# Different [back] [round]'s: Lexical origin and vowel harmony in Turkish

Michael Dow, Jeffrey Lamontagne · Université de Montréal, Indiana University

**Introduction**    Turkish vowel harmony processes (with respect to anteriority and roundness) are often described as regular between roots and suffixes but no longer productive within roots (e.g., Bennink, 1992), due in part to the introduction of Arabic and Persian loanwords since at least the 11th century (hereafter "QUASI-BORROWINGS") (e.g., Lanfranca, 2012; Özcelik, 2014). More recent borrowings from other languages such as English and French are observed to harmonize with suffixes (e.g. Baturay, 2012), whether harmonic or not, but to our knowledge, no large-scale, quantitative study has been conducted. In this paper, we examine vowel harmony in a Twitter corpus of Turkish with respect to both lexical origin and root (dis)harmony. Results suggest the non-productivity in roots is increasingly extending to derived forms.

**Methodology**    We extracted all nouns with IPA transcriptions and etymologies from the Turkish Wiktionary lexicon (Ylonen, 2022). Words were classified as BORROWINGS (English, French or Italian origin), QUASI-BORROWINGS (Persian or Arabic origin), or NATIVE. The rightmost vowel of each word was extracted and classified according to its anteriority and rounding. Vowel-final roots and roots containing four syllables or more were excluded. Using raw frequency data from the 2016 Open Subtitles database, we limited roots to a maximum of the 10 most frequent words by harmonic features governed (root-final anteriority and rounding), origin and number of syllables, for a total of 391 roots. Each root was then combined with all possible variants of the first-person possessive marker (*-im* [im], *-üm* [ym], *-ım* [ɯm] and *-um* [um]).

We then gathered tweets and associated metadata for Turkish-language Twitter using the then-available academic API access and the `academictwitteR` package (Barrie and ting Ho, 2021). BORROWINGS were exhaustively gathered for a 10-year window of April 2013 to April 2023. After an initial test of the same timeframe, QUASI-BORROWINGS and NATIVE WORDS were limited to 1 year (April 2022 to April 2023) and 1 month (November 2022), respectively, due to the sheer volume of results. This yielded nearly 7.2 million tweets, the majority of which target borrowings (5.8 million). The current analysis uses mixed-effects logistic regression, with roots as random intercepts, to predict whether anteriority and rounding harmony will affect the vowel quality of the possessive suffix in 26,182 derived words.

**Results**    Harmonising with the root's final vowel in both anteriority ($p < 0.0001$) and especially rounding ($p < 0.0001$) is typical, but neither harmony is categorical. Harmonising rounding increases the probability of harmonising anteriority ($p < 0.0001$) and vice versa ($p < 0.0001$). Root anteriority alone does not affect harmony likelihood, but unrounded back roots exhibit more anterior harmony ($p < 0.0001$) whereas unrounded front roots exhibit more rounding harmony ($p = 0.010$). Monosyllabic roots are most likely to exhibit disharmony ($p = 0.0248$; $p = 0.0002$) compared to both harmonic and disharmonic roots, which pattern similarly. We find no significant difference in suffix harmony between BORROWINGS and NATIVE WORDS, but QUASI-BORROWINGS have a lower likelihood of rounding harmony ($p = 0.0004$).

**Discussion**    Phonological integration may correlate with morphological integration (Bessett, 2016), as expected from the typological pattern that phonological processes are more often categorical in derived words (e.g. Kiparsky, 1993; Chong, 2019). However, harmony is not categorically applied in our data, neither in roots nor in derived words, with root disharmony not predicting suffix harmony variability. QUASI-NATIVE words exihibit lower rates of rounding harmony, suggesting that certain lexical patterns may be replicated in suffix behaviour, but also that not all borrowings are conform to the same tendencies. Results thus far suggest the Turkish lexicon may be organised in strata, but unusually those strata would not conform to a strict hierarchy predicted by the time period when each stratum was introduced into the lexicon (cf. LaCharité and Paradis, 1993; Itô and Mester, 1999). Further, monosyllabic words' lower rate of harmony may be explained by their need for greater distinction, particularly relative to neighbours in the native lexicon, and for greater identifiability as borrowings in early stages of their adaptation (see e.g. Poplack et al., 2020).

# References

Barrie, Christopher and Justin Chun ting Ho. 2021. academictwitteR: an R package to access the Twitter Academic Research Product Track v2 API endpoint. *Journal of Open Source Software* 6(62): 3272. doi:10.21105/joss.03272. URL `https://github.com/cjbarrie/academictwitteR`.

Baturay, Semra. 2012. Loan word adaptation and vowel harmony in Turkish: A government phonology account. In *Proceedings of ConSOLE XX*. 1–22.

Bennink, Clemens. 1992. Harmony and disharmony in Turkish. *Linguistics in the Netherlands* 9(1): 25–36.

Bessett, Ryan M. 2016. The role of phonology in discerning lone other-language items as borrowing or codeswitches. In *New waves of analyzing variation (NWAV) 45*. Vancouver, BC.

Chong, A. J. 2019. Exceptionality and derived-environment effects: A comparison of Korean and Turkish. *Phonology* 36(4): 543–572.

Itô, Junko and Armin Mester. 1999. The phonological lexicon. In *The handbook of Japanese linguistics*, ed. Natsuko Tsumijura. Oxford: Blackwell Publishers Ltd, 62–100.

Kiparsky, Paul. 1993. Blocking in nonderived environments. In *Phonetics and phonology 4: Studies in lexical phonology*, ed. E. Kaisse and S. Hargus. San Diego: Academic Press, 277–313.

LaCharité, Darlene and Carole Paradis. 1993. Introduction: The emergence of constraints in generative phonology and a comparison of three current constraint-based models. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 38(2): 127–153.

Lanfranca, Mark. 2012. *An acoustic study of underspecified vowels in Turkish*. Ph.D. thesis, University of Kansas.

Özcelik, Öner. 2014. Turkish language. In *Encyclopedia of Islam and the Muslim world*, ed. M. Richard. New York: Macmillan, 1197–1199.

Poplack, Shana, Suzanne Robillard, Nathalie Dion, and John C Paolillo. 2020. Revisiting phonetic integration in bilingual borrowing. *Language* 96(1): 126–159.

Ylonen, Tatu. 2022. Wiktextract: Wiktionary as machine-readable structured data. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. 1317–1325.