE OF EXTERNAL SELECTION (PXS):
The SYNSEM value of a word is identical to the XSEL value of each element on its ARG-ST list.
In a head-adjunct phrase and in a head-marker phrase the SYNSEM value of the HEAD-DTR is identical to the XSEL value of the nonhead daughter.

Ideally, we could have formulated the PXS in a uniform way on all headed structures with a selectional relation between the daughters. Then, the PXS would enforce identity between the nonhead's XSEL and the head's SYNSEM value. However, in argument-raising constellations, a *synsem* object may occur on several COMPS lists, while still only occuring on one ARG-ST list.[2] Therefore, we refer to the ARG-ST list for complements, instead of giving a structural account of the XSEL requirements of complements. Head-filler structures are exempt from the PXS, as there is no selectional relation between head and filler.

Using the attribute XSEL, a UNC can select a particular preposition by the LEXEME feature. The introduction of a mechanism to select particular lexemes will be deferred to section 12.5. What we get by these adaptations has already been illustrated in (8), where *Anhieb* (*first go*) selects the preposition *auf* (*at*) through the XSEL feature. Additionally, we give the lexical entry of *Anhieb* in (10). Note that there is a restriction on the content: the CONTENT of the complement of the preposition must be identical to the CONTENT of *Anhieb* itself. This excludes the occurrence of modifiers with *Anhieb*.

(10)
$$\left[\begin{array}{l} \textit{word} \\ \text{PHON} \left\langle \textit{Anhieb} \right\rangle \\ \text{SS LOC} \left[\begin{array}{l} \text{CAT} \left[\begin{array}{l} \text{HEAD} \left[\begin{array}{l} \textit{noun} \\ \text{CASE } \textit{acc} \\ \text{XSEL} \left[\text{LOC} \left[\begin{array}{l} \text{CAT} \left[\begin{array}{l} \text{HEAD } \textit{preposition} \\ \text{VAL COMPS} \left\langle \text{NP}\left[\text{LOC CONT } \boxed{1}\right] \right\rangle \end{array}\right] \\ \text{CONT INDEX LEXEME } \textit{auf} \end{array}\right]\right] \end{array}\right] \\ \text{VAL COMPS } \textit{e-list} \end{array}\right] \\ \text{CONT } \boxed{1} \end{array}\right] \end{array}\right]$$

In (11) we have sketched the XSEL value of *Daffke* for comparison. *Daffke* simply selects the preposition *aus*. As there are no further restrictions on the complement NP of this preposition, it follows from the normal selection mechanism that modification is possible, like in *aus reiner Daffke* (*just for pure fun*).

[2]Following Meurers (2000), we assume that argument raising takes place only via the COMPS list.

(11) $$\begin{bmatrix} \textsc{xsel} & \begin{bmatrix} \textsc{loc} & \begin{bmatrix} \textsc{cat} & \begin{bmatrix} \textsc{head}\ \textit{preposition} \\ \textsc{val comps}\ \langle \text{NP} \rangle \end{bmatrix} \\ \textsc{cont index lexeme}\ \textit{aus} & \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

In principle, the content of complement NPs are available for marking semantic restrictions. This allows further specification of the character of modification in the lexical entry. Ultimately, it is possible for the UNC to select its preposition.

12.5 Selection of Lexemes

In the previous section we have examined how PPs with unique nominal complements can be accounted for. Some of these expressions go together with one particular verb, as in *zu <u>Potte</u> kommen* (*get to the point*) or *in <u>Mitleidenschaft</u> ziehen* (*to affect*). With our PXS we can account for these data, too. We simply specify a *synsem* object in the XSEL value, in this case of *Potte*, that triggers the preposition *zu*. This preposition has another XSEL value, requiring the verb *kommen* (see 12).[3] To select the verb *kommen*, we lack one important tool, namely the possibility to select a particular lexeme.

(12) $$\begin{bmatrix} \textsc{xsel} & \begin{bmatrix} \textsc{loc cat head} & \begin{bmatrix} \textsc{pform} & \textit{zu} \\ \textsc{xsel} & \begin{bmatrix} \textsc{loc} & \begin{bmatrix} \textsc{cat head}\ \textit{verb} \\ \textsc{cont index lexeme}\ \textit{komm-} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Krenn and Erbach (1994) have proposed a way of lexeme selection, introducing an attribute LEXEME below the CONTENT INDEX feature. This has three advantages: the LEXEME value is below SYNSEM and thus available for subcategorization; it percolates along the head line (i.e., a projection has the same LEXEME value as its lexical head); and a pronoun (always sharing the INDEX with its antecedent according to the HPSG Binding Theory) gets also the same LEXEME value as its antecedent.[4] Nevertheless, their approach has one major disadvantage: Only nouns have an INDEX value of their own, which leaves open the question of what to do with verbs (lacking an INDEX) or adjectives (sharing the INDEX with the modified noun).

The solution we propose is a new definition of the INDEX feature for all parts of speech. We separate LEXEME from the traditional INDEX features PERSON, NUMBER and GENDER, pooled together below a new feature PHI (see 13). As subsorts of *lexeme*, we introduce a new atomic sort for each lexeme.

[3]The PXS makes correct predictions for cases of argument raising, as assumed for German verbal complexes (Meurers, 2000). In (i) the PP *zu Potte* is realized as a complement of the modal verb *sollte*. Still, it only occurs on the ARG-ST list of the verb *kommen*, whose LEXEME value is as required in the XSEL value in the lexical entry of *Potte* in (12).

(i)	dass	Peter	zu	<u>Potte</u>	kommen	sollte	
	that	Peter	to	pot	come	should	‘that Peter should get to the point’

[4]See (16) for illustration.

(13)
$$\left[\text{INDEX}\left[\begin{array}{ll}\text{LEXEME} & \textit{lexeme}\\ \text{PHI} & \left[\begin{array}{ll}\text{PERSON} & \textit{person}\\ \text{NUMBER} & \textit{number}\\ \text{GENDER} & \textit{gender}\end{array}\right]\end{array}\right]\right]$$

These changes have no consequences for nouns. For verbs, we leave the PHI values underspecified, nevertheless LEXEME receives a non-trivial value. In the lexical entries of adjectives, we define the PHI values to be identical with those of the modified noun, but give each adjective its individual LEXEME value. We also adapt the SEMANTICS-PRINCIPLE in a way in which INDEX values are inherited along the syntactic head (whereas all other values like NUCLEUS percolate along the semantic head). This facilitates a selection of individual lexemes and extends our analysis of UNC to combinations with fixed verbs as in (12).

12.6 Some Empirical Predictions

In the preceeding sections, we presented a selectional analysis of UNCs. A combination of three factors made this possible: (i) the generalized use of selection by nonheads (by XSEL), (ii) the fact that the CONTENT value of a sign is part of its SYNSEM value (see 10), and (iii) the possibility of lexeme selection (by LEXEME). In this section, we will continue to explore the potential of this approach. Specifically, we will show that the interaction of these three factors accounts for the phenomena that Sailer and Richter provide as evidence for their unconstrained collocational mechanism: cranberry words (Richter and Sailer, 2002) and polarity items (Sailer and Richter, 2002). Furthermore, a set of non-trivial constraints on potential collocational restrictions follow from our approach which have been empirically confirmed by the above-mentioned corpus of approximately 300 distributionally restricted lexical elements.[5]

12.6.1 Restrictions on Lexeme Selection

Richter and Sailer (2002) discuss so termed *cranberry words* (CW), i.e., lexical elements that can only occur together with a particular other lexeme. Clearly, this subsumes the UNCs considered so far. They distinguish three possible cases, depending on the selectional relation between the CW and its collocate: either the CW selects its collocate (14a), or the collocate selects the CW (b), or there is no direct selectional relation (c).

(14) a. jemandem **Angst/** einen **Schrecken/** *Entsetzen einjagen
someone fear/ a fright/ horror in.chase 'to frighten sb.'

b. Tacheles **reden/** *sprechen
goal talk/ speak 'talk straight'

[5] See Webelhuth and Ackermann (1994) and Everaert and Kuiper (1996) for other applications of this method.

c. **essen/** *futtern wie ein <u>Scheunendrescher</u>
eat/ chow down like a barn.thresher ‘eat like a horse’

Richter and Sailer (2002) claim that the (a)-case is problematic, because HPSG does not provide the means for selecting individual lexemes. With the new treatment of lexeme selection illustrated in (13), this argument loses ground. The case in (14b) is resolved by the XSEL mechanism, because we have allowed all nonheads to impose selectional restrictions on the heads with which they combine. Together with the possibility of lexeme selection, the data conform to our proposal as well.

Finally in (14c), there is no direct selection between the CW *Scheunendrescher* and its collocate, the lexeme *essen* (*eat*). Nonetheless, the NP headed by the CW is selected by *wie* (*like*), whose projection acts as an adjunct to the verb *essen*. Therefore, the preposition *wie* occurs on the XSEL value of the CW and the verb *essen* on the XSEL value of the preposition. Thus, we are in the same situation as with *zu <u>Potte</u> kommen* in (12).

Our approach can handle the discussed data correctly. Moreover, we predict that there are certain restrictions on lexeme selection. In particular, a head cannot select a certain adjunct lexeme. This follows from the fact that we have generalized the potential of nonheads to select heads by XSEL, but we have not enlarged the selectional potential of heads. Our collection of 300 German expressions with CWs does not contain a single counter-example to this restriction, and there are abundant examples of lexeme selection of and by complements, and of adjuncts that select particular heads.

We also incorporate a locality restriction. A selector has access to the LEXEME value of a directly selected element, and, as shown for *zu Potte kommen*, to the LEXEME value of signs that are externally selected by these. We predict that there are no CWs which ask for a specific **complement** of a verb (instead of a specific verb). The hypothetical constellation is sketched in (15), where the bold face NP is a complement to the verb V, and the collocate of the CW. Here, the information about the lexeme of the complement NP is not available at the VP level, i.e. part of the SYNSEM value of the VP. The prediction is fully confirmed in our corpus.

(15) [$_{\text{VP}}$ V **NP**] [$_{\text{PP}}$ P <u>CW</u>]

On the other hand, the apparent non-locality of lexeme selection is not a problem. The data show that lexeme information must be present at the phrasal projection, (16a), and must be shared between a pronoun and its antecedent, (16b,c).

(16) a. Das hat mir einen riesigen **Schrecken** <u>eingejagt</u>
this has me a huge fright in.chased ‘This frightened me a lot.’

b. Der **Schrecken**$_i$, **den**$_i$ mir das <u>eingejagt</u> hat, war riesig.
the fright which me this in.chased has was huge

‘The fright which he caused me was huge.’

c. Das war ein riesiger **Schrecken**$_i$ für mich. Und ausgerechnet du hast
this was a huge fright for me and just you have
ihn$_i$ mir eingejagt.
it me in.chased

'This was a huge shock for me and you were the one to cause it to me.'

These non-local effects, which can cross root sentences (see 16c), result from the inclusion of the lexeme information as part of the INDEX value. In contrast, Richter and Sailer (2002) would be required to assume that collocations are possible across root sentences in general.

We have shown two restrictions on potential lexeme-specific co-occurrence requirements. They are a direct consequence of the XSEL mechanism and appear to be empirically robust. In the analysis of Richter and Sailer (2002), it is not possible to express these constraints, let alone have them be the independent outcome of the way their collocational mechanism is built.

12.6.2 CONTENT Restrictions

In Sailer and Richter (2002) the collocational requirements of the German verb *fackeln* (*dither*) are discussed. The authors base their study on a corpus investigation, according to which the verb occurs together with *nicht lange* (*not for long*) in 86% of the cases (i.e. 241 out of 280). See (17):

(17) Der neue Geschäftsführer fackelte nicht lange.
the new manager dithered not long

'The new manager didn't dither very long.'

They demonstrate that the adverb *lange* must be considered a modifier, i.e., as not selected by the verb *fackeln*. Thus, this seems to be a counter-example to the restrictions on XSEL, because heads cannot impose lexeme restrictions on their modifiers. However, Sailer and Richter (2002) give a more refined characterization of the occurrence restrictions of *fackeln*: In 95% of the corpus data, and in all of the introspectively grammatical data, a negation and a durational modifier, that have scope over the verb *fackeln*, are present, but not necessarily *nicht* and *lange*, see (18). This variation is expected in the XSEL approach: again, there is no lexeme restriction imposed by the head verb on its modifier.

(18) Keiner hat einen Moment gefackelt, bevor er ins Wasser gesprungen
nobody has a moment dithered before he into the water jumped
ist.
has 'Nobody dithered a moment before jumping into the water.'

Before we can turn to the analysis of *fackeln*, we should address a remaining issue in connection with the XSEL value. The PXS in (9) and the HFP determine the XSEL values of all signs that are part of a phrase that will eventually occur as a nonhead daughter in a sentence. Nothing is said, however, about the XSEL value of signs which occur along the head projection line of a root sign. To avoid spurious ambiguity, we can simply assume that the SYNSEM value of the root sign is the XSEL value in these cases. The notion of a "root sign" is needed independently in grammar (Richter, 1997).[6] Following Riehemann (2001, p. 189f.), we assume a subsort of *sign* for this purpose, which we will call *root-sign*. The XSEL value of root signs is then determined by the constraint in (19).

(19) $$\textit{root-sign} \rightarrow \begin{bmatrix}\text{SYNSEM } \boxed{1} \begin{bmatrix}\text{LOC CAT HEAD } \begin{bmatrix}\text{XSEL } \boxed{1}\end{bmatrix}\end{bmatrix}\end{bmatrix}$$

As an effect of the constraint in (19) the CONTENT value of an utterance is in principle accessible to the XSEL specification of every word in the utterance, by recursively accessing the XSEL value of the *synsem* object in the XSEL value. We can use this property to account for the *fackeln* data of Sailer and Richter, because there, the distributional restrictions of the verb *facklen* (*dither*) are characterized in terms of the particular CONTENT values of the utterances in which it may occur.

Ignoring details of the semantic representation, Sailer and Richter give a description of a CONTENT value which contains a negation and a durational modifier, both having scope over the semantic contribution of the verb *fackeln*. Let δ be this description. Then, the occurrence restrictions on *fackeln* can be captured in the following lexical entry.

(20) $$\begin{bmatrix} \text{PHON } \langle \textit{fackeln} \rangle \\ \text{SYNS LOC } \begin{bmatrix} \text{CAT } \begin{bmatrix} \text{HEAD } \begin{bmatrix} \textit{verb} \\ \text{XSEL } \begin{bmatrix}(\text{LOC CAT HEAD XSEL})^{*} \text{ LOC CONT } \delta\end{bmatrix} \end{bmatrix} \\ \text{ARG-ST } \langle \text{NP} \rangle \end{bmatrix}\end{bmatrix}\end{bmatrix}$$

where δ describes a *content* object iff it contains a negation, a durative operator and the lexical content of *fackeln*, where the negation and the durative operator both have scope over the latter.

In (20) we use a Kleene-star operator with paths in the AVM. The notation is to be interpreted as describing the existence of a path of the given form. Even though this is not standard in HPSG, a formalization of HPSG, as given in Richter (2000), will provide a way to formalize the idea expressed in this rather intuitive notation.

It can be seen that the XSEL requirements of *fackeln* are met in sentence (17). Here, the XSEL value of the sentence is identical to its own SYNSEM value. Since *fackeln* is the syntactic head of the sentence, this very same *synsem* object occurs in its XSEL value. Thus, the XSEL LOC CONT value of *fackeln* is identical to the CONT of the root sentence. This content was described by δ in (17). The absence of a negation or a durational modifier would lead to a violation of the lexical specification of the XSEL value of *fackeln*.

If we consider occurrences of *fackeln* in embedded sentences, the non-trivial effects of the Kleene-star operator in (20) can be observed.

[6] Root properties include the presence of illocutionary force and the absence of unfilled gaps.

(21) Niemand behauptet, dass Peter lange fackelt.
nobody claims that Peter long dithers

'Nobody claims that Peter dithers for long.'

In (21) it is only the CONT value of the root clause that contains the negation, the durational modifier and the verb *fackeln*. This CONT value appears as the XSEL LOC CONT value of the highest verb, *behauptet* (*claims*). *Fackeln* is the lexical head of the complement clause to the verb *behauptet*. Thus, the XSEL value of *fackeln* is identical to the SYNSEM value of *behauptet*. Consequently, the XSEL LOC CAT HEAD XSEL LOC CONT value of *fackeln* is identical to the CONT value of the root clause. As this path is part of the specification in the lexical entry of *fackeln*, the occurrence restriction of the verb is captured correctly.

This example shows that the non-local effect of the occurrence restriction on *fackeln* is fully accounted for in our approach, while we correctly exclude the possibility of constraints specific to the particular lexemes that occur, to the phonology of the required context elements or to details of the syntactic structure. On the other hand, the CONTENT of the utterance is available for the imposing of constraints.

12.7 Conclusion

In this paper, we have offered a leaner generalization of the selectional mechanism of HPSG. We have shown that this allows us to handle the phenomenon of PPs with unique nominal complements. The same mechanism extends naturally to other phenomena of distributional restrictions. Our proposal crucially relies on the overall architectural design of HPSG. This design restricts selection to *synsem* objects, where lexeme information and the content are part of *synsem*. The discussion in section 12.6 revealed that the interaction of these components leads to interesting and empirically valid predictions concerning what kinds of collocational restrictions are possible. Nonetheless it deserves further discussion whether it is adequate to include lexeme information and the entire content within *synsem* (see Sailer (to appear) for some thoughts regarding this question).

Bibliography

Dobrovol'skij, D. (1988). *Phraseologie als Objekt der Universallinguistik*. Verlag Enzyklopädie, Leipzig.

Dobrovol'skij, D. and E. Piirainen (1994). Sprachliche Unikalia im Deutschen: Zum Phänomen phraseologisch gebundener Formative. *Folia Linguistica*, **27**(3–4):449–473.

Everaert, M. and K. Kuiper (1996). Theory and data in idiom research. In L. McNair et al., ed., *CLS32: the Parasession on Theory and Data in Linguistics*, pp. 43–57. Chicago Linguistic Society.

Fillmore, C., P. Kay, and M. O'Connor (1988). Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language*, **64**:501–538.

Fleischer, W. (1989). Deutsche Phraseologismen mit unikaler Komponente — Struktur und Funktion. In G. Gréciano, ed., *Europhras 88*, pp. 117–126.

Fleischer, W. (1997). *Phraseologie der deutschen Gegenwartssprache*. Niemeyer, Tübingen, 2nd, revised edition edition.

Krenn, B. and G. Erbach (1994). Idioms and support verb constructions. In J. Nerbonne et al., ed., *German in Head-Driven Phrase Structure Grammar*, pp. 365–396. CSLI Publications. Lecture Notes 46.

Meurers, W. D. (2000). Lexical generalizations in the syntax of German non-finite constructions. Arbeitsberichte des SFB 340, Nr. 145, Universität Tübingen.

Pollard, C. and I. A. Sag (1994). *Head-Driven Phrase Structure Grammar*. University of Chicago Press, Chicago and London.

Richter, F. (1997). Die Satzstruktur des Deutschen und die Behandlung langer Abhängigkeiten in einer Linearisierungsgrammatik. In E. Hinrichs et al., ed., *Ein HPSG-Fragment des Deutschen. Teil 1: Theorie*, Arbeitspapiere des SFB 340, Nr. 95, pp. 13–187. Universität Tübingen.

Richter, F. (2000). A mathematical formalism for linguistic theories with an application in Head-Driven Phrase Structure Grammar. Dissertation, Universität Tübingen.

Richter, F. and M. Sailer (2002). Cranberry words in formal grammar. In C. Beyssade et al., ed., *Empirical Issues in Formal Syntax and Semantics*, volume 3. Presses Universitaires de Paris-Sorbonne, Paris.

Riehemann, S. Z. (2001). *A Constructional Approach to Idioms and Word Formation*. Ph.D. thesis, Stanford University.

Sag, I. A. (1997). English relative clause constructions. *Journal of Linguistics*, **33**:431–483.

Sailer, M. (2000). Combinatorial semantics and idiomatic expressions in Head-Driven Phrase Structure Grammar. Dissertation, Universität Tübingen.

Sailer, M. (to appear). Distributionsidiosynkrasien: Korpuslinguistische Erfassung und grammatiktheoretische Deutung. In *Tagungsband der 39. Jahrestagung des IDS*. Mannheim.

Sailer, M. and F. Richter (2002). Not for love or money: Collocations! In G. Jäger et al., ed., *Proceedings of Formal Grammar 2002*, pp. 149–160.

Soehn, J.-P. (2003). Von Geisterhand zu Potte gekommen. Eine Analyse von PPs mit unikalen Komponenten. Magisterarbeit, Seminar für Sprachwissenschaft, Universität Tübingen.

Tseng, J. L. (2000). *The Representation and Selection of Prepositions*. Ph.d. thesis, University of Edinburgh.

Webelhuth, G. and F. Ackermann (1994). German idioms: An empirical approach. *Studies in the Linguistic Sciences*, **24**:455–471.