

# Exploiting embeddings for analyzing the role of emotion in lexical change

Andreas Baumann

Emotional factors like valence, concreteness, and arousal have been shown to influence lexical processing, in that they exhibit non-neutrality, positivity, and negativity biases, respectively (e.g., Kuperman et al. 2014). Evidently, even weak cognitive biases like this can yield strong tendencies on a larger time scale. In this study, we investigate diachronic long-term effects of these emotional variables on the reproductive success of English words, and we exploit pre-trained word-embedding models to make our analysis more robust.

We operationalize reproductive success in three ways: by means of (1) lexical prevalence, (2) age of acquisition, and (3) diachronic growth. For (1-2), we use crowd-sourced prevalence and age-of-acquisition norms (Brysbaert et al. 2019; Kuperman et al. 2012). For (3), we analyze frequency trajectories extracted from COHA. Crucially, we control for semantic shifts through comparing word embeddings across periods to ensure that only words with relatively stable semantics end up in our diachronic evaluation. For this, we simply use pre-trained (temporally aligned) embeddings from the NLPL embedding repository (Kutuzov et al. 2017).

Measures of reproductive success are then combined with emotional norms (Warriner et al. 2013, Brysbaert et al. 2019). Generalized additive models are used to estimate main effects and interactions. We show that long-term effects of valence and concreteness largely mimic cognitive short-term biases. However, arousal, quite surprisingly, exhibits a clearly positive effect on lexical reproduction. Taking interactions among valence, concreteness, and arousal into account, it is shown that, on average, positive, concrete, and arousing words show the highest reproduction rates across all measures (1-3).

In a post-hoc cluster analysis, we again use word embeddings to infer which semantic fields belong to the positive, concrete, and arousing domain. We find that it is the semantic fields of ‘food’, ‘money’, and ‘sex’, which seem to be associated with this emotional setup. It is concluded, in a somewhat speculative manner, that neuro-cognitive processes responsible for managing rewards and semantic processing might be relevant to our findings (Arsalidou et al. 2020). On a methodological level, it is argued that word-embedding techniques should, in the long run, belong to the analytic standard repertoire of linguists that work in cognitive diachronic linguistics, and that embedding repositories like NLPL represent a major step towards this goal.

## References

- Arsalidou, M., Vijayarajah, S., & Sharaev, M. (2020). Basal ganglia lateralization in different types of reward. *Brain imaging and behavior*, 14(6), 2618-2646.
- Brysbaert, M., Mandera, P., McCormick, S. F., & Keuleers, E. (2019). Word prevalence norms for 62,000 English lemmas. *Behavior research methods*, 51(2), 467-479.
- Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior research methods*, 46(3), 904-911.
- Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior research methods*, 44(4), 978-990.
- Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: valence and arousal affect word recognition. *Journal of Experimental Psychology: General*, 143(3), 1065.
- Kutuzov, A., Fares, M., Oepen, S., & Velldal, E. (2017). Word vectors, reuse, and replicability: Towards a community repository of large-text resources. In *Proceedings of the 58th Conference on Simulation and Modelling* (pp. 271-276). Linköping University Electronic Press.
- Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior research methods*, 45(4), 1191-1207.