An ‘unnatural’ pattern of variation in vowel harmony: a frequency-based account

Péter Rebrus − Miklós Törkenczy
Research Institute for Linguistics, Hungarian Academy of Sciences (MTA)
Eötvös Loránd University (ELTE), Budapest {rebrus, tork}@nytud.hu

Variation and transparency in Hungarian front/back harmony is known to be sensitive to the height of the neutral vowels in the language /i(ː)/, /eː/, /ɛ/ (dubbed the Height Effect by Hayes & Cziráky Londe 2003 (H&C)) and the proximity of a back trigger to a target separated from it by neutral vowels (called the Count Effect by H&C). Various analyses of these effects have been proposed but very little attention has been given to the *interaction* of these two effects. The extant approaches (Bowman 2013, H&C) assume, implicitly or explicitly, that the two effects are ‘additive’ in the sense that they reinforce each other when both can apply. In this paper we want to show that this assumption is empirically incorrect in some cases and propose an analysis that accounts for the non-additive nature of the interaction.

The Height Effect (HE) consists in the fact that higher neutral vowels are easier to skip (i.e. are more transparent) than lower ones. Accordingly, there is a hierarchy of vowel transparency in Hungarian (below X>Y means that, when suffixed with a harmonically alternating suffix, a stem of the type X is more likely to co-occur with a back suffix alternant than a stem of the type Y; B is a back vowel; consonants are not indicated):

(1) HE Bi(ː) > Beː > Bɛ

Thus, in a [Bi(ː)]_ context (where [,] are morpheme boundaries) the harmonic value of a harmonically alternating suffix is back (*B*): e.g. *kotʃi-nɒk/*nɛk* ‘car+DAT’; in a [Beː]_ context there is (lexically conditioned) variation, but the harmonic value in the suffix is more likely to be *B* than front (*F*): e.g. *kaːveː-nɒk/*nɛk* ‘coffee+DAT’, *ɒrzeːn-nɒk/nɛk* ‘arsenic+DAT; in a [Bɛ]_ context there is variation, but the value is more likely to be *F* than *B*, e.g. *kontsɛrt-*ok/ɛk* ‘concert+PL’ and *fotɛl-ok/ɛk* ‘armchair+PL’.

The Count Effect (CE) means that a sequence of more than one neutral vowel (N) is more difficult to skip (i.e. less transparent) than a single one. This again results in a hierarchy in transparency:

(2) CE $BN > BNN^+$

Thus, for instance, a single high N in a [Bi(ː)]_ context is fully transparent so a following harmonically alternating suffix is *B* (in accordance with HE), but variation occurs after more than one high N, in the context [Bi(ː)i(ː)]_, e.g. *ɒlibi-nɒk/nɛk* ‘id.+DAT’.

The question we address here is whether CE and HE are additive, i.e. whether it is true that a sequence of N vowels less transparent by CE is necessarily less transparent than a sequence (of the same number) of N vowels which are more transparent by CE, as in (3):

(3) Additive interaction

$$\text{If } BN_i > BN_j \text{ then } BN_iN_i > BN_iN_j,\ BN_jN_i > BN_jN_j$$

(3) seems to hold true in some cases, e.g. there is variation between *F* and *B* suffix alternants in the context [Bii]_ *ɒlibi-nɒk/nɛk* ‘id.+DAT’, but in the context [Biɛ]_ suffix alternants are practically always *F*: *kɒlibɛr-*nɒk/nɛk* ‘calibre+DAT’. This is to be expected by (3) since Bi > Bɛ. However, [Bii] vs. [Bieː]_ (especially when eː is root final) do not conform to (3). The suffix alternants in the latter context are predominantly *B* although Bi > Beː, e.g. *mɒtineː-nɒk/?nɛk* ‘matinée+DAT’.

(4) CE+HE

a. Bii > Biɛ (additive interaction)
b. Bii < Bie: (non-additive interaction)

Why is it that [Bieː]_ stems do not follow a general strategy like (3) for suffix harmony when this strategy *is* otherwise available, cf. (4a)? This ‘unnatural’ behaviour (i.e. ungrounded in markedness or phonetics, cf. Hayes *et al* 2009) of the harmonic context [Bieː]_ is puzzling if we want to derive it from the *inherent* properties of the context itself but can be explained with reference to its connections to other, partially similar contexts.

In an analogy-based approach (e.g. Bybee 2007) the behaviour of a pattern (analogical target) is related to that of other similar patterns (analogical sources), the strength of whose influence depends on their frequency and their degree of similarity to the target. The greater these are, the stronger the connection is between the source and the target. Variation in a target pattern is the result of conflicting sources with approximately equal strengths (e.g. Kálmán *et al.* 2012). In this spirit, the variation in the context [BNN]_ can be interpreted as a result of conflicting harmonic behaviour in partially similar contexts, namely [BN]_ and [NN]_ (relativised to the specific N segments). As [Bi] stems have *B* suffixes, (in line with HE) and [ii] stems always get *F* suffixes (e.g. *kifli-*nɒk/nɛk* ‘crescent roll+DAT’), this conflicting behaviour of the analogical sources results in variation in the case of [Bii] stems.

Frequency asymmetries in the sources result in strength differences between the source-target connections and therefore different target behaviour. We argue that this approach to variation *can* explain the unnatural harmonic behaviour of the [Bieː]_ context. The predominance of *B* suffixes in this context is due to the relative weakness of the [NN] (=[ieː]) analogical source: although the context [ieː]_ invariably has an *F* suffix (e.g. *file:-*nɒk/nɛk* ‘filet+DAT’), the frequency of these forms is strikingly low and cannot ‘counterbalance’ the conflicting influence of the other sources, the [BN] contexts [[Bi/e:]_, where predominantly *B* suffix alternants occur. By contrast, *F*-suffixed forms of [ii] stems are very frequent and therefore exert a greater influence on the [Bii] context resulting in near-free variation in suffix harmony. (5) shows these connections for the special case of vowel-final roots with the number of lemmas and the token frequency of suffixed forms from the Szószablya webcorpus containing 103k lemmas and 541M word tokens (Halácsy *et al.* 2004). Note that the frequency data of the type [[ii]F] are greater than those of [[ie:]F] by an order of magnitude.

(5) Analogical sources of [BNN]-targets with frequencies (lemma; token)

a. [[Bi]**B**] frequent → [[Bii]**B/F**] ← [[ii]**F**] **frequent** (71; 21,800)

[[Bii]**B/F**] B≈F (16; 666≈720)

b. [[Bi/e:]**B**] frequent → [[Bie:]**B**/(F)] ⇠ [[ie:]F] **rare** (8; 2,200)

[[Bie:]**B**/(F)] B>>F (7; 1869>>41)

We conclude by arguing that the significance of the phenomenon is that an unnatural pattern may be not simply exceptional but may have an explanation that lies in its frequency-sensitive connections to patterns which are sufficiently similar to it.

References

Bowman, Samuel R. 2013. Two arguments for a positive vowel harmony imperative. Ms. [ROA 1181]

Bybee, Joan. 2007. *Frequency of use and the organization of language*. Oxford: OUP.

Halácsy, P., A. Kornai, L. Németh, A. Rung, I. Szakadát & V. Trón. 2004. *Creating open language resources for Hungarian.* In Proceedings of LREC04, 203–210. LREC.

Hayes, Bruce & Zsuzsa Cziráky Londe (2006). Stochastic phonological knowledge: The case of Hungarian vowel harmony. *Phonology* 23. 59–104.

Hayes, Bruce, Kie Zuraw, Péter Siptár, & Zsuzsa Londe. (2009). Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85. 822-863.

Kálmán, László, *et al*. 2012. Possible and impossible variation. In: F. Kiefer et al. (eds.) *Current Issues in Morphological Theory. (Ir)regularity, Analogy and Frequency.* Amsterdam: John Benjamins. 23–49.