Morphological variability in linguistic generalization: The Turkish aorist

Serkan Uygun[1] and Harald Clahsen[1]

[1]Potsdam Research Institute for Multilingualism

serkanuygun2014@gmail.com

A controversial issue in psycholinguistic literature concerns the question of which mechanisms are employed in morphological generalization to nonce words. Different mechanisms have been suggested for morphological generalization, in particular associative extensions of existing patterns, and/or generalization based on morphological rules or rule-like operations; see Veríssimo and Clahsen (2014) for a review. The present study employs the Gradient Symbolic Computation (GSC) framework (Smolensky et al., 2014; Goldrick et al., 2016) to investigate morphological generalization in the Turkish aorist. GSC is a constraint-based approach that brings together symbolic grammatical constraints and graded representations. It defines grammar via a set of violable constraints that assigns a numerical well-formedness value (Harmony) to each of the possible outputs by generating a probability distribution.

By using a nonce-word production experiment, 50 Turkish native speakers were tested in the Turkish aorist. The aorist form encodes habitual aspect or general present tense and unlike most inflectional exponents in Turkish, it is not completely regular but involves a restricted set of irregular forms, specifically for monosyllabic stems. 78 nonce-words were created using rhyme similarity, in four conditions: (i) nonce words similar to existing irregular verbs, (ii) similar to existing regular verbs, (iii) rhymes that exist both with existing regular and irregular verbs, and (iv) nonce words that are phonotactically legal but do not rhyme with existing verbs of Turkish. Participants had to complete sentences by forming an irregular or regular aorist form of a nonce-word presented in its infinitive form. The results showed that Turkish native speakers most commonly relied on the regular aorist in all conditions. The highest proportion of irregular aorist forms was found for irregular rhymes (49%), followed by the mixed condition (iii) (39%), with very low proportions for the non-rhyme condition (7.9%) and for regular rhymes (4.8%).

Two constraints were determined for GSC modeling of the Turkish aorist: strength of the rhyme (how prototypical is a given rhyme for a regular vs. an irregular aorist form) and applying the rule (forming the nonce-word in regular aorist form). The model was fed with the corpus frequencies of a large data base with 4.950.407 word types and 491.360.398 tokens of written and spoken existing Turkish verbs (Sezer, 2017) to yield the estimated probabilities for the four types of nonce words we tested with the human participants. Comparison of the model's output to the participants' nonce-word productions revealed that while the model correctly estimated the proportions of regular rhymes in condition (ii), it overestimated the proportions of irregular aorist forms in the other three conditions.

These results indicate that a model that was trained on the frequency distribution of the similarity clusters amongst existing verb forms only provides a partial account of

Uygun, S., & Clahsen, H. (2019). *Morphological variability in linguistic generalization: The Turkish aorist*. Paper presented at the International Morphological Processing Conference (MoProc 2019), University of Tübingen, Tübingen, Germany. 4 - 7 November.

morphological generalization. We conclude that morphological generalization in the Turkish aorist is largely driven by the regular pattern and less so by similarity to existing forms.

Paper presentation

References:

Goldrick, M., Putman, M., & Schwarz, L. (2016). Coactivation in bilingual grammars: A computational account of code mixing. *Bilingualism: Language and Cognition, 19*(5), 857-876.

Sezer, T. (2017). TS corpus project: An online Turkish dictionary and TS DIY corpus. *European Journal of Language and Literature,9*(1), 18-24.

Smolensky, P., Goldrick, M., & Mathis, D. (2014). Optimization and quantization in gradient symbol systems: A framework for integrating the continuous and the discrete in cognition. *Cognitive Science, 38*(6), 1102-1138.

Veríssimo, J., & Clahsen, H. (2014). Variables and similarity in linguistic generalization: Evidence from inflectional classes in Portuguese. *Journal of Memory and Language, 76*, 61-79.