

# Vowel Harmony Loss in West Rumelian Turkish

Andrew Dombrowski (adombrow@uchicago.edu)

LSA 2010: Extended Abstract

## 1 Introduction

The synchronic modeling of vowel harmony has been a topic of recurring interest for phonologists (e.g. Ringen and Heinämäki 1999, Goldsmith and Riggle 2007 on Finnish, Ringen and Vago 1998 on Hungarian, Beckman 1997 on Shona). However, from a diachronic point of view, the mechanisms leading to the evolution and disintegration of harmony systems are less well-understood; agent-based models have been applied successfully to the emergence of harmony systems, but not to their breakdown (Harrison et al. 2002). This paper provides a quantitative analysis of harmony in the lexicon of the Ohrid dialect of Turkish, which has lost productive vowel harmony (Kakuk 1972). As expected, harmony in the lexicon of this dialect is less robust than what has been reported for languages with productively functioning harmony systems. However, it is shown that (a) the lexicon of Ohrid Turkish is still moderately harmonic and (b) correcting for recent loanwords, the invariant morphemes *-miş* and *-çe*, and the merger of final high vowels does not significantly improve the harmony in the lexicon. On the basis of these findings, it is argued that the loss of productive vowel harmony in Ohrid Turkish is best explained through grammatical interference from neighboring Indo-European languages.

Ohrid Turkish belongs to the West Rumelian group of Turkish dialects, which are spoken to the west of a boundary line that runs N-S through western Bulgaria. These dialects are characterized by extensive structural influence from neighboring Balkan languages; see Friedman 2003 for a comprehensive discussion of phonological and especially morphosyntactic innovations in West Rumelian Turkish due to its participation in the Balkan Sprachbund. Language shift seems to have been a mechanism important in the evolution of West Rumelian Turkish; in areas like Kosovo and the Republic of Macedonia, Turkish functioned as an urban prestige language as late as 1945. As a result of this, there are in fact few recent loanwords from other Balkan languages present in pre-1945 West Rumelian Turkish, although grammatical interference seems to be considerable.

In general, these dialects do not have productive vowel harmony. This can be seen in data involving the pluralizing suffix *-lar*. Examples from Ohrid Turkish include: *saç-lar* ‘hair’, *yolcu-lar* ‘travellers’, *anlar-lar* ‘they understand’, *olar-lar* ‘they are’, *can-ler* ‘souls’, *yaş-ler* ‘tears’, *soyar-lar* ‘they peel’, *koyar-ler* ‘they put’, *aga-lar* and *aga-ler* ‘aghas’, *dag-lar* and *dag-ler* ‘mountains’,

*mal-lari* and *mal-leri* ‘their fortune’, *yapmaz-lar* and *yapmaz-ler* ‘they do not do/make it’, *bakti-lar* ‘they saw’, *oturdi-lar* ‘they were seated’, *bozdi-lar* ‘they spoiled’, *çardi-ler* ‘they called’, *yapmış-ler* ‘they did’, *konusmuş-ler* ‘they talked’, *olmuş-lar* ‘they were’, *kaşık-ler* ‘spoons’, *ufacik-ler* ‘little ones’, *kari-ler* ‘women’, *kınalı-ler* ‘people painted with henna’, *komşi-ler* ‘neighbors’, *kuzi-ler* ‘sheep’, *ev-ler* ‘houses’, *kim-ler* ‘who?’, *gider-ler* ‘they go’, *yedi-ler* ‘they ate.’

Some common innovations in the West Rumelian dialect group result in a more disharmonic lexicon, in particular the generalization (or possible preservation) of the invariant evidential marker *-miş*, the enhanced productivity of the diminutive suffix *-çe*, and most crucially, the merger of all word-final high vowels (/i/, /ɯ/, /y/, and /u/) to /i/. Since these factors are known to be common to West Rumelian Turkish and since they all militate against harmony, a null hypothesis would be that lexemes affected by these factors, along with disharmonic loanwords, accumulate in the lexicon to the point where vowel harmony is no longer generalizable.<sup>1</sup> This hypothesis, which may be termed the lexical hypothesis, is argued against in this paper on the basis of the quantitative analysis described below.

## 2 Quantitative Analysis

A quantitative analysis was carried out on the lexicon of Ohrid Turkish as reflected in the Kakuk 1972 dialect description, which is itself based on fieldwork from 1963.<sup>2</sup> Kakuk 1972 contains a grammar outline, poetry, and transcriptions of oral narratives in the Ohrid dialect. These were analyzed separately. In order to assess the impact of recent loanwords, disharmonic *-miş* / *-çe*, and final high vowel merger, each corpus was analyzed both in its original form (yielding “baseline harmony”) and with either the words or syllables containing the above known disharmonic factors discarded (yielding “adjusted harmony”). In order to test for frequency effects in the lexicon, each corpus was analyzed both with and without duplicate tokens.

Tests were performed both on the word level and on the level of bisyllabic domains ( $\sigma\sigma$ ). Turkish vowels were divided based on the features [+/- front], [+/- round], and [+/- high]. Words are said to be harmonic if all vowels in a word agree for [+/- front] and [+/- round].<sup>3</sup> A bisyllabic domain of two vowels ( $\sigma\sigma$ ) is said to be harmonic if both syllables agree for [+/- front] and [+/- round]. A further level of granularity was obtained over  $\sigma\sigma$  sequences by

<sup>1</sup>Harrison et al. 2002 incorporate similar factors as motivations for vowel harmony loss in their attempt to use agent-based modeling to simulate harmony breakdown. Since this did not succeed, they explicitly suggest that “harmony decay is an entirely different mechanism” (Harrison et al. 2002: 15)

<sup>2</sup>As such, Kakuk 1972 is based on speakers who learned Turkish before widespread education either in modern standard Turkish or in standard Macedonian.

<sup>3</sup>This is not equivalent to harmony in standard Turkish, in which rounding harmony operates only over high vowels. However, four-way harmony in high vowels is recent in Turkish (Kerlake 1998: 185), and therefore this metric was chosen as more directly reflective of the robustness of harmony in the lexicon.

ascertaining whether a  $\sigma\sigma$  sequence is harmonic for both, one, or none of the relevant distinctive features.

None of the factors controlled for above (genre, duplicates, word-counting vs.  $\sigma\sigma$  counting) yielded significant differences. Across all 18 trials,  $\%(\text{word}) < \%(\sigma\sigma)$  consistently, but the percentages correlated closely ( $r = .85403$ ). Controlling for duplicates had almost no effect; the results with and without duplicate tokens correlated almost completely ( $r = .98918$ ). Genre effects were more pronounced when counting words than when counting  $\sigma\sigma$  sequences;  $(\mu)\%(\text{word}) = 7.51$  whereas  $(\mu)\%(\sigma\sigma) = 5.74$ .

A comprehensive sampling of baseline harmony in the Ohrid data yields the following values:  $\%(\text{word}) = 44.02$ ,  $\%(\sigma\sigma) = 56.95$ . This result is not surprising; Harrison et al. 2002 define the robustness of backness harmony as 53.8% in Uzbek, which has likewise lost productive vowel harmony.<sup>4</sup> Controlling for known anti-harmonic features does not significantly improve the percentage of harmony in the lexicon; the corresponding figures for adjusted harmony are:  $\%(\text{word}) = 49.58$ ,  $\%(\sigma\sigma) = 59.98$ . The fact that adjusted harmony is only marginally higher than baseline harmony suggests that at least synchronically, the accumulation of non-harmonic elements in the lexicon is not sufficient to account for vowel harmony loss in Ohrid Turkish.

Despite the relatively low levels of baseline and adjusted harmony described above, a closer look at  $\%(\sigma\sigma)$  figures reveals that on a granular level, the lexicon is nonetheless moderately harmonic.  $\sigma\sigma$  sequences can be broken down as harmonic according to both backness and rounding ( $\%\sigma\sigma(\text{B,R})$ ), backness alone ( $\%\sigma\sigma(\text{B})$ ), rounding alone ( $\%\sigma\sigma(\text{R})$ ), or neither feature ( $\%\sigma\sigma(\text{none})$ ). Based on a comprehensive sampling of the data (i.e., including duplicates and anti-harmonic factors,  $n = 3538$ ), the breakdown is as follows:  $\%\sigma\sigma(\text{B,R}) = 57.77$ ,  $\%\sigma\sigma(\text{B}) = 11.61$ ,  $\%\sigma\sigma(\text{R}) = 25.44$ ,  $\%\sigma\sigma(\text{none}) = 5.17$ . This is extremely unlikely to reflect a random distribution of vowels; given the standard Turkish eight-vowel phoneme system, the null hypothesis for random distribution would be to expect 25% in each category. Working with this null hypothesis,  $p < 0.0001$ . Therefore, it can be concluded that harmony is still observable on an extremely statistically significant level in the lexicon of Ohrid Turkish.

### 3 Conclusions

The above results show that the erosion of morphologically productive vowel harmony in Ohrid Turkish cannot be synchronically accounted for by the gradient accumulation of disharmonic lexemes. It is furthermore demonstrated that the lexicon of Ohrid Turkish displays extremely statistically significant vowel harmony on the level of bisyllabic sequences. These findings are not fully compatible with the lexical hypothesis for the breakdown of vowel harmony. The wide presence of contact-induced morphosyntactic innovations in West Rumelian Turkish, combined with the likelihood of widespread language shift in the development

---

<sup>4</sup>However, methodological differences between this study and Harrison et al. 2002 mean that the percentages should not be compared directly.

of West Rumelian Turkish, suggests an alternative hypothesis in which the lack of grammaticalized vowel harmony was introduced into West Rumelian Turkish through grammatical interference from neighboring Indo-European languages, including most saliently Macedonian, Serbian, and Albanian.

If correct, this proposal not only provides a better model for harmony breakdown in West Rumelian Turkish itself, but also furthers our understanding of the breakdown of harmony systems by providing at least one instance in which vowel harmony loss can be shown to have been induced by contact-induced grammatical change rather than through gradient lexical developments. A promising direction of further research would be to determine how many (if any) demonstrable instances there are of harmony being eroded through lexical change. Further elucidation of this question would have clear consequences for the study of the diachrony of vowel harmony systems, as well as possibly raising relevant issues for the synchronic modeling of vowel harmony. These findings also are relevant for the field of contact linguistics insofar as vowel harmony breakdown has not previously been documented as a result of language shift.

## 4 References

Beckman, Jill. 1997. Positional faithfulness, positional neutralisation and Shona vowel harmony. *Phonology* 14: 1-46.

Friedman, V. A. 2003. *Turkish in Macedonia and beyond: studies in contact, typology, and other phenomena in the Balkans and the Caucasus*. Harrassowitz: Wiesbaden.

Goldsmith, John A., and Jason Riggle. 2007. *Information theoretic approaches to phonology: The case of Finnish vowel harmony*. Chicago: University of Chicago, MS. <http://hum.uchicago.edu/~jagoldsm//Papers/boltzmann.pdf>

Harrison, K. David, Mark Dras and Berk Kapioglu. 2002. Agent-based modeling of the evolution of vowel harmony. *NELS* 32.

Kakuk, Suzanne. 1972. Le dialecte turc d'Ohrid. *Acta Orientalia Academiae Scientiarum Hungaricae* 26: 227-282.

Kerslake, Celia. 1998. "Ottoman Turkish." in Johanson, L. and Csató, E., eds. *The Turkic Languages*. London-New York: Routledge: 179-202

Ringen, C. and R. Vago. 1998. Hungarian vowel harmony in Optimality Theory. *Phonology* 15: 393-416.

Ringen, C. and O. Heinämäki. 1999. Variation in Finnish vowel harmony: An OT account. *Natural Language and Linguistic Theory* 17, 303-337