Morphological Priming in German: The Word is Not Enough (Or Is It?)

Sebastian Padó[*] **Britta D. Zeller**[*] **Jan Šnajder**[†]
[*]Stuttgart University, Institut für maschinelle Sprachverarbeitung
Pfaffenwaldring 5b, 70569 Stuttgart, Germany
[†]University of Zagreb, Faculty of Electrical Engineering and Computing
Unska 3, 10000 Zagreb, Croatia
{pado,zeller}@ims.uni-stuttgart.de jan.snajder@fer.hr

Abstract

Studies across multiple languages show that overt morphological priming leads to a speed-up only for transparent derivations but not for opaque derivations. However, in a recent experiment for German, Smolka et al. (2014) show comparable speed-ups for transparent and opaque derivations, and conclude that German behaves unlike other Indo-European languages and organizes its mental lexicon by morphemes rather than lemmas. In this paper we present a computational analysis of the German results. A distributional similarity model, extended with knowledge about morphological families and without any notion of morphemes, is able to account for all main findings of Smolka et al. We believe that this puts into question the call for German-specific mechanisms. Instead, our model suggests that cross-lingual differences between morphological systems underlie the experimentally observed differences.

1 Semantic and Morphological Priming

Priming is a general property of human language processing: it refers to the speed-up effect that a stimulus can have on subsequent processing (Meyer and Schvaneveldt, 1971). This effect is assumed to result from an activation (in a broad sense) of mental representations, and priming is a popular method to investigate properties of the mental lexicon. The original study by Meyer and Schvaneveldt established *lexical priming* (*nurse* → *doctor*), but priming effects have also been identified on other linguistic levels, such as syntactic priming (Bock, 1986) and morphological priming (Kempley and Morton, 1982).

A recent study by Smolka et al. (2014) investigated overt *morphological priming* on prefix verbs in German, where the base verb and derived verb can be semantically related (transparent derivation: *schließen – abschließen (close – lock)*) or not (opaque derivation: *führen – verführen (lead – seduce)*). Experiment 1, an overt visual priming experiment (300 ms SOA) involved 40 six-tuples that paired up a base verb with five prefix verbs of five prime types (see Figure 1). The verbs were normed carefully, e.g., for association, to exclude confounding factors. The authors reported three main findings: (a), no priming for Form and Unrelated; (b), no priming for Synonymy; (c), significant priming of the same strength for both Transparent and Opaque Derivation.

These findings suggest that morphological priming on German prefix verbs use a mechanism that is different from lexical priming, which assumes that the strength of the semantic relatedness is the main determinant of priming – i.e., lexical priming would predict finding (a), but neither (b) nor (c). The findings by Smolka et al. are also at odds with overt priming patterns found in similar experimental setups for other languages such as French (Meunier and Longtin, 2007) and Dutch (Schriefers et al., 1991), where patterns were found to be indeed consistent with lexical priming. Smolka et al. (2014) interpret this divergence as evidence for a German *Sonderweg*: the typological properties of German (separable prefixes, morphological richness, many opaque derivations) are taken to suggest a *morpheme*-based organization of the mental lexicon more similar to Semitic languages like Hebrew or Arabic than to other Indo-European languages.

Our paper investigates this claim on the computational level. We present a simple model of corpus-based word similarity, extended with a database of *morphological families*, that is able to predict the three main findings by Smolka et al. outlined above. The ability of the model to do so, even though it operates completely at the word level without any notion of morphemes, may put into question Smolka

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

Target	*binden (bind)*
1 Transparent Derivation	*zubinden (tie)*
2 Opaque Derivation	*entbinden (give birth)*
3 Synonym	*zuschnüren (tie)*
4 Form	*abbilden (depict)*
5 Unrelated	*abholzen (log)*

Figure 1: Smolka et al. (2014)'s five prime types

et al.'s call for novel morpheme-level mechanisms for German.

2 Modeling Priming

We model the priming effects shown in Smolka et al. by combining two computational information sources: A distributional semantic model, and a derivational lexicon.

Distributional Semantics and Priming. Distributional semantics builds on the distributional hypothesis (Harris, 1968), according to which the similarity of lemmas correlates with the similarity of their linguistic contexts. The meaning of a lemma is typically represented as a vector of its contexts in large text collections (Turney and Pantel, 2010; Erk, 2012), and semantic similarity is operationalized by using a vector similarity measure such as cosine similarity. Traditional models construct vectors directly from context co-occurrences, while more recent models learn distributed representations with neural networks (Mikolov et al., 2013), which can be seen as advanced forms of dimensionality reduction.

A classical test case of distributional models is exactly lexical priming, which has been modeled successfully in a number of studies (McDonald and Lowe, 1998; Lowe and McDonald, 2000). The assumption of this model family, which we call DISTSIM, is that the cosine similarity between a prime vector $\vec{p}$ and a target vector $\vec{t}$ is a direct predictor of lexical priming:

$$priming_{\text{DISTSIM}}(p,t) \;\propto\; cos\left(\vec{p},\vec{t}\right)$$

Regarding morphological priming, this model predicts the result patterns for French and Dutch but should not be able to explain the German results.

Derivational Morphology in a Distributional Model. In Padó et al. (2013), we proposed to extend distributional models with morphological knowledge in the form of *derivational families* $\mathcal{D}$, that is, sets of lemmas that are derivationally (either transparently or opaquely) related (Daille et al., 2002), such as:

> *knien$_V$ (to kneel$_V$), beknien$_V$ (to beg$_V$), Kniende$_N$ (kneeling person$_N$)*, kniend$_A$ (kneeling$_A$), Knie$_N$ (knee$_N$)

While our motivation was primarily computational (we aimed at improving similarity estimates for infrequent words by taking advantage of the shared meaning within derivational families), these families can be reinterpreted in the current context as driving *morphological generalization* in priming. More specifically, consider the following model family, which we call MORGEN and which is an asymmetrical version of the "Average Similarity" model from Padó et al. (2013):

$$priming_{\text{MORGEN}}(p,t) \;\propto\; \frac{1}{N}\sum_{p'\in\mathcal{D}(p)} cos\left(\vec{p'},\vec{t}\right)$$

This model predicts priming as the average similarity between the target t and *all* lemmas p' within the derivational family of the prime p. It operationalizes the intuition that the prime "activates" its complete derivational family, no matter if transparently or opaquely related. Each of the family members then contributes to the priming effect just like in standard lexical priming.

The MORGEN model should have a better chance of modeling Smolka et al.'s results than the DISTSIM model. Note, however, that it remains completely at the word level, with derivational families as its only source of morphological knowledge.

3 Experiment

Setup. We compute a DISTSIM model by running `word2vec` (Mikolov et al., 2013), a system to extract distributional vectors from text, with its default parameters, on the lemmatized 800M-token German web corpus SdeWaC (Faaß and Eckart, 2013). To build MORGEN, we use the derivational families from DERIVBASE v1.4, a semi-automatically induced large-coverage German lexicon of derivational families (Zeller et al., 2013).[1]

[1] DERIVBASE defines derivational families through a set of about 270 surface form transformation rules. MORGEN does not use information about rules, only family membership. Nevertheless, it is a question for future research to assess the potential criticism that the rule-based induction method implicitly introduces morpheme-level information into the families.

Prime Type	Exp. 1 (RT in ms) (Smolka et al.)	DISTSIM (cos sim) (our model)	MORGEN (cos sim) (our model)
1 Transparent Derivation	563**	**0.44*****	**0.35*****
2 Opaque Derivation	566**	0.28	**0.35*****
3 Synonym	580	0.41***	**0.30***
4 Form	600	**0.24**	**0.26**
5 Unrelated	591	0.25	0.27

Table 1: Average Reaction Times and cosines, respectively. Significance results compared to level *Unrelated*. Correct contrasts shown in boldface. Legend: *: p<0.05; **: p<0.01; ***: p<0.001

Following Smolka et al., we analyze the predictions with a series of one-way ANOVAs (factor Prime Type with reference level Unrelated). As appropriate for multiple comparisons, we adopt a more conservative significance level (p=0.01).

Results. Table 1 reports the experimental results and model predictions (average experimental reaction times, cosine model predictions, and significance of differences). Model contrasts that match experiment contrasts are marked in bold.

As expected, DISTSIM predicts the patterns of classical lexical priming: we observe significant priming effects for Transparent Derivation and Synonymy, and no priming for Opaque Derivation. This is contrary to Smolka et al.'s experimental results.

Our instance of the MORGEN model does a much better job: It predicts highly significant priming effects for both Transparent and Opaque derivations (p<0.001) while priming is not significant at p<0.01 for Synonyms (p=0.04). These predictions correspond very well to Smolka et al.'s findings (cf. Table 1). We tested for two additional contrasts analyzed by Smolka et al.: the difference in priming strength between Transparent and Opaque Derivation (not significant in either experiment or model) and the difference between Transparent Derivation and Synonym (highly significant in both experiment and model).

4 Discussion

In sum, we find a very good match between MORGEN and the experimental results, while the DISTSIM model cannot account for the experimental evidence. Recall that the main difference between the two models is that MORGEN's includes all members of the prime's derivational family into the prediction of the priming strength. This leads to the following changes compared to DISTDIM:

1. For Opaque Derivation, MORGEN typically predicts stronger priming than DISTSIM, since prime and target are typically members of the same derivational family (assuming that there are no coverage gaps in DERIVBASE), and the average similarity between the target and the words in the family is higher than the similarity to the prime itself. Taking Figure 1 as an example, the Opaque Derivation pair *entbinden (give birth) – binden (bind)* is relatively dissimilar, and the similarity increases when other pairs like *binden (bind) – zubinden (tie)* are taken into consideration.

2. For Synonymy, MORGEN typically predicts weaker priming than DISTSIM, since the average similarity between target and all members of the prime's family tends to be lower than the similarity between target and original prime. Again considering Figure 1, the Synonym pair *binden (bind) – zuschnüren (tie)* is relatively similar, while including terms derivationally related to the prime *zuschnüren (tie)* like *schnurlos (cordless)* introduces low-similarity pairs like *schnurlos (cordless) – binden (bind)*.

MORGEN is not the only model that takes a distributional stance towards morphological derivation. Marelli and Baroni (2014) propose a compositional model that computes separate distributional representations for the meanings of stems and affixes and is able to compute representations for novel, unseen derived terms. The morpheme-level approach of Marelli and Baroni's model corresponds more directly to Smolka et al.'s claims and might also be able to account for the experimental patterns.

However, our considerably simpler model, which only has knowledge about distributional families, is also able to do so. This at the very least means that morpheme-level processing is not an *indispensable* property of any model that explains Smolka et al.'s experimental results and that the evidence for a special organization of the German mental lexicon, in contrast to other languages, must be examined more carefully.

In fact, our model provides a possible alternative source of explanations for the cross-lingual differences: Since the MorGen predictions are directly influenced by the size and members of the derivational families, German opaque morphological priming may simply result from the high frequency of opaque derivations. In the future, we plan to apply the model to Dutch and French to check this alternative explanation.

Acknowledgments

We gratefully acknowledge funding by Deutsche Forschungsgemeinschaft through Sonderforschungsbereich 732, project B9.

References

J. Kathryn Bock. 1986. Syntactic persistence in language production. *Cognitive Psychology*, 18:355–387.

Béatrice Daille, Cécile Fabre, and Pascale Sébillot. 2002. Applications of computational morphology. In Paul Boucher, editor, *Many morphologies*, pages 210–234.

Katrin Erk. 2012. Vector space models of word meaning and phrase meaning: A survey. *Language and Linguistics Compass*, 6(10):635–653.

Gertrud Faaß and Kerstin Eckart. 2013. SdeWaC – a corpus of parsable sentences from the web. In Iryna Gurevych, Chris Biemann, and Torsten Zesch, editors, *Language Processing and Knowledge in the Web*, volume 8105 of *Lecture Notes in Computer Science*, pages 61–68. Springer Berlin Heidelberg.

Zellig Harris. 1968. *Mathematical Structures of Language*. Wiley.

Steve T. Kempley and John Morton. 1982. The effects of priming with regularly and irregularly related words in auditory word recognition. *British Journal of Psychology*, pages 441–445.

Will Lowe and Scott McDonald. 2000. The direct route: Mediated priming in semantic space. In *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*, pages 675–680, Philadelphia, PA.

Marco Marelli and Marco Baroni. 2014. Dissecting semantic transparency effects in derived word processing: A new perspective from distributional semantics. In *9th International Conference on the Mental Lexicon*, pages 7–8, Niagara on the Lake, Canada. Abstract.

Scott McDonald and Will Lowe. 1998. Modelling functional priming and the associative boost. In *Proceedings of the 20th Annual Conference of the Cognitive Science Society*, pages 675–680, Madison, WI.

Fanny Meunier and Catherine-Marie Longtin. 2007. Morphological decomposition and semantic integration in word processing. *Journal of Memory and Language*, 56:457–471.

David E. Meyer and Roger W. Schvaneveldt. 1971. Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90(2):227–234.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems 26*, pages 3111–3119.

Sebastian Padó, Jan Šnajder, and Britta Zeller. 2013. Derivational smoothing for syntactic distributional semantics. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 731–735, Sofia, Bulgaria.

Herbert Schriefers, Pienie Zwitserlood, and Ardi Roelofs. 1991. The identification of morphologically complex spoken words: Continuous processing or decomposition? *Journal of Memory and Language*, 30:26–47.

Eva Smolka, Katrin H. Preller, and Carsten Eulitz. 2014. 'verstehen' ('understand') primes 'stehen' ('stand'): Morphological structure overrides semantic compositionality in the lexical representation of German complex verbs. *Journal of Memory and Language*, 72:16–36.

Peter D. Turney and Patrick Pantel. 2010. From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, 37(1):141–188.

Britta Zeller, Jan Šnajder, and Sebastian Padó. 2013. DErivBase: Inducing and evaluating a derivational morphology resource for German. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 1201–1211, Sofia, Bulgaria.