Collocation Extraction: Needs, Feeds and Results of an Extraction System for German

Julia Ritz
Institut für Maschinelle Sprachverarbeitung (IMS)
Universität Stuttgart
Azenbergstr. 12
70174 Stuttgart
Germany
Julia.Ritz@ims.uni-stuttgart.de

Abstract

This paper provides a specification of requirements for collocation extraction systems, taking as an example the extraction of noun + verb collocations from German texts. A hybrid approach to the extraction of habitual collocations and idioms is presented, aiming at a detailed description of collocations and their morphosyntax for natural language generation systems as well as to support learner lexicography.

1 Introduction

Since Firth first described collocations as habitual word combinations in the 1950ies (cf. Firth, 1968), a number of papers focusing on collocation extraction have been published (see the overviews in (Evert, 2004; Bartsch, 2004)). Most studies concentrate on the extraction from English. However, the procedures proposed in these studies cannot necessarily be applied to other languages as English stands out, e.g. with respect to configurationality. They rely on the fact that the syntax of English (and of all configurational languages) provides positional clues to the grammatical function of noun phrases, and they exploit this concept by means of window-based, adjacency-based or pattern-based extraction, combined with association measures to identify co-occurrences that are more frequent than statistically expectable. What these procedures do not cover is semantic-oriented definitions like (a) and (b).

a. A collocation is a combination of a free ('autosematic') element (the *base*) and a lexically determined ('synsemantic') element (the *collocate*, which may lose (some of) its meaning in a collocation) (adapted from (Hausmann, 1979; Hausmann, 1989; Hausmann, 2003)).

b. A collocation is a word combination whose semantic and/or syntactic properties cannot be fully predicted from those of its components, and which therefore has to be listed in a lexicon (Evert, 2004).

We argue that linguistic knowledge could not only improve results (Krenn, 2000b; Smadja, 1993) but is essential when extracting collocations from certain languages: this knowledge provides other applications (or a lexicon user, respectively) with a fine-grained description of how the extracted collocations are to be used in context.

Additional requirements resulting from the needs of dictionary users are described in (Hausmann, 2003; Heid and Gouws, 2005) and are of interest not only in lexicography but can also be transferred to the field of natural language generation. These requirements influence the development of collocation extraction systems, which motivates this paper.

The structure of the paper is as follows: In chapter 2, the requirements, depending on factors like the targeted language, are presented. We then discuss and suggest methods to meet the given needs. A documentation of ongoing work on the extraction of noun + verb collocations from German texts is given in chapter 3. Chapter 4 gives a conclusion and an outlook on work still to be done.

2 Collocation Extraction Tools: Requirements

The development of a collocation extraction tool depends on the following conditions:

1. properties of the targeted language

2. the targeted application
3. the kinds of collocations to be extracted
4. the degree of detail

Whereas issues 1 to 3 deal with the collocation itself, issue 4 is focused at the collocation in context, i.e. its behaviour (from a syntagmatic analysis point of view) or, respectively, its use (from a generation perspective).

2.1 Language factors

One of the most important factors is, of course, the targeted language and its main characteristics with respect to word formation and word order. Depending on word and constituent order, the pros and cons of positional vs. relational extraction patterns need to be considered. Positional patterns (based on adjacency or a 'window') are adequate for configurational languages, but in languages with rather free word order, words belonging to a phrase or collocation do not necessarily occur within a predefined span[1].

Extracting word combinations using relational patterns (represented by part of speech (PoS) tags or dependency rules) offers a higher level of abstraction and improves the results (cf. (Krenn, 2000b; Smadja, 1993)). However, this requires part of speech tagging and possibly partial parsing. A system extracting word combinations by applying relational patterns, obviously profits from language specific knowledge about phrase and sentence structure and word formation. One example is the order of adjective + noun pairs: in English and German, the adjective occurs left of the noun, whereas in French, the adjective can occur left or right of the noun. Another example is compounding, handled differently in different languages: noun + noun in English, typically separated by a white space (e.g. *death penalty*) vs. noun + prepositional phrase in French (e.g. *peine de mort*) vs. compound noun in German (e.g. *Todesstrafe*). Consequently, language specific word formation rules need to be considered when designing extraction patterns. For languages with a rich inflectional morphology where the individual word forms are rather rare, frequency counts and results of statistical analyses are little reliable. To allow a grouping of words sharing the same lemma, lemmatisation is crucial.

2.2 Application factors

Other important factors are the targeted application (i.e. analysis vs. generation) and, to some extent resulting from it, factors (3.) and (4.), above. Depending on the purpose of the tool (or lexicon, respectively), the collocation definition chosen as an outline may vary, e.g. including transparent and regular collocations (cf. (Tutin, 2004)) for generation purposes, but excluding them for analysis purposes. In addition, a more detailed description of the use of collocations in context (e.g. information about preferences with respect to the determiner, etc.) is needed for generation purposes than for text analysis.

2.3 Factors of collocation definition

Collocations can be distinguished on two levels: the formal level and the content level. On the formal level, a collocation can be classified according to the structural relation between its elements. Typical patterns are shown in table 1[2] (taken from (Heid and Gouws, 2005)).

On the content level, there are regular, transparent, and opaque collocations (according to (Tutin, 2004)) and, taking definition (b) into account, idioms as well. However, as a classification at the content level needs detailed semantic description, we see no means of accomplishing this goal other than manually at the moment.

2.4 Contextual factors

(Hausmann, 2003; Heid and Gouws, 2005; Evert et al., 2004) argue that collocations have strong preferences with respect to their morphosyntax (see examples (1) and (2)) and may be combined (see example (3)). The collocation in example (1) ('to charge somebody') is restricted with respect to the determiner (*null determiner*) of the base, whereas the same base shows a strong preference for a (*definite or indefinite*) determiner when used

[1] In German, e.g., in usual verb second constructions with a full verb in the left sentence bracket (topological field theory see (Wöllstein-Leisten et al., 1997)), particles of particle verbs appear in the right sentence bracket. The middle field (containing arguments and possibly adjuncts of the verb) is of undetermined length.

[2] Abbreviations in table 1:
advl - adverbial
prd - predicative
subj - subject
obj - object
pobj - prepositional object
dat - dative case
gen - genitive case
quant - quantifying

No.	Type	Example
1	N + Adj	*tiefer Schlaf*
2	Adj + Adv	*tief rot*
3	V + Adv	*tief schlafen*
4	V + NP_{advl}	*Bauklötze staunen*
5	V + N_{subj}	*Frage + sich stellen*
6	V + N_{dat}	*Anforderungen + genügen*
7	V + N_{obj}	*Frage + aufwerfen*
8	V + PP_{pobj}	*zu + Darstellung + gelangen*
9	V + Adj_{prd}	*verrückt spielen*
10	N + N_{gen}	*Einreichung des Antrags*
11	N_{quant} + N	*ein Schwarm Heringe*
the category containing the base is underlined.		

Table 1: Collocational patterns

with a different collocate (example (2), 'to drop a lawsuit'). Example (3) shows two collocations sharing the base can form a collocational sequence (example taken from (Heid and Gouws, 2005)).

(1) ø Anklage erheben

(2) die/eine Anklage fallenlassen

(3) *Kritik* üben + scharfe *Kritik*
scharfe *Kritik* üben

For both natural language generation systems and lexicography, such information is highly relevant. Therefore, the extraction of contextual information (called 'context parameters' in the following) should be integrated into the collocation extraction process.

3 Extracting noun + verb collocations from German

The standard architecture for collocation extraction systems contains three stages (cf. (Krenn, 2000)): a more or less detailed linguistic analysis of the corpus text (preprocessing), an extraction step and a statistic filtering of the extracted word combinations. We follow this architecture (see figure 1). However, our hypothesis differs from other approaches. Collocations are often restricted with respect to their morphosyntax. We test to what extent they can be identified via these restrictions.

3.1 Approach

In an experiment, we extracted relational word combinations (verb + subject/object pairs) from German newspaper texts.

The syntactic patterns for the extraction of these combinations concentrate on verb-final constructions as in example (4) and verb second constructions with a modal verb in the left sentence bracket according to the topological field theory (see (Wöllstein-Leisten et al., 1997)) as in example (5). The reason is that, in these constructions, the particle forms one word with the verb (see example (6)), as opposed to usual verb second constructions (see example (7)). Thus, we need not recombine verb + particle groups that appear separatedly.

(4) ... wenn Wien einen *Antrag* auf Vollmitgliedschaft *stellt*.
('if Vienna an *application* for full membership *puts*')
(if Vienna *applies* for full membership)

(5) ... kann Wien einen *Antrag* auf Vollmitgliedschaft *stellen*.
('might Vienna an *application* for full membership *put*.')
(Vienna might *apply* for full membership.)

(6) ..., daß er ein Schild *auf*stellt.
('that he a sign *up*puts')
(that he puts up a sign)

(7) Er stellt ein Schild *auf*.
('He puts a sign *up*.')
(He puts up a sign.)

Preprocessing

As data, we used a collection of 300 million words from German newspaper texts dating from 1987 to 1993. The corpus is tokenized and PoS-tagged by the Treetagger (Schmid, 1994), then chunk annotated by YAC (Kermes, 2003). The chunker YAC determines phrase boundaries and heads, and disambiguates agreement information as far as possible. It is based on the corpus query language cqp (Christ et al., 1999)[3], which can in turn be used to query the chunk annotations.

Data Extraction

The syntactic patterns used to extract verb + subject/object combinations are based on PoS tags and chunk information. These patterns are represented using cqp macros (see figure 2). The cqp syntax largely overlaps with regular expressions.

[3] http://www.ims.uni-stuttgart.de/projekte/CorpusWorkbench/

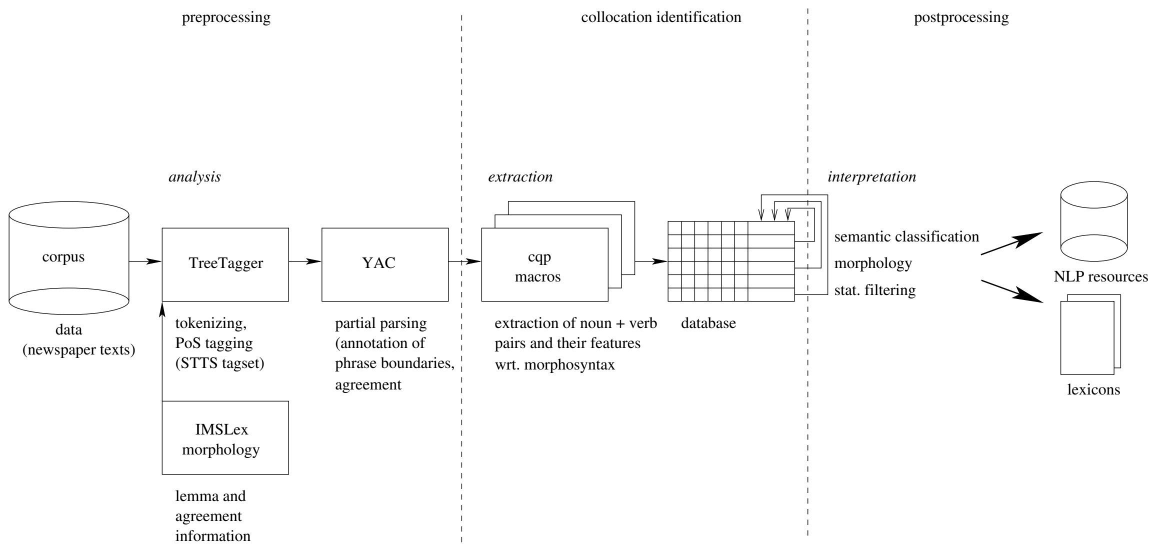

Figure 1: Tool architecture

```
( 1) MACRO n_vfin(0)
( 2) (
( 3) [pos = "(KOUS|VMFIN)"]
( 4) []*
( 5) <np>
( 6)   [!pp
( 7)    & _.np_f not contains "ne"
( 8)    & _.np_f not contains "pron"
( 9)    & _.np_f not contains "meas"
(10)    & _.np_h != "@card@"]+
(11)</np>
(12)[pos != "($.|KOUS|VMFIN)"]*
(13)[pos = "V.*"]+
(14)[pos = "($.|KON)"]
(15))
(16);
```

Figure 2: sample macro

Line (1) of figure 2 contains the name of the macro and the number of its parameters. In line (3), a word PoS tagged `KOUS` (subordinating conjunction) or `VMFIN` (finite modal verb) is requested, followed by an arbitrary number ('`*`') of words without any restrictions (line (4)). Line (5) indicates the start of a nominal phrase (`np`), line (11) its end. The elements within this np (one or more words, as indicated by '`+`') must not be part of a prepositional phrase (`pp`) to avoid the extraction of pp + verb (line (6), see example (8)). In addition, the np must be neither a named entity (`ne`, see line (7)) nor a pronoun (`pron`, line (8)) nor an np of measure (`meas`, line (9), see example (9)), nor must its head be a cardinal number (`card`, line (10), see example (10)). An arbitrary number of words may follow the np (punctuation marks (PoS tagged `$.`), subordinating conjunctions and finite modal verbs excluded). At least one verb is required (line (13), all PoS tags for verbs start with a capital '`V`'[4]). Line (14) indicates the end of the subclause or sentence.

(8) ... kann [zur *Verfügung*]$_{pp}$ gestellt werden.

(9) ... weil davon jährlich [3,5 *Tonnen*]$_{np\ meas}$ eingeführt werden.

(10) ... obwohl er [*1989*]$_{np\ card}$ noch dort arbeitete.

By applying the macro to the corpus, all sequences of words matching the pattern are extracted.

From these sequences, the following information is made explicit (cf. (Heid and Ritz, 2005)):

- lemma of the noun (potential base)
- lemma of the verb (potential collocate)
- number of the noun (singular, plural)
- case of the noun

[4]The search condition is underspecified with respect to the finiteness and the role of the verb (auxiliary, modal or full verb). Thus, line (13) matches verbal complexes. It also covers cases where full verbs are accidentally PoS tagged modal or auxiliary verbs.

- determination of the noun (definite, indefinite, null, demonstrative, quantifier)
- modification of the noun (adjective, cardinal number, genitive np, compound noun etc.)
- negation (yes/no)
- auxiliaries and modal verbs
- original phrase from the corpus

For each instance found, the lemmas of noun and verb along with all the context parameters mentioned above are stored as feature value pairs in a relational data base. The database can be queried via SQL. See figure 3 for a sample query asking for distinct lemma pairs, ordered by frequency (in descending order), and figures 5 and 4 for more specific queries and some of their results.

```
SELECT COUNT(*) AS f,
n_lemma, v_lemma
FROM comfea1
GROUP BY n_lemma, v_lemma
ORDER BY f DESC;
```

Figure 3: sample query

Filtering

The instances extracted in the previous step are grouped according to noun and verb lemmas, i.e. instances of the same lemma pair form one group. Within these groups, a relative frequency distribution is computed for each of the features. For queriability reasons, the results of this postprocessing are also stored in the database, as shown in figure 1. A word combination is chosen as a collocation candidate if a preference (specified by a threshold of e.g. 60% of the occurrences) for a certain feature value (singular / plural, presence / absence of a determiner, definite / indefinite / demonstrative / possessive / quantifying determiner, presence of modifying elements) is discovered.

3.2 Results

From 300 million words, we extracted more than 1.3 million noun + verb combinations, the instances of 726,488 different lemma pairs. 10,934 of these lemma pairs appeared with a minimum

```
SELECT COUNT(*) AS f,
n_lemma, v_lemma
FROM comfea1
WHERE neg = '+'
GROUP BY n_lemma, v_lemma
ORDER BY f DESC;
```

f	**n_lemma**	**v_lemma**
1152	Rede	sein
748	Angabe	machen
322	Einigung	erzielen
228	Chance	haben
217	Forderung	erfüllen
188	Problem	lösen
151	Rolle	spielen
131	Auskunft	geben
127	Stellungnahme	abgeben
120	Alternative	geben
110	Interesse	haben
110	Angabe	bestätigen
102	Geld	haben

Figure 4: sample query: word combinations from negated phrases

frequency of 10. Sample results are shown in figure 6[5].

We evaluated collocation candidates with a frequency of at least 100. Within the 323 most frequent collocation candidates, we found 213 collocations (including 11 idioms). This corresponds to a precision of 66% (see table 2[6]). As a comparison, a window-based study was carried out on the same (PoS-tagged) data. In this study, the window was defined in a way that up to two tokens (excluding sentence boundaries and finite full verbs) were allowed to appear between a noun (PoS tagged `NN`) and a finite full verb (PoS tagged `VVFIN`). Log-likelihood[7] was used as an association measure. The precision of this approach is 41%[8].

[5]Abbreviations in figure 6:
c - rated as a collocation in evaluation
i - rated as an idiom in evaluation.
For chosing collocation candidates, a threshold of 60% is used. However, additional preferences are displayed for values greater than 50%.

[6]Abbreviations in table 2:
log-l - window-based approach using log-likelihood
feat - pattern-based approach using morphosyntactic features

[7]`www.collocations.de`

[8]Note that partial matches, such as *Verfügung* + *stellen*

```
SELECT COUNT(*) AS f,
n_lemma, v_lemma
FROM comfea1
WHERE cas = 'Akk'
GROUP BY n_lemma, v_lemma
ORDER BY f DESC;
```

f	n_lemma	v_lemma
507	Beitrag	leisten
237	Antrag	stellen
173	Eindruck	erwecken
173	Weg	finden
167	Umsatz	steigern
145	Hut	nehmen
140	Bericht	vorlegen
135	Betrieb	aufnehmen
121	Sprung	schaffen
120	Ausschlag	geben
116	Mut	haben
111	Sitz	haben
106	Weg	ebnen
105	Zuschlag	erhalten
104	Platz	finden
100	Anspruch	haben
94	Tod	feststellen
93	Zusammenhang	geben
90	Vertrag	unterzeichnen
90	Riegel	vorschieben

Figure 5: sample query: verb + accusative object

However, the evaluation criteria from definitions (a) and (b) remain vague or even contradictory for some of the results. First, there is the problem of semantic equivalence: does the combination express more than its elements (consider example (11))? Secondly, definitions (a) and (b) may judge the same example differently: *Anteil nehmen* (example (12)) is usually agreed upon to be a support verb construction, but the distinction of the noun *Anteil* as the base (making the main contribution to the meaning) is questionable. On the other hand, its unpredictable syntactic properties (e.g. null determiner) and semantics (partial loss of meaning of the collocate *nehmen*) make it clear that this combination has to be listed in a lexicon.

(without the corresponding preposition), have been treated as correct matches in 72 cases.

	log-l	feat
collocation candidates	700	323
collocations (manually verified)	290	213
precision	41%	66%

Table 2: evaluation results

For evaluation purposes, combinations judged collocations by either (or both) of the definitions were marked as correct matches. In cases like example (11), combinations were marked as correct matches if no alternative collocate existed for describing the denoted situation or event.

(11) *Chance* + *haben* ('*to have a/the chance*')

(12) *Anteil* + *nehmen* ('*to commiserate*')

4 Conclusion and Outlook

We presented a system for collocation extraction that takes into account the behaviour or use of collocations in context. Profiting from linguistic information (PoS tagging, chunking), the tool reaches a precision of 66% on the top 323 candidates by frequency. On the same data, a window-based approach relying only on PoS information reached a precision of 41%.

As the extracted word combinations as well as their context parameters (including the original evidence from the corpus) are stored in a database, the tool also supports explorative research in lexicography.

However, there are some enhancements worth doing: Especially when dealing with low frequencies, relative frequencies lack reliability. Therefore, we suggest computing a confidence interval as proposed in (Evert, 2004b; Heid and Ritz, 2005; Ritz, 2005).

As indicated in figure 1, several postprocessing steps can be added to the system, e.g. enabling a sorting of collocation candidates with compound nouns by the morphological heads of their base.

In order to get more data, the extraction from verb first and verb second constructions is also possible. To complete the tool, extraction patterns for collocations of different syntactic relations (cf. table 1) could be designed.

	n_lemma	v_lemma	total	restrictions
-	Polizei	mitteilen	5689	sg(100%), det(99.49%), def(99.49%), modif(53.84%)
c	Rede	sein	2144	sg(99.81%), det(99.16%), quant(53.59%)
-	Sprecher	mitteilen	1401	sg(94.15%), det(99.21%), indef(93.36%), modif(68.31%)
c	Fall	sein	1233	sg(99.03%), det(97.32%), def(96.03%)
-	Kantonspolizei	mitteilen	1094	sg(99.82%), det(59.60%), def(59.60%), modif(100%)
-	Behörde	mitteilen	952	pl(82.14%), det(99.89%), def(99.37%), modif(63.13%)
c	Stellung	nehmen	831	sg(99.88%), no_det(88.21%)
c	Angabe	machen	802	pl(98.50%), det(93.64%), quant(93.27%)
-	Polizeisprecher	mitteilen	737	sg(90.91%), det(92.67%), indef(90.77%), modif(100%)
i	Rolle	spielen	724	sg(98.62%), det(97.93%), indef(65.61%)
c	Problem	lösen	690	det(89.13%), def(64.06%), modif(51.88%)
-	Zeitung	berichten	670	sg(94.78%), det(94.78%), def(91.79%), modif(83.28%)
-	Nachrichtenagentur	melden	667	sg(98.95%), det(98.20%), def(97.90%), modif(100%)
c	Rechnung	tragen	661	sg(100%), no_det(98.94%)
-	Unternehmen	mitteilen	655	sg(98.16%), det(98.32%), def(98.32%), modif(66.26%)
c	Chance	haben	614	sg(85.67%), det(86.31%)
c	Beitrag	leisten	575	sg(96.70%), det(95.65%), indef(52.70%), modif(62.54%)
-	Polizei	berichten	564	sg(100%), det(98.40%), def(98.40%)
c	Einigung	erzielen	551	sg(99.82%), det(92.20%), quant(58.26%)
-	Sprecher	sagen	508	sg(76.38%), det(97.83%), indef(76.18%), modif(56.69%)
c	Arbeit	aufnehmen	492	sg(98.37%), det (98.37%), poss(76.62%)
c	Ziel	erreichen	476	sg(78.78%), det(96.22%)
-	Nachrichtenagentur	berichten	454	sg(100%), det(98.90%), def(98.90%), modif(100%)
c	Druck	ausüben	451	sg(100%), no_det(88.70%), modif(74.72%)
c	Erfolg	haben	438	sg(99.32%), no_det(78.54%)
-	Frau	sein	425	sg(54.59%), det(50.12%), modif(61.41%)
-	Land	verlassen	421	sg(99.76%), det(98.57%), def(86.70%)
c	Frage	stellen	419	sg(71.60%), det(79.00%), def(68.26%)

Figure 6: sample results

References

Sabine Bartsch. 2004. *Structural and functional properties of collocations in English*. A corpus study of lexical and pragmatic constraints on lexical cooccurrence Narr, Tübingen.

Oliver Christ, Bruno M. Schulze, Anja Hofmann and Esther König. 1999. *The IMS Corpus Workbench: Corpus Query Processor (CQP). User's manual*. Institut für maschinelle Sprachverarbeitung, Universität Stuttgart.

Jonathan Crowther, Sheila Dignen and Diana Lea. 2002. *Oxford Collocations Dictionary for students of English*. Oxford University Press.

Stefan Evert. 2004. *The Statistics of Word Coocurences - Word Pairs and Collocations*. PhD thesis. Institut für Maschinelle Sprachverarbeitung (IMS), Universität Stuttgart.

Stefan Evert, Ulrich Heid and Kristina Spranger. 2004. *Identifying morphosyntactic preferences in collocations*. In M. T. Lino, M. F. Xavier, F. Ferreira, R. Costa and R. Silva (eds.): *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, p. 907 - 910. Lisbon, Portugal.

Stefan Evert. 2004. *The statistical analysis of morphosyntactic distributions*. In M. T. Lino et al. (eds.): Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004), p. 1539 - 1542. Lisbon, Portugal.

John R. Firth. 1968. F.R. Palmer (ed.): *Selected Papers of J.R. Firth 1952-59*. Longman, London.

Franz Josef Hausmann. 1979. *Un dictionnaire de collocations est-il possible?* Travaux de Linguistique et de Litterature XVII(1), p. 187-195. Centre de philologie et de littérature romanes de l'université de Strasbourg.

Franz Josef Hausmann. 1989. *Le dictionnaire de collocations*. In F. J. Hausmann et al. (eds.): *Wörterbücher, Dictionaries, Dictionnaires*, p. 1010 - 1019. De Gruyter, Berlin.

Franz Josef Hausmann. 2003. *Was sind eigentlich Kollokationen?* In K. Steyer (ed.): *Wortverbindungen - mehr oder weniger fest. Jahrbuch des Instituts für Deutsche Sprache 2003*:309 – 334. De Gruyter, Berlin.

Ulrich Heid and Rufus H. Gouws. 2005. *A model for a multifunctional electronic dictionary of collocations*. Institut für Maschinelle Sprachverarbeitung (IMS), Universität Stuttgart, submitted to EURALEX 2006.

Ulrich Heid and Julia Ritz. 2005. *Extracting collocations and their contents from corpora*. F. Kiefer et al. (eds.): *Papers in Computational Lexicography: Complex 2005*. Hungarian Academy of Sciences, Budapest.

Hannah Kermes. 2003. *Off-line (and On-line) Text Analysis for Computational Lexicography*. Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung (AIMS):9(3). Stuttgart.

Brigitte Krenn. 2000. *The Usual Suspects: Data Oriented Models for the Identification and Representation of Lexical Collocations*. PhD thesis. DFKI und Universität des Saarlandes, Saarbrücken.

Brigitte Krenn. 2000. *Collocation Mining: Exploiting Corpora for Collocation Identification and Representation*. W. Zühlke and E. G. Schukat-Talamazzini (eds.): *Proceedings of KONVENS 2000*, p. 209-214. Ilmenau, Deutschland.

Julia Ritz. 2005. *Entwicklung eines Systems zur Extraktion von Kollokationen mittels morphosyntaktischer Features*. Diploma Thesis. Institut für Maschinelle Sprachverarbeitung (IMS), Universität Stuttgart.

Helmut Schmid. 1994. *Probabilistic part-of-speech tagging using decision trees*. D. Jones and H. Somers (eds.): *Proceedings of the International Conference on New Methods in Language Processing (NeMLaP)*. Manchester, U.K.

Frank Smadja. 1993. *Retrieving Collocations from Text: Xtract*. Computational Linguistics (19), p. 143-177. Manchester, U.K.

Agnès Tutin. 2004. *Pour une modélisation dynamique des collocations dans les textes*. G. Williams and S. Vessier (eds.): *Proceedings of the Eleventh EURALEX International Congress*. Lorient, France.

Angelika Wöllstein-Leisten, Axel Heilmann, Peter Stepan and Sten Vikner. 1997. *Deutsche Satzstruktur*. Stauffenburg, Tübingen.