Identifying predicatively used adverbs by means of a statistical grammar model

Heike Zinsmeister & Ulrich Heid
Institut für Maschinelle Sprachverarbeitung
Universität Stuttgart
Azenbergstraße 12, 70174 Stuttgart, Germany
zinsmeis,uli@ims.uni-stuttgart.de

1 Introduction

This paper presents work on the corpus-based acquisition of predicatively used German adverbs by means of a statistical grammar model. The predicative use of adverbs is lexically restricted to local adverbs, temporal adverbs, and to a limited number of other adverbs, some of which are part of idiomatic phrases. Besides listing local and temporal adverbs we are mainly interested in identifying the list of 'other' adverbs. These adverbs cannot be subsumed under a common semantic class. Their predicative function is an idiosyncratic property of the lemma. The identification task was pursued by extracting predicatively used adverbs and prepositional phrases with an adverb complement from a statistically trained grammar model. The results were sorted by frequency and ranked according to a statistical association score. We then manually classified them into four classes: (i) local adverbs, (ii) temporal adverbs, (iii) other adverbs, and (iv) parsing errors. We discuss the extraction results from a linguistic point of view .

2 Motivation

Much effort is being put into the creation of detailed, machine-readable, broad-coverage lexicons (cf. e.g. Braasch et al. 1998, Lezius et al. 2000). Manual creation is very time-consuming and therefore methods of (semi-)automatic lexicon creation by exploiting large corpora are pursued (e.g. Manning 1993, Eckle 1999, Heid and Kermes 2002). In this paper we describe an approach that is based on corpus investigation by means of a statistical grammar model (for a general overview of this approach see Schulte im Walde et al. 2001). We concentrate on the acquisition of a narrow set of data, German adverbs that occur in a predicative function. We are interested not only in bare adverbs in this function but also in prepositional phrases with an adverb complement ('adverb PP'). It is desirable to list predicatively used adverbs, to avoid unnecessary ambiguities in parsing. General language defining dictionaries, such as *Duden - Deutsches Universalwörterbuch* are not an ideal source of information, since the predicative use is not described explicitly, but at most given in examples, under the respective adverb entries.

3 Predicative construction

The predicative construction is characterised by a copular verb, like *sein* 'be', *werden* 'become', or *bleiben* 'remain', in combination with a non-verbal predicative which cannot be dropped without changing the meaning of the verb (cf. e.g. Quirk et al. 1985: §16.12). The predicative is realised by different categories, most prominently by noun phrases (NP), adjective phrases (AdjP), and prepositional phrases (PP).

(1)	a.	Der Gärtner ist der Mörder. The gardener is the murderer.	[NP]
	b.	Der Gärtner ist verdächtig. The gardener is suspicious.	[AdjP]
	c.	Der Gärtner ist im Hof. The gardener is in the yard.	[PP]
	d.	Der Gärtner ist dort. The gardener is there.	[AdvP]

The predicative use of adverbs (AdvP), cf. (1-d), is less prominent, and it is sometime ignored in descriptive grammars. Helbig and Buscha (1998) are an exception to this. They analyse the predicative use as one of the core syntactic functions of (German) adverbs.

4 Predicative adverbs

Helbig and Buscha (1998: 338f.) define three syntactic frames for adverbs in German: (i) adverbial, i.e. modification of the clause or modification of an adjective or another adverb, (ii) postnominal attributive, and (iii) predicative. Based on these frames, they distinguish four distributional classes, two of which include the predicative use (Helbig and Buscha 1998: 342f.). They give some examples for each class. The first class (adverbial, postnominal attributive, and predicative use) consists of local and temporal adverbs (*dort* 'there', *hier* 'here', *da* 'there', *draußen* 'outside', *drinnen* 'inside', *drüben* 'over there', *damals* 'then', *gestern* 'yesterday', *morgen* 'tomorrow', *heute* 'today'), whereas the examples for the second class (adverbial and predicative use, only) are manner adverbs (*anders* 'differently', *so* 'so', *ebenso* 'just so'). The following examples illustrate the different semantic types of predicative adverbs and also the predicative occurrence of adverb PPs.

(2)	a.	Sie werden bei Tagesanbruch *hier* sein. 'They will be here by dawn.'	[local
	b.	Das Fest ist *heute*. 'The party is today.'	[temporal]
	c.	Seine Bemühungen waren *vergebens*. 'His efforts were to no avail.'	[other]
(3)	a.	Der Junge ist *von gegenüber*. 'The boy is from across the road.'	[P+local]
	b.	Das Brot ist *von heute*. 'The bread is from today.'	[P+temporal]

Our goal is to acquire predicatively used adverbs from a corpus. Besides listing local and temporal adverbs we are mainly interested in identifying the list of 'other' adverbs which are not restricted to manner adverbs. These adverbs cannot be subsumed under a common semantic class which means that their predicative function is an idiosyncratic property of the lemma.

In traditional linguistics, 'adverb' is treated as a kind of default word class. "It is tempting to say simply that the adverb is an item that does not fit the definitions for other word classes" (Quirk et al. 1985: §7.46). In German, especially the distinction between adverbs, on the one hand, and adverbially used adjectives, on the other hand, is blurred. In our experiments we make use of DMOR (Schiller 1995), a morphological analyser, the output of which corresponds to the STTS tagging guidelines (Stuttgart-Tübingen Tagset) for German corpora (see Schiller et al.1999: 56f.). Table 1 summarises the distributional and semantic criteria which are used in STTS to differentiate between adverbs and adjectives.

Table 1: STTS - adverbs vs. adjectives

adverbial use	predicative use	attributive use	example
ADV	—	—	*lediglich* 'only'
ADV	ADV	—	*vergebens* 'in vain, to no avail'
ADV	—	ADJ different meaning	*nämlich* 'namely, actually' vs. 'same'
ADV different meaning	ADJ	ADJ	*eben* 'just' vs. 'flat, even'
ADJ	ADJ	ADJ	*wahrscheinlich* 'probable, probably'
—	ADJ	ADJ	*schuldig* 'guilty'
—	—	ADJ	*obere* 'top'
uninflected form		inflected form	

We add one criterion to the STTS classification and require that elements, which subcategorise for an internal argument, are never be classified as adverbs but as adjectives. This even holds for elements such as *imstande, etwas zu tun* 'able to do something', which never occurs in attributive function, cf. **der (dazu) imstande Mensch*. Items as *rechtens* 'legally' in *Es ist rechtens, das zu tun* 'It is legal to do that' are still treated as adverbs. The infinitval clause is not an object but the extraposed subject which comes together with correlate *es*.

We assume that pronominal adverbs (also 'R-pronouns') such as *davor* 'before this/in front of it' or *hierin* 'in this' are prepositional phrases (cf. e.g. van Riemsdijk 1978) and treat them in the same way as adverb PPs[1]. Adverbs such as *hier* 'here' or *so* 'so' have a context-dependent meaning and may be analysed as proforms which substitute for a phrasal category in syntax (cf. Helbig and Buscha 1998: 347f.). If all adverbs in predicative function behaved that way we could assign them a phrasal tag, such as PP or NP, and restrict the predicative construction to the main categories NP. AdjP, and PP, after all. But even then, we would need a list of appropriate adverb candidates to mark them correspondingly. We do not follow this approach. We argue instead that the predicative construction has to allow for predicative adverbs in addition to the other categories, since there are cases which cannot be analysed as proforms, like e.g. *nirgends* 'nowhere' or *rechtens* 'legally'.

In predicative clauses with more than one adverb it is sometimes not clear whether the adverb functions as predicative or as modifier. As a rule of thumb, the rightmost adverb functions as the predicative. But there are exceptions to this rule: the non-predicative adverb might be extraposed and then occur at the right edge of the clause; either adverb, the predicative or the non-predicative, might be topicalised and occur to the left of the finite verb. In both scenarios, it largely depends on the context which adverb is understood as the predicative and which as the modifier.

(4) a. Die Probe ist morgen hier.
b. Die Probe ist hier morgen.
c. Morgen ist die Probe hier.
'Tomorrow, the rehearsal will be here./Here, the rehearsal will be tomorrow.'

All examples in (3) are potentially ambiguous. Their interpretation depends on the context and is normally indicated by the intonation pattern. In written text the ambiguity cannot be resolved on sentence level. Examples like this should thus be excluded from the data to be analysed[2].

We explicitly excluded ambiguous items such as adverb/adjective, e.g. *langsam* 'slowly/slow', adverb/participle, e.g. *ausgenommen* 'exceptionally/exclude' and adverb/substitutive indefinite pronoun, e.g. *viel* 'a lot'. They rarely allow the adverb reading in predicative function and would therefore create a lot of noise in the extracted data.

5 Acquisition technique

Our goal is to identify the subset of adverbs that may function as a predicative. Instead of mere pattern matching, we make use of a fully-fledged probabilistic grammar model that encodes predicate argument structures. Due to ambiguities (cf. the discussion above) and to the relatively free word order in German, it is not sufficient to simply analyse adjacent words to determine whether a given phrase functions as a predicative. Although the distribution of predicative and verbal elements is more restricted than the distribution of nominal arguments, it still allows for variation. We combine different types of information: (i) distributional information: statistically estimated frequencies of adverbs in predicative function; and (ii) relevance information: ranking of predicative occurrence according to a statistical association measure; and, for secondary purposes, (iii) selectional information: collection of the nominal heads of the corresponding subject phrases. Our grammar provides the described features: it recognises predicate argument structure independently from linearisation; it estimates the frequency of a

[1] Pronominal adverbs are labelled with the specific tag PROADV in DMOR. Our results feature some adverbs in the class of 'simplex' adverbs that might as well be analysed as pronominal adverbs. These items are either ambiguous between a lexicalised reading and a reading as transparent pronominal adverb, or they are simply not captured in DMOR. We did not manipulate the lexical input.
[2]Even though the ambiguity cannot be resolved, the analysis provided by the grammar spanning the whole sentences allows to identify such cases.

grammatical structure with respect to its lexical heads; furthermore, it learns the cooccurrence frequency of pairs of lexical heads with respect to a grammatical structure. The frequency information forms the input for subsequent relevance ranking by means of the association measure t-score.

6 Grammar model

We use a probabilistic grammar that models linguistic knowledge and provides full sentence parses (see Schulte im Walde et al. 2001). It is based on a manually established context-free grammar which was not developed specifically for this particular extraction task. It is a general model which encodes a large variety of syntactic and lexical information (Schulte im Walde 2003 gives a general overview of extraction possibilities; see e.g. Zinsmeister and Heid 2003 for a particular application).

The predicative construction is encoded in the context free grammar as follows. The projection of a copula verb expands to a predication phrase. The predication phrase in turn expands to a non-verbal phrase, e.g. to an adverbial phrase *PRED → ADVP'*. The mother category *PRED* is given on the left hand side of the rule and the daughter category *ADVP* is given on the right hand side. The apostrophy marks the head, which becomes relevant if there is more than one daughter on the right hand side of the rule. In our experiments we make use of the morphological analyser DMOR (Schiller 1995) for lexicon creation,see also section 4. The output is mapped onto part of speech tags which also function as terminal grammar tags. They might include feature specifications such as the case feature in the nominal tag *NN.Gen(enitive)*. Each lexicon entry includes an (inflected) token and a list of triples (part of speech, frequency, lemma).

(5) a. lexical entry of *vorbei* 'over'
vorbei ADV vorbei 0.00 *VPRE 0.00 vorbei* [VPRE=verb prefix]
b. lexical entry of *Jahres* 'year'
Jahres NN.Gen 0.00 Jahr

The grammar was trained by a statistical parser (LoPar, Schmid 2000). The parameters of the grammar, i.e. the unknown frequency values, which are easily translated into probabilities, are iteratively estimated with an instance of the Expectation-Maximisation (EM) algorithm (Baum 1972). The estimated frequencies correspond to parsing probabilities, therefore the numbers are not discrete counts, but gradual fractions. The grammar is lexicalised, which means that each rule is multiplied by all potential lexical heads and the probability mass of each rule is spread over the lexicalised rule variants. Lexical heads are determined by terminal symbols. The respective lemma is propagated as head feature to the mother category. In a non-terminal rule context, the mother category inherits the head feature of the daughter category which is marked as head (by an apostrophe). Lexicalisation has the effect that common structures might be 'unlearned' for specific lexical heads such that lexically determined structural preferences surface in the analysis. Lexicalisation does not only affect the validation of grammar rules with respect to the lexical head of the structure. It also allows the grammar to learn selectional preferences, i.e. head-head relations between mother nodes and their non-head daughter nodes, for example the relation between the predicative head of a clause and the nominal head of its subject. The context-free grammar was trained on about 4,000,000 sentences of 5 to 20 tokens which were taken from a newspaper corpus.

7 Extraction Procedure

The distributional information is read off the lexicalised grammar rules in which every rule is instantiated for each potential lexical head together with its estimated frequency.

(6) lexicalised grammar rule

	freq	head	mother category→daughter category'
a.	*132.25*	*vorüber*	*PRED→ADVP'*
b.	*2.02*	*gegenüber*	*PRED→PP.von'*

The selectional information, respectively, is given in the lexicalised cooccurrence rules which encode a list of all mother categories together with their non-head daughter categories annotated with the lexical heads of both and the estimated frequency of the particular selectional configuration. (7) encodes the

frequency information of *Wurm* as subject to predicative *drin*, as in e.g. *Heute ist im Spiel der Wurm drin*. 'There is something wrong in the game, today.'

(7) lexicalised cooccurrence rule

freq	non-head daughter	mother
10.95	*Wurm NP.Nom*	*drin VPK*

Due to ambiguities as described above, in section 4, and also due to general parsing errors, we expect a certain degree of noise in the extracted results, especially with low frequency data. It is important to balance precision and recall in a reliable way such that the resulting lists can be presented to a human evaluator without excluding too much valid information yet not overextending the task of manual inspection by including too much incorrect material. A simple way to reduce noise is to set a cut-off at a certain frequency value. According to the rule of thumb precision improves and recall decreases if the cut-off value is raised. In the task of lexicon creation in general, higher recall is more important than better precision and a lower cut-off value is preferred. Filtering of adverbs that are ambiguous between adverb, on the one hand, and adjective, participle, or indefinite pronoun, on the other hand (see also section 4), reduces the candidate set to 319 adverbs. For our classification task, we did not use a frequency cut-off but reranked the candidate list according to a statistical association measure, see section 8 below. The resulting was manually checked in a substitution test and furthermore classified into three semantic classes (local, temporal, and other). If an item fails the substitution test and does not constitute part of an idiomatic expression, it is classified as 'error'

(8) Substitution test

Ich bin ein ADV. 'I am an ADV.'	examples:	*durcheinander, (von) hier*
Das war ein ADV. 'This was an ADV.'		*gestern, vorher*
Der X ist ein ADV. 'The X is an ADV.'		*(mit) inbegriffen, vergebens*

8 Sorting results with an association measure

We employ the t-test for reranking the frequency list. The resulting t-score is a statistical association measure which is used, for example, to determine collocations in a corpus (see e.g. Church and Hanks 1989). Given a pair of words in a corpus, the t test calculates the deviation of the observed frequency of the pair from its expected frequency which is determined under the assumption that all words of the corpus occur independent from each other. A great deviation means a high t-score value. which in turn means that the pair has a strong correlation and that the independence hypothesis can be rejected. We apply the association measure to the pair (adverb, predicative adverb phrase). For our task we expect high t-score values for adverbs that occur relatively often in predicative use with respect to their overall occurrence. This helps to suppress the unwanted listing of parsing errors of high frequent adverbs. We implemented a version of t-score that refers to the partitions P_{ij} of a contingency table, see Table 2 and compares observed (our 'estimated') frequencies O_{ij} with expected frequencies E_{ij}. Taking the analysed adverbs as one parameter ('adv') and the predicative use of adverbs as the other ('pred'), t-score is calculated as follows from the frequencies (cf. Evert 2002):

Table 2: contingency table

t-score = O11-E11/√O11	predicative use (pred)	other uses of adverbs (non-pred)	
adverb (adv)	E11 = (R1 * C1)/N O11 = f (adv,pred)	O11= f (adv,non-pred)	R1 = O11 + O12
other adverbs (adv')	O21= f (adv',pred)	O22= f (adv',non-pred)	R2
	C1 = O11 + O21	C2	N = C1 + C2 = R1 + R2

For the manually inspection we set a cut-off at t-score 0.00. The resulting list comprises 138 candidates, including some very low frequency items such as *solcherart* (frequency: 1.00) or nütze (frequency: 0.57).

9 Results

We extracted the estimated frequencies of adverbs as head of a predication phrase. The grammar assigned a frequency larger that 0.5 to 319 adverbs. In addition, there were ambiguous cases which we ignored, namely 35 adverb/adjective, three adverb/participle, and seven adverb/indefinite pronoun. We sorted the adverb list according to t-score values. This has the effect that the most prominent error candidates are moved to the end of the list, e.g. *auch* 'also' (frequency rank: 6, t-score rank: 317), *noch* 'still' (frequency rank: 11, t-score rank: 297), and *nur* 'only'(frequency rank: 12, t-score rank: 309). Table 3 shows the 20-best predicative adverbs sorted by t-score. The second column gives the manual classification: *loc(al)*, *temp(oral)*, *other*, and *error*. The third column lists the t-score value, and finally the fourth column gives the estimated frequency.

Table 3: 20-best predicative adverbs sorted by t-score

Adverb	type	t-score	frequency	adverb	type	t-score	frequency
Da	loc	41.22	2397.13	je	error	11.30	181.87
Vorbei	temp	39.72	1648.69	vorüber	temp	11.28	132.25
Anders	other	37.73	1572.29	her	(temp)	11.25	176.52
So	other	29.95	1988.53	zurück	loc	10.11	294.73
Genug	other	26.53	780.70	raus	loc	10.07	125.86
Unterwegs	loc	25.60	680.23	allemal	error	10.07	115.52
Soweit	temp	25.16	660.81	hier	loc	9.70	534.23
Weg	loc	24.79	673.39	denn	(other)	9.33	252.07
Draußen	loc	12.09	170.67	umsonst	other	8.90	92.58
Vonnöten	other	11.97	146.99	allein	other	8.69	233.41

Within the 138-best adverbs according to the t-score ranking, we get a precision of 77% . There are 38% local adverbs, 17% of type 'other', 9% temporal adverbs, and finally 7% of a mixed type. For theoretic reasons, we suggest to reanalyse the items in (9) as adjectives. They subcategorise for an internal argument or require obligatory modification, such as. *her* 'from' (t-score: 11.25) which requires obligatory modification, as in *lange/drei Tage her* 'long/three days ago'.

(9) candidates for reanalysis as adjective
außerstande + zu-infinitive, *her* + modifier, *imstande* + zu-infinitive, *nütze* + modifier, *unschwer* + zu-infinitive, *wohl* + dative, *zumute* + modifier, *zuteil* + dative, *zuwider* + dative

23% of the candidates are classified as errors (e.g. *je* 'each', *allemal* 'any time', *kaum* 'hardly', *keineswegs* 'no way', *keinesfalls* 'on no account', *zueinander* 'to each other'), whereby 41% of the errors come with a frequency of 1.00. This means that setting an additional frequency cut-off would have suppressed them. *Denn* is a special case. It might be part of the idiomatic phrase *es sei denn* 'unless', which is not a transparent predicative construction as such but still a highly relevant combination. Therefore, we did not count it as an error. There are several items that also function as (separable) verb particles, e.g. *vor* 'forward', which leads to an additional ambiguity. This is a problem, we mostly ignored in our investigation. We can only sketch it here. *Vor* is a local adverb, alternatively it functions as prefix of the particle verb *vorsein*. It might also be analysed as a case of verb ellipsis in which the context-dependent motion verb such as *vorgehen/-rennen/-fliegen* is elided. Furthermore, *vor* might be part of a split pronominal adverbial, e.g. *davor*, as is common in Dutch: *Hij leest in het book > Hij leest erin > Hij leest er graag in*. In German the construction seems marginal and maybe regional.

We are specificly interested in the group of 'other' adverbs. They cannot be subsumed under a common semantic class and are therefore of particular interest for both, lexicographers as well as computational linguists. (10) lists the manually classified result, based on 138 candidates (threshold: t-score value larger than 0.0). Items that are marked with an asterisk are part of a fixed or idiomatic expression.

(10) predicative adverbs of type 'other'
allein, alleine, alltags, allzuviel, anders, andersherum, andersrum, auseinander, bestens, dahin*, denn*, dran*, durcheinander, genauso, genug, hin*, hinüber*, inbegriffen, obenauf*,*

rechtens, so, solcherart, soweit, umsonst, unrechtens, vonnöten, wohlauf, vergebens, zusammen

Adverb PPs are extracted with a precision of only 69% (out of 70 candidates). We find a large proportion of local and temporal items in adverb PPs and only few of the type 'other', like *mit inbegriffen* as in *Es ist im Preis mit inbegriffen* 'It is included in the price'. Sample results are given in Table 4. The type refers to the type of the adverb and not to the prepositional phrase as a whole. For example, *von gestern* 'from yesterday', which is number 3 in the t-score ranking, is also part of the idiomatic phrase *Er ist nicht von gestern* 'He wasn't born yesterday', which is not marked in the table. A similar example is *gegen rechts* 'against right'. It has an additional lexicalised meaning as in *Er ist gegen rechts* 'He is against the right wing'. In addition to t-score and estimated frequency, the table includes the t-score value of the bare adverb in predicative function.

Table 4: predicative adverb PPs

prep+ adverb	Type	t-score	freq	t-score adv	prep+ adverb	typ	t-score	freq	t-score adv
mit dabei	Loc	7.44	70.20	-0.22	von heute	temp	2.91	15.591	15.59
an dabei	Error	4.64	27.42	-0.22	nah dran	loc	2.27	6.53	4.29
von gestern	Temp	4.04	21.53	-27.24	von drüben	loc	1.68	4.00	2.81
für immer	Temp	2.64	10.47	-14.00	...				
von hier	Loc	2.30	15.84	9.70	von vor-vorgestern	temp	0.78	1.00	—

The grammar model does not only provide distributional information but also frequencies of lexical cooccurrences, see section 7. We expected to gain additional idiomatic material from this type of data. Sample results are given in table 5.

Table 5: selectional information – subject of predicative adverb

subject+adv	t-score	freq	comment
Geld da	6.90	66.53	
Zeit vorbei	6.47	52.94	
Spuk vorbei	5.00	30.00	figurative
Luft raus	4.52	20.99	idiomatic 'has gone flat'
Situation anders	4.20	19.85	
Wurm drin	3.17	10.95	idiomatic 'There is something wrong'
Jahr vorbei	2.91	11.00	
Krieg vorbei	2.69	9.99	
Stimmung anders	2.49	7.00	
Fahrer unterwegs	2.46	7.00	

10 Conclusion and Outlook

We discussed the phenomenon of predicatively used adverbs in German. Furthermore, we introduced a general acquisition technique for lexical information which makes use of a probabilistic grammar model. We extracted candidates for predicative adverbs and their model frequencies. The frequency-based candidate list was reranked after calculating the association measure t-core. We extracted candidates for predicative adverbs and adverb PPs in predicative use. Both types allow for local and temporal adverbs in general. The former type is also realised by manner adverbs and other, sometimes idiomatic, items. In the case of adverb PPs almost all adverbs belong to the local or temporal type. Some idiomatic phrases were found, as well. The inspection of selectional information, i.e. the nominal heads of the subjects, brought out some further idiomatic phrases. These are interesting data especially for

lexicographers. The frequency list as such is a valuable result, as well. Subsequent ranking by t-score emphasises relevant data. In our case it also reduced the noise by suppressing parsing errors of highly frequent adverbs.

We did preliminary experiments on clustering the predicative adverbs on the basis of their subjects. The goal was to automatically differentiate between the different semantic types of adverbs. The results were not satisfactory. It might be necessary to extend the domain of investigation to a broader range of constructions; but this is left for future work. The same holds for the investigation of multi word adverbs and coordinated phrases in predicative function such as *auf und davon* 'away' or *aus und vorbei* 'over', which we did not consider here.

References

Baum L E 1972 An Inequality and Associated Maximization Technique in Statistical Estimation for Probabilistic Functions of Markov Processes. Inequalities III: 1-8.

Braasch A, Buhr Christensen A, Olsen S, and Pedersen B 1998 A large scale lexicon for Danish in the information society. In Proceedings from the First International Conference on Language Resources & Evaluation, LRAA. ELRA.

Church K W, Hanks P 1989 Word associtation norms, mutual information and lexicography. In Preceedings of the 27th Annual Meeting of the Association of Computational Linguistics, pp 76-83.

DUDEN - Deutsches Universalwörterbuch 2001 4th edition. Mannheim, Dudenverlag.

Eckle-Kohler J 1999. Linguistisches Wissen zur automatischen Lexikon-Aquisition aus deutschen Textcorpora. Berlin, Logos (PhD thesis).

Evert S 2002 Mathematical Properties of AMs. Handout, Workshop Computational Approaches to Collocations. Vienna, cf. http://www.collocations.de/AM.

Heid U, Kermes H 2002. Providing Lexicographers with Corpus Evidence for Fine-grained Syntactic Descriptions: Adjectives Taking Subject and Complement Clauses. In Braasch A and Povlsen C (eds), Proceedings of theTenth EURALEX International Congress, volume1, pp 119-128.

Helbig, G, Buscha J 1991 Deutsche Grammatik - ein Handbuch für den Ausländerunterricht. Leipzig and Berlin and München, Langenscheidt.

Lezius W, Dipper S, Fitschen A 2000 IMSLex -- representing Morphological and Syntactical Information in a Relational Database. In Heid U, Evert S, Lehmann E, and Rohrer C (eds), Proceedings of the 9th EURALEX International Congress, pp 133-139.

Manning, C 1993 Automatic acquisition of a large subcategorization dictionary from corpora. In proceedings of the 31th Annual Meeting of the Association of Computational Linguistics, pp 235-242.

Manning, C D., Schütze H 1999 Foundations of Statistical Natural Language Processing, 1st edition. Cambridge MA, MIT Press.

Quirk R, Greenbaum S, Leech G, Svartvik J 1985 A comprehensive grammar of the English language. London, Longman.

van Riemdijk H 1978 A Case Study in Syntactic Markedness. The Binding Nature of Prepositional Phrases. Lisse, The Peter de Ridder Press.

Schiller A 1995 DMOR: Entwicklerhandbuch. Institut für Maschinelle Sprachverarbeitung, Stuttgart, Universität Stuttgart.

Schiller, A, Teufel S, Stöckert C, Thielen C 1999 Guidelines für das Tagging deutscher Textcorpora mit STTS. Technical report, Institut fur Maschinelle Sprachverarbeitung, Universität Stuttgart.

Schmid H 2000 Lopar: Design and Implementation. Arbeitspapiere des Sonderforschungsbereichs 340 \textitLinguistic Theory and the Foundations of Computational Linguistics 149, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.

Schulte im Walde S 2003. A collocation database for German verbs and nouns. In Proceedings of Complex 2003, Budapest.

Schulte im Walde S, Schmid H, Rooth M, Riezler S, and Prescher D 2001 Statistical Grammar Models and Lexicon Acquisition. In Rohrer C, Rossdeutscher A, Kamp H (eds), Linguistic Form and its Computation. Stanford, CSLI publications, pp. 387-440.

Zinsmeister H, Heid U 2003 Significant Triples: Adjective+Noun+Verb Combinations. In Proceedings of Complex 2003, Budapest.