
Orthography-informed phonology and phonology-informed spelling in the lexicon

Jennifer Keller & Ingo Plag

Heinrich-Heine-Universität Düsseldorf

[jennifer.keller,ingo.plag]@hhu.de

The Discriminative Lexicon is a theory of the mental lexicon which implements a computational architecture using high-dimensional vectors to establish a system of form-meaning relations by mapping form onto meaning, and meaning onto form (Baayen et al. 2019). To date, either orthographic or phonological representations have typically been used as representations of form, although it is uncontroversial that the two kinds of representations are not independent of each other (e.g. Brysbaert 2022). The present study explores whether the Discriminative Lexicon theory can be used to combine orthography and phonology by producing orthography-informed phonological representations and phonology-informed orthographic representations. It is tested whether such combined representations provide better predictions of language behavior than models based on only orthography or phonology, using existing data sets from previous studies of German number (e.g. Plag et al. 2024).

To derive orthography-informed phonological representations and phonology-informed spelling representations, we map a matrix of orthographic word representations onto a matrix of phonological word representations (and vice versa). This will provide us with transformation matrices that can be used to derive for each word a phonological representation as predicted on the basis of orthography, and an orthographic representation as predicted on the basis of the phonological representation.

We then compute Discriminative Learning models in which we map our combined form representations onto word embeddings (as meaning representations) to model comprehension (mapping form onto meaning) and production (mapping meaning onto form). Finally, we compare and explore the performance of the models based on an orthography-informed phonology and a phonology-informed spelling with models based on orthography-only or phonology-only form representations.

References: • Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E. & J. P. 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* 2019(1), 1–39. • Brysbaert, M. (2022). Word recognition II: Phonological Coding in Reading. In M. J. Snowling, C. Hulme & K. Nation (eds.), *The Science of Reading: A Handbook*, 79-101. Wiley. • Plag, I., Heitmeier, M. & F. Domahs (2024). German nominal number interpretation in an impaired mental lexicon: A naïve discriminative learning perspective. *The Mental Lexicon* 18(3), 417-445.