A Word Database for Natural Language Processing

Brigitte Barnett
Hubert Lehmann
Magdalena Zoeppritz

IBM Scientific Center, Tiergartenstraße 15, 6900 Heidelberg, Federal Republic of Germany

Abstract: The paper describes the design of a fair sized lexical database that is to be used with a natural language based expert system with German as the language of interaction. Sources for entries and tools for constructing and maintaining the database are discussed, as well as the information needed in the lexicon for the purposes of syntactic and semantic processing.

1 *Introduction*

The intent of this paper is to show some aspects of a computer dictionary geared towards the natural language component of an expert system. The dictionary is organized as a database to integrate the various aspects of lexicographic work and, at the same time, enable fast access from a parser. Work on the lexicon was long neglected - both in theoretical linguistics and natural language processing projects - so we felt that a principled approach was overdue (cf. Sedelow (1985) for a survey of related work). In the past two years, we concentrated therefore on the formulation of criteria for establishing syntactic features which have to be coded in the lexicon, and we will report here on some of our findings. This will be preceded by a brief overview of the aims of our overall project and a short description of the prototype system we are building. We will then describe the design of our lexicographic database including the criteria for selecting sources of the vocabulary and some of our tools for editing and querying.

The main objectives of the project Linguistics and Logic Based Legal Expert System, which is a Joint Research Project between the University of Tübingen and the IBM Scientific Center Heidelberg, are to design and implement a natural language based knowledge acquisition and query system and to build a legal expert system on its basis. It consists of the following components:

- The *dialog component* controls the interaction with users and contains among other things an *explanation component* and a component for preparing system output for display and for eventually generating natural language explanatory texts.

  In a so-called user profile, as much information about a user is kept as necessary: to improve answers and explanations, one must know certain things about the user, mainly about her or his knowledge in current sessions. For example, one may want to avoid explanations about details the user already knows.

- The *deductive component* is activated by user queries, by input of new knowledge, and by requests of the Natural Language Analyzer.

- The *knowledge manager* administers the actual knowledge base in the working area as well as its permanent version in the database. It is the only component allowed to update the permanent knowledge base. It loads knowledge from the database into storage and requests consistency checks for new knowledge. With the exception of the lexical database, the Knowledge Manager also accesses the database on behalf of other system components.

- The *SQL Data System* (IBM (1983)) maintains the database which is a repository of facts and rules:
  - linguistic knowledge, e.g. dictionary and grammar
  - common sense knowledge, including a thesaurus
  - legal knowledge (law, rules from commentaries and decisions, legal strategies)
  - cases
  - user profiles

- The *natural language analyzer* and its *dictionary* are extensions and modifications of the existing *User Specialty Languages* system (USL) developed at the Heidelberg Scientific Center (Lehmann, H., N. Ott, M. Zoeppritz (1985), M. Zoeppritz (1984)). USL is a natural language front end to SQL/DS (IBM 1983) operational in six languages. Within the scope of this project, it will be enhanced to suit the requirements of a natural language (German) based expert system. This means that it must be able to deal with both running texts and queries and to translate them into their corresponding logical form.

  The Natural Language Analyzer consists of the following parts:
  - a sentence separator splitting texts into sentences,
  - a pre-parser for dictionary look-up,
  - the parser and the routines for semantic analysis,
  - routines for the generation of the logical form from intermediate structures (cf. Guenthner and Lehmann (1984) for a description),
  - routines for semi-automatic generation of thesaurus extensions (Wirth, R. (1984)).

As a specific application, the area of German traffic law was chosen for the expert system which shall be used in two modes: for consultation by a legal expert and as a tutor for law students (cf. Alschwee et al. (1985) for details).

2 *Description of the Dictionary*

Within such an environment, a fairly large-sized and detailed dictionary is needed. Aspects of its design, the structure in the database, and the editing and querying facilites will be discussed (cf. also Barnett (1985)). The expected size of the dictionary within the scope of the project is estimated to be some 20,000 entries. Its current size is some 12,000 entries.

2.1 Word Database

Because we must be able to handle a large number of words in this project, we felt that it would be necessary to administrate them in a more appropriate form than the usual file organization and that a relational database would be the best tool for dealing with lexical information because of the following advantages:

- excerpting grammatical information according to specific features;
- links to related information not necessarily kept in the same table;
- easier control of updates;
- many types of integrity checks;
- automatic backup so that, in case of a systems breakdown, a consistent status remains available;
- another great advantage of database technology is concurrency capabilities which preventusers working on the same table from getting in each other's way.;
- and, within the realm of this project, the possibility to link to the Natural Language Analyzer.

2.2 Scope

The scope of the information contained in our dictionary is geared towards the processing of natural language by computer. Lexical information must therefore be more detailed and more explicit than in standard dictionaries intended for humans. Also, a computer dictionary is of no value unless it matches the grammar and the needs of the semantic processing.

We started with the coding of morphological and syntactic information, since we felt to be on rather stable ground there. We will report on some of the difficulties we encountered - many of them not unknown to theoretical linguistics - in the next section.

Semantic information is coded primarily in the form of meaning rules, but we have not included these in our lexical database yet, as we are still experimenting with different kinds of information and representations before we go to large-scale coding. We also hope that, at least to some extent, the acquisition of such information can be automated (cf. the approach taken by Wirth (1984)).

2.3 Sources

For the purpose of our particular application, we need to cover the vocabulary occurring in German traffic law. However, to meet the goal of general applicability, it is also necessary to include the core of the general German vocabulary. We will try therefore to code the relevant legal words based on texts from this very domain. In addition - and this is the greater problem - we must try to define which words pertain to the common vocabulary.

- As a first step, we have compiled a preliminary list of the 4000 most frequent German words from an existing frequency dictionary (Meier (1967)) and news texts. Later, we will make frequency counts of representative samples of texts to arrive at a more reliable list of words.
- IBM Germany has a dictionary of 70,000 entries containing morphological and hyphenation information.
- The vocabulary of the application area, i.e. from the legal domain, stems from the following sources:
  - A collection of relevant court decisions (from our study partner),
  - A number of accident descriptions collected from newspapers,
  - A few word lists used for document retrieval from both the Legal and Public Relations departments of IBM Germany.
- We plan to investigate to what extent machine-readable dictionaries or legal texts can be used for an automatic or semi-automatic acquisition of lexical and grammatical information and of common-sense knowledge.

2.4 Layout of the Dictionary Relation

In our word database, every word constitutes an entry, and most columns in the entry contain information concerning a particular word. Even though semantic aspects are not coded in this particular version, one may regard the codes as a representation of a word's morphological and syntactic meaning. Some words have more than one entry: to code multiple entries becomes necessary when different grammatical feature sets have to be assigned to one lemma.

All words are contained in a single table or relation. One could also envisage a separate table for every part of speech; however, this would be rather inconvenient, as it would be impossible to compare grammatical phenomena across different categories. Also it may be desirable to look at words of the same root but belonging to different parts of speech. With this necessity in mind, we designed an overall, general relation which would contain all words. In order to treat the words individually and according to their specific needs, a so-called "view" was defined for each part of speech. The present structure of the relation is described in Figure 1.

2.5 Tools and Aids

To facilitate coding and to ensure its accuracy, we use the following tools:

Editing: A Dictionary Editor (a menu-driven program running under ISPF (IBM 1982) interacting with the SQL/DS database) was developed to facilitate adding, updating, deleting, and checking of entries at the terminal.

Under this editor, a specific set of menus and help panels was implemented for nouns, verbs, and adjectives. Whereas the main menus contain only short hints to the grammatical information as a sort of reminder to the lexicographer, help menus give more detailed examples for the individual codes. Subpanels, as extensions to the main panel for input, and error messages also assist the lexicographer. Codes are verified by the Dictionary Editor to keep down the error rate.

Queries, Reports, and Files: Independently of the Dictionary Editor, the lexicographers work with the standard database interfaces ISQL (IBM 1983) and QMF (IBM 1983, 1984) to query, extract, recover, and to view the contents of the word database. QMF is used to select information and to format, display, and write reports. - The style of data display is easy to understand, so that persons who are not experts in data processing but are competent linguists can examine and alter the linguistic description according to

General Description	Field in Database	Noun	Verb	Adjective
Lemma (may not be empty)	WORT	Infinitive	Nominative	Positive
Current number of entry	CONT	#	#	#
Part of Speech	CAT	N	VERB	ADJ
MORPHOLOGY:				
Morphological codes	MORPH1	Declension code	Sets of suffixes	Scope of stem form
Morphological codes	MORPH2	Alternative declension	Syntax and scope of stems	Declension class
SYNTAX:				
Possible complement(s)	SYNTA1	--	Valency	Valency
Prepositions governed: alternating or cumulative	SYNTA2-5	Preposition	Preposition	Preposition
Up to 4 adverbials	SYNTB1-4	place/temporal/modal	place/temporal/modal	place/temporal/modal
Complement Clauses introduced by "daß"	SYNTB5	case/preposition	case/preposition	case/preposition
Complement Clauses: Infinitive introduced by "zu"	SYNTB6	case/preposition	case/preposition	case/preposition
Logical Subject of infinitive clause	SYNTB7	case/preposition	case/preposition	case/preposition
Complement Clauses introduced by "ob"	SYNTB8	case/preposition	case/preposition	case/preposition
Different for each part of speech:	SYNTB9	Gender	Separable Prefix	not used attributively
Different for each part of speech:	SYNTB10	Alternate Gender	Reflexivity	not used adverbially
Different for each part of speech:	SYNTC1	--	Impersonal Subject only	not used predicatively
Different for each part of speech:	SYNTC2	--	Usage of Participle	--
Obligatory Complements	SYNTC3	--	x	x
Subject Area	SUBJECT	x	x	x
Source: Userid of who coded; origin of entry	SOURCE	x	x	x
Frequency Count	FREQ	x	x	x
Source of Frequency Count	FREQS	x	x	x
Level of Style	STYLE	x	x	x
Stem for 1st Entry:	XREF	Nominative Singular	Infinitive without "-(e)n", without separable prefix	Positive
Secondary Stems:	XREF (cont'd)	Plural Form	Imperfect, Past Participle	Gradation
Hyphenation	TRENN	x	x	x
Date of last Update	DATUM	x	x	x
5 additional, not yet used fields	F1,F2,F3,F4,F5	--	--	--

Fig.1: Structure of the Word Relation

their particular needs.

It is part of the work of a lexicographer to account for all grammatical constructions within which a word may appear. To achieve this, the lexicographers consult standard dictionaries such as Duden (1976-1981) and Brockhaus Wahrig (1980-1984), but most importantly, they consult the texts mentioned above in the form of a concordance which we generate dynamically.

3 *Syntactic Information*

The restrictions on co-occurrence with other words (or phrases) is what we consider to be syntactic information. Here we include information on government (or valency) and on adverbials (or attributes) which serve to subclassify the various parts of speech. Our work is based on the work of Fillmore (1968), Gross (1984), Heidolph et al. (1980), Steinitz (1969), Bierwisch (1963), Helbig and Schenkel (1975), Sommerfeldt and Schreiber (1977, 1980), and on Zoeppritz (1984).

For the practical work on a dictionary, it is of utmost importance to make fully explicit the criteria for the different classifications used. Such criteria are notoriously difficult to extract from theoretical as well as practice-oriented works.

3.1 Government

German verbs can be classified according to the objects they govern (accusative (A), dative (D), genitive (G), prepositional object (P), predicate noun or adjective (N)). We decided to include also the subject (N) among the complements governed by the verb. A similar classification can be carried out for adjectives (they govern cases as well as prepositions) and for nouns (which govern prepositional attributes; genitive attributes are not coded but admitted for every noun). While some of the complements may be missing in a sentence, others must be regarded as obligatory, and this must be coded in the dictionary as well. So we code two features indicating the maximum and the minimum of complements of a given word. This is illustrated by the adjective *überlegen* (superior) which gets the maximal code DP and the minimal code ZZ (nil):

(D) (P)
Paul ist uns im Weitsprung überlegen.
(Paul is superior to us in the long jump.)
Paul ist im Weitsprung überlegen.
Paul ist uns überlegen.
Paul ist überlegen.

There are a number of problems in determining what complements can be governed by a given verb. We will discuss here the problem with datives, and in the section on adverbials we will discuss the problem of how to distinguish between adverbials and prepositional objects.

It was noted by case grammarians (but also to some extent in traditional grammar) that datives perform different functions ("semantic roles"), and that only some of them should be regarded as subcategorizing the class of verbs; the others are sometimes called "free" datives which are exemplified by

ethical: Wirf *mir* die Vase nicht weg.
(Be sure not to throw the vase away)

possession: Paul brach *mir* den Arm.
(Paul broke my arm)

benefactive: Paul übersetzte *mir* den Brief
(Paul translated the letter *for me*)

responsibility: Die Vase ist *mir* zerbrochen.
(The vase broke during the time I had it)

(cf. also Heidolph et al. (1980) for a discussion and Wegener (1985)).

Free datives are never obligatory, but all other criteria so far are only semantically motivated and are - particularly in the case of the benefactive - not very well defined. But still, free datives can be taken into account by the grammar and can thus be attached to any suitable verb.

3.2 Complement Clauses

In accordance with Bierwisch (1963) and Heidolph et al. (1980), we consider complement clauses as filling the positions of nominal complements. The complement clauses we consider are *daß* (that) clauses, *ob* (whether) clauses, and infinitive clauses (pure infinitives and infinitive clauses introduced by *zu* (to)). Bierwisch was first in subcategorizing German verbs according to the implied subjects of the infinitive clauses they govern. Consider the following examples in English

John permitted Paul to leave
John persuaded Paul to leave

This is problematic, however, with some verbs in German when the infinitive clause contains modal verbs. Consider

Er flehte sie an zu gehen
(he begged her to leave)

with the implied subject *sie*, and

Er flehte sie an, gehen zu dürfen
(He begged her to be permitted to leave)

with the implied subject *er*. This shift of implied subject must be coded in the lexicon (for verbs like *dürfen*), so that the code can be used by the syntax rules for complex verbs.

A second problem concerns cases where an implied subject cannot be found in the matrix clause as in

Paul ordnete an, den Saal zu räumen
(Paul ordered the room to be cleared)

Es ist verboten, den Rasen zu betreten
(it is forbidden to walk on the lawn)

There are two different phenomena involved: The dative governed by *verbieten*, would be the implied subject, but happens to be omitted, whereas *anordnen* does not govern a candidate for implied subject. If there is no suitable candidate in the context, we get a generic interpretation, and we code our complement features accordingly.

3.3 Adverbials

Our approach to adverbials is closely related to the one taken by Steinitz (1969) and Heidolph et al. (1980), which - to us - seems far better motivated than e.g. the classification in Brockhaus Wahrig (1980-1984). Certain types of adverbials we consider to be governed by certain verbs, nouns and adjectives, and these are hence used for subcategorization. They include adverbials of

place: Paul wohnt *in Heidelberg*
(Paul lives in Heidelberg)

direction: Paul geht *nach Heidelberg*
(Paul goes to Heidelberg)

modality: Paul benimmt sich *schlecht*
(Paul behaves badly)

measure: der Vortrag dauert *eine Stunde*
(the lecture lasts one hour)

It has been our tendency to code adverbials only when they are obligatory, but this certainly does not cover all the information necessary.

A further problem concerns the decision between adverbial and prepositional object. Criteria for this distinction have been described by Steinitz (1969) and Heidolph (1980): they mainly involve observations on the role prepositions play - their variability and whether they have retained their meaning. Consider

Paul stood on the table
Paul insisted on the table

In the first case, we could have *near*, *under*, *by*, etc. instead of *on* whereas in the second case, we do not have a choice.

3.4 Coding Example

The following example shows the test questions and corresponding coding decisions for verbs and adjectives. The sample form is *überlegen*, that appears as verb with separable prefix in the meaning 'to cover', as verb with inseparable prefix in the meaning 'to reflect', and as an adjective meaning 'superior'.

Tests for *überlegen* 'to cover':

Prefix: separable or not?
legt über -- hat übergelegt -- überzulegen

Full government:

Paul legt Maria eine Jacke über

Dative can be left out:
Paul will eine Jacke überlegen

Accusative cannot be left out:
*Paul legt der Maria über *Paul legt über

Coding for *überlegen* 'to cover'

Stem: *leg*

Prefix *über*

Word class: VERB

Government: nominative dative accusative

Obligatory: nominative accusative

Testing for *überlegen* 'to reflect':

Prefix: separable or not?
er überlegt -- hat überlegt -- zu überlegen

Full government:
Paul überlegt sich eine Frage

Dative can be left out:
Paul überlegt eine Frage

Accusative can be left out as well:
Paul überlegt

Does the dative have to be reflexive?
*Paul überlegt uns eine Frage

The accusative can be replaced by *zu*-infinitive, *daß* and *ob*-clauses:
Paul überlegt (sich), Maria zu besuchen
Paul überlegt (sich), daß er Maria besuchen will
Paul überlegt (sich), ob er Maria besuchen will

The implied subject of infinitive clauses is the main clause subject:
Paul überlegt sich, Maria zu besuchen
Paul besucht Maria

Coding for *überlegen* 'to reflect':

Word class: VERB

Stem: *überleg*

Government: nominative dative accusative

Clauses: infinitive as accusative
implied subject is nominative
daß-clause as accusative
ob-clause as accusative

Reflexive: dative

Obligatory: nominative accusative

Testing for *überlegen* 'superior':

Full government:
Paul ist Maria im Weitsprung überlegen
der uns allen im Weitsprung überlegene Paul

Prepositional can be left out:
Paul ist Maria überlegen
der uns allen überlegene Paul

Dative can be omitted as well:
Paul ist überlegen -- der überlegene Paul

Zu-infinitives (marginally) and *daß*-clauses are possible in lieu of the prepositional. Clauses introduced by *ob* are not allowed, not even with negation.
Paul ist Maria darin überlegen,
daß er weiter springen kann
*Paul ist Maria (nicht) darin überlegen,
ob er weiter springen kann
Diese Schrift ist anderen darin überlegen,
leichter lesbar zu sein

The subject of infinitive clauses is the head of the adjective:
der darin, springen zu können, überlegene Paul
Paul ist überlegen
Die Schrift ist der anderen darin überlegen,
besser lesbar zu sein
Die Schrift ist besser lesbar

The preposition may not be omitted:
*Paul ist Maria überlegen,
weiter springen zu können

The adjective can be used in predicative and attributive position:
Paul ist überlegen -- der überlegene Paul

Coding for *überlegen* 'superior'.

Word class: ADJECTIVE

Stem: *überlegen*

Government: dative prepositional

Prepositions: *in bei*

Clauses: infinitive as prepositional
daß-clause as prepositional

Restrictions: no restrictions to predicative or attributive use

4 Semantic Information

When dealing with semantic information we should distinguish between the information needed for obtaining a (if possible) disambiguated logical form and the information needed to draw inferences from this logical form, even though - at least in part - this information may be identical. Among the former we include a *concept lattice* (or *hierarchy*) and *selection restrictions* which both are special cases of the *meaning rules* we use to represent *common sense* and *domain* knowledge. (For a detailed discussion of our approach to knowledge representation, cf. Guenthner / Lehmann / Schönfeld, 1986). All of this information we encode as Prolog terms, and we also store these in SQL/DS, but separate from the word relation described above.

We give an example here of the concept hierarchy currently used in our Natural Language Analyzer. ("bt" stands for broader term, the period in the third rule indicates a compound term):

```
bt(angeklagt,mensch).
bt(fahrzeug,fortbewegungsmittel).
bt(fortbewegungsmittel,hergestellt.objekt).
```

This hierarchy is used in conjunction with the selection restrictions listed below to disambiguate sentences, to recover ellipses, and to resolve anaphoric references. (The format of the selection restrictions is: Lemma, reading, state vs. event, list of restrictions for the respective complements, including an indication whether the verb is distributive (dist) or collective on a given complement):

```
verb(abbremsen,1,event,
    nom(dist,fortbewegungsmittel).nil).
verb(abbremsen,2,event,
    nom(dist,mensch).
    acc(dist,fortbewegungsmittel).nil).
```

Wirth (1984) has described a procedure to extend a concept hierarchy and selection restrictions from text on the basis of given sentences. In his procedure, human intervention is still required, and it seems doubtful at this point whether a fully automatic procedure is feasible. Further, one observes a certain discrepancy between linguistic usage and logical behavior of certain words. We are investigating ways to overcome these problems, but a discussion of them has to be left to forthcoming publications.

5 Conclusions

We have described the design of a lexical database to be used with a natural language based expert system, discussed a number of problems we encountered when coding syntactic information for words, and also mentioned where addtional work needs to be done in order to achieve a comprehensive dictionary for language processing.

By November 1985, we coded morphological and syntactic information for some 5,500 nouns, approximately 3,000 verbs, and 3,500 adjectives. Our next steps are to fully integrate the lexical database with the rest of our prototype to improve our concordance programs, and to continue the development of criteria for syntactic subcategorization.

Acknowledgement

We want to thank B. Endres, A Franzke, S. Goeser, D. Jacob, E. Latniak, W. Pietschke, S. Ritzenfeld, G. Schülzke, A. Storrer, and especially K. Horländer for the work they did in developing the Dictionary Editor, coding dictionary entries, and for their help and suggestions in formulating criteria for coding. Without them, this paper would not have been possible.

References

Alschwee. B., A. Blaser, He. and Hu. Lehmann, W. Schönfeld (1985): "Ein juristisches Expertensystem mit natürlichsprachlichem Dialog - ein Projektbericht", Proceedings of the International GI Congress '85, Munich, W. Brauer, B. Radig (eds.): Wissensbasierte Systeme, Informatikfachberichte 112, Springer Verlag, Berlin Heidelberg New York Tokyo.

Barnett, B. (1985): "A German Dictionary for a Linguistics and Logic Based Legal Expert System", E.O. Lippmann (ed.): Proceedings of the ELS Conference on Natural-language Applications, IBM Danmark A/S, September 10-15, 1985, section 11:1-19.

Bierwisch, M. (1963): Grammatik des deutschen Verbs, Studia Grammatica II, Akademie Verlag, Berlin.

Brockhaus Wahrig (1980 (Vol.1) - 1984 (Vol.6)): Deutsches Wörterbuch in sechs Bänden, Wiesbaden, Stuttgart.

DUDEN (1976 (Vol.1) - 1981 (Vol.6)): Das große Wörterbuch der deutschen Sprache, Bibliographisches Institut Mannheim/Wien/Zürich, Dudenverlag.

Fillmore, C. J. (1968): The Case for Case, Universals in Linguistic Theory, Holt, Rinehart and Winston, Inc., New York Chicago San Francisco Atlanta Dallas Montreal Toronto London.

Gross, M. (1984): "Lexicon-Grammar and the Syntactic Analysis of French", Proceedings 10th International Conference on Computational Linguistic, 22nd Annual Meeting of the Association for Computational Linguistics, Stanford University, California, July 2-6, 1984. Morristown: Association for Computational Linguistics (order from Dr. D. E. Walker, Bell Communications Research, 445 South Street, Morristown NJ 07960), 275-282.

Guenthner, F., Hu. Lehmann (1984): "Automatic Construction of Discourse Representation Structures", Proceedings 10th International Conference on Computational Linguistic, 22nd Annual Meeting of the Association for Computational Linguistics, Stanford University, California, USA, July 2-6, 1984. Morristown: Association for Computational Linguistics (order from Dr. D. E. Walker, Bell Communications Research, 445 South Street, Morristown NJ 07960), 398-401.

Guenthner, F., H. Lehmann, W. Schönfeld (1986): "A theory of representing knowledge", IBM Journal of Research and Development, Vol. 30 No. 1, pp. 39-56

Heidolph, K.E., Flämig, W. und Motsch, W. (1980): Grundzüge einer deutschen Grammatik, Akademie-Verlag, Berlin, second edition (1984).

Helbig, G.- Schenkel, W. (1975): Wörterbuch zur Valenz und Distribution deutscher Verben, VEB Bibliographisches Institut, Leipzig.

IBM (1983): SQL Data System, General Information, GH24-5013, International Business Machines Corporation, Endicott, New York.

IBM (1982): Interactive System Productivity Facility, General Information, GC34-2181, International Business Machines Corporation, Cary, North Carolina.

IBM (1983, 1984): Query Management Facility, General Information GC26-4071, International Business Machines Corporation, San Jose, California.

Lehmann, H., N. Ott, M. Zoeppritz (1985): "A Multilingual Interface to Databases", IEEE Database Engineering Bulletin, Vol. 8, No. 3, New York.

Meier (1967): Deutsche Sprachstatistik, Georg Olms Verlagsbuchhandlung, Hildesheim.

Sedelow, S.Y. (1985): "Computational Lexicography", Computers and the Humanities 19, 97-161.

Sommerfeldt, K.-E., Schreiber H. (1977): Wörterbuch zur Valenz und Distribution deutscher Adjektive, VEB Bibliographisches Institut, Leipzig.

Sommerfeldt, K.-E., Schreiber H. (1980): Wörterbuch zur Valenz und Distribution deutscher Substantive, VEB Bibliographisches Institut, Leipzig.

Steinitz, R. (1969): Adverbial-Syntax, Studia Grammatica X, Akademie-Verlag, Berlin.

Wegener, H. (1985): Der Dativ im heutigen Deutsch, Narr, Tübingen.

Wirth, R. (1984): Halbautomatische Erweiterung eines bestehenden Thesaurus, diploma thesis, Heidelberg University and Fachhochschule Heilbronn.

Zoeppritz, M. (1984): Syntax for German in the User Specialty Languages System, Niemeyer, Tübingen.