A segment-specific metric for quantifying participation in harmony

Avery Ozburn, University of British Columbia

Despite a recent trend of investigating gradient co-occurrence patterns in harmony systems (e.g. Pierrehumbert 1993, Frisch et al. 2004), there remains no systematic way of calculating the degree to which a segment participates in a harmony system. Nonetheless, segment-specific participation is a complex phenomenon that could benefit from a more nuanced treatment as gradient and quantifiable. For example, it is well known that certain vowels are typologically favoured as triggers and targets of certain types of harmony (e.g. Kaun 1995 on rounding harmony), and non-participation in harmony seems to be correlated with specific vowel qualities independent of inventory contrast (e.g. low vowels in ATR harmony, which are contrastive but neutral in Mayak; Andersen 1999). If there is a universal basis for these patterns, we might expect them to also be apparent on a gradient level cross-linguistically. Moreover, there is a great deal of variation and idiosyncratic behaviour in harmony, which suggests that neutrality and participation are not clear-cut categories (e.g. 'neutral' [e:] in Hungarian alternates harmonically with [a:]; Siptár and Törkenczy 2000). Combined, the typological patterns in and the variability of harmony participation suggest that a segment-specific, gradient measure of harmony participation could illuminate interesting language-internal and cross-linguistic patterns that are not apparent from categorical descriptions or existing quantificational measures.

Observed/Expected (O/E) values give a measure of the gradient behaviour of each individual pair of segments, but not a comprehensive picture of how harmonic a segment is, across its co-occurrences with all segments. Moreover, it is insufficient to measure participation as the percentage of the segment's co-occurrences that are harmonic, because that fails to take into account the relative frequency of the two harmonic categories. Instead, I adopt *relative risk* or *risk ratio* (RR) to compute segment harmony participation. RR is a statistical measure used to determine the relative probability of an event in one condition versus another (Agresti 2013). Interpreting the event as a specific vowel, and the conditions as its harmonic versus non-harmonic contexts, RR provides a frequency-corrected measure of the probability that a vowel occurs in harmonic contexts, and therefore of how harmonic it is. RR has previously been used in linguistics for speech errors (Tupper and Alderete in prep). Since all relevant counts are known in the present case, the simple formula in (1) can derive the RR for vowel V from category C1.

$$\text{(1) } RR(V) = \frac{(\#\text{ V in context C1}) \text{ x } (\#\text{ anything in context C2})}{(\#\text{ V in context C2}) \text{ x } (\#\text{ anything in context C1})}$$

To test the validity and usefulness of RR(V), I examine front/back harmony in corpora of Chuvash, Tatar, and Mari (Luutonen et al. 2007; Luutonen et al. 2016). The corpora are dictionary lists that include derived forms, providing a picture of broad harmonic behaviour throughout the lexicon. Using Phonological CorpusTools (Hall et al. 2015), counts of all pairs of syllable-adjacent vowels were obtained for all corpora, which were used to compute O/E values for all vowel pairs, as well as RR(V) for each vowel. Results show a high degree of correspondence between RR(V) and the informal impression of how much V participates in harmony, based on its O/E values. For example, in Chuvash, [y] almost never occurs before back vowels, with O/E values ranging from 0 to 0.17; RR(y) is very high, at 66.85, indicating that the relative probability that [y] will occur in front (harmonic) contexts is very high. In contrast, [i] occurs fairly regularly before back vowels in Chuvash, with O/E values ranging from 0.24 to 0.91; correspondingly, RR(i) is low at 2.30, indicating that it does have a greater relative probability of occurring in harmonic contexts, but not by nearly as much as [y]. The vowel [ɯ] is intermediate between [y] and [i], with RR(ɯ)=10.31 and O/E values before front vowels ranging from 0.02 to 0.35. Similar correspondences exist for other vowels and languages, suggesting that RR(V) does capture the intuitive concept of participation in harmony.

With the concept validated, I compare the RR values across vowels within a language. Most notably, in all three languages, RR(i) is very low, meaning that [i] has a low degree of participation in the harmony systems. This result is particularly interesting because, across front/back harmony systems, [i] is a very common transparent vowel, and this pattern has often been connected to the lack of a back counterpart to [i] in languages like Finnish (e.g. Kiparsky and Pajusalu 2003). In Chuvash, Tatar, and Mari, [i] is not described as being neutral and has a back counterpart, yet RR(i) shows that it nonetheless participates less in the harmony system, in a gradient way. This suggests that the cross-linguistic behaviour of [i] as a neutral vowel is due to reasons beyond harmonic pairing and beyond categorical properties, and that we need to re-examine more carefully the factors involved in harmony participation.

In summary, this paper develops a method of calculating segment-specific harmony participation. By comparing RR(V) to O/E values in three corpora, I show how RR(V) captures the informal notion of harmony participation. I then examine the implications of gradient participation. I show that [i] has a low degree of participation in these harmony systems, which corresponds to the fact that [i] is often neutral cross-linguistically, but is interesting because this neutrality is often contributed to factors not present in the examined languages. I argue that RR(V) provides a more nuanced, cross-linguistic way of looking at participation in harmony, and that this approach provides significant new insights into harmony systems.

References

Agresti, Alan. 2013. *Categorical data analysis*. John Wiley & Sons.

Andersen, Torben. 1999. Vowel harmony and vowel alternation in Mayak (Western Nilotic). *Studies in African Linguistics*, 28(1), 1-30.

Frisch, Stefan A., Janet B. Pierrehumbert, & Michael B. Broe. 2004. Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22: 179–228.

Hall, K. C., Allen, B., Fry, M., Mackie, S., and McAuliffe M. 2015. Phonological CorpusTools, Version 1.1.0. [Computer program]. Available from https://sourceforge.net/projects/phonologicalcorpustools/.

Kaun, Abigail R. 1995. *The Typology of rounding harmony: an optimality theoretic approach*. Los Angeles, CA: UCLA dissertation.

Kiparsky, Paul & Karl Pajusalu. 2003. Toward a typology of disharmony. *The Linguistic Review*, 20, 217-241. DOI: 0167–6318/03/020-0217

Luutonen, Jorma, et al. (ed.). 2007. *Electronic Word Lists: Mari, Mordvin and Udmurt. With SFOu WordListTool 1.3*. Lexica Societatis Fenno-Ugricae XXXI:1. ISBN 978-952-5150-98-8. Helsinki.

Luutonen, Jorma, et al. (ed.). 2016. *Electronic Word Lists: Komi, Chuvash and Tatar. With SFOu WordListTool 1.4*. Lexica Societatis Fenno-Ugricae XXXI:2. ISBN 978-952-5667-79-0. Helsinki.

Pierrehumbert, Janet B. 1993. Dissimilarity in the Arabic verbal roots. *Proceedings of NELS 23*: 367–381.

Siptár, Peter, and Miklos Törkenczy. 2000. *The phonology of Hungarian*. Oxford: Oxford University Press.

Tupper, Paul, and John Alderete. In prep. Risk ratios for linguists.