A word-based account of comprehension and production of Turkish

Denız Ekin Yavaş, Jessica Nieder, Ruben van de Vijver
Heinrich-Heine-Universität
Linguistics

For agglutinative languages, such as Turkish, it has been claimed that a word-based account of their morphology is unfeasible (Hankamer, 1989). This is reflected in morphophonological accounts of vowel harmony in Turkish, which all, as far as we know, assume that words in Turkish consist of a string of morphemes, and that the vowels in the suffixes share phonological properties with the vowels of the root (Aksënova et al., 2020; Kabak, 2011). Yet, word-based accounts of the morphology of agglutinative languages are not only available (Ainsworth, 2019), but even considered better since they better predict relations among words within and across paradigms (Blevins, 2016; van de Vijver & Uwambayinema, 2022, Accepted). For Turkish, a word-based account of its morphology raises the question as to whether and how a word-based account of Turkish morphology account for vowel harmony?

We answer this question by analyzing production and comprehension of Turkish computationally within the framework of the Discriminative Lexicon (Baayen et al., 2019; Nieder et al., 2023), a word-based theory of the mental lexicon, rooted in the idea that language has a communicative function (Van Valin Jr, 2005). In this theory comprehension is modeled as a mapping of the phonology onto semantics, and production is modeled as a mapping of semantics onto phonology. In our study the phonology will be represented by words taken from a corpus (Sezer & Sezer, 2015), and their semantics as word embeddings from `fastText` (Mikolov et al., 2018). We have created the datasets and (as of July 14th) we plan to test 100089 words (our hardware permitting).

In our computational modeling we will pursue three goals. The first is to investigate whether the model is able to correctly comprehend words. This will be assessed by predicting the meaning of a word as predicted by the model to the actual meaning of the word. The second goal is to investigate whether the model is able to correctly produce words. This will be tested by comparing the predicted word form with the targeted word form. The third goal is to analyze the errors made by the model. Focusing on the comprehension and productions errors of the model not only provides us with insights into the ability of the computational model to recreate Turkish vowel harmony but also allows us to assess how far a word-based model of morphology can take us when focusing on agglutinative languages.

Ainsworth, Z.-J. (2019). The Veps illative: The applicability of an abstractive approach to an agglutinative language. *Transactions of the Philological Society*, *117*(1), 58–78.
Aksënova, A., Rawski, J., Graf, T., & Heinz, J. (2020). The computational power of harmony. *Oxford Handbook of Vowel Harmony, Oxford University Press, under review*.
Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., & Blevins, J. P. (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*.
Blevins, J. P. (2016). The minimal sign. In A. Hippisley & G. Stump (Eds.), *The cambridge handbook of morphology* (pp. 50–69). Cambridge University Press. https://doi.org/10.1017/9781139814720.003
Hankamer, J. (1989). Morphological parsing and the lexicon. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 392–408). MIT Press.
Kabak, B. (2011). Turkish vowel harmony. In *The blackwell companion to phonology* (pp. 1–24). John Wiley Sons, Ltd. https://doi.org/https://doi.org/10.1002/9781444335262.wbctp0118
Mikolov, T., Grave, E., Bojanowski, P., Puhrsch, C., & Joulin, A. (2018). Advances in pre-training distributed word representations. *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*.
Nieder, J., Chuang, Y.-Y., Vijver, R. v. d., & Baayen, R. H. (2023). A Discriminative Lexicon Approach to Word Comprehension, Production and Processing: Maltese Plurals. *Language*, *99*(2), 242–274. https://doi.org/10.31234/osf.io/rkath
Sezer, T., & Sezer, T. (2015). *Ts wikipedia ldc2015t15*. Linguistic Data Consortium. Philadelphia. https://doi.org/10.35111/mem6-4951
van de Vijver, R., & Uwambayinema, E. (2022). A word-based account of comprehension and production of Kinyarwanda nouns in the Discriminative Lexicon. *Linguistics Vanguard*.
van de Vijver, R., & Uwambayinema, E. (Accepted). A word-based account of comprehension and production of Kinyarwanda nouns in the Discriminative Lexicon. *Linguistics Vanguard*.
Van Valin Jr, R. D. (2005). *Exploring the syntax-semantics interface*. Cambridge University Press.