Derivational morphology in flux: A corpus study of word-formation change in German

The availability of large annotated corpora has opened up new perspectives on the development of word-formation patterns in German (see e.g. Author 2016; Kempf 2016; Kopf 2017). At the same time, recent in-depth studies of the diachronic development of several German word-formation patterns have also shown that some long-standing issues in corpus-based (historical) morphology are still far from resolved. For instance, it is still an open question how the productivity of a pattern can best be assessed. From a diachronic perspective, this question is all the more important, given that word-formation change can be defined as change in productivity (e.g. Scherer 2006, 2007).

This paper draws on the example of the German nominalization pattern [V-ung], e.g. Landung ‘landing’, Bildung ‘education’, to demonstrate how a mixed-methods approach as advocated by e.g. Author & Co-Author (2016) can help us gain a comprehensive, nuanced, and data-driven picture of how a word-formation pattern changes. Demske (2000) has already pointed out that the word-formation pattern of ung-nominalization experiences a steep decline in productivity from the Early New High German (ENHG) period onwards. This can partly be explained by the lexicalization of highly frequent word-formation products, which leads to a decrease in “schema strength” (Bybee 1995), i.e. the extant word-formation products cannot serve as “templates” for new formations any more to the same extent as they could in the past.

The present study combines data from three different corpora: The Mainz Early New High German Corpus (Kopf 2016), the GerManC corpus (Durrell et al. 2007), and DTA baby (Author 2018), a balanced subset of the German Text Archive (DTA, Geyken & Gloning 2015). To assess changes in the word-formation pattern’s productivity, a number of measures is used, including Baayen’s hapax-based measure of potential productivity (Baayen 2009), extrapolation techniques using a finite Zipf-Mandelbrot model (Evert 2004), and methods drawing on new types rather than hapax legomena, as discussed in e.g. Cowie (1999), Berg (forthc.), and Author & Co-Author (forthc.). The combination of different methods lends support to an overall decrease in productivity but also allows for identifying trends that have been overlooked in previous research, such as a slight surge in productivity in the 18th century, when many French loan verbs in -ieren entered the German language (see e.g. Schmidt 2007), which are quite eligible as bases for ung-nominalization to the present day.

In addition, the data have been annotated for the (syntactic) context in which the derivatives appear, which allows for a detailed quantitative analysis of their usage profiles – and especially of the way their use changes over time. Both the analyses of individual variables and an aggregated analysis using Multi-Dimensional Scaling (see e.g. Wheeler 2005) lend support to Demske’s (2000) hypothesis that ung-nominals tend to occur in (semantically and syntactically) less “verby” contexts. This in turn substantiates the idea that the diachronic development of the word-formation pattern in question can be explained in terms of decreasing schema strength, as sketched above. At the same time, however, the corpus-based and data-driven approach taken here shows that the diachronic development of the pattern is less straightforward and much more multi-faceted than was previously acknowledged.
References


