A Corpus Study on Turkish Vowel Harmony: Confusing Observations

In this study, I will report the results of an on-going corpus study on Turkish Vowel Harmony (TVH), which can be described as a process in which a non-initial vowel agrees with the preceding vowel in terms of frontness/backness and, if it is high, also in rounding. After having conducted a dictionary survey and another corpus study (using TS corpus, Sezer & Sezer 2013) based on frequency of use (Avar 2015), the rate of disharmony in Turkish was found to be substantial, especially when we look at loanwords -as already reported in the literature- (cf. Clements & Sezer 1982 among others). In order to test for the psychological reality of TVH, I also conducted psycholinguistic experiments where certain regularities were found with respect to how Turkish native speakers deal with disharmony. In particular, it was found that the majority of speakers tend to use irregular –i.e. non-final- stress patterns when reading out loud disharmonic nonce words embedded in a Turkish narrative (Avar 2016). Upon looking at the phonological shape of frequently used disharmonic words I already had found that the significant majority of them either had an irregular stress and/or had long vowels. Thus I arrived at a hypothesis that stress and vowel length play a role in TVH. In order to further test my hypothesis, I started looking at the rate and nature of disharmony in children's speech.

 I analyzed the data provided by Aksu Corpus at CHILDES. The corpus contains 54 files in the form of transcriptions of conversations with children with an age range of 2;0 to 4;8. Leaving out the adults' utterances, onomatopoetic words, proper names, compounds, and monosyllabic words, the corpus had a total number of 5,557 types and 25,674 tokens (as word-forms). Among those, the rate of harmonic word-forms was 76%. As a next step, I analyzed the nature of disharmonic words. In 792/3,333 (types/tokens) of them, disharmony was caused by a non-alternating suffix (and, in particular, 3,218 due to the progressive suffix –(I)yor which also has an irregular stress-assignment). A closer look at the remaining 557/2,392 revealed a total number of 279 distinct disharmonic roots, the most frequently used ones being *an:e* 'mother' (523 times), and *ta:ne* 'piece' (161 times).

 While the analysis continues, the observations of the data so far are as follows: Among the disharmonic sequences, (i) 24.7% have an irregular stress pattern, (ii) 29.8% have a long vowel, and (iii) the overwhelming majority of *a-u* sequences (7.9% of all disharmonic sequences) can be explained by labial attraction rule (Lewis 1967). The fact that there was a sonorant segment dividing up the disharmonic sequences in the most frequently used words lead me to a further analysis, which revealed that 48.4% of all disharmonic sequences are divided up by *m, n, l* or *r* – the four consonants in Turkish that are known to have an effect on the quality of a preceding /e/ (cf. Operstein & Kütükçü 2004 among others).

 Moreover, if we embrace unary elements rather that binary features as phonological primes, and assume that all root vowels can be lexically specified, it can be argued that true disharmony occurs when an element fails to spread even though the conditions for its spreading are met. From this point of view, for instance, an *a-i* sequence does not violate harmony, while the reverse does. Approaching the data this way, only 20% of the sequences remain to be disharmonic in the strong sense.

 The study is far from being complete, but it aims to shed light on the unsolved issues on TVH by suggesting that the process is more complicated than considered so far in that it is affected by -rather- external factors such as quality of neighboring consonants, stress and vowel quantity.

Selected List of References

Avar, B. (2016). Aliens vs. Allies: The Heterogeneity of Disharmony in Turkish. Paper presented at the *13th Old World Conference in Phonology*, Budapest, 13-16 January 2016.

Avar, B. (2015). Save Harmony: Rebellious Loans Troubling GP. Paper presented at the *12th Old World Conference in Phonology*, University of Barcelona, 27-30 January 2015.

Clements, N. & Sezer, Engin (1982). Vowel and consonant disharmony in Turkish. In H. van der Hulst & N. Smith (ed.s), *The Structure of Phonological Representations* (Part II). Lewis, G. L. (1967). *Turkish Grammar*. New York: Oxford University Press.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk.* Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Operstein, N. & Kütükçü, T. (2004). Türkiye Türkçesi'ndeki /e/ ünlüsünün sesbirimcikleri üzerine [On the allophones of /e/ in Turkish spoken in Turkey]. *Dilbilim Araştırmaları Dergisi, 2004*: 29-50. İstanbul: Boğaziçi Üniversitesi Yayınevi.

Sezer, B. Sezer, T. (2013). TS Corpus: Herkes için Türkçe Derlem. *Proceedings of the 27th National Linguistics Conference*, Antalya, pp. 217\225. Hacettepe University, Linguistics Department.

Slobin, D. (1982). Universal and particular in the acquisition of language. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art*, pp. 128-172. New York: Cambridge University Press.

TDK Türkçe Sözlük (1988). Ankara: Türk Tarih Kurumu Basım Evi.