Exploiting Fine-grained Syntactic Transfer Features to Predict the Compositionality of German Particle Verbs

Stefan Bott and Sabine Schulte im Walde
Institut für Maschinelle Sprachverabeitung, Universität Stuttgart
{stefan.bott,schulte}@ims.uni-stuttgart.de

Abstract

This article presents a distributional approach to predict the compositionality of German particle verbs by modelling changes in syntactic argument structure. We justify the experiments on theoretical grounds and employ GermaNet, Topic Models and Singular Value Decomposition for generalization, to compensate for data sparseness. Evaluating against three human-rated gold standards, our fine-grained syntactic approach is able to predict the level of compositionality of the particle verbs but is nevertheless inferior to a coarse-grained bag-of-words approach.

1 Introduction

In German, particle verbs (PVs) such as *aufessen* (*to eat up*) are a frequent and productive type of multi-word expression composed of a base verb (BV) and a prefix particle. We are interested in predicting the degrees of compositionality of German PVs, which exhibit a varying degree of compositionality with respect to their base verbs, as illustrated in (1) vs. (2). The meaning of the highly compositional PV *nachdrucken* (*to reprint*) is closely related to its BV *drucken* (*to print*), while the PV *nachgeben* (*to give in*) has little in common with the BV *geben* (*to give*).

(1) *Der Verlag druckte das Buch nach.*
The publisher printed the book again-PRT.
'The publisher reprinted the book.'

(2) *Peter gab ihrer Bitte nach.*
Peter gave her request in-PRT.
'Peter gave in to her request.'

In previous work we demonstrated that the compositionality level of PVs can be predicted by using a simple Word Space Model which represents local word contexts as a bag-of-words extracted from a symmetric window around the target PV instances (Bott and Schulte im Walde, 2014a). The approach worked well because compositional PVs tend to co-occur locally with the same words as their corresponding base verbs.

The compositionality of German PVs is, however, also influenced by syntactic factors. While semantically similar verbs in general tend to have similar subcategorization frames (Merlo and Stevenson, 2001; Joanis et al., 2008), PV-BV pairs may differ in their syntactic properties, even if the PV is highly compositional. We refer to this linguistic phenomenon as "syntactic transfer problem". We understand transfers as regular changes in subcategorization frames of PVs by transfer, incorporation or addition of complements in comparison to the BV (Stiebels, 1996; Lüdeling, 2001). For example, the semantic role expressed by the subject of the BV *leuchten* in (3) is "transferred" to an instrumental PP of the highly compositional PV *anleuchten* in (4). In addition, the patient of *anleuchten* (i.e., the direct object) has no correspondence for *leuchten*. We call this a case of argument extension. The opposite case (i.e., a PV does not realize a semantic role used by its BV) is called argument incorporation.

Proceedings of the 11th International Conference on Computational Semantics, pages 34–39,
London, UK, April 15-17 2015.

(3) *Die Lampe leuchtet.*
'The lamp-SBJ shines.'

(4) *Peter leuchtet das Bild mit der Lampe an.*
Peter-SBJ shines the picture-OBJ$_{ACC}$ with the lamp-PP$_{DAT}$ at-PRT.
'The man beams at the picture with the lamp.'

Our hypothesis is that the degree of reliability of the prediction of such syntactic transfers represents an indirect indicator of semantic transparency: If many of the complements of a PV correspond to a complement of its BV, the PV is regarded as highly compositional, even if the PV complements are not realized as the same syntactic argument types. Conversely, if few of the PV complements correspond to BV complements, this is an indicator of low compositionality.

To explore our hypothesis, we rely on the distributional similarity between PV–BV complements, to model argument correspondences in order to predict PV compositionality. For example, identifying strong distributional similarity between the instrumental PPs of *anleuchten* and the subjects of *leuchten* (see examples (3) and (4) above) would allow us to predict strong PV compositionality, even though the distributional similarity of identical complement types (e.g., the subjects) is low.

Our novel approach exploits fine-grained syntactic transfer information which is not accessible within a window-based distributional approach, while it should preserve an essential part of the information contained in context windows, since the head nouns within subcategorization frames typically appear in the local context. To compensate for the inevitable data sparseness, we employ the lexical taxonomy *GermaNet* (Hamp and Feldweg, 1997), *Topic Models* (Blei et al., 2003) and *Singular Value Decomposition (SVD)* to generalize over individual complement heads. All of them have proven effective in other distributional semantics tasks (Joanis et al. (2008), Ó Séaghdha (2010), Guo and Diab (2011), Bullinaria and Levy (2012), among others).

The variants of our fine-grained syntactic approach are able to predict PV compositionality, but even though our model is (a) theoretically well-grounded, (b) supported by sophisticated generalization methods and (c) successful, a conceptually much simpler bag-of-words approach to the distributional representation of PVs cannot be outperformed.

2 Related Work & Motivation

The problem of predicting degrees of PV compositionality is not new and has been addressed previously, mainly for English (Baldwin et al., 2003; McCarthy et al., 2003; Bannard, 2005). For German, Schulte im Walde (2005) explored salient features at the syntax-semantics interface that determined the semantic nearest neighbors of German PVs. Relying on the insights of this study, Kühner and Schulte im Walde (2010) used unsupervised clustering to determine the degree of compositionality of German PVs. They hypothesized that compositional PVs tend to occur more often in the same clusters with their corresponding BVs than opaque PVs. Their approach relied on nominal complement heads in two modes, (1) with and (2) without explicit reference to the syntactic functions. The explicit incorporation of syntactic information (mode 1) yielded less satisfactory results, since a given subcategorization slot for a PV complement does not necessarily correspond to the same semantic type of complement slot for the BV, thus putting the syntactic transfer problem in evidence, again.

In our previous approach, we relied on word window information with no access to syntactic information (Bott and Schulte im Walde, 2014a), with a focus on PV frequency and ambiguity. For the current work, we started out from the idea that syntactic information should be more useful than window information if the distributional similarity is measured over individual salient slot correspondences rather than across all slots as in earlier approaches. Therefore, a pre-processing step automatically determines the distributionally most similar complement slot pairs for a given PV-BV pair and their subcategorization frames, in order to measure the similarity between PVs and their BVs. In Bott and Schulte im Walde (2014b) we already showed that the prediction of syntactic transfers with distributional methods is feasible. In the present work we exploit the prediction of syntactic transfer patterns as an intermediate step for the assessment of compositionality levels.

Through dividing up the local context among different subcategorization slots we expected a problem of data sparseness more severe than for window-based approaches which represent all the context words in the same vector and are less likely to result in sparse representations. For this reason, we apply a series of generalization techniques utilizing a lexical taxonomy and Topic Models, as well as SVD as a dimensionality reduction technique.

3 Experiments

3.1 Syntactic Slot Correspondence

In order to build a model of syntactic transfer to predict PV compositionality, a pre-processing step determined a measure of *syntactic slot correspondence*. We selected the 5 most common subcategorization frames of each PV and each BV induced from dependency parses of the German web corpus *SdeWaC* containing approx. 880 million words (Bohnet, 2010; Faaß and Eckart, 2013). From these 5 most probable verb frames, we used all noun and prepositional phrase complement slots with nominal heads, except for adjuncts. Each PV slot was compared against each BV slot, by measuring the cosine between the vectors containing the complement heads as dimensions, and head counts[1] within the slots as values. E.g. (see examples (3) and (4)), we found the nouns *Licht* and *Taschenlampe* (among others) both as instrumental PP (DAT-mit)[2] of *anleuchten* and as subject (SBJ) of *leuchten*, and the cosine of this slot correspondence over all nouns was 0.9898.

3.2 Syntactic Transfer Strength

In order to use the syntactic slot correspondence scores to predict the degree of PV-BV compositionality, we first selected the best matching BV slot for each PV complement slot, as suggested in Bott and Schulte im Walde (2014b) and then calculated the average score over these best matches across all PV slots. This average value is considered as a confidence measure for the assumption that the PV–BV complement slots correspond to each other and realize the same semantic roles. Regarding our hypothesis, we rely on the average cosine value to predict the degree of PV compositionality.

To account for possible null correspondences in argument incorporation and argument extension cases, we applied a variable threshold on the cosine distance ($t = 0.1/0.2/0.3$, and $t = 0$ referring to no threshold). If the best matching BV complement slot of a PV complement slot had a cosine below this threshold, it was not taken into account.

3.3 Generalization

The major problem of this approach is data sparseness. We thus applied three generalization techniques to the head nouns:

1. ***GermaNet (GN)*** is the German version of WordNet (Hamp and Feldweg, 1997). We use the n^{th} topmost taxonomy levels in the GermaNet hierarchy as generalizations of head nouns. In the case of multiple inheritance the counts of a subordinate node are distributed over the superordinated nodes.
2. ***LDA***: We use the MALLET tool (McCallum, 2002) to create LDA topic generalizations for the head nouns, in a similar way as Ó Séaghdha (2010). While LDA is usually applied over text documents, we consider as document the set of noun heads in the same subcategorization slot.
3. ***SVD***: We use the DISSECT tool (Dinu et al., 2013) to apply dimensionality reduction to the vectors of complement head nouns.

[1] We used *Local Mutual Information (LMI)* (Evert, 2005).

[2] PP slots are marked with case and preposition.

3.4 Evaluation

We evaluated our models against three gold standards (GS). Each of them contains PVs across different particles and was annotated by humans for the degree of compositionality:
GS1: A gold standard collected by Hartmann (2008), consisting of 99 randomly selected PVs across 11 particles, balanced over 8 frequency ranges and judged by 4 experts on a scale from 0 to 10.
GS2: A gold standard of 354 randomly selected PVs across the same 11 verb particles, balanced over 3 frequency ranges while taking the frequencies from three corpora into account. We collected ratings with Amazon Mechanical Turk on a scale from 1 to 7.[3]
GS3: A subset of 150 PVs from GS2, after removing the most frequent and infrequent PVs as well as prefix verbs, because we concentrate on the empirically challenging separable PVs. [4]

In the actual evaluation, we compared the rankings of the system-derived PV–BV cosine scores against the human rankings, using Spearman's ρ (Siegel and Castellan, 1988).

4 Results & Discussion

In the following, we describe and discuss our results across methods, across cosine threshold values, and across gold standards. Figure 1 presents the ρ values for the threshold $t = 0.3$ (which in the majority of cases outperformed the other threshold levels) and across gold standards. Across all syntactic models, we obtained the best results when evaluating against GS3. This was expected given that this gold standard excludes prefix verbs and very infrequent and very frequent PVs which are hard to assess in terms of PV-BV compositionality: Infrequent verbs are highly affected by data sparseness; highly frequent verbs have a tendency towards lexical ambiguity(Bott and Schulte im Walde, 2014a). In the same vein, the particularly low results[5] obtained with GS1 can be explained by its large proportion of low-frequent and high-frequent PVs.

Figure 1 also shows that the syntactic approach (a) provides poor results when it relies on raw frequency counts or LMI values; (b) is better for GermaNet level 2 than level 1 and the levels >2;[6] (c) provides the best results with SVD and (d) relying on LDA is most robust against low and high frequency and obtains the best results for GS2, which are however outperformed by GermaNet and SVD models.

Finally, Figure 1 demonstrates that, against our expectations, our new approach was not able to perform better than our previous bag-of-words models extracted from local windows. Even if the window models are conceptually simple, they seem to carry a lot of salient information which is also more robust against low frequency and ambiguity (obtaining better results for GS1 vs. GS2 vs. GS3). The virtues of bag-of-words models can apparently not even be outperformed by generalizing over nouns or by dimensionality reduction. Hoping that our novel syntactic information is in some way complementary to window information, we carried out an additional experiment where we computed a weighted average of the cosine values obtained from both feature types. Comparing the combined predictions with the human rankings, the system was however still beaten by window information alone.

Figure 2 provides a deeper look into our results across thresholds, now focusing on GS3. The plot shows that for the most successful generalization models (GN level 2 and SVD), the results improve with an increasing threshold. Excluding subcategorization complement slots of PVs that do not correspond to a distributionally similar subcategorization slot of its BV thus seems to support the identification of PV syntactic argument changes. This is an interesting theoretical result because it corroborates the influence of argument incorporation and argument extensions.

Error analysis in combination with theoretical considerations revealed that, overall, data sparseness appears to remain a central problem. The representation of each verb as a series of vectors, one for each

[3] `https://www.mturk.com`

[4] We do not treat non-separable prefix verbs like *ver|lieben*, but note that a series of verbs, such as *um|fahren* do exists as PVs and prefix verbs, with different readings.

[5] Negative ρ values are omitted in the plot.

[6] GN results for levels >3 are omitted for space reasons.

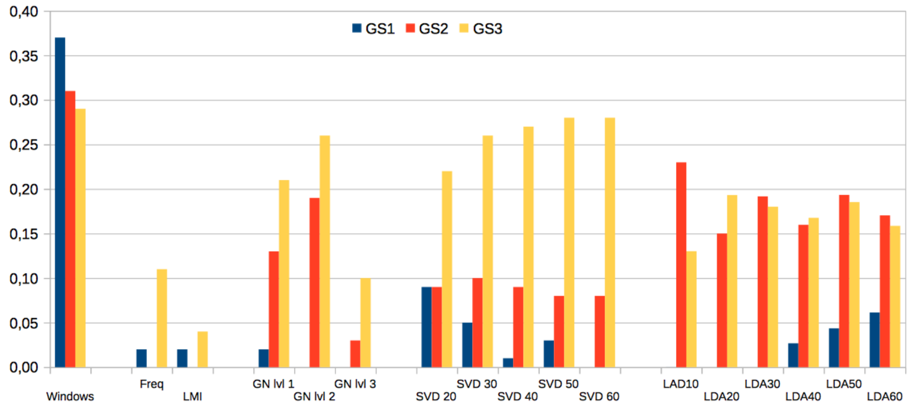

Figure 1: Results across gold standards, for t=0.3.

Figure 2: Results across thresholds, for GS3.

subcategorization complement, splits up the mass of counts in comparison to a verb window vector. Our syntax-based approach may need much more data to perform on an equal level as the window approach.

5 Conclusions

In this article we described a novel distributional approach to predict the degree of compositionality of German particle verbs. Our approach exploited syntactic information and involved a direct modeling of the syntactic transfer phenomenon. Relying on various gold standards, and varying complement similarity thresholds and generalization methods, we successfully predicted PV compositionality. Threshold variation indicated that we indeed capture PV-BV syntactic argument changes, and generalization by GermaNet high taxonomy levels and SVD helped with the apparent data sparseness. Nevertheless, information provided by context windows outperforms our fine-grained syntactic approach.

Acknowledgments

The research was supported by the DFG Research Grant SCHU 2580/2 (Stefan Bott) and the DFG Heisenberg Fellowship SCHU-2580/1 (Sabine Schulte im Walde).

References

Baldwin, T., C. Bannard, T. Tanaka, and D. Widdows (2003). An Empirical Model of Multiword Expression Decomposability. In *Proceedings of the ACL Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, Sapporo, Japan, pp. 89–96.

Bannard, C. (2005). Learning about the Meaning of Verb–Particle Constructions from Corpora. *Computer Speech and Language 19*, 467–478.

Blei, D., A. Ng, and M. Jordan (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research 3*, 993–1022.

Bohnet, B. (2010). Top Accuracy and Fast Dependency Parsing is not a Contradiction. In *Proceedings of the 23rd International Conference on Computational Linguistics*, Beijing, China, pp. 89–97.

Bott, S. and S. Schulte im Walde (2014a). Optimizing a Distributional Semantic Model for the Prediction of German Particle Verb Compositionality. In *Proceedings of the 9th International Conference on Language Resources and Evaluation*, Reykjavik, Iceland, pp. 509–516.

Bott, S. and S. Schulte im Walde (2014b). Syntactic Transfer Patterns of German Particle Verbs and their Impact on Lexical Semantics. In *Proceedings of the 3rd Joint Conference on Lexical and Computational Semantics*, Dublin, Ireland, pp. 182–192.

Bullinaria, J. A. and J. P. Levy (2012). Extracting Semantic Representations from Word Co-Occurrence Statistics: Stop-Lists, Stemming, and SVD. *Behavior Research Methods 44*, 890–907.

Dinu, G., N. The Pham, and M. Baroni (2013). DISSECT – DIStributional SEmantics Composition Toolkit. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, Sofia, Bulgaria.

Evert, S. (2005). *The Statistics of Word Co-Occurrences: Word Pairs and Collocations*. Ph. D. thesis, IMS, Universität Stuttgart.

Faaß, G. and K. Eckart (2013). SdeWaC – a Corpus of Parsable Sentences from the Web. In *Proceedings of the International Conference of the German Society for Computational Linguistics and Language Technology*, Darmstadt, Germany, pp. 61–68.

Guo, W. and M. Diab (2011). Semantic Topic Models: CombiningWord Distributional Statistics and Dictionary Definitions. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Edinburgh, UK, pp. 552–561.

Hamp, B. and H. Feldweg (1997). GermaNet – A Lexical-Semantic Net for German. In *Proceedings of the ACL Workshop on Automatic Information Extraction and Building Lexical Semantic Resources for NLP Applications*, Madrid, Spain, pp. 9–15.

Hartmann, S. (2008). Einfluss syntaktischer und semantischer Subkategorisierung auf die Kompositionalität von Partikelverben. Studienarbeit. IMS, Universität Stuttgart.

Joanis, E., S. Stevenson, and D. James (2008). A General Feature Space for Automatic Verb Classification. *Natural Language Engineering 14*(3), 337–367.

Kühner, N. and S. Schulte im Walde (2010). Determining the Degree of Compositionality of German Particle Verbs by Clustering Approaches. In *Proceedings of the 10th Conference on Natural Language Processing*, Saarbrücken, Germany, pp. 47–56.

Lüdeling, A. (2001). *On German Particle Verbs and Similar Constructions in German*. CSLI.

McCallum, A. K. (2002). MALLET: A Machine Learning for Language Toolkit.

McCarthy, D., B. Keller, and J. Carroll (2003). Detecting a Continuum of Compositionality in Phrasal Verbs. In *Proceedings of the ACL Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, Sapporo, Japan, pp. 73–80.

Merlo, P. and S. Stevenson (2001). Automatic Verb Classification Based on Statistical Distributions of Argument Structure. *Computational Linguistics 27*(3), 373–408.

Ó Séaghdha, D. (2010). Latent Variable Models of Selectional Preference. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Uppsala, Sweden, pp. 435–444.

Schulte im Walde, S. (2005). 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*, Borovets, Bulgaria, pp. 608–614.

Siegel, S. and N. J. Castellan (1988). *Nonparametric Statistics for the Behavioral Sciences*. Boston, MA: McGraw-Hill.

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