Modeling Prefix and Particle Verbs in GermaNet

Christina Hoppermann
Department of Linguistics
University of Tübingen, Germany
`christina.hoppermann@uni-tuebingen.de`

Erhard Hinrichs
Department of Linguistics
University of Tübingen, Germany
`erhard.hinrichs@uni-tuebingen.de`

Abstract

Verbal word formation processes involving prefixes and particles are highly productive in Germanic languages. The compositional semantics of such prefix and particle verbs requires an in-depth analysis of the interdependence of their constituent parts for adequately representing these types of complex verbs in lexical-semantic networks. The present paper introduces modeling principles that account for such language-specific phenomena in the German wordnet *GermaNet* (Hamp and Feldweg, 1997; Henrich and Hinrichs, 2010), considering the continuum between full semantic transparency and highly lexicalized meanings as well as the semantic contribution of the prefix or particle to the meaning of the complex verb as a whole.

1 Introduction

This paper addresses the question how morphologically complex words can be adequately modeled in a wordnet and focuses on two classes of such verbs in German: (i) prefix verbs such as *entladen* 'unload' and *zerstören* 'destroy', and (ii) particle verbs such as *übergehen* 'bypass someone' and *losfahren* 'start driving'. Both types of verbs consist of a word-initial element followed by a host constituent. In the case of prefix verbs, the word-initial element is a bound morpheme such as *ent-* or *zer-*, while for particle verbs it is typically a free[1] morpheme such as *über* or *los*, which can be separated from its host

constituent depending on the clause type[2] that the particle verb appears in. The host constituent of a prefix or particle verb can either be a simplex (or: base verb) as in the examples above or a nominal or adjectival base as in *bedachen* 'put on a roof' or *erblassen* 'grow pale'.

A systematic treatment of prefix and particle verbs in a wordnet setting is desirable and significant for at least the following reasons:

1. The word formation processes involved in the two classes of verbs are highly productive for all Germanic languages.
2. The host constituent of a prefix or particle verb can be derived from an adjectival or nominal base. Therefore, an adequate treatment of these verbs has to include suitable morphological and semantic relations among the word classes involved. What makes such an account particularly interesting in a wordnet setting is the fact that nouns, verbs, and adjectives are the very word classes modeled in a wordnet.
3. The lexical semantics of prefix and particle verbs crucially involves a continuum between full semantic transparency on the one hand and highly lexicalized meanings on the other hand. Verbs such as *entladen* 'unload' and *losfahren* 'start driving' are fully transparent: Their semantics can be compositionally derived from the meanings of their parts, as the preverbs[3] *ent-* and *los* contribute the meanings of removal and initiation of the actions denoted by the simplex. By con-

---

[1] There are also occurrences of inseparable particles (e.g., *umfáhren* 'bypass sth.'), which are always unstressed (Dewell, 2011) in contrast to separable particles (e.g., *úmfahren* 'run into so.').

[2] Free particles are separated in verb-first and verb-second clauses. They are only inseparable as infinitives or in subordinate clauses in clause-final position (Dewell, 2011).

[3] The term *preverb* is used as cover term for both prefixes and particles (Booij and Kemenade, 2003; Los et al., 2012).

trast, *zerstören* 'destroy' and *übergehen* 'bypass someone' are highly lexicalized, since their base verbs do not make a semantically transparent contribution to the meaning of the expression as a whole in present-day language use.

The continuum of semantic transparency and lexicalization is not restricted to the lexical semantics of German prefix and particle verbs. It has also been observed with respect to other word formation processes such as nominal compounding and is, thus, of wider interest. A case in point is the contrast between *Hauswand* 'house wall', which is compositionally derived from its parts, and *Bahnhof* 'train station', which according to a simple composition of its constituent parts should denote a yard for trains, but which actually refers to a building.

What lexicalized meanings of morphologically complex words have in common is that the meaning of the complex word is not a hyponym of the meaning of its host or head constituent: *zerstören* 'destroy' is not a hyponym of *stören* 'disturb' and *Bahnhof* 'train station' is not a hyponym of its head constituent *Hof* 'yard'. This finding also indicates that a simple account that establishes a hyponymic relation between a particle or prefix verb and its host constituent will not provide a satisfactory account of the phenomena in question.

In the remainder of this paper, we will argue that an adequate account of prefix and particle verbs has to be based on the following two main considerations: (i) the distinction between semantic transparency and lexicalization, and (ii) the way in which the word-initial element contributes to the meaning of the complex verb as a whole. These considerations will lay the foundation for defining general principles of hypernym selection for modeling complex verbs in the German wordnet GermaNet (GN).

2 Prefix and Particle Inventory

The inventory of prefixes considered in the present study includes all *native* (Los et al., 2012) inseparable prefixes in German: *be-*, *ent-*, *er-*, *miss-*, *ver-*, and *zer-* (Eisenberg, 1998; Fleischer and Barz, 1995; Mungan, 1986; Stiebels, 1996). Prefixes with a Latinate origin, such as *de-*, *dis-*, *re-*, or *trans-* (Fleischer and Barz, 1995), are not within the scope of this study. In contrast to the closed set of prefixes, the particle inventory is more extensive and comprises particles such as

ab, an, auf, aus, bei, durch, ein, los, nach, über, um, unter, voll, vor, wider, wieder, and *zu* (Dewell, 2011; Eisenberg, 1998; Fleischer and Barz, 1995; Mungan, 1986; Stiebels, 1996).

The present analysis makes use of existing semantic classifications of preverbs (e.g., Augst, 1998; Dewell, 2011; Donalies, 2005; Fleischer and Barz, 1995; Mungan, 1986; Stiebels, 1996) and develops them further in a wordnet setting. At the time of writing the paper, GermaNet contains 94273 nouns, 12111 adjectives and 14333 verbs, of which 3040 are prefix verbs and 5171 are particle verbs. Out of the total number of prefix verbs, the frequency distribution is as follows: *ver-* (45%), *be-* (25%), *er-* (14%), *ent-* (11%), *zer-* (4%), and *miss-* (1%).

3 Modeling Complex Verbs in GN

Although it seems natural that the host constituent of a complex verb could be used as its hypernym, the subsequent analysis of the continuum between lexicalization and semantic transparency will demonstrate that this solution is not viable in all cases. Rather, the continuum requires a distinction between various classes, which differ in the selection and in the number of hypernyms.

3.1 Lexicalization

Highly lexicalized verbs are at one end of the continuum between full semantic transparency and highly lexicalized meanings. Both German prefix and particle verbs are subject to lexicalization. As pointed out in section 1, it is not possible to assign lexicalized prefix and particle verbs as hyponyms to their host constituents, since the semantics of the host constituent is no longer part of the meaning of the complex verb. As a consequence, this lack of semantic transparency requires finding an appropriate hypernym that takes account of the meaning of the lexicalized verb as a whole.

For the majority of lexicalized complex verbs, the semantic contribution of the word-initial element is not apparent so that the hypernym selection is to be conducted in the same way as for simplex verbs (Vossen, 2002). This is the case for particle verbs such as *aufnehmen* 'record', which is modeled as hyponym of the synset *aufzeichnen/mitschneiden* 'record', as it cannot be linked to its base verb *nehmen* 'take'.

Nevertheless, there are cases in which semantic classifications of the word-initial element can be used as indicator for choosing an appropriate hypernym. This mainly applies to lexicalized

complex verbs such as *zerstören* 'destroy', for which the meaning of the prefix *zer-* expresses 'destroying or damaging something' (Augst, 1998; Fleischer and Barz, 1995; Mungan, 1986). Thus, the stand-alone transparent semantics expressed by *zer-* is used as indicator for finding an appropriate hypernym (*"materielle Zustandsveränderung"* 'material change of state'), as a relation to the contemporary meaning of the simplex *stören* 'disturb' is not possible.

Although there is no conceptual relation to the simplex, the information on the individual word-internal components of the complex lexicalized verb is still available in GN in the form of a morpho-syntactic analysis, which separates the preverb from its simplex.

3.2 Semantic Transparency

In contrast to highly lexicalized verbs, semantically transparent complex verbs form the opposite end of the continuum. What these transparent verbs have in common is that there is always either a conceptual (i.e., hypernymic/hyponymic) or lexical (e.g., antonymic) relation to the respective base verb. However, there are two interrelated factors that vary along the continuum: (i) the degree of semantic transparency, and (ii) the semantic contribution of the word-initial element to the complex verb as a whole. On the basis of these two factors, three different classes can be distinguished and will be introduced below.

Class 1: Full Transparency, Light Contribution
The meaning of complex verbs within this class is fully transparent and is always represented by the respective simplex as the exclusive hypernym. This can be ascribed to the interaction of the preverb with its base verb: The semantics of the complex verb can be compositionally derived from the meaning of its parts. Thus, the simplex keeps its original meaning while the semantic contribution of the preverb is light, fulfilling one of the following two core functions: (a.) indication of a direction or (b.) intensification of the meaning denoted by the simplex.

a. Indicator of a Direction
The majority of German particle verbs indicate a direction. Particles are typically free morphemes that are frequently used as adpositions or adverbs without being part of a complex verb (Los et al., 2012). In combination with a verbal base, they usually retain the meanings they have in isolation (Brinton and Closs Traugott, 2005), such as path expressions (Dewell, 2011). Thereby, they only

add further directional information to the simplex, whose meaning remains highly transparent. As a consequence, the simplex always serves as the exclusive hypernym of the respective complex verb. This applies, e.g., to the verb *laden* 'load', which has, inter alia, the following directional hyponyms in GN: *aufladen* 'load up', *einladen* 'load into', and *umladen* 'reload'. These particle verbs all denote a specific direction by the particles *auf* ('up', i.e., upward movement), *ein* ('into', i.e., inward movement), and *um* (i.e., movement from one location to another), sharing the semantics of the corresponding adposition.

b. Intensifier
The second core function within class 1 refers to the use of word-initial elements as intensifiers of the meanings denoted by their host constituents. The word-initial element only has a light semantic contribution so that the entire complex verb remains highly transparent and is thus assigned as hyponym to its simplex. This is, e.g., the case for *verärgern* 'annoy', which has a hyponymic relation to its simplex *ärgern* 'tease'.

Class 2: Full Transparency/High Contribution
This class represents an exceptional case that is only valid for a limited number of complex verbs such as prefix verbs with *miss-* as negator of the meaning denoted by the simplex (Fleischer and Barz, 1995). Consequently, the simplex cannot function as hypernym, as shown below for the synset *missgönnen/neiden* 'begrudge'.
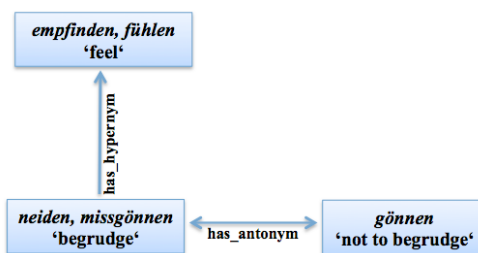

Figure 1. Conceptual and lexical relations.

Instead, another hypernym is chosen that takes account of the semantics of the complex verb (i.e., the synset *empfinden/fühlen* 'feel'). As for all transparent complex verbs, the relation to the simplex *gönnen* 'not to begrudge' is still present and is indicated by an antonymic relation.

The relation to the simplex can also be implicit, as some verbs with *ent-*, which refer to the inversion of an action denoted by the base (Fleischer and Barz, 1995), are antonyms of another complex verb sharing the same simplex. This is the case for *entladen* 'discharge', whose antonym is the particle verb *aufladen* 'charge'.

Class 3: Low Transparency/High Contribution
The third class displays the highest semantic contribution of the word-initial element while the meaning of the complex verb as a whole still remains transparent. Accounting for this predominant semantics requires treating verbs within this class both as hyponyms of their base verbs and of an additional hypernym, which expresses the prevailing semantic contribution of the preverb. The two hypernyms thus jointly account for the semantic contributions of preverb and base verb and lead to a more precise definition of the verb classes in question (cf. Bosch et al., 2008). This is, for instance, the case for one of the meanings of the prefix *ver-* 'make a mistake' (Mungan, 1986). This meaning is contained, e.g., in the reflexive prefix verbs represented in Figure 2 as a selection of hyponyms of both the *artificial concept*[4] *"falsch machen/Fehler machen"* 'make a mistake' and of each corresponding base verb: *sich versprechen* 'make a slip of the tongue', *sich verfahren* 'get lost while driving', and *sich verrechnen* 'miscalculate'.
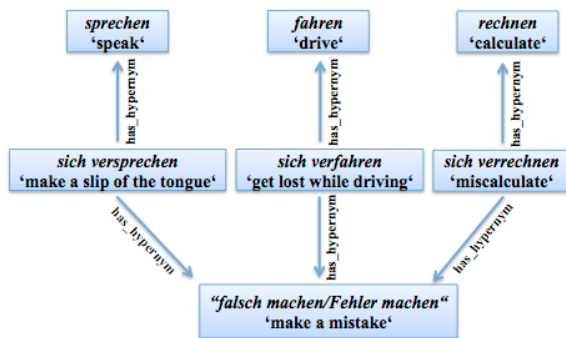

Figure 2. Selected verbs with two hypernyms.

Complex verbs in class 3 do not only include prefix but also particle verbs. Thus, the same approach can be applied to the verb *aufschrauben* 'unscrew', which has the following two hypernyms: its base verb *schrauben* 'screw' and the verb *öffnen* 'open'.

Another type of word-initial elements, which can be systematically modeled in class 3, represents preverbs indicating lexical aspect or *Aktionsart* 'manner of action'. On the one hand, this includes *ingressive* markers such as the prefix *er-* (e.g., *erklingen* 'start to sound') and the particle *los* (e.g., *loslaufen* 'start running'). On the other hand, the prefix *ver-* (e.g., *verglühen* 'burn out') as well as some word formations with the particles *auf* and *aus* characterize *egressive* verbs

(e.g., *aufessen* 'eat up', *auslesen* 'finish reading'), which express the termination or accomplishment (Vendler, 1957) of an action or state denoted by the base verb (Donalies, 2005; Stiebels, 1996; Helbig and Buscha, 1987). Both types of Aktionsart markers are modeled as hyponyms of two verb forms: of their respective simplex as well as of a verb denoting the particular aspectual meaning.

3.3 Principles of Hypernym Selection

The decision tree in Figure 3 summarizes the principles of hypernym selection, which specify the number of hypernyms to be chosen (i.e., one versus two), the synsets to be selected as hypernyms (i.e., simplex or not), and the use of further relations. Following the decision tree from top to bottom, it first needs to be determined whether the complex verb has a verbal, nominal, or adjectival base. If the base is verbal, the left branch of the tree needs to be passed through, deciding whether the complex verb is lexicalized or transparent. While lexicalized verbs only have one hypernym that does not equal the simplex, transparent verbs always have either a conceptual or (implicit) lexical relation to the simplex and are distinguished into three classes (cf. section 3.2).

The topmost right branch of the decision tree considers verbs with a nominal or adjectival base. As there is consequently no verbal base that could be used as hypernym for the respective complex verb, another verb form is to be chosen that expresses the semantics of the complex verb as a whole. Thus, the semantic contribution of the word-initial element is of prime importance for selecting an adequate hypernym. For instance, the meaning *to equip sth. with a/an <base noun>* is expressed by the prefixes *be-* and *ver-* as well as by the particle *um*. This can be represented by the synset *versehen/ausrüsten/ausstatten/ausstaffieren*. The hyponyms for this synset include the following entries, where the base noun is indicated in angle brackets: *be<dach>en* 'equip sth. with a <roof>', *ver<glas>en* 'enclose sth. with <glass>', and *um<mantel>n* 'surround with a <sheath>'. In order to account for the relation to the host constituent, a new derivational relation needs to be introduced that creates a connection to the base noun. This way, it is possible to tighten the wordnet by establishing relations that cross the line of word classes.

---

[4] In GermaNet, artificial concepts are not only used for filling lexical gaps. Similar to the verb classes defined by Levin (1993), they also serve the purpose of classifying semantically related concepts together by means of co-hyponymy.
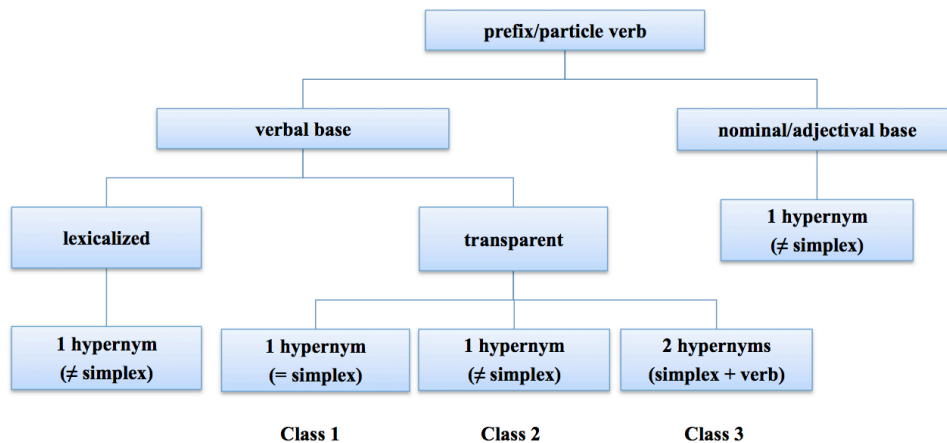
Figure 3. Principles of hypernym selection.

In the case of deadjectival verbs, the meaning *to become/to make <base adjective>* is often denoted by the preverb. An example provides the artificial concept *"materielle Zustandsveränderung"* 'material change of state', which is used as hypernym of deadjectival verbs such as *verflüssigen* 'liquefy', *verdicken* 'thicken', or *erwärmen* 'warm up'. If applicable, the causative meaning expressed by these preverbs is explicitly modeled by the *causes*[5] relation, which refers to the base adjective being the result of the process denoted by the complex verb (e.g., *<erblassen>* 'grow pale' *causes <blass>* 'pale').

4   Related Work

The use of multiple hypernyms for representing the compositional semantics of complex verbs can be identified in the Dutch wordnet project (Vossen et al., 1999). As in GN, the Dutch complex verb *opendraaien* 'open by turning' has two hypernyms (Vossen et al., 1999): its simplex *draaien* 'turn' and the verb *openmaken* 'open'.

In contrast, complex verbs in the Princeton WordNet (Fellbaum, 1999) only make use of one hypernym: The phrasal verb *to blow sth. up* is only a hyponym of the verb *expand*. The hypernym of its German equivalent *aufblasen* expresses the same semantics (i.e., *vergrößern* 'expand'), but the particle verb additionally has the simplex *blasen* 'blow' as second hypernym.

Regarding the different kinds of relations used in wordnets, Šojat et al. (2012) discuss the inclu-

sion of morphosemantic relations in the Croatian WordNet (CroWN). These relations e.g. group the meanings of preverbs into the class *location*, which indicates the directions of movements (e.g., *loc_bott_up* for upward movement).

Other wordnets dealing with (morpho-) semantic or derivational relations include the Polish wordnet (Maziarz et al., 2012) and the Czech wordnet (Bosch et al., 2008). They make fine-granular distinctions between various relation types, such as *inchoativity* and *derivationality*, which have also been addressed in this paper.

5   Conclusion

The present paper has established criteria for modeling morphologically complex verbs in the lexical-semantic network GermaNet, focusing on German prefix and particle verbs and accounting for their compositional semantics. Two main factors have been identified that provide the basis for their representation: (i) the continuum between full semantic transparency and highly lexicalized meanings, and (ii) the semantic contribution of the word-initial element to the meaning of the complex verb as a whole.

It has been demonstrated that a compositional analysis of the word-initial element and its host constituent enables a rule-based derivation of general modeling principles, which can systematically be applied in order to achieve a consistent depiction of complex verbs in the wordnet.

---

[5] The use of the *causes* relation is not restricted to complex verbs with an adjectival base. It is generally used for denoting resultative states for both simplex and complex verbs complying to the pattern *<causative transitive verb> causes <resultative intransitive verb>*, thereby signifying the *causative-inchoative alternation* (Levin, 1993), e.g., *zerbrechen* 'sb. breaks sth. to pieces' *causes zerbrechen* 'sth. breaks to pieces' (Bohnemeyer, 2007).

References

Gerhard Augst. 1998. *Wortfamilienwörterbuch der deutschen Gegenwartssprache*. Max Niemeyer Verlag, Tübingen.

Jürgen Bohnemeyer. 2007. Morphological Transparency and the argument structure of verbs of cutting and breaking. *Cognitive Linguistics*, 18(2):153-177.

Geert Booij and Ans van Kemenade. 2003. Preverbs: An Introduction. In Geert Booij and Jaap van Marle, editors, *Yearbook of Morphology 2003*. Kluwer, Dordrecht, pages 1-12.

Sonja Bosch, Christiane Fellbaum, and Karel Pala. 2008. Enhancing WordNets with Morphological Relations: A Case Study from Czech, English and Zulu. In *Proceedings of the Fourth Global WordNet Conference 2008*, Szeged, Hungary, pages 74-90.

Laurel J. Brinton and Elizabeth Closs Traugott. 2005. *Lexicalization and Language Change*. Cambridge University Press, Cambridge.

Robert D. Dewell. 2011. *The Meaning of Particle/Prefix Constructions in German*. John Benjamins Publishing Company, Amsterdam/Philadelphia.

Elke Donalies. 2005. *Die Wortbildung des Deutschen: Ein Überblick*. Gunter Narr Verlag, Tübingen.

Peter Eisenberg. 1998. *Grundriß der deutschen Grammatik: Das Wort*. Metzler, Stuttgart/Weimar.

Christiane Fellbaum. 1999. *WordNet: An Electronic Lexical Database*. MIT Press, Camebridge.

Wolfgang Fleischer and Irmhild Barz. 1995. *Wortbildung der deutschen Gegenwartssprache*. Niemeyer, Tübingen.

Birgit Hamp and Helmut Feldweg. 1997. GermaNet - a Lexical-Semantic Net for German. In *Proceedings of ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, Madrid.

Gerhard Helbig and Joachim Buscha. 1987. *Deutsche Grammatik: Ein Handbuch für den Ausländerunterricht*. Verlag Enzyklopädie, Leipzig.

Verena Henrich and Erhard Hinrichs. 2010. GernEdit – The GermaNet Editing Tool. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010)*, Valletta, Malta, pages 2228-2235.

Beth Levin. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. The University of Chicago Press, Chicago.

Bettelou Los, Corrien Blom, Geert Booij, Marion Elenbaas, and Ans van Kemenade. 2012. *Morphosyntactic Change: A Comparative Study of Particles and Prefixes*. Cambridge University Press, Cambridge.

Marek Maziarz, Maciej Piasecki, and Stan Szpakowicz. 2012. An Implementation of a System of Verb Relations in plWordNet 2.0. In *Proceedings of the Sixth Global WordNet Conference 2012*, Matsue, Japan, pages 181-188.

Güler Mungan. 1986. *Die semantische Interaktion zwischen dem präfigierenden Verbzusatz und dem Simplex bei deutschen Partikel- und Präfixverben*. Peter Lang, Frankfurt am Main.

Krešimir Šojat, Matea Srebačić, and Marko Tadić. 2012. Derivational and Semantic Relations of Croation Verbs. *Journal of Language Modelling*, 0(1):111-142.

Barbara Stiebels. 1996. *Lexikalische Argumente und Adjunkte: Zum semantischen Beitrag von verbalen Präfixen und Partikeln*. Akademie Verlag, Berlin.

Zeno Vendler. 1957. Verbs and Times. *The Philosophical Review*, 66(2):143-160.

Piek Vossen, Laura Bloksma, and Paul Boersma. 1999. *The Dutch WordNet*. Version 2, Final, July 12, 1999, University of Amsterdam.

Piek Vossen. 2002. *EuroWordNet: General Document*. Version 3, Final, July 1, 2002. http://hdl.handle.net/1871/11116.