

A Constraint-based Representation Scheme of Collocational Structures

Dirk Heylen† André Schenk† Marc Verhagen‡
(heylen@let.ruu.nl) (schenk@let.ruu.nl) (verhm@sx.ac.uk)

† OTS – Trans 10 – 3512 JK Utrecht (NL)

‡ CLMT Essex University – Wivenhoe Park – C04 3SQ Colchester (UK)

For collocational cooccurrence restrictions the direction of selection is the reverse of that found in other cooccurrence restriction structures (subcategorization, selection-restrictions, and theta-role assignment, for instance). In this paper we provide a Mel'čukian analysis of collocational structures in unification based theories of language. We show how the use of partial linguistic descriptions and the use of a hierarchically structured lexicon (common in such constraint-based analyses of natural language) provide the means to account elegantly for collocational restrictions.

1 Defining Collocations

Before we come to the analysis of collocational constructions, a brief word about the definition of the constructions themselves is appropriate. Indeed, the word ‘collocation’ seems to mean many things to many people. This is partly due to the fact that often different criteria for defining collocations are put into one definition. For instance, [Smadja, 1993] lists amongst others the following properties of collocations:

- They are recurrent word combinations.
- They are cohesive lexical clusters¹.
- They can be idiomatic to varying degrees.
- They cannot be translated literally.

The first property concerns an observable directly related to language *use*. The second relates to the result of word association tests as used by *psycholinguists* and *lexicographers*. The third and fourth may relate to some specific *linguistic* properties. This example is prototypical for many definitions found in the literature. It illustrates the many approaches to collocations that exist.

We will distinguish between three modes of defining collocations:

- the statistical mode (performance-based — concerned with patterns in the products of language use),
- the lexicographic (incompetence-based² — concerned with ‘what people don’t know’ about language) and
- the linguistic (competence-based — concerned with ‘knowledge of language’).

It is not surprising therefore that the resulting set of expressions which people consider collocations is often very heterogeneous. As a first attempt to make things clearer it seems appropriate to associate each mode with a specific definition and to separate the different sets of expressions accordingly.

¹‘The presence of one or several words of the collocations often [...] suggests the rest’.

²No offence intended.

Statistical Statistical definitions of collocations are presented in terms of the probability of the co-occurrence of word combinations. Within computational linguistics and corpus linguistics there is a long standing tradition in fine-tuning the algorithms for extracting ‘collocations’ from texts: [Choueka *et al.*, 1983], [Church and Hanks, 1989], [Smadja, 1993]. The main concern of these studies is to find the appropriate formula that picks out constructions which are intuitively felt to be collocations concentrating on the assumption that collocations are combinations of words that co-occur more often than expected by chance. As is clearly illustrated in [Smadja, 1993] the intuitive notion of collocation of which the statistical data is taken to be a reflection, is seldom made very precise. It is often a conglomerate of properties found in other literature, mostly lexicographic.

Lexicographic However, dictionary makers do not all use the same defining characteristics. Given the nature and the purpose of dictionaries, a lexicographic perspective on collocations would take them to be a combination of words which is ‘arbitrary’ or *unpredictable* from the ordinary rules of grammar and which therefore has to be listed in a dictionary. One source of unpredictability is the figurative meaning of one part of the combination (e.g. *heavy smokers* need not have problems with their weight). Another source of unpredictability is the preference of native speakers to always choose *x* in combination with *y* instead of the synonymous *x'*. In defining ‘collocation’ lexicographers differ in the sources of unpredictability they include³. It is quite difficult to find out what exact criteria are used by lexicographers to call something a ‘collocation’. Introductions of dictionaries explicitly admit that implicit lexicographer’s intuitions may play an important role.

Linguistic The notion of frequency of occurrence is relevant in a paradigm in which language-use is studied, but not in a linguistic paradigm that studies the notion of competence. Some of the properties used by lexicographers to decide on the collocational status of a combination also constitute defining characteristics from a linguistic point of view. But whereas the lexicographer is primarily concerned with the things speakers/hearers may not know about, linguists are concerned with what speakers/hearers know. The linguistic perspective defines collocations not in terms of ‘occurrence’ nor in terms of ‘preferences’ on how native speakers phrase things but in terms of how the grammar and the lexicon (syntax, semantics, morphology, phonology) of collocations differs from other constructions.

A tentative definition After all these (sketchy) introductory notes on the definition of collocations we are in a better situation to make explicit some background assumptions of our approach. In this paper we present one possible *linguistic* perspective on collocations, using amongst others concepts derived from a particular (rather) lexicographic analysis of collocations. In particular, we show how the Mel’čukian analysis of collocations is implemented in a constraint-based formalism making use of partial linguistic descriptions and a hierarchically structured lexicon. We will not go into a detailed linguistic analysis as such or the data supporting this analysis. Instead, we will focus on the mechanisms used in the formalization of these analyses.

Examples of the collocational structures we have in mind are *heavy smoker*, (adjective-noun) and *give a demonstration* (verb-noun, or support verb construction). In the Ex-

³An inspiring analysis of different factors that could lead to qualifying a combination as a ‘collocation’ is given in [Allerton, 1984].

planatory Combinatory Dictionary, [Mel'čuk *et al.*, 1984] uses lexical functions — which we interpret as general semantic operations — in the description of collocations.

The expressions that are analyzed by means of these kinds of lexical functions, exhibit a typical kind of cooccurrence restriction. In the case of adjective noun combinations such as *une ferme intention*, *une résistance acharnée*, *un argument de poids*, *un bruit infernal*, *un désir ardent*, *une envie folle*, *une règle stricte*, *une vérité incontestable*⁴, the adjectives all express (more or less) the same meaning (= lexical function), although they are not interchangeable. Each adjective occurs only in restricted contexts. It is claimed that the nouns select these specific adjectives. Furthermore, the selection is purely lexical. No semantic factors are involved.

To provide for this direction of selection, the dictionary entry for *smoker* (the base of the collocation) lists all the lexical functions appropriate for it and its 'lexicalised' values (the collocates): **Magn**(*smoker*) = *heavy*. For *give a demonstration* the analysis would be **Oper**(*demonstration*) = *give*.

In many collocations, the collocate has a (sometimes figurative) meaning which it only has in this specific combination (or in a restricted set of contexts). We identify this meaning with a lexical function⁵. However, other properties of the collocate are shared by its free occurring counterpart. The essential characteristics of this approach that will be accounted for in the analysis presented in this paper are then the following.

- The collocational cooccurrence restriction.
- The dependence of the meaning of the collocate on the combination.
- The relation between the 'collocate' occurrence of a word and its free occurrence.

One of the design criteria which has guided the representation is to be 'minimally redundant'. This has led to the following solution.

- In the lexicon, it is specified on the base which collocates it *selects* for a particular lexical function.
- The collocate is given by a partial lexical entry, but as a *generalization* of its 'free-occurring' counterpart. For instance its meaning is left unspecified.

2 A Constraint based implementation

Important in this analysis is the fact that the collocational selection may run counter to the syntactic selection. In the LF analysis, we can speak of 'bases' and 'collocates': the arguments and the values of the lexical functions respectively. Now in terms of selection, the basis (argument) is the selector and the collocate (value) is the selected. However, in syntactic terms, the collocate may be the head of the construction and the base a complement⁶.

In this section we point out how this direction of selection is implemented in an HPSG-like theory. We focus on the lexical mechanisms used in our implementation of this Mel'čuk inspired architecture. We first present possible mechanisms for some non-collocational restrictions briefly.

⁴Roughly: a firm intention, a dedicated resistance, a weighty argument, an infernal noise, an ardent desire, a foolish want, a strict rule, an incontestable truth.

⁵Notice that we are simplifying matters. In any case Mel'čuk would not want to claim that the meaning is reduced to this lexical function. In the project we take the possibility of this reduction as one of the central issues to be investigated. In this paper we will not be concerned with the problems involved.

⁶This poses interesting problems for generation. For more detail we refer the reader to [Heid and Raab, 1989].

2.1 Other kinds of cooccurrence restrictions

Linguistic theories use a number of mechanisms to express restrictions on the combination of parts in some syntagm. We discuss three types: subcategorization requirements, selection restrictions and our cross-referencing analysis of collocations. In HPSG, subcategorization requirements specify the number and type of arguments the head of a phrase combines with to form a ‘saturated constituent’. They typically refer to the syntactic properties of the arguments. The restrictions that predicates put on the kind of object an argument may denote, are expressed by selection restrictions, at least in some versions of some linguistic theories.

Subcategorisation Requirements Grammatical collocations (*look at*) are not collocations in the ‘proper sense’. From our perspective, there is no relevant connection or some unifying property between grammatical and lexical collocations. The mechanisms responsible for describing the two types of expressions are therefore also taken to be quite different.

Most (if not all) grammatical collocations can be accounted for in HPSG by subcategorization restrictions. This is the actual proposal of [Pollard and Sag, 1987]. An example:

```
$look at:
[ phon      look
  syn|subcat < PP[pform at], NP > ]
```

The specification of the *pform* of the argument together with the Subcategorization principle account for the correct analysis of grammatical collocations.

Selection Restrictions In the OALD, the definition of ‘rancid’ mentions the restriction “(of fat)”. From a lexicographic or statistical point of view the combination *rancid butter* is often called a collocation. From our perspective we would call this a case of selection restrictions and deal with it differently⁷. In this case the selection resides in the predicate. It seems part of the definition of ‘rancid’ that it can only be used for oily things.

In HPSG terms we can use the *context* attribute introduced in [Pollard and Sag, 1992]⁸.

```
$rancid:
[ phon      rancid
  syn      [ head [ mod N' : [ var [1]
                                rest {[3]} ] ] ]
  subcat   <>
  sem|cont [ var [1]
            rest { [ reln rancid ]
                  [ inst [1] ] }, [3] ]
  cntxt|back { [ reln oily ]
               [ inst [1] ] }
```

⁷Some people argue that many collocations are extreme cases of selection restrictions. Conceptually speaking, the difference would be that with collocations the selection is one of ‘lexical items’ whereas with selection restrictions, the selection is of ‘semantic classes’.

⁸Notice that we only want to point out the general mechanisms, not a specific analysis.

2.2 Collocations

Cross-reference For cases such as the adjective-noun construction “heavy smoker”, and the support verb construction “give a demonstration”, we propose an analysis inspired on Mel’čuk and his co-authors. For these cases, we introduce a new feature in the lexical entries for the base of the collocation. Let’s call this ‘col’. The type of value this attribute can take is (simplifying somewhat) a set of lexical entries. To illustrate we describe the grammatical apparatus needed for *heavy smoker*. The Lexical Function MAGN means “very”, “intense” or “to a high degree”⁹.

\$smoker:

$$\left[\begin{array}{l} \textit{phon} \textit{ smoker} \\ \textit{sem} \left[\begin{array}{l} \textit{cont} \left[\begin{array}{l} \textit{var} [1] \\ \textit{rest} \left\{ \begin{array}{l} \textit{reln} \textit{ smoker}' \\ \textit{inst} [1] \end{array} \right\} \end{array} \right] \end{array} \right] \\ \textit{col} \{ \$\textit{heavy} \left[\begin{array}{l} \textit{cont} \left[\begin{array}{l} \textit{var} [1] \\ \textit{rest} \left\{ \begin{array}{l} \textit{reln} \textit{ MAGN}' \\ \textit{inst} [1] \end{array} \right\}, \textit{NounRest} \end{array} \right] \end{array} \right] \} \end{array} \right] \end{array} \right]$$

In contrast with the selectional restriction, where, in the case of adjective noun phrases, the restriction is imposed by the adjective, the information regarding the correct collocation is put in the entry for the noun. The key (variable) \$heavy refers to some feature structure in the lexicon. It is not, however, the entry of the word *heavy* which appears in *heavy weight*. For one, they don’t share the same meaning attribute. They do share a lot of other things, though. To express the relatedness we use the hierarchical structure of the HPSG lexicon. Simplifying for expository reasons, we assume that the lexicon consists of a hierarchy of typed feature structures. At the bottom we find the lexical entries for individual words. These contain the information that is specific to the individual word. The information which is shared is inherited from the entries of the supertypes, which are also indicated (at least the immediate supertypes) at the entry.

Collocates If we follow the idea that the meaning of the collocate in some collocation is determined by the combination or the base, rather than that it is independently given, we might want to have only a partial specification of the information concerning the collocate in the lexicon. The solution we propose is to introduce ‘partial entries’ which take their proper place in the hierarchical structure. The entry (\$heavy) referenced in the base (\$smoker) is not the entry for *heavy* as it appears in the combination with *weight*, but some feature structure which subsumes the latter, as it does not contain a specification for *content*¹⁰. They do share quite a lot of other features though¹¹. The same procedure can be applied to some other phenomena. For instance, different levels can be created to deal with morphologically deficient collocates as well. For subcategorization requirements of support verbs see below.

Combination Ordinary rules of combination generate the modifier/ head structures of which adjective-noun collocations are instances. As far as the semantics of modifier constructions is concerned we can follow the suggestions as made in [Pollard and Sag,

⁹Notice that other types of intensification are possible with “smoker”. In general, to accommodate different of intensification, Lexical Functions can be superscripted as in $\text{MAGN}^{\text{temp}}$, where the superscript *temp* refers to the temporal degree of intensification.

¹⁰These entries should only be used in those contexts in which they can be made ‘complete’, i.e. for instance licensed by a base selecting for them.

¹¹This kind of sharing of information between entries is reminiscent of the notion of ‘generic word’ as discussed in [Pollard and Sag, 1987, p.214].

1992] and others. However, in the case of collocates, the *cont* value is not filled in by the lexical entry but by the *col* field in the noun.

The representation for 'heavy smoker' could look like this.

$$\left[\begin{array}{l} \text{sem|cont} \left[\begin{array}{l} \text{var} \quad [1] \\ \text{rest} \quad \{[4],[3]\} \end{array} \right] \\ \\ \text{dtrs} \left[\begin{array}{l} \text{hdtr} \left[\begin{array}{l} \text{phon} \quad \text{smoker} \\ \text{sem} \quad \left[\text{cont} \quad [5] \right] \\ \text{col} \quad \{[\$heavy]\} \end{array} \right] \\ \text{adjdtrs} \quad \{[\$heavy]\} \end{array} \right] \end{array} \right]$$

Where,

$$[5] \left[\begin{array}{l} \text{var} \quad [1] \\ \text{rest} \quad \{[3] \left[\begin{array}{l} \text{reln} \quad \text{smoker}' \\ \text{inst} \quad [1] \end{array} \right] \} \end{array} \right]$$

and $\$heavy$ is partly specified by the noun (from the collocations field and also from the partial entry for $\$heavy$ itself (comparable to the entry given above for *rancid*, but without the *cont* attribute).

Information specified in *col* field of *smoker*

$$\left[\text{cont} \left[\begin{array}{l} \text{var} \quad [1] \\ \text{rest} \quad \{[4] \left[\begin{array}{l} \text{reln} \quad \text{MAGN}' \\ \text{inst} \quad [1] \end{array} \right], \text{NounRest} \} \right] \right]$$

$\$heavy$, the lexical entry for the collocate 'heavy':

$$\left[\begin{array}{l} \text{phon} \quad \text{heavy} \\ \\ \text{syn} \left[\begin{array}{l} \text{head} \left[\text{mod} \quad N' : \left[\begin{array}{l} \text{index} \quad [1] \\ \text{rest} \quad \{[3]\} \end{array} \right] \right] \\ \text{subcat} \quad <> \end{array} \right] \end{array} \right]$$

This latter structure is a generalization of the ordinary (full) entry for *heavy* as it appears in *heavy weight*.

Not only can we apply the ordinary syntactic rules, but also the same semantics principle apply to both collocational and non-collocational modification. The only difference between the two derives from the sources of the relevant information. Within a constraint-based theory, where information is added monotonically this is not a problem. What we still need is a principle which allows us to unify the correct structures, in this case an element of the collocations-list and the adjunct-daughter. For this purpose we could introduce some sort of collocations-principle. Abstracting from the details of the previous example this could look as follows.

$$\left[\begin{array}{l} \text{head_dtr} \left[\text{col} \quad \{...[1]...\} \right] \\ \text{adj_dtrs} \quad \{...[1]...\} \end{array} \right]$$

We would like to note that in many cases a number of analyses may overlap: a certain construction may exhibit strict collocational restrictions together with selectional restrictions. This means that the description of some construction does not merely involve a choice as to the kind of construction we are dealing with, but a choice as to what parts of the construction are dealt with in what ways.

Support Verbs The representation of support verb constructions is similar to that of adjective-noun collocations in the sense that the support verb takes over its semantic

properties from the predicative noun. In addition to this in most cases the support verb takes some of its syntactic properties from the noun, notably its subcategorisation requirements. So a support verb like *give* is represented only with features expressing that it is a verb and some form features. It can combine with all predicative nouns that are specified to take *give* as a support verb.

The specific subcategorisation requirements depend on the noun. For example, at the lexical entry for *demonstration* it is specified that the support verb *give* subcategorises for the subject daughter of *demonstration*, a complement daughter headed by *demonstration* and (optionally) the second complement daughter of *demonstration*. This is necessary to account for the observation that in *John gives a demonstration of the system to the students*, *John* is interpreted as subject of *demonstration* and realized as subject of *give* and that *to the students* is interpreted as argument of *demonstration* but realized either as a complement of *demonstration* or as a complement of *give*.

Support verb constructions consist of a verb and a predicative noun, for example *give demonstration*. The verb has little or no meaning and takes its arguments from the noun. The predicative noun does have arguments; in this example *demonstration* is a three-place predicate.

Globally speaking, the representation of support verb constructions (SVCs) is similar to that of adjective-noun collocations in the sense that the support verb (SV) is specified in the set of collocations (COL). However, there are also differences between the representations of SVCs and the Adjective-Noun collocations, notably that the SV takes over the subject and the subcategorisation list of the predicative noun. To illustrate the lexical entry of *demonstration* is given.

The elements in the subcat list of *demonstration* are optional, cf. *John gave a demonstration (of the system) (to the students)*. The subcat list of the collocate contains two elements, i.e. an NP, which is the NP headed by *demonstration*, and it takes the second PP ([2]) from the subcat list of the noun. This is necessary to account for eg. *The demonstration was given to the students*. It does not take over the first PP ([1]), since *The demonstration was given of the system* is unwellformed. Also the subject [3] of *give* is taken from the noun.

The semantics of the collocate is specified in *rest*, i.e. OPER1 applied to [5], which is the *sem* of the noun.

$$\left[\begin{array}{l} \textit{phon} \quad \textit{demonstration} \\ \textit{syn|loc} \quad \left[\textit{subcat} \quad < (\textit{PP}[4]), (\textit{PP}[2]), \textit{NP}[3] > \right] \\ \textit{sem} \quad [5] \quad \left[\begin{array}{l} \textit{reln} \quad \textit{demonstration} \\ \textit{arg1} \quad [3] \\ \textit{arg2} \quad [4] \\ \textit{arg3} \quad [2] \end{array} \right] \\ \textit{col} \quad \{ \$\textit{gives} \left[\begin{array}{l} \textit{sem|cont|ind}[6] \quad \left[\begin{array}{l} \textit{var} \quad [1] \\ \textit{rest} \quad \{ [8] \left[\begin{array}{l} \textit{reln} \quad \textit{OPER}_1 \\ \textit{inst} \quad [5] \end{array} \right] \} \right] \\ \textit{syn|loc} \quad \left[\textit{subcat} \quad < \textit{NP}, [2], [3] > \right] \end{array} \right] \} \end{array} \right. \end{array} \right]$$

Analogously to the representation of adjectival collocates the SV is partly specified by the noun, but also from the underspecified entry of the SV itself, cf. below. The entry for the SV *gives* is similar to non-collocational entries of *gives*, but it does not have a *cont* or a *subcat* attribute.

$$\left[\begin{array}{l} \textit{phon} \quad \textit{gives} \\ \textit{syn} \quad \left[\textit{loc|head} \quad \left[\begin{array}{l} \textit{maj} \quad \textit{verb} \\ \textit{vform} \quad \textit{fin} \end{array} \right] \right] \end{array} \right]$$

The combination of the lexical entry of the collocate *give* and the structure represented at *demonstration* would yield, omitting some detail, the structure below.

$$[7] \left[\begin{array}{l} \textit{phon} \\ \textit{syn} \\ \textit{sem|cont|ind}[6] \end{array} \begin{array}{l} \textit{gives} \\ \left[\begin{array}{l} \textit{loc} \left[\begin{array}{l} \textit{head} \left[\begin{array}{l} \textit{maj} \quad \textit{verb} \\ \textit{vform} \quad \textit{fin} \end{array} \right] \\ \textit{subcat} \quad \langle \textit{NP}, [2], [3] \rangle \end{array} \right] \\ \textit{var} \quad [1] \\ \textit{rest} \quad \left\{ \begin{array}{l} \textit{reln} \quad \textit{OPER}_1 \\ \textit{inst} \quad [5] \end{array} \right\} \end{array} \right] \end{array} \right]$$

The representation of *John gives a demonstration to the students* could look like the following structure. For the sake of readability we have omitted details.

$$\left[\begin{array}{l} \textit{sem} \\ \textit{dtrs} \\ C \end{array} \left[\begin{array}{l} \textit{cont|ind} \left[\begin{array}{l} \textit{var} \quad [1] \\ \textit{rest} \quad [8] \end{array} \right] \\ H|dtrs \left[\begin{array}{l} H \quad [7] \\ C \quad \langle [4], \\ \left[\begin{array}{l} \textit{phon} \quad \textit{a demonstration of the system} \\ \textit{syn} \quad \dots \\ \textit{dtrs|H...} \left[\begin{array}{l} \textit{phon} \quad \textit{a demonstration} \\ \dots \textit{subcat} \quad \langle [2], [4], [3] \rangle \end{array} \right] \\ \textit{dtrs|C...} \quad [2] [\textit{phon of the system}] \end{array} \right] \end{array} \right] \end{array} \right] \end{array} \right]$$

Implementation The representation scheme presented above has been implemented in the ALEP formalism developed by the European Commission.

2.3 Other proposals

Our representation structures above are a combination of Mel'čuk's analysis of collocations on the one hand, and a HPSG-like theoretical framework on the other.

Mel'čuk's proposal to use lexical functions and have the collocates listed at the entries for the bases, has also been taken up by [Heid and Raab, 1989, p. 132]: "We propose that every lexeme entry in the lexicalization dictionary contains slots for lexical functions, whose fillers are possible collocates [...]."

They further discuss the possibility to generalize over some regularities in collocation formation for members of semantically homogeneous lexeme classes and some issues concerning generation and lexicalization. They do not go into much detail on how this lexical information fits into the general linguistic theory or how the information is distributed over the lexical entries of bases and collocates and rules and principles. They do not consider using a hierarchical lexicon or partial information structures in this paper either.

Independently from our research, a proposal for treating support verbs in HPSG was developed in [Erbach, 1992] and [Krenn and Erbach, To appear]. However, their solution contrasts with ours in as follows. They do not take up Mel'čuk's architecture in which the base selects the collocate. In their proposal the support verbs subcategorize for specific predicative nouns. They do not consider other types of collocations either.

3 Further Research

What we have done so far is to suggest different ways to specify the linguistic analysis of three types of lexical cooccurrence phenomena: grammatical collocations, selectional restrictions and collocations (proper), within an HPSG like framework. As far as the analysis of collocations is concerned, some of the defining characteristics (collocational restriction, meaning dependence of the collocate on the base etc.) were coded making explicit use of the power of the unification formalism: the use of partial entries and the spreading out of information on the combination over parts etc.

In the previous paragraphs we have not dealt with variation, preferences and blocking phenomena. For the time being we do not want to specify the details of the mechanisms that need to be introduced to tackle these properties.

The objective of the introduction of the above analyses is to allow for the correct analysis and generation of collocational structures. It does not inhibit an awkward interpretation of collocations, nor the generation of non-collocational structures which normally seem to be 'blocked' because of the existence of a preferred collocational structure. We believe this would not be a proper way to go about. In some cases, such as *heavy smoker*, the literal (heavy weight) reading is equally possible. It might be less plausible, but this is not a matter for the grammar to explain. In other cases, such as *bitter winter*, the literal reading is even less plausible. In this case one can argue that the reason is that selection restrictions are violated. This is a mechanism which works orthogonally on the collocation restriction.

Acknowledgements The research reported in this paper is part of the project "Collocations and the Lexicalisation of Semantic Operations" (ET-10/75). Financial contributions are by the Commission of the European Community, Association Suissetra (Geneva) and Oxford University Press.

References

- [Allerton, 1984] D.J. Allerton. Three (or four) levels of word cooccurrence restriction. *Lingua*, 1984.
- [Choueka *et al.*, 1983] K. Choueka, S.T. Klein, and E. Neuwitz. Automatic retrieval of frequent idiomatic and collocational expressions in a large corpus. *ALLC Journal*, 1983.
- [Church and Hanks, 1989] Kenneth Ward Church and Patrick Hanks. Word association norms, mutual information and lexicography. In *27th Annual Meeting of the Association for Computational Linguistics*, Vancouver, 1989.
- [Erbach, 1992] Gregor Erbach. Head-driven lexical representation of idioms in HPSG. In *Proceedings of Idioms*, pages 11-23, Tilburg, 1992.
- [Heid and Raab, 1989] U. Heid and S. Raab. Collocations in multilingual generation. In *Fourth Conference of the European Chapter of the ACL*, pages 130-136, Manchester, 1989.
- [Krenn and Erbach, To appear] Brigitte Krenn and Gregor Erbach. Idioms and support verb constructions. In John Nerbonne, Klaus Netter, and Carl Pollard, editors, *German Grammar in HPSG*. CSLI Lecture Notes, To appear.

- [Mel'čuk *et al.*, 1984] Igor Mel'čuk, Nadia Arbatchewsky-Jumarie, Léo Elnitsky, Lidija Iordanskaja, and Adèle Lessard. *Dictionnaire explicatif et combinatoire du français contemporain*. Les Presses de l'Université de Montréal, Montreal, 1984.
- [Pollard and Sag, 1987] Carl Pollard and Ivan Sag. *Information Based Syntax and Semantics*. Center for the Study of Language and Information, Stanford, 1987.
- [Pollard and Sag, 1992] Carl Pollard and Ivan Sag. An information based approach to syntax and semantics: vol 2: Agreement, binding and control. Draft, June 1992.
- [Smadja, 1993] Frank Smadja. Retrieving collocations from text: Xtract. *Computational Linguistics*, 19(1):143–177, 1993.