

The Representation of German Prepositional Verbs in a Semantically Based Computer Lexicon

Rainer Osswald, Hermann Helbig, Sven Hartrumpf

Intelligent Information and Communication Systems
Department of Computer Science, FernUniversität in Hagen
{firstname}.{lastname}@fernuni-hagen.de

Abstract

We describe the treatment of verbs with prepositional complements in HaGenLex, a semantically based computer lexicon for German. Prepositional verbs such as *bestehen auf* ('insist on') subcategorize for a prepositional phrase where the preposition usually has no independent meaning of its own. The lexical semantic information in HaGenLex is specified by means of MultiNet, a full-fledged knowledge representation formalism, which proves to be particularly useful for representing the semantics of verbs with prepositional complements. We indicate how the semantic representation in HaGenLex can be used to define semantic classes of prepositional verbs and briefly discuss the relation of these classes to Levin's verb classes. Moreover, we present first results on the automatic identification of prepositional verbs by corpus-based methods.

1. Introduction

Lexical resources suitable for advanced NLP applications should provide detailed morphosyntactic and semantic information. One of the difficult tasks that has to be lexically supported is the semantic interpretation of prepositional phrases. In this context, an adequate lexical representation of prepositional verbs such as *consist of* (German: *bestehen aus*, French: *consister en*) and *refer to* (German: *verweisen auf*, French: *faire référence à*) is of particular importance since the semantics of prepositional verbs is typically non-compositional in that the preposition does not have the regular interpretation it has in prepositional phrases occurring as adjuncts. In the following, we report on the treatment and coverage of German prepositional verbs in the computer lexicon HaGenLex, which is employed in various NLP systems for text retrieval and question answering; see e.g. (Hartrumpf, 2005).

2. Prepositional verbs in German

A first characterization of prepositional verbs in German, as in other European languages, requires them to subcategorize for a prepositional phrase (PP) where the preposition is determined by the verb.¹ It is common to speak of a *prepositional object* in this case. A closer look at this phenomenon reveals a whole variety of different cases; see also (Lerot, 1984) and (Breindl, 1989). In its most stringent version, the above characterization requires the subcategorization to be *strict* and the preposition to be *uniquely* determined. Examples are

- (1) *verhelfen zu* ('help to get')
 abzielen auf ('aim/get at')
 beruhen auf ('be based on')

¹Since Baldwin (2005) draws some attention to the problem of distinguishing prepositional verbs from particle verbs in English, it is worth mentioning that in German, prepositional verbs can be easily distinguished from prefix and particle verbs since particles appear either immediately in front of the verb, usually realized as a prefix, or are stranded on the right of the clause, whereas stranding of the prepositions of prepositional complements is not possible.

None of the (German) verbs in (1) allows an alternative reading.

The question of alternative readings arises if a verb can be used with and without a prepositional object because then one has to decide whether or not these uses come along with different readings; if not, the subcategorization for the PP is not strict. For instance, the standard reading of *zählen* ('count') should surely be regarded as different from the reading of the (transitive) verb *zählen zu* ('count among'), because counting an object among a certain group does not mean to count that object. But there are less clear examples such as *sich freuen auf* ('look forward to'). In general, we have here the notorious problem of identifying different verb readings. Nevertheless, there are clear cases of prepositional verbs where the subcategorization is not strict in that the prepositional object can be elliptical. Examples are *warten auf* ('wait for') and *sich verlieben in* ('fall in love with'). Note that the situation is different for verbs of social interaction like *streiten mit* ('quarrel with'), where the dropping of the PP complement goes along with the so-called reciprocal alternation by which the subject NP becomes collective as in *Peter and Mary quarreled*.

Besides examples with optional PP complement but uniquely determined preposition (such as *warten auf*), there are also prepositional verbs like *stimmen für/gegen* ('vote for/against'), where the PP complement is obligatory but the preposition can vary to a certain degree. Such a paradigmatic variation of prepositions, which does not affect the core meaning of the verb, is often taken as an indication that the prepositions in question are not semantically empty; cf. (Lerot, 1984). NP alternatives to PP complements, in contrast, as in dative alternations such as *etwas an jemanden schicken* vs. *jemandem etwas schicken* ('send sth to sb' vs. 'send sb sth'), are to be seen as mere variants of each other.

Prepositional objects are to be distinguished from strictly subcategorized PPs that have the regular interpretation of PP adjuncts. Examples for verbs with locative PP interpretations are *wohnen* ('live') and *sich aufhalten* ('stay'). It is characteristic for these verbs that the preposition is not de-

terminated by the verb but can vary in the same way as in the corresponding adverbial use of the PP. Exceptions are particle verbs like *hineingelangen in* ('get into') and *hineinfallen in* ('fall into'), where the preposition is determined by the direction that is incorporated by the particle *hinein* ('into').

Besides prepositional objects and obligatory adverbials there is a third type of obligatory PP described in the literature, namely predicative complements as exemplified in (2).

- (2) a. Peter galt als Aufschneider.
'Peter was regarded as a boaster.'
b. Mary hielt Peter für einen Aufschneider.
'Mary took Peter for a boaster.'

The predication underlying predicative constructions is typically relativized to a belief or opinion as in (2) or to the result of an action or event as for the verb *ernennen zu* ('appoint'). The predication can be either about the subject (2-a) or the object (2-b) of the matrix sentence. Note that the PP complement of the verbs in (2) may also contain an adjectival phrase instead of an NP.

Although prepositional objects by definition do not have the regular interpretation of prepositional adjuncts, there are subregularities in the sense that the preposition selected by the prepositional verb can be predicted on the basis of a subclass the verb belongs to. Two such subclasses, taken from (Breindl, 1989), are given in (3): The verbs of social interaction (3-a) select the preposition *mit* ('with'), while the verbs of obtaining select the preposition *von* ('from') (see Section 5. for additional examples).

- (3) a. sich unterhalten, streiten, sich versöhnen mit
'talk', 'quarrel', 'become reconciled with'
b. leihen, borgen, erhalten, bekommen von
'lean', 'borrow', 'receive', 'obtain from'

Note that a given preposition can be predicted by several semantically distinct subclasses; witness the examples in (4) for the preposition *an* ('at', 'on', ...), which are again adopted from (Breindl, 1989).

- (4) a. denken, sich erinnern an ('think of', 'recall')
b. leiden, sterben an ('suffer from', 'die of')
c. teilnehmen, teilhaben an ('participate in')

So, even if the prepositional object has a reasonably stable interpretation within a subclass, this interpretation strongly depends on that subclass.

3. The computer lexicon HaGenLex

HaGenLex (Hagen German Lexicon) is a general domain lexicon for German comprised of about 25,000 lexical units with rich morphological, syntactic, and semantic information.² The lexical semantic description in HaGenLex is based on the MultiNet knowledge representation formalism. The MultiNet (Multilayered Extended Semantic Networks) formalism provides classificational means including a hierarchy of ontological sorts (such as *substance*, *action*,

²See (Hartrumpf et al., 2003) for a more detailed description of HaGenLex.

| Relation | Short description |
|----------|------------------------------------|
| AGT | Agent |
| ARG1/2 | Argument specification (metalevel) |
| DIRCL | Local direction or goal |
| DUR | Duration |
| MCONT | Mental content |
| MEXP | Mental experiencer |
| OBJ | Unaffected, neutral object |
| ORNT | Orientation towards something |
| SUB | Subordination (of objects) |
| SUBR | Subordination (of relations) |
| SUBS | Subordination (of situations) |

Table 1: Set of MultiNet relations used in this paper

property, etc) and a set of binary semantic features (such as ANIMATE, HUMAN, ARTIFICIAL, and INFO(RMATION)) as well as relational means consisting of a predefined and well documented repertoire of more than hundred semantic relations including a set of semantic case roles (Helbig, 2006). (The MultiNet relations used in this paper are listed in Table 1.)

All HaGenLex entries are semantically classified by their ontological sorts and semantic features. Their semantic valency can be characterized by role frames consisting of semantic roles provided by MultiNet; the semantic features can be used for specifying selectional restrictions. (5) sketches the semantic description of the verb *beantworten* ('answer') in HaGenLex, plus its syntactic case frame.

- (5) *action* MENTAL –
AGT NP[nom] HUMAN +
ORNT NP[dat] HUMAN + optional
OBJ NP[acc] INFO +

In addition to simple role frames, HaGenLex allows more expressive semantic specifications in terms of general MultiNet expressions. As we will see in Section 4., this is especially useful in the case of prepositional verbs; for prepositional complements often occur with verbs of higher arity which in turn typically describe complex relationships between participants not expressible by semantic roles alone. In particular, this is the case for predicative complements.

Since many adjuncts (such as directional and duration adverbials) are restricted to certain types of verbs, the semantic relations (such as DIRCL and DUR) compatible with the semantics of a verb are explicitly listed in the entry. For instance, DIRCL and DUR are excluded for state and punctual verbs, respectively. The regular adjunct semantics of prepositions is treated within HaGenLex by context dependent interpretation rules (Hartrumpf, 1999; Hartrumpf et al., 2006).

4. Prepositional verbs in HaGenLex

The representation of prepositional verbs in HaGenLex follows the general outline sketched in the previous section. The preposition of a prepositional complement is part of the syntactic specification of the corresponding slot in the verb's subcategorization frame. For instance, the HaGen-

Lex entry for the prepositional verb *halten für* ('take for') is just a certain reading of the verb *halten*, whose valency frame is sketched by (6). (The frame is of course rather incomplete since the accusative NP can be replaced by infinitival and sentential complements and the PP may contain an adjectival phrase as well.)

- (6) x_1 : MEXP NP[nom]
 x_2 : OBJ NP[acc]
 x_3 : — PP[für NP[acc]]

The MultiNet relations MEXP (mental experiencer) and OBJ (neutral object) in (6) are the semantic roles of the arguments x_1 and x_2 , respectively. Argument x_3 lacks a direct role specification because there is no sensible interpretation of this argument as a participant of the situation described by the verb. In this case, the semantic relation between the verb concept and its arguments is characterized within HaGenLex by a MultiNet expression. For the above example, the expression in question has the form (7-a), with c standing for the concept of the given entry, whereas the role semantics specified by (6) is given by (7-b).

- (7) a. MCONT(c, y) & SUBR(y, SUB')
 & ARG1(y, x_2) & ARG2(y, x_3)
 b. MEXP(c, x_1) & OBJ(c, x_2)

Both expressions together indicate that the verb in question denotes a mental state of the experiencer x_1 whose mental content y (MCONT) is a (reified) subordination relation (SUB') between the arguments x_2 (ARG1) and x_3 (ARG2); in short: x_1 thinks/believes that x_2 is (an) x_3 . This is apparently nothing else than a formalization of the object predication expressed by *halten für* relativized to a belief of the subject.

At present, HaGenLex contains more than 2,000 prepositional verb entries, which have been manually compiled based on various German dictionaries. Due to its considerable size and its semantic orientation, this collection of prepositional verbs can be exploited as a resource for investigating regularities between verb semantics and prepositional case marking. Revealing such regularities is not only of interest for theoretical studies at the syntax-semantics interface but also highly useful for the hierarchical semantic organization of HaGenLex entries: knowledge about how the argument realizations of prepositional verbs (and verbs in general) depend on the semantic class of the verbs can be encoded into (default) inheritance rules, which in turn allow to reduce redundancies in the lexicon to a minimum. Such an inheritance-based approach strongly supports the consistency of the lexicon, simplifies its maintenance, and allows to generate default hypotheses for prepositional verbs not yet covered by HaGenLex.

It is worth mentioning that HaGenLex also contains prepositional adjectives (*wütend auf* / 'angry at') and nouns (*Wut auf* / 'anger at') and that in many of these cases the choice of the preposition is subject to derivational rules; for instance, deverbal nouns inherit prepositional complements from their base verbs (Osswald and Helbig, 2005), witness *Teilnahme an* ('participation in') vs. *teilnehmen an* ('participate in').

5. Semantic classes of prepositional verbs

A first analysis of the prepositional verbs in HaGenLex has led to about 150 semantic classes, which have been created by abstracting over the formal semantic specification of the entries, by taking into account the selected preposition, and by additional semantic grouping. Although at a preliminary stage, our classification allows to reveal certain subregularities concerning the interpretation of the preposition. The prepositional verb *halten für* ('take for'), for example, belongs to the semantic class of verbs with object predicatives defined by expression (7). Other members of this class are *ansehen als/für*, *auffassen als*, *betrachten als*, *einschätzen als*, *verstehen als*, etc ('regard as', 'look upon as', 'think of sth/sb as', etc). In this class, the preposition *für* ('for') is obviously less regular than *als* ('as').

It is tempting to compare our classes to Levin's verb classes (Levin, 1993).³ The above class of verbs is (unsurprisingly) covered by Levin's class 29 of verbs with predicative complements, more precisely, by the class 29.2 of *characterize* verbs. Similarly, take the verbs *ernennen zu*, *erklären zu*, *machen zu*, *krönen zu*, etc. This class of German prepositional verbs roughly corresponds to Levin's class 29.1 of *appoint* verbs, which is also a subclass of the verbs with predicative complements. Note that in English, the *appoint* verbs can occur with two postverbal NPs and undergo the *as* alternation, whereas in German, the preposition *zu* ('to') is obligatory. Another instructive example is given by Levin's class 13.4 of verbs of providing which resembles a fairly homogenous class of German prepositional verbs including *ausstatten mit*, *versehen mit*, *versorgen mit*, *bedenken mit*, and *beliefern mit*.

Despite these promising observations, some qualifying remarks concerning the comparability with Levin's classes may be in order. There is a natural limitation in coverage with respect to prepositional verbs since the underlying motivation for Levin's verb classes is that their members participate in diathesis alternations. In particular, Levin's classification neglects idiosyncratic selections of prepositions. Moreover, one has to keep in mind that the selection of prepositions is language specific to a certain degree. (8) shows two subregular classes of German prepositional verbs not covered by Levin's classes (in the sense that their English counterparts are not listed).

- (8) a. *bestehen auf*, *beharren auf*, *pochen auf*, *insistieren auf* ('insist on')
 b. *beruhen auf*, *basieren auf*, *füßen auf*, *sich stützen auf* ('be based on')

Other typical gaps include prepositional verbs like *warten auf* ('wait for'), where *wait* is only listed in the class 47.1 of *exist* verbs, and *glauben an* ('believe in'), where *believe* is only listed as a verb with predicative complement (on a par with *assume* and *declare*). All in all, a systematic comparison with Levin's classification may be useful to some

³In (Villavicencio, 2003), Levin's classes have been successfully employed for predicting verb particle constructions in English. Here, in contrast, we are not interested in the classes as a basis for productivity but in the prepositional complements occurring in these classes.

extent in order to stabilize our own classification but has its intrinsic limitations.

6. Automatic identification of prepositional verbs

Prepositional verbs are frequent and central lexical phenomena in German. As it is important to cover most cases in NLP applications and as manually compiled lists are incomplete, a corpus-based, automatic identification method was developed. Similar problems have been tackled by calculating significance lists of verb-preposition pairs based on different association measures like dice coefficient and mutual information and/or linguistic tests; see e.g. (Baldwin, 2005). As our lexicon already contains many prepositional verbs the top N verb-preposition pairs delivered by approaches based on significance lists are not very useful for extending the lexicon. Instead, a classifier is needed that works even for rare verbs with say 20 occurrences in the corpus.

Our classification method employs the memory-based learning tool TiMBL (Daelemans et al., 2004) and builds mainly on frequency features derived from PP attachments and interpretations during syntactico-semantic parsing of a corpus, here the German news corpus from CLEF. The features characterize how often a PP with a given preposition P (and a given syntactic case, if several cases are possible for P) is attached to the verb and how often a specific interpretation of P is used for this attachment. Further features are the separable prefix of the verb (if any) as specified in the lexicon and the frequency of clause correlates as determined by heuristic analysis of collocations in the corpus.

To evaluate the correctness of the classifier, a sample of 93 new, manually derived prepositional verbs that contain only one preposition in their valency frame and a sample of 93 new non-prepositional verbs were randomly determined. All verbs had to have only one reading and at least 5 occurrences in the corpus parses. The parser only knew the morphology of the verbs and assumed a transitive valency frame. The classifier had to assign the class no-pv if the verb has no prepositional verb reading and the class pv- P if the verb is a prepositional verb with a PP headed by the preposition P . The classifier achieved 49% correctness (in 10-fold cross-validation); for the easier task of classifying into two classes (pv and no-pv) correctness increased to 65%. The latter percentage is well above the most-frequent baseline, but clearly there is room for improvements. The corpus of 90 million words seems to be too small and too imbalanced for many verbs; the automatic separation of readings involving different prepositional verb readings (and non-prepositional verb readings) will reduce noise for the classifier. In an ablation study, different feature groups were omitted. This always resulted in a small correctness drop, but the samples must be enlarged to see which differences are statistically significant.

7. Future work

The next step will be a more systematic generation of semantic classes of prepositional verbs and a detailed analysis as to which prepositions correspond to which classes up to

which degree of regularity. Moreover, the automatic identification method must be investigated with larger corpora and more test and training examples.

8. References

- Timothy Baldwin. 2005. Looking for prepositional verbs in corpus data. In *Proceedings of the 2nd ACL-SIGSEM Workshop on the Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications*, Colchester, UK, April.
- Eva Breindl. 1989. *Präpositionalobjekte und Präpositionalobjektsätze im Deutschen*. Niemeyer, Tübingen.
- Walter Daelemans, Jakub Zavrel, Ko van der Sloot, and Antal van den Bosch. 2004. TiMBL: Tilburg memory based learner, version 5.1, reference guide. Technical Report 04-02, ILK.
- Sven Hartrumpf, Hermann Helbig, and Rainer Osswald. 2003. The semantically based computer lexicon HaGen-Lex – Structure and technological environment. *Traitement automatique des langues*, 44(2):81–105.
- Sven Hartrumpf, Hermann Helbig, and Rainer Osswald. 2006. Semantic interpretation of prepositions for NLP applications. In *Proceedings of the 3rd ACL-SIGSEM Workshop on Prepositions*, Trento, Italy, April.
- Sven Hartrumpf. 1999. Hybrid disambiguation of prepositional phrase attachment and interpretation. In *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP/VLC-99)*, pages 111–120, College Park, Maryland.
- Sven Hartrumpf. 2005. Question answering using sentence parsing and semantic network matching. In C. Peters, P. Clough, J. Gonzalo, G. J. F. Jones, M. Kluck, and B. Magnini, editors, *Multilingual Information Access for Text, Speech and Images: 5th Workshop of the Cross-Language Evaluation Forum, CLEF 2004*, volume 3491 of *Lecture Notes in Computer Science*, pages 512–521. Springer, Berlin.
- Hermann Helbig. 2006. *Knowledge Representation and the Semantics of Natural Language*. Springer, Berlin.
- Jacques Lerot. 1984. Die verbregierte Präposition in Präpositionalobjekten. In Werner Abraham, editor, *Satzglieder im Deutschen. Vorschläge zur syntaktischen, semantischen und pragmatischen Fundierung*, pages 261–291. Niemeyer, Tübingen.
- Beth Levin. 1993. *English Verb Classes and Alternations. A Preliminary Investigation*. University of Chicago Press, Chicago.
- Rainer Osswald and Hermann Helbig. 2005. Derivational semantics in HaGenLex – an interim report. In Stefan Langer and Daniel Schnorbusch, editors, *Semantik im Lexikon*, number 479 in *Tübinger Beiträge zur Linguistik*, pages 87–127. Narr, Tübingen.
- Aline Villavicencio. 2003. Verb-particle constructions and lexical resources. In *Proceedings of the ACL'03 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, pages 57–64, Sapporo, Japan.