Understanding the Morphological Productivity : internal vs external factors

This study provides a new analysis of the productivity of German and Ukrainian derived verbs. The productivity of a derivational rule has traditionally been approached either from a quantitative or quantitative angle. Faced with the problem that neither qualitative (Plag, 1999) nor quantitative (Baayen, 2005) approaches provide sufficient criteria for predicting whether a derivational rule is productive, this project explores a new approach to morphology, using distributional semantics. My current research indicates that there are two fundamentally different ways in which a word formation process can be productive. These ways offer very different solutions to two central mysteries of word formation:

> *(1) How can a word formation process, serving the function of providing names for new ideas and concepts, be productive, given that these concepts and ideas are themselves not predictable or compositional?*
>
> *(2) If a word formation process gives rise to "names" that are unavoidably to a greater or lesser extent semantically idiosyncratic, how can it be productive at the same time?*

In recent research on productivity, corpus-based embeddings (i.e., high-dimensional numeric vectors representing words' meanings) are playing an increasingly important role in addressing these mysteries. Here, further advances in understanding productivity are anticipated.

The present work focuses on the correlation between the productivity of German and Ukrainian complex verbs and their semantic transparency, measuring (a) the particle and its base, (b) the particle and the corresponding particle verb, (c) the base and the corresponding particle verb, (d) semantic correlation of all pairs of complex verbs sharing a given particle; (e) semantic correlation of all pairs of complex verbs sharing a given suffix, and (f) semantic transparency between suffix verbs and their base.

The relation between morphological productivity and semantic transparency investigated building on previous studies of German derivation (Stupak&Baayen 2022), using a Gaussian locations scale GAM (Wood, 2017), visualization techniques such as t-distributed stochastic neighbor embedding (Maaten and Hinton, 2008), skip-gram model (Bojanowski et al. 2017) word2vec (Mikolov et al., 2013).

To answer the questions: Does the productivity of German complex verbs correlate with their transparency and How does productivity reflect semantic transparency, we conducted distributional semantics. We noted the correlation between semantic transparency and the category-conditioned degree of productivity (Figure 1):

> (i)   *the positive correlation between semantic transparency and the productivity;*
> (ii)   *the negative correlation of variability and the productivity*;
> (iii)   *the positive correlation between similarity and productivity reveals all complex verbs sharing the same particle*.

Thus, the predictable category-conditioned degree of productivity might depend upon the following factors: a) the transparency of complex verbs and their constituents; b) the morphological length (complex, simple) of a particle; c) semantic variability of particles.
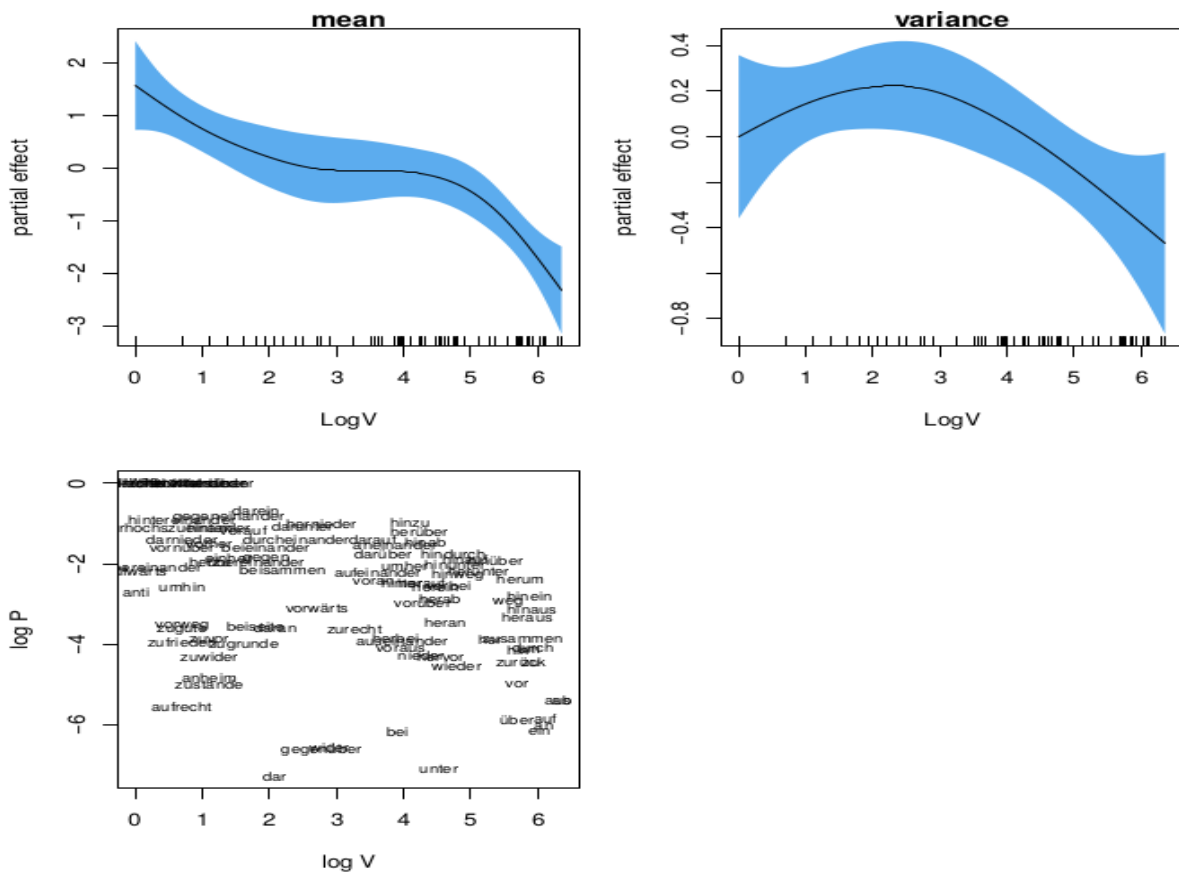
Figure 1: Log category-conditioned productivity P as a function of log number of Types V for particle verbs.

References

Baayen, R. H. (2005). Corpus linguistics in morphology: morphological productivity. In Lüdeling, A., Kytö, M., and McEnery, T., editors, *Handbook of Corpus Linguistics* (Handbücher zur Sprach- und Kommunikationswissenschaft). De Gruyter.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. Transactions of the association for computational linguistics, 5, 135-146.

Maaten, L. v. d. & Hinton, G. (2008). Visualizing data using t-sne. *Journal of machine learning research*, **9,** 2579–2605.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. Advances in neural information processing systems, 26.

Plag, I. (1999). Morphological productivity. Structural constraints in English derivation. Berlin: Mouton de Gruyter.

Stupak, I., & Baayen, R. H. (2022). "An Inquiry into the Semantic Transparency and Productivity of German Particle Verbs and Derivational Affixation." PsyArXiv. October 7. doi:10.31234/osf.io/ebspt.

Wood, S. N. (2017). Generalized Additive Models. Chapman & Hall/CRC, New York.

https://uni-tuebingen.de/fakultaeten/philosophische-fakultaet/fachbereiche/neuphilologie/seminar-fuer-sprachwissenschaft/arbeitsbereiche/quantitative-linguistik/korpora/#c734223