

Morphological productivity – individual and community-level factors

Kristian Berg

Universität Oldenburg, kristian.berg@uol.de

Some word formation patterns can be extended, some not, and some only sometimes. For example, when a new noun enters the German language, it can be used to form new compounds like *TikTok-Sucht*, or new derived words with patterns like *-isieren* (*TikTokisieren*). Other patterns are not extensible in the same way (e.g., **TikToksam*). Most word formation patterns are not fully productive.

The central question that has occupied word formation research for the past four decades is, how can we explain the many differences in productivity? Some of the factors isolated as relevant for productivity in word-formation are linguistic factors in the narrow sense: type and token frequency, semantic and phonetic transparency of the word formation product (Hay & Baayen 2005), and a sufficient number of un-derived bases (Aronoff 1976). Some factors are linguistic, but in a wider sense, like the medium of the text (Plag et al. 1999) and its register (Baayen 1994).

Some factors, however, are located at the individual or group level. De Smet (2020) shows that the proclivity to form new words varies from person to person. Productivity may be unevenly distributed among language users, with only some coining new words – and many others reading, storing, and repeating them. This shifts the focus to individuals’ production of words: Which ones are new, which ones are repeated, and what are the factors determining both processes?

In a case study addressing these questions, I focus on one very prolific German writer, Thomas Mann. With the help of his diaries and letters as well as secondary literature, I am currently constructing two corpora: The set of all texts he ever wrote, and (an approximation of) the set of all texts he ever read. These two corpora allow for a description of the relation between input and output for word-formation patterns like *-ieren* and *-sam*, using both established productivity measures and semantic homogeneity measures on the basis of word embeddings. It is also possible to model the input over time, and to compare it to Mann’s output. To this end, I use Linear Discriminative Learning (Baayen et al. 2019).

Literatur

- Aronoff, M. (1976). *Word Formation in Generative Grammar*. MIT Press, Cambridge, Mass.
- Baayen, R. H. (1994). Derivational productivity and text typology. *Journal of quantitative linguistics*, 1(1), 16-34.
- Baayen, H., Y. Chuang, E. Shafaei-Bajestan and J. Blevins (2019). The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de) composition but in linear discriminative learning. *Complexity*. <https://onlinelibrary.wiley.com/doi/epdf/10.1155/2019/4895891>
- De Smet, H. (2020). What predicts productivity? Theory meets individuals. *Cognitive Linguistics* 31(2). 251-278.
- Hay, J. B. and R. H. Baayen (2005). Shifting paradigms: gradient structure in morphology. *Trends in Cognitive Sciences*, 9:342–348.
- Plag, I., C. Dalton-Puffer and R. H. Baayen (1999). Morphological productivity across speech and writing. *English Language & Linguistics*, 3(2), 209-228.