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**)
{rebrus, tork}@nytud.hu Eötvös Loránd University (**ELTE**), Budapest

plan

1 variation in Hungarian front/back harmony
2 interactions
3 background: earlier analyses
4 a frequency-based explanation
5 unnatural constraints

1 variation in vowel harmony

Hungarian vowel harmony

- harmonic vowels: B={u uː o oː ɒ aː}, F={y yː ø øː}
- neutral vowels: N={i eː ɛ}
- transparency: generally BN+B,

but it is more complicated

effects modifying transparency

- **height effect** (HE): lower neutral vowels are less transparent than higher one, ie. in frontness scale:

Bi/i: < **Be**: < **Bɛ**

  - ***categorical***:
    back (Bi+B) vs. vacillating Be+F/B and Bɛ+F/B
  - ***gradual*** effect in variation:
    back majority (**Be**+B/F) vs. front majority (**Bɛ**+F/B),
    frontness ratios:

word types:	**12.4%**	vs.	**74.7%**
word tokens:	1.4%		94.1%

- **count effect** (CE): A sequence of more than one neutral vowel is less transparent than a single one, ie. the frontness scale:

BN < **BNN(N)**

  - ***categorical***:
    back (Bi+B) vs. vacillating (Bii+F/B) vs. front (Bɛɛ+F)
  - ***gradual*** effect in variation:
    back majority (**Be**+B/F) vs. front majority (**Bie**+F/B)
    frontness ratios:

word types:	**12.4%**	vs.	**57.9%**
word tokens:	1.4%		58.0%

2 interactions

- **simple additivity**: Only the height ranking of *N closest to the suffix* V and the distance of the suffix V from the B source matters, ie. the frontness scale:

(1) **BNi** < **BNe** < **BNɛ** for all Ns
but BiN = BeN = BɛN for a specific N

- **complex additivity**: The height ranking of *both Ns* and the distance of the suffix V and each N from the B source matters, ie.

(2) **BNi** < **BNe** < **BNɛ** *for a specific* N
(3) **BiN** < **BeN** < **BɛN** for a specific N

- **unidirectionality** the combined effect of CE&HE is unidirectional with the effect of CE and HE under both simple and complex additivity

* This work has been supported by National Scientific Grant OTKA-104897 'Variation in Phonology'.

interaction in Hungarian

- **additive interaction**

	Bii	vs.	Bei
word types:	56.0%	vs.	67.4%
word tokens:	9.4%		86.7%

	Bɛi	vs.	Bɛe
word types:	72.6%	vs.	86.4%
word tokens:	69.2%		95.6%

- non-additive interaction?

	Be	vs.	Bie
word types:	**56.0%**	vs.	**57.9%**
word tokens:	9.4%		58.0%

factors in non-additivity? consider the very different harmonic behaviour of C-final and V-final roots:

	BieC#	vs.	Bie#
word types:	**75.3%**	vs.	**33.3%**
word tokens:	92.1%		2.9% (**!**)

V-final Bii vs. Bie show non-additive interaction

- **non-additive interaction**

	Bii#	vs.	Bie#
word types:	**63.4%**	vs.	**33.3%**
word tokens:	95.9%		2.9%

while C-final Bii vs Bie show additive interaction

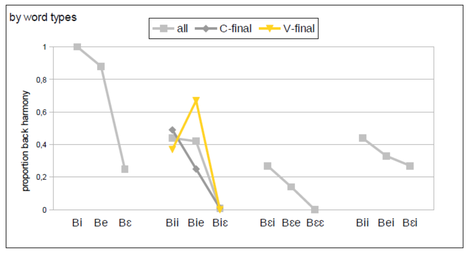
- **additive interaction**

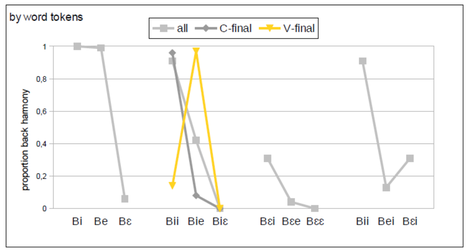
	BiiC#	vs.	BieC#
word types:	50.6%	vs.	75.3%
word tokens:	3.7%		92.1%

the whole picture

	all	i	e	ɛ	C# V#	i	e	ɛ
--		0.00	0.12	0.75		0.00 0.00	0.15 0.03	0.75 --
i		**0.56**	**0.58**	0.99		**0.51** **0.63**	**0.75** **0.33**	0.99 --
e		0.67	(0.93)	1.00		0.70 0.61	-- (0.93)	1.00 --
ɛ		0.73	0.86	1.00		0.70 0.77	0.86 --	1.00 --

frontness ratios of BN(N) stems counted in word types (from Szószablya webcorpus, Halácsy et al. *2006)*





3 background (simple additivity)

Hayes and Londe (2006)

- ***stochastic OT analysis***
- *Hungarian parameters set to model test results*
- google type frequencies, 2 speakers' judgements and wug test of BN and BNN word-forms with B and F alternants of a single suffix (dative)

backness ratios:

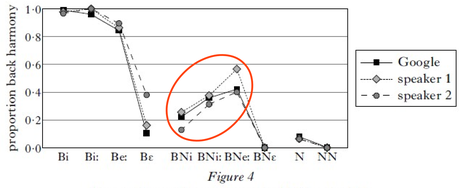


Figure 4
Two native speakers compared with Google data.

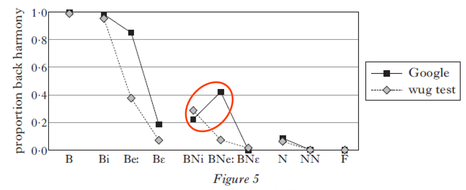


Figure 5
Wug-test data compared with Google data.

- ***based on the wug test results H&L assume that the interaction is additive:***

> "[...] in the wug data [...] BNi stems took more back suffixes than BNe: stems, which in turn took more back suffixes than BNɛ stems. The data from the Google survey contradicted this pattern, with more back responses for BNe: than for BNi. In our view, it is the Google data that most likely are aberrant." (p73)

- H&L constraints and rankings ***disregard the potential effect of N that is not the closest to the sfx***
- ***H&L analysis for the combined effect of CE+HE is simply additive***
  - its effect is always unidirectional with CE and HE: **not true of Bie# roots**
  - the order of N_1 and N_2 in a BN_1N_2+V matters: **true of Bei# vs. Bie# roots**

Bowman (2013)

- ***trigger competition analysis*** (serial harmonic grammar with positive constraints)
- Hungarian parameters set to model (simplified) literature data
- TS trigger strength TS≥1 B=4; ε=2; i,e=1
  DM distance multiplier DM<1 DM=0.5
  d distance *d*≥0 number of Vs from suffix

$$f(x) = TS \cdot DM^{d}$$

- formula always compares 2 triggers and determines if the the harmonic influence of a given trigger vowel wins over that of another
- ***Bowman formula for the combined effect of CE and HE is simply additive***
  - its effect is always unidirectional with CE and HE **not true of Bie# roots**
  - the order of N_1 and N_2 in a BN_1N_2+V matters **true of Bei# vs. Bie# roots**

4 a frequency-based explanation

- **analogical source and target**: the behaviour of a pattern (*analogical target*) is related to other similar patterns (*analogical sources*), the strength of whose influence depends on their frequency and their degree of similarity to the target
- patterns of ***low frequency*** (targets) are influenced by patterns of ***high frequency*** (sources)
- the greater the frequency of the source is and the greater the similarity between the source and the target is, the stronger the influence of the source is on the target
- ***variation*** in a target pattern is the result of ***conflicting sources*** with approximately equal and great enough strengths

why can BNN stems vacillate?

- **similarities**: the set of potential sources may be numerous, but we only consider the two strongest sources and disregard those that have low frequency and weak similarity to the target, e. g. for target Bii:

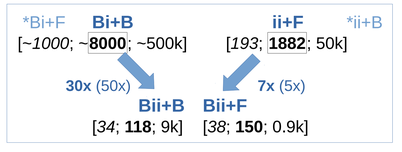
a.	**B i i** ~ **i i**	last two vowels
b.	**B i i** ~ **B i**	harmonic vowel & last vowel
c.	**B i i** ~ **i**	last vowel
d.	**B i i** ~ **B**	harmonic V
e.	**B i i** ~ **B i e**	harmonic V & other neutral Vs
	etc.	

there may be differences in similarity to the target between the two relevant sources because the source in (a) contains both vowels closest to the suffix eg. a>b:

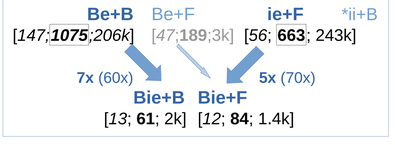
sim(**ii**, Bii) > sim(**Bi**, Bii)

- **frequencies** of harmonically suffixed forms
[*number of lemmas*; **number of word types**; number of tokens]

plain vacillation: Bii+B/F (F-ratio: 56.0%; 9.4% in tokens)



plain vacillation: Bie+B/F (57.9%; 58.0% in tokens)

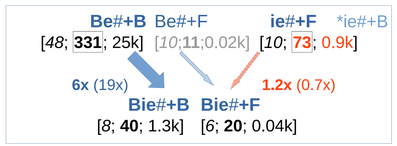


why is there a difference between the V-final and the C-final patterns? Bie# vs. BieC#

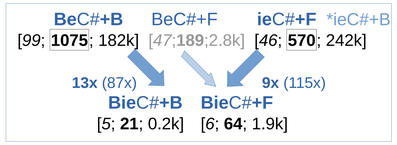
difference in *similarity* and *frequency.*

- **similarity**: sim(ie#, Bie#) > sim(ieC#, BieC#)
because the **greater diversity of consonants among C-final stems reduces the similarity**, thus the analogical influence of a V# source on a V# target is greater (other things being equal) than a VC# source on a VC# target (in the case of different C's)
- **frequency**: for BieC# stems the analogical influences are in equilibrium while for Bie# stems **the extreme low frequency of one of the sources results in asymmetry**

back majority: Bie# (33.3%; 2.9% (!) in tokens)



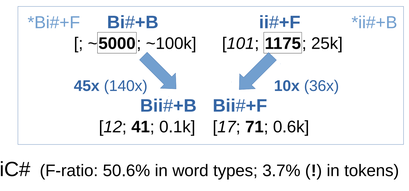
front majority: BieC# (75.3%; 92.1% in tokens)



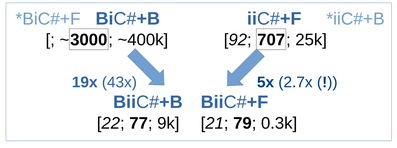
why we do not have this effect among V-final (and C-final) Bii-stems?

- no extreme ***frequency asymmetry*** between the sources
- reverse relation of the F-ratios: Bii# > BiiC#

Bii# (F-ratio: 63.4% in word types; 85.9% in tokens)



BiiC# (F-ratio: 50.6% in word types; 3.7% (!) in tokens)



To sum up

- unexpected deviations from otherwise pervasive patterns can be explained by **asymmetries in frequency**
- although vowel harmony is based on the properties of vowels, in the case of variation, the harmonic patterns may be influenced by some **properties of stem-final segments** that are independent of the harmonizing feature(s). differences in these properties result in **differences in similarities** between forms thereby causing differences in analogical influences.

5 unnatural constraints

comparison with Hayes et. al. (2009)

Hayesian unnatural constraints (e.g. stem-final sibilants promote F alternant in vacillation)

- are based on asymmetries in frequency of forms and are not grounded in phonetics or markedness
- reduce variation by promoting the frequency of the F suffix alternant
- Hayesian unnatural constraints are ***unidirectional*** with CE and HE and additive CE+HE interaction
- non-additive interaction of CE & HE in Bie# roots is unnatural and ***counter-directional***: it promotes the B alternant