

German affixed words: morphological productivity and semantic transparency.

Inna V. Stupak¹, R. Harald Baayen²

¹University of Tuebingen, inna.stupak@philosophie.uni-tuebingen.de

²University of Tuebingen, harald.baayen@uni-tuebingen.de

Keywords: Morphological productivity, Semantic transparency, German affixed words, Embeddings

This present paper discusses the relation between morphological productivity and semantic transparency. Many studies have proposed criteria for establishing whether a derivational rule is productive, and what factors co-determine its productivity (Baayen, 2005; Plag, 2003; Dressler, 2003; Fernández-Domínguez, 2009). What emerges from these studies is that, qualitatively, productivity decreases as the number of restrictions and conditions on a word formation process increases. Quantitative studies of productivity have led to several measures that provide an overall numeric assessment of different aspects of productivity (e.g., Baayen, 2005), but these measures by themselves cannot answer the question of to what extent the different qualitative factors contribute to the actual values of the quantitative measures.

The paper addresses one specific qualitative factor that has been argued to co-determine productivity, semantic transparency (see, e.g., Aronoff, 1976; Baayen, 1993; Bonami and Paperno, 2018). The present study reports on exploratory research that follows up on the studies of Mandarin compounding (Shen and Baayen, 2022), but now addressing the productivity and transparency of German particle verbs and German derivational suffixes. Using distributional semantics, we compare German word formation using particles with derivational word formation. We observed that derivational suffixes, but not particles, tend to make strong independent semantic contributions to their carrier words (Figure 1). In two-dimensional t-SNE maps, complex words show clustering by affix, but not by particle (Figure 2). Furthermore, the semantic vectors of suffixed words are predictable from their base words with higher accuracy than is possible for particle verbs. For particle verbs, but not affixed verbs, semantic similarity within the set of complex words correlated negatively with the number of types (Figure 3). Furthermore, only for particle verbs, a greater number of observed types predicted a reduced probability of observing unseen types. We propose that particle verbs primarily serve the onomasiological function of labeling, resulting in relatively idiosyncratic semantic vectors. By contrast, words sharing derivational affixes form distinct clusters in semantic space while maintaining strong and consistent semantic relations with their base words. This enables these words to serve not only as labels, but also allows them to be used with an anaphoric function in discourse.

References

- Aronoff, Mark. 1976. *Word Formation in Generative Grammar*. MIT Press, Cambridge, Mass.
- Baayen, R. Harald. 1993. On frequency, transparency, and productivity. In Booij, G. E. and van Marle, (ed.), *Yearbook of Morphology 1992*, 181–208. Kluwer Academic Publishers, Dordrecht.
- Baayen, R. Harald. 2005. Data mining at the intersection of psychology and linguistics. In Cutler, A. (ed.), *Twenty-first century psycholinguistics: Four cornerstones*, 69–83. Erlbaum, Hillsdale, New Jersey.
- Bonami, Olivier and Paperno, Denis. 2018. Inflection vs. derivation in a distributional vector space. *Lingue e Linguaggio*, 17(2).173–195.
- Plag, Ingo. 2003. *Word Formation in English*. Cambridge University Press, Cambridge, UK.
- Dressler, Wolfgang U. 2003. Degrees of grammatical productivity in inflectional morphology: 1119. *The Italian Journal of Linguistics* 15 (2003): 31.
- Fernández-Domínguez, Jesús. 2009. *Productivity in English Word-formation: An approach to N+N compounding*. European University Studies, Series 21: Linguistics. Vol. 341. Peter Lang, Bern.
- Shen, Tian and Baayen, Harald. R. 2022. Productivity and semantic transparency: An exploration of word formation in Mandarin Chinese. *The Mental Lexicon*.

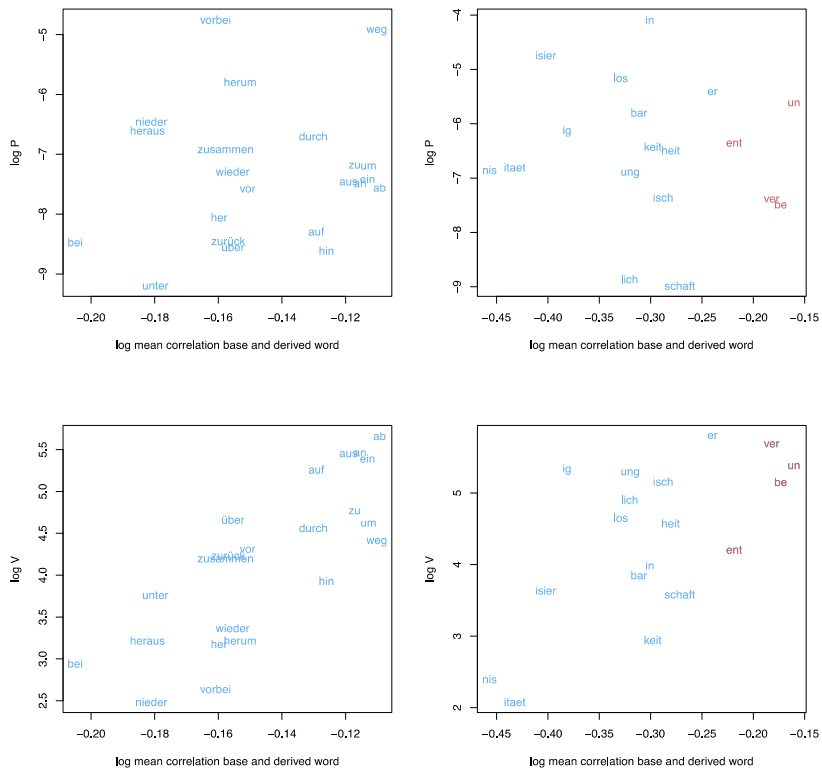


Figure 1: Log mean correlation of the semantic vectors of base word and complex word (horizontal axis) for particles verbs (left panels) and affixed words (right panels), by log potential productivity (P(N), upper panels) and extent of use (V (N), lower panels). Prefixes are highlighted in red.

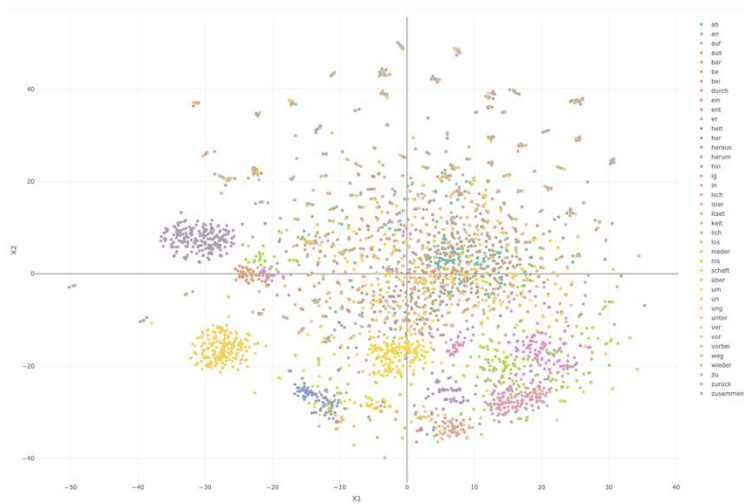


Figure 2: Locations of the shift vectors for affixed words and particle verbs in their joint t-SNE shift space. The well-defined clusters represent affixed words, excluding the prefixed verbs. Prefixed verbs tend to appear in base-word defined clusters together with particle verbs.

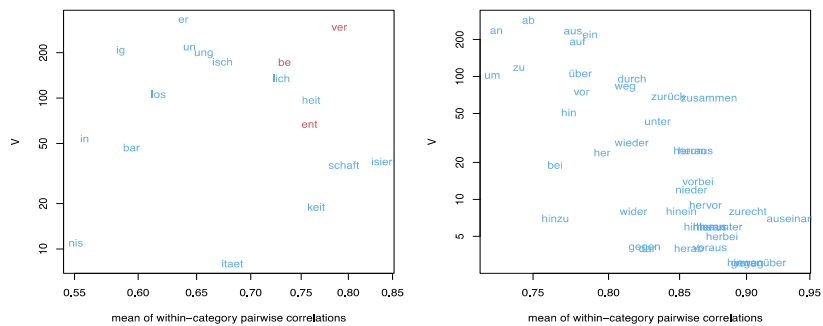


Figure 3: Mean of within-category pairwise correlations as predictor of the number of types. Left: affixed words; right: particle verbs.