How to measure incomplete harmony

Péter Rebrus & Miklós Törkenczy; Research Institute for Linguistics, Hungarian Academy of Sciences

Harmony is typically not complete in vowel harmony systems, i. e. they usually show some degree of disharmony (e.g Kiparsky & Pajusalu 2003). In root controlled systems with affix harmony (Rose & Walker 2011) the following types of disharmony can be distinguished: (i) root-internal disharmony (mixed roots), (ii) disharmony involving harmonically invariant affixes, (iii) disharmony involving harmonically alternating affixes (transparency/opacity, anti-harmony). Abstract phonological analyses focusing on harmonic regularity at underlying level dispense with some of these disharmonies. They typically employ two – sometimes cooccurring – strategies: they *reanalyse* some surface disharmonies as underlyingly harmonic, and/or they mark some instances and/or types of surface disharmony as *irregular*. The problem with these strategies is that they are highly theory-dependent and often arbitrary/stipulative. As a result they often obscure the overall degree of harmony of a system and make the different types of disharmony within a language and the harmony systems of different languages difficult to compare or incomparable. Furthermore, they tend to disregard static phonological patterns of harmony/disharmony (types i and ii above) and overemphasize harmony as manifested in alternations (type iii above).

In this paper we propose an approach in which harmony can be measured quantitatively and we present an overview of Hungarian front/back harmony using this measure. In the analysis we will carefully distinguish the effect of each disharmonic sequence (type i-iii above) on the overall harmonic character of Hungarian. All the three types of disharmony occur in Hungarian. Disharmony involving harmonically mixed roots (i) and non-harmonizing suffixes (ii) can both occur with neutral and non-neutral vowels as well. These types of disharmonies have been identified in the literature (e.g. Siptár & Törkenczy 2000), but the extent of their influence of the overall pattern has not been explored (especially when non-neutral vowels are involved). The extent of disharmony involving alternating suffixes (type iii) is difficult to assess without a quantitative study because of the high level of variation associated transparency and anti-harmony in some forms. For instance, the lexical statistics of anti-harmonic and harmonic roots are known, but the frequency of the suffixed word forms of these stems and their concomitant effect on the overall harmonic pattern has been uncharted by previous studies.

Our quantitative study is based on the corpus statistics of vowel combination in word forms (source: SzóSzablya webcorpus: 5,4 bn tokens, 2,3M word types; cf. Halácsy *et al.* 2004). We have calculated the observed/expected (O/E) ratios for all bisyllabic combinations of vowel types (Archangeli *et al.* 2012). This approach is more refined than lexical statistics and calculations by pure token frequency, and can potentially (dis)prove the rarity of certain types of disharmony.

(a) within word (86,6k word types)			(b) stem+suffix (37,7k word types)			(c) stem+harmonizing suffix (30,5k word types)		
BB 1.55	BN 0.52	BF 0.10	[B]B 1.80	[B]N 0.41	*[B]F 0.00	[B]$B_{\sim F}$ 1.76	*[B]$N_{\sim B}$ 0.00	*[B]$F_{\sim B}$ 0.00
NB 0.51	NN 1.55	NF 1.39	[N]B 0.18	[N]N 1.74	[N]F 1.63	[N]$B_{\sim F}$ 0.18	[N]$N_{\sim B}$ 2.24	[N]$F_{\sim B}$ 1.62
FB 0.08	FN 1.46	FF 3.81	[F]B 0.00	[F]N 1.48	[F]F 3.45	*[F]$B_{\sim F}$ 0.00	[F]$N_{\sim B}$ 1.82	[F]$F_{\sim B}$ 3.61

O/E values for vowel combinations in bisyllabic words by vowel types (Back, Front, Neutral)

The following observations can be made from the chart above: (a) *within word*: ▸ disharmonic sequences with non-neutral vowels (BF, FB) are very rare; ▸ disharmonic sequences with neutral vowels are more frequent (BF<<BN, FB<<NB), but much less

frequent than the comparable front vowel combinations (BN<FN, NB<NF); ▸ these patterns are symmetric: approximately the same number of disharmonic words occur in each type independently of their order (BF≈FB, BN≈NB); ▸ the high O/E-value for FF is caused by front/back and rounding harmony together. (b-c) *suffixed stems*: ▸ there are no invariant F-suffixes and anti-harmonic B-stems ([B]F=0), but invariant N-suffixes occur frequently [B]N>>0); ▸ invariant B-suffixes do occur, but their frequency is statistically negligible (F[B]≈0); ▸ anti-harmonic N-stems occur not very frequently ([N]B is low); the symmetry between [B]N and [N]B does not hold because of the different statistical effects of invariant suffixes and anti-harmony.

One of the most puzzling phenomena concerning neutral vowels is the Height Effect (Hayes *et al* 2006, 2009), which expresses the graduality of neutralness of front unrounded vowels: the higher the vowel, the more neutral behaviour it exhibits. This effect can be quantified by calculating the O/E values of disharmonic sequences containing each individual neutral vowel.

(a) within word							(b) stem+suffix					
BB	Biː	Bi	Beː	Bɛ	BF		[B]B	[B]iː	[B]i	[B]eː	[B]ɛ	*[B]F
1.55	0.85	1.09	0.50	0.16	0.10		1.80	(--)	1.04	0.56	0.00	0.00
BB	iːB	iB	eːB	ɛB	FB		[B]B	[iː]B	[i]B	[eː]B	[ɛ]B	[F]B
1.55	1.00	0.86	0.30	0.34	0.08		1.80	1.06	0.22	0.04	0.00	0.00

O/E values for combinations with back vowels in bisyllabic words

The following salient patterns can be identified by the values above: ▸ the Height Effect is verified both *within the word and for suffixed forms*, and both for NB and for BN combinations by the monotonically decreasing values for different Ns (*i*(ː)>*e*ː>*ɛ*); ▸ the only real neutral vowels are *i* and *i*ː within the word they occur in the same number as expected (Bi(ː)≈1, i(ː)B≈1) while the other neutral+back combinations have much lower O/E values; ▸ type Bɛ is very rare compared to the other NB and BN values. *In suffixed forms only (b)*: ▸ *ɛ* very rarely occurs in invariant suffixes and anti-harmonic stems, hence statistically [B]ɛ=0 and [ɛ]B=0; ▸ the contribution of *e*ː in anti-harmony is very low ([eː]B≈0); ▸ the most striking (and unexpected) difference is between the anti-harmonicity of *i* and *i*ː, the former is rather weakly anti-harmonic, while for the latter the frequency of occurrence in anti-harmonic stems is equal to the expected value ([i]B<<[iː]B≈1).

We will discuss how the quantification of these patterns can help explain some notoriously elusive quirks in Hungarian backness harmony: e.g. the variability and the high degree of vacillation in harmonic suffixation after type Bɛ (e.g. *hotɛl-ok/ɛk* 'hotels') is due the rarity of the type and the related fact that the relevant stems are typically relatively recent loans; the "half-productive" character of antiharmony with recent loans in long *i*ː (%*fiːd-ol* 'feed') vs. short *i (klikk-ɛl* 'click') is related to the asymmetry between *i*ː and *i* shown above.

References. Archangeli D, Mielke J & Pulleyblank D. 2012. Greater than noise: frequency effects in Bantu height harmony, in Phonological Explorations: Empirical, Theoretical and Diachronic Issues, eds Botma B, Noske R. Berlin: Mouton de Gruyter; 191–222. ♦ Siptár, Péter & Miklós Törkenczy. 2000. *The phonology of Hungarian*. Oxford: OUP. ♦ Halácsy P, Kornai A, Németh L, Rung A, Szakadát I & Trón V. 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. ♦ Kiparsky, Paul & Karl Pajusalu. 2003. Towards a typology of disharmony. *The Linguistic Review* 20. 217–241. ♦ Rose, Sharon & Rachel Walker. 2011. Harmony Systems. In J Goldsmith, J Riggle & A Yu (eds.), *Handbook of Phonological Theory.* 2nd ed., 240–290. Cambridge, MA: Blackwell.