Towards a dynamic lexicon:

Predicting the syntactic argument structure of complex verbs

Nadine Aldinger

University of Stuttgart
Institute for Natural Language Processing
Azenbergstr. 12
70174 Stuttgart
nadine.aldinger@ims.uni-stuttgart.de

Abstract

The presentation summarizes the results of a recent (2002) corpus-based study on the syntactic argument structure of German separable verbs, in particular the question of its predictability from the argument structure of their bases. The study shows that the argument structure effects of verb formation with separable preverbs are astonishingly regular, even if the meaning of the resulting complex verb itself is opaque. Alternation patterns can be formulated for single preverbs or, in some cases, preverb classes as well.

The results can be used in various ways in NLP and linguistic resource building: An implementation, e.g. in the form of LFG-style lexical rules, within a parser for German text will allow sensible guesses at argument structure frames for newly encountered complex verbs, even if the argument structure of the base verb is not known. Similarly, the patterns can be used for the acquisition of argument structure information for a (static) NLP lexicon.

When integrated with a static dictionary, the patterns thus allow to deal with the productive formation of complex verbs and its impact on syntactic argument structure.

Preverbs and prefixes

In German, new verbs can be productively created from other verbs by means of prefixation. There are two kinds of verbal 'prefixes': separable 'prefixes', which we will call *preverbs*, and inseparable prefixes. Preverbs (2) differ in their behaviour from 'real' (inseparable) prefixes (1): They bear the main accent of the complex verb (2a in contrast to 1a, where the main accent remains on the verbal base), and they are separated from the verbal base in verb-second and verb-first clauses, where the base moves from clause-final position to the front, leaving the preverb behind (2b):

(1) a. … *dass sie ihren Vater <u>besucht</u>.* [b@'zu:xt]
… that she her father be-seeks
"that she goes to see her father"
b. *Sie <u>besucht</u> ihren Vater.*
she be-seeks her father
"She goes to see her father"

(2) a. ... *dass sie ihren Anwalt <u>aufsucht</u>.* ['awfzu:xt]
... that she her lawyer on-seeks
"that she goes to see her lawyer"
b. *Sie <u>sucht</u> ihren Anwalt <u>auf</u>.*
she seeks her lawyer on
"She goes to see her lawyer"

Roughly speaking, the preverbs can be divided into two main classes:

- closed-class preverbs derived from prepositions (e.g. *auf-* "on, up", *durch-* "through") and adverbs (*hinein-* "into", *zurück-* "back"):
jdn. auf#suchen "to seek on s.b.", i.e. "to go to see s.b.", *etw. auf#schneiden* "to cut s.th. up", *(durch etw.) durch#kommen* "to come through s.th.", *etw. durch#lesen* "to read through s.th.";
etw. zurück#fordern "to demand s.th. back", *etw. zurück#stellen* "to put / set s.th. back".
A subset of these preverbs forms what is traditionally called the set of '(verbal) particles'; their complex verbs are called 'particle verbs'. These are e.g. the preverbs *auf-*, *aus-* ("out"), *ein-* ("in"), *los-* ("start to" – e.g. *los#lachen* "to start laughing" – note that this is only one of several readings of *los#*).

- open-class preverbs, i.e. deadjectival (*frei-* "free"), denominal (*berg-* "mountain") and deverbal (*stehen-* "stand") ones:
etw. frei#schaufeln "to shovel s.th. free", *frei#stehen* "to stand free, unprotected";
berg#steigen "to mountain-climb", i.e. "to climb mountains"; *stehen#bleiben* "to remain standing".
Deadjectival prefixes are frequently used to form resultative predicates, e.g. *etw. kochen* ("to cook s.th.") – *etw. weich#kochen* ("to cook s.th. soft").

New preverbs can be added to the second class by taking almost any noun, adjective or verb as basis – naturally, it is not possible to form new closed-class preverbs. In contrast, a typical denominal or deverbal preverb occurs with one or two base verbs only, whereas the 'verbal particles' of the first class abound in German and combine productively with (in many cases) hundreds of different verbs. To summarize, we observe productivity in both classes, but productivity of different kinds. In both cases it is of linguistical interest to capture syntactic and/or semantic regularities accompanying the formation of new complex verbs.

Historically, 'particle verbs' were probably formed out of adverb-verb contractions, which then underwent further semantic change.
In consequence, the meaning of some complex verbs can (still) be compositionally derived from the semantic contents of preverb and base verb today (like

auf#schneiden), others are more or less opaque in present-day German (like *auf#suchen*). In some of the transparent cases – which have been investigated in the past, e.g. by Stiebels (1996), Stiebels and Wunderlich (1994) – complex verb formation also has a clearly describable effect on the argument structure of the verb: only certain classes of base verbs (e.g. transitive verbs) combine with the particular preverb, and the semantic as well as syntactic argument structure of the complex verb can be derived from its compositional meaning and be predicted for each verb class.
Syntactic argument structure effects of complex verb formation have, to our knowledge, never been analyzed for a larger set of preverbs and/or verbs. Thus, it remained unclear whether and how semantic change can be accompanied by a change in syntactic argument structure as well. It has mostly been assumed more or less implicitly that non-compositional or lexicalized meaning of a complex verb corresponds with non-predictability of its syntactic argument structure. In consequence, the lexicalized cases were not considered further.

State of research

Because of their strong tendency to lexicalize, the class of so-called 'particle verbs' has been distinguished from the rest of preverb-verb constructions in the literature, e.g. by Fleischer&Barz (1995) who treat particle verbs as results of a derivational process but other preverb-verb constructions as compounds.
Olsen (1996) states that verbal particles, similar to verbal prefixes, saturate a PP argument position of the base verb. Along with prefixation, the argument structure of the verb is reordered as the former PP role-player is lifted to direct object position and the former direct object can only be expressed as an adjunct PP – a fact illustrated by the well-investigated locative alternation:

(3) a. Er *schreibt* Formeln *an die Tafel*.
b. Er *beschreibt* die Tafel (mit Formeln).
he be-writes the blackboard (with formulas)
"He writes formulas on the blackboard."

Verbal particles, in contrast, do not reorder the argument structure of the verb:

(4) a. Sie *graben* die Kartoffeln *aus der Erde*.
"They dig the potatoes out of the earth."
b. Sie *graben* die Kartoffeln *aus*.
"They dig the potatoes out."

Stiebels&Wunderlich (1994) and Stiebels (1996) provide analyses of particle verbs and their semantic argument structure in terms of decompositional semantics. They give a set of templates by which the semantic structures of preverb and base verb can be brought together to form the semantic structure of a complex verb.

Lüdeling (1999) argues that the division between 'particle verbs' and other preverb-verb constructions cannot be upheld on either structural or semantic grounds. Moreover, the class of 'particle verbs' seems to have never been really defined in the literature. Lüdeling proposes a phrasal analysis for separable complex verbs including most 'particle verbs' and excluding only denominal and deverbal preverbs. In particular, preverbs can function as
- resultative predicates (as in *weich#kochen* "to cook soft", *durch#kämpfen* "to fight through"),
- depictive predicates (as in *leer#stehen* "to stand empty")
- or adverbial modificators (as in *etw. an#lesen* "to read s.th. on", i.e. "to start reading s.th.").

The resultatives are further divided into ones that predicate over the direct object of the base verb (Lüdeling calls them "control resultatives") and ones that introduce their own argument ("raising resultatives").
Structural quirks of 'particle verbs' commonly used to demonstrate their non-phrasal nature, e.g. the inability of the preverb to occur in the *Vorfeld* of a sentence without the verbal base, are explained by Lüdeling as effects of the semantic opacity of most 'particle verbs'.
In the present study, Lüdeling's analysis and terminology were adopted to express the argument structure alternation patterns which we derived from a large sample of verbs and preverbs. This confirms Lüdeling's claim that her analysis can explain the same set of data as the traditional 'single-word' analyses of 'particle verbs' can, but with less effort and without having to introduce special grammatical features.

Setup of the study

We analyzed the syntactic argument structure of a large set of complex verbs from all of the preverb classes mentioned above in relation to their base verbs, based upon the following data:

i. a list of about 400 separable preverbs and their respective base verbs automatically extracted from a corpus of 200 million words of newspaper text (1987-93),
ii. the corpus itself,
iii. and an automatically generated large lexicon of verb subcategorization frames.[1]

The list (i) contains the number of attested verb forms with each preverb as well as the verbs themselves and the number of occurrences of each complex verb form. Parsers are usually not very good at identifying V2-separated verb forms of complex verbs like (2b) – after all, the great majority of the preverbs and base verbs also occur freely. Thus, extraction was limited to infinite forms, i.e. infinitives, *zu*-infinitives, and participles – which can never be syntactically separated – and homophonous, non-separated finite forms.

For the study, the preverbs were first grouped manually according to the part of speech of their freely occurring counterpart (see the classes above). A few preverbs that have more than one possible reading were classified twice, e.g. *auf-* as preposition (= "on, onto") and adverb (= "open"). Where common features of preverbs within a group could be detected, subgroups were formed (e.g. 'colour adjectives' or 'PPs ending in *-einander*').

[1] The preverb-verb list (i) was compiled by Arne Fitschen. The verb subcategorization frame lexicon (iii) is in use at the IMS and is based on work by Judith Eckle (Eckle-Kohler, 1999).

We ended up with eight major groups – the groups mentioned above and three additional ones:
- prepositional phrases (mostly pronominal adverbs like *darauf-* "on s.th."),
- preverbs without counterpart in present-day German (*dar-*, e.g. *jdm. etw. dar#reichen* "to administer s.th. to s.b."),
- and combinations of two preverbs or of a preverb and an inseparable prefix (*etw. wiedergut#machen* "to compensate s.th.").

The largest group in our system is formed by the deadjectival preverbs (116 preverbs total), followed by the de-adverbial ones (103). The de-prepositional preverbs (which include most of the 'particles') contribute by far the largest number of complex verbs per preverb (on average 238.5, compared to 84.9 for de-adverbial and only 4.3 for denominal preverbs – all numbers include false positives, which were especially frequent with denominal preverbs).

To investigate the argument structure effects of complex verb formation, a random sample of preverbs from most of the groups was taken. The size of each sample ranges from 2 to 9 per group, depending on group size and number of verbs found with the respective preverb. For each preverb, the argument structure of up to 50 verbs was compared to that of their base verbs.

The comparisons were performed manually but supported by extensive corpus research. Consulting the corpus was necessary e.g. if instances of an argument pattern provided by the verb subcategorization frame lexicon (iii) could not be readily identified or if the verb was not present in the lexicon at all. Corpus evidence was also added to check speakers' intuitions.

The results of the study were captured in semi-formal argument frame alternation patterns resembling the lexical rules of LFG. How the patterns were grouped depends on argument structure of the complex verb, argument structure of the base verb, and semantic content of the preverb.

The concept of 'argument frame alternation' is inspired by previous research like Levin's (1993) on argument structure alternations in English; in our case, the morphological change of the verb and further semantic change triggered by the preverb also has to be taken into account. A further difference is that our alternation patterns are not so much semantically motivated.

Results

The argument structure effects of verb formation with separable preverbs are astonishingly regular, even if the meaning of the complex verb itself is opaque. Alternation patterns can be formulated for single preverbs or, in some cases, whole preverb groups as well.

For each analyzed preverb, we found up to 5 alternation patterns, some of them differing in the kinds of base verbs they combine with. (5 is the maximum for the more polysemous closed-class preverbs and 'particles'; for the other preverb groups, the number is closer to 2, and the patterns differ less within the preverb groups.)

The choice of possible patterns for a complex verb is mostly determined by two properties of the syntactic argument structure of the base verb:
- transitivity (presence / absence of a direct object)
- directionality (indicated syntactically by presence / absence of a PP with either concrete or abstract / metaphorical directional meaning; in some cases even a prepositional object can take over this role).

Here is an example – the 'typical' pattern already mentioned where the preverb saturates a PP argument (or adjunct) position of the base verb, cf. (4):

Base verb argument structure:
<subj (pp-dir)> |
<subj obj (pp-dir)>
Argument structure effect of the preverb:
pp-dir → 0

This means that the rule applies to intransitive as well as transitive base verbs with an obligatory or optional directional PP, and that the preverb 'deletes' the directional PP in the argument frame of the verb. Many preverbs fall under this pattern with at least one of their readings, e.g. *gegen-* ("against", with 3 attested complex verbs), *bei-* ("by", 29), *durch-* ("through", 26), *zu-* ("to", 58), *drunter-* ("under s.th.", 2), *darauf-* ("on s.th.", 7), *dagegen-* ("against s.th.", 10), *entzwei-* ("apart", 3), *vorüber-* ("by, past; gone", 9), *los-* (26), *auf-* (22), *hinüber-* ("over there", 30), *hinab-* ("down there", 33), *frei-* ("free", 39). In some of these cases the directional PP can be reinstantiated by a new directional component with a certain pre-selected preposition, e.g.
jdn. (pp-dir) *schicken* "to send s.o. (to a place)" →
jdn. (pp:auf / pp:zu / pp:in) *hinab#schicken* "to send s.o. down (to a place)"
The base verb *schicken* allows for a rather arbitrary directional component; for the complex verb *hinab# schicken*, however, the choice of preposition seems to be limited to *auf*, *zu* or *in*. (An optional path argument can be instantiated as an NP with lexical accusative case for the complex verb, but not the base verb.)

In addition, many of the verbs in this pattern do not have a compositionally derivable meaning, but are lexicalized or at least used metaphorically (e.g. most verbs with *bei-*: *etw. bei#tragen* ("to carry s.th. by", i.e. "to contribute s.th."), *etw. bei#steuern* ("to steer s.th. by", i.e. "to contribute s.th."), *jdn. bei#ziehen* ("to pull s.o. by", i.e. "to consult s.o.").

The most common argument structure patterns can be described as the effect of secondary predication on verb phrases, often resultative constructions:

- A number of preverbs form result predicates with both intransitive and transitive base verbs, saturating their directional argument or adjunct position (often combined with further semantic change, resulting in opacity of meaning but still regular argument structure); this is typical of de-prepositional and de-adverbial preverbs with spatial meaning:
  rutschen "to slide" – *hinein#rutschen* "to slide into";
  nehmen "to take" – *hinein#nehmen* "to take into"
  If a preverb is derived from a preposition which can only assign dative case like *zu-* ("to") and *entgegen-* ("towards"), the complex verb may inherit the dative argument from the preposition:

lächeln "to smile" – *jdm. entgegen#lächeln, jdm. zu#lächeln* "to smile at s.b."
etw. in e. Richtung werfen "to throw s.th. in a certain direction" – *jdm. etw. entgegen#werfen, zu#werfen* "to throw s.th. at/to s.b."
There are some more argument structure effects in this area which are too complex to discuss here.

- Some patterns of de-prepositional, deadjectival and de-adverbial preverbs apply to transitive verbs, changing their result state (or introducing one, if the base verb is atelic) and/or the selectional restrictions for the direct object, usually making the complex verbs inherently reflexive:
fressen "to eat" – *sich (in etw.) hinein#fressen* "to eat o.s. into s.th."

- Some patterns of de-prepositional, deadjectival and de-adverbial preverbs apply to intransitive verbs only, introducing a new result state and adding a direct object to the argument frame:
kämpfen "to fight" – *sich (durch etw.) durch#kämpfen* "to fight o.s. through (s.th.)"

- Some deadjectival and de-adverbial preverbs form depictive secondary predicates or first-order subject or object predicatives. This is also typical of deverbal preverbs and verbs like *bleiben* "to remain" as base verbs:
liegen "to lie" – *frei#liegen* "to lie unprotected";
bleiben – stehen#bleiben, sitzen#bleiben, hängen# bleiben „to remain standing, sitting (i.e. to stay down), hanging (i.e. to get caught)" …

- A further major argument structure pattern found is adverbial modification:
etw. lesen "to read s.th." – *etw. an#lesen* "to start reading s.th."; *fahren* "to drive" – *los#fahren* "to start driving"
Adverbial modificators do not change the argument frame of the verb – but note that verbs like *los#fahren* are necessarily atelic and sound odd with a directional PP, which is probably due to the focus shift in the verb's meaning.

- Denominal preverbs develop from two more complex morphological processes: incorporation of a noun into the base verb (e.g. *(eine) Zeitung lesen* "to read (a) newspaper" → *zeitung#lesen*, to newspaper-read) or rederivation from a verb-derived and subsequently compounded noun (e.g. *landen* "to land" → *Landung* "landing" → *Notlandung* "emergency landing" → *not#landen* "to emergency-land").
Because of these rather special origins, denominal preverbs usually combine with very few base verbs – with the exception of some adverb-like preverbs like *probe-* "test" which combine more productively – see *probe#fahren* "to test-drive", *probe#fliegen* "to test-fly", *probe#sitzen* "to test-sit" etc.

The argument structure patterns of denominal preverbs can also be explained by these processes: a canonical direct object – or, in some cases like *berg#steigen* "to mountain-climb", a directional argument – is replaced by the preverb and thus deleted from the argument frame of the verb. *probe-*, on the other hand, is another example of an adverbial modificator.

Applications

The results of the study suggest two possible uses in NLP and linguistic resource building:
An implementation within a parser for German text will allow sensible guesses at argument structure frames for newly encountered complex verbs, even if the argument structure of the base verb is not known. If it is known, the number of possibilities is even further reduced.
Similarly, the patterns can be used for the acquisition of argument structure information for a (static) NLP lexicon.

When integrated with a static dictionary, the patterns thus allow to deal with the productive formation of complex verbs and its impact on syntactic argument structure.

References

Eckle-Kohler, J. (1999). Linguistisches Wissen zur automatischen Lexikon-Akquisition aus deutschen Textcorpora. Berlin: Logos.

Fleischer, W. & Barz, I. (1995). Wortbildung der deutschen Gegenwartssprache. Tübingen: Niemeyer.

Levin, B. (1993). English verb classes and alternations: A preliminary investigation. Chicago et al.: Chicago University Press.

Lüdeling, A. (1999). On particle verbs and similar constructions in German. Doctoral dissertation, University of Stuttgart.

Olsen, S. (1996). Partikelverben im deutsch-englischen Vergleich. In Lang, E. & Zifonun, G. (Eds.), Deutsch – typologisch (pp. 261-288), Berlin: de Gruyter.

Stiebels, B. & Wunderlich, D. (1994). Morphology feeds syntax: The case of particle verbs. Linguistics 32.6, 913-968.

Stiebels, B. (1996). Lexikalische Argumente und Adjunkte: Zum semantischen Beitrag von verbalen Präfixen und Partikeln. Berlin: Akademie Verlag.