Shifting from phonological conditions to morphological ones: a corpus study on the entropy-based predictability in the morphologization of Umlaut

Junyu Ruan

The Ohio State University

Alternation patterns between sounds under certain conditions are widely attested in the history of various languages, and the development of German Umlaut is one of the best-studied sound alternations that have their triggering condition shifted from phonological to morphological ones. In Germanic linguistics, the term *Umlaut* refers to the fronting process of certain vowels. It first occurs as a purely phonological alternation that is triggered by a following unstressed syllable containing *i, î* or *j* in Old High German (OHG). Since some suffixes of certain morphological categories start with unstressed syllables with *i, î* or *j*, there is an overlapping between the phonological and the morphological conditions, and when the phonological condition is obscured because of the dropping of *j* and the weakening of *i, î* to a schwa in Middle High German (MiHG), the Umlaut alternation becomes a pattern that is completely conditioned by the morphological categories of the umlauting forms.

It has long been debated how morphologization of phonological processes such as Umlaut is motivated. This paper proposes a hypothesis with regards to morphologization process in general: whereas the predictability of a phonologically conditional alternation is reduced because of the loss of the triggering condition, the condition of the alternation might be re-associated to morphological factors, so that the decrease in its phonological predictability could be counterbalanced by an increase in morphological predictability.

To test the hypothesis, we conduct a quantitative investigation on OHG and MiHG corpora with different regional varieties, and, following the practice in Ackerman et al. (2009) and Ackerman & Malouf (2013), use conditional entropy as a quantitative measurement of predictability. We collect verb tokens with umlautable stem vowels from annotated corpora of OHG and MiHG, and the conditional entropy of the Umlaut alternation is calculated based on their type and token frequency in the corpora, conditioned by the phonological context (defined as the quality of the vowel that immediately follows the stem syllable) and the morphological category (mood, tense, person, number and the combination of the four) of the tokens.

The results support the abovementioned hypothesis to a certain extent. The entropy values conditioned by phonological environments is roughly the same as the entropy conditioned by morphological factors among OHG tokens, but the conditional entropy under morphological factors is lower than under phonological factors among MiHG tokens, indicating that the role morphological factors play in predicting Umlaut becomes more important relatively to phonological ones in MiHG than in OHG. Moreover, the token-based entropy is lower than the type-based entropy in general. It can be assumed from this study that phonological regularity and morphological regularity might be subject to a predictability that is more general, and the decrease in phonological regularity could be somewhat compensated by an increase in morphological regularity, so that a sound alternation pattern would not become

completely unpredictable.

Reference

Ackerman, Farrell, James P. Blevins, and Robert Malouf. 2009. Parts and Wholes: Implicative Patterns in Inflectional Paradigms. In *Analogy in Grammar*, 54–82. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199547548.003.0003.

Ackerman, Farrell, and Robert Malouf. 2013. Morphological Organization: The Low Conditional Entropy Conjecture. *Language* 89: 429–64.