

Predicting the Degree of Compositionality of German Particle Verbs based on Empirical Syntactic and Semantic Subcategorisation Transfer Patterns

Silvana Hartmann, Sabine Schulte im Walde, Hans Kamp
Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart
{hartmasa,schulte,hans}@ims.uni-stuttgart.de

German particle verbs are productive compositions of a base verb and a prefix particle, whose part-of-speech varies between open-class nouns, adjectives, and verbs, and closed-class prepositions and adverbs. Particle verb senses may be transparent (i.e. compositional) or opaque (i.e. non-compositional) with respect to their base verbs. For example, *abholen* ‘fetch’ is transparent with respect to its base verb *holen* ‘fetch’, *anfangen* ‘begin’ is opaque with respect to *fangen* ‘catch’, and *einsetzen* has both transparent (e.g. ‘insert’) and opaque (e.g. ‘begin’) verb senses with respect to *setzen* ‘put/sit (down)’.

Even though German particle verbs constitute a significant part of the verb lexicon, recent work is mostly devoted to theoretical investigations, such as Stiebels (1996); Lüdeling (2001). To our knowledge, so far only Aldinger (2004) and Schulte im Walde (2004, 2005) have addressed German particle verbs from a corpus-based perspective. The reason for this lack of attention is probably that German particle verbs represent a challenge to the automatic acquisition of lexical knowledge. For example, they show specific patterns of behaviour at the syntax-semantics interface, as they may change the behaviour of their base verbs; the particle can saturate or add an argument to the base verb’s argument structure, cf. example (1) from Lüdeling (2001).

- (1) Sie *lächelt*.
‘She smiles.’
*Sie *lächelt* [_{NP_{acc}} ihre Mutter].
‘She smiles her mother.’
Sie *lächelt* [_{NP_{acc}} ihre Mutter] *an*.
‘She smiles her mother at.’

Theoretical investigations (Stiebels, 1996) as well as corpus-based work (Aldinger, 2004) have demonstrated that such changes are quite regular, independent of whether a particle verb sense is compositional or not. This fact, however, raises a difficulty to standard approaches in data-intensive lexical semantics: they often rely on the long-standing linguistic hypothesis which asserts a tight connection between the lexical meaning of a word and its distributional behaviour (Harris, 1968; Pinker, 1989; Levin, 1993). For example, even though *anlächeln* in example (1) is strongly compositional, its subcategorisation properties differ from those of its base verbs; thus, automatic means that rely on distributional cues might not recognise that *anlächeln* is semantically related to its base verb.

This work continues on the question of how to exploit the empirical syntax-semantics interface for particle verb semantics. More specifically, we ask whether taking the changes in subcategorisation behaviour into account enables us to predict the degree of compositionality of particle verbs. For example, if we determine automatically that particle verbs with the particle *an-* and base verbs such as *gähnen*, *grinsen*, *lächeln*, *starren* are typically not used as intransitives (in contrast to their base verbs) but can subcategorise for an accusative object (also in contrast to their base verbs), we can induce a “subcategorisation transfer pattern” that encodes these changes. The transfer patterns then can be taken into account to determine the semantic relatedness between particle and base verbs, and the degree of compositionality of the particle verbs.

We present experiments that automatically establish subcategorisation transfer patterns, and check whether the regularity of a transfer pattern for a specific particle verb corresponds to its degree of compositionality. Our experiments are performed on two levels of subcategorisation information, (i) a frame-based, syntactic level, and (ii) an argument-based, semantic level. Experiment (i) is deemed to fail, given that previous work on German particle verbs already demonstrated that the syntactic variation is independent of the degree of compositionality. For example, the two particle verbs *abbestellen* and *aberkennen* undergo similar transfer patterns, but the former is transparent, the latter is opaque. This experiment can therefore be considered as a baseline experiment. Experiment (ii) then builds on experiment (i) and in addition considers the nominal fillers of verb arguments within the subcategorisation frames of particle and base verbs. Here, we expect that a greater degree of agreement on lexical argument heads corresponds to the degree of particle verb compositionality. For example, the nominal heads of direct objects of *abbestellen* should be more similar to those of *bestellen* than the direct objects of *aberkennen* in comparison to those of *erkennen*.

Both experiments use corpus-based information on verb subcategorisation from Schulte im Walde (2002), induce an empirical description of the subcategorisation transfer patterns, and calculate the divergence of specific particle verbs from the average transfer pattern per particle by a standard similarity measure, the *skew divergence* (Lee, 2001). The experiments are performed for a random choice of 99 German particle verbs, which is balanced across 10 corpus frequency bands, and 11 particles. The divergence scores of the two experiments are compared with a gold standard of human judgements on particle verb compositionality, where the 99 verbs have been judged by 4 humans,¹ following McCarthy *et al.* (2003).

We believe that next to the actual results of our two experiments it is important to gain more insight into what corpus-based information offers to an automatic approach towards particle verb compositionality. A focus of our work is thus on a detailed manual description and assessment of the empirical data, addressing the role of individual particles, the ambiguity of particles, the ambiguity of particle verbs, and frequency effects, as those aspects seem crucial to us for further steps in empirical particle verb compositionality.

¹Kendall Concordance $W = 0.75$; $\chi_2 = 274.65$, $df = 91$, significant correlation, $p \leq 0.000001$

References

- Nadine Aldinger. Towards a Dynamic Lexicon: Predicting the Syntactic Argument Structure of Complex Verbs. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, Lisbon, Portugal, 2004.
- Zellig Harris. Distributional Structure. In Jerold J. Katz, editor, *The Philosophy of Linguistics*, Oxford Readings in Philosophy, pages 26–47. Oxford University Press, 1968.
- Lillian Lee. On the Effectiveness of the Skew Divergence for Statistical Language Analysis. *Artificial Intelligence and Statistics*, pages 65–72, 2001.
- Beth Levin. *English Verb Classes and Alternations*. The University of Chicago Press, 1993.
- Anke Lüdeling. *On German Particle Verbs and Similar Constructions in German*. Dissertations in Linguistics. CSLI Publications, Stanford, CA, 2001.
- Diana McCarthy, Bill Keller, and John Carroll. Detecting a Continuum of Compositionality in Phrasal Verbs. In *Proceedings of the ACL-SIGLEX Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, Sapporo, Japan, 2003.
- Steven Pinker. *Learnability and Cognition: The Acquisition of Argument Structure*. MIT Press, Cambridge, MA, 1989.
- Sabine Schulte im Walde. A Subcategorisation Lexicon for German Verbs induced from a Lexicalised PCFG. In *Proceedings of the 3rd Conference on Language Resources and Evaluation*, volume IV, pages 1351–1357, Las Palmas de Gran Canaria, Spain, 2002.
- Sabine Schulte im Walde. Identification, Quantitative Description, and Preliminary Distributional Analysis of German Particle Verbs. In *Proceedings of the COLING Workshop on Enhancing and Using Electronic Dictionaries*, pages 85–88, Geneva, Switzerland, 2004.
- Sabine Schulte im Walde. Exploring Features to Identify Semantic Nearest Neighbours: A Case Study on German Particle Verbs. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing*, pages 608–614, Borovets, Bulgaria, 2005.
- Barbara Stiebels. *Lexikalische Argumente und Adjunkte. Zum semantischen Beitrag von verbalen Präfixen und Partikeln*. Akademie Verlag, Berlin, 1996.