

Vowel ~ zero alternations in Hungarian inflectional and derivational nominal:

An analogy-based statistical approach

Lilla Magyar

*Theoretical Linguistics Department
Eötvös Loránd University, Budapest*

1. The phenomenon: vowel ~ zero alternations in Hungarian nominal paradigms

Hungarian vowel ~ zero stem alternations:

- | | | |
|-----|--|---|
| (1) | <i>tor<u>o</u>ny</i> 'tower' nom. sg. | <i>tor</i> Ø <i>nyok</i> nom.pl. |
| | | <i>tor</i> Ø <i>nyot</i> acc.sg. |
| | <i>kor<u>o</u>m</i> 'soot' nom.sg. | <i>kor</i> Ø <i>mos</i> 'sooty' |
| | | <i>kor</i> Ø <i>mot</i> acc.sg. |
| | <i>baj<u>u</u>sz</i> 'moustache' nom.sg. | <i>baj<u>u</u>szt</i> ~ <i>baj</i> Ø <i>szot</i> 'moustache' acc.sg. |
| | | <i>baj<u>u</u>szok</i> ~ <i>baj</i> Ø <i>szok</i> 'moustache' nom.pl. |

Hungarian vowel ~ zero stem alternations, where the stem vowel is present in the unsuffixed stem as opposed to the suffixed ones, have always posed a difficulty for traditional generative models as such alternations cannot be analysed either as cases of insertion or deletion for widely discussed reasons.

- | | | |
|-----|---|-----------------------|
| (2) | <i>tor<u>o</u>ny</i> (*tor Ø ny) | |
| | <i>tor</i> Ø <i>nyok</i> (*tor <u>o</u> nyok) | |
| | 'tower' nom. sg.-nom.pl. | <i>szur<u>o</u>ny</i> |
| | | <i>szur<u>o</u>ny</i> |
| | | 'bayonet' |

Furthermore, these stem alternations exhibit some inter-speaker or even intra-speaker vacillation: some of the stems may or may not exhibit vowel ~ zero alternation (c.f. (1) *bajuszt* 'moustache' acc.sg. vs. *bajszot*).

2. Uniformity in inflexional and derivational paradigms

- McCarthy (2002):
 - o Only derivational paradigms have bases (as derivational forms are clearly derived from the base form).
 - o Inflectional paradigms do not have an analogical source in the same way: there is a correspondence relation from every paradigm member to every other paradigm member, i.e. members of the paradigm equally influence each other.
- Bybee and Brewer (1980) and Benua (1997):
 - o Morphologically derived surface forms stand in an Output-Output relation with their common base form.
 - o Inflectional forms are of equal complexity, therefore the base form is not immediately apparent.
- Albright (2002, 2004):
 - o Derivational paradigms are determined on the basis of morphological/morphophonological markedness
 - o Inflectional paradigms have bases, too. Inflectional paradigms are determined not on the basis of morphological / morphophonological

markedness but on that of the informativeness of the relevant lexical properties.

3. The objective of the analysis is to answer the following questions:

- Is there a paradigm uniformity effect in the inflectional and derivational paradigms of (potentially) vowel-zero alternating stems?
- If so, is the analogical source selecting process different in the case of inflectional and derivational paradigms (as predicted by previous studies)?
- Does frequency play a role in the selection of the base in paradigms?
- Are inflectional and derivational paradigms strictly categorical or rather gradient?

4. The data

- Collected from the Hungarian Webcorpus and the Szószablya frequency dictionary (Halácsy et al., 2004).
- The samples contain all the possible inflectional and derivational forms of nouns which are potentially capable of vowel ~ zero alternation
- Inflectional suffixes which trigger vowel ~ zero alternation (Rung 2010)
- Inflectional suffixes which do not or partly trigger vowel ~zero alternation
- Vowel ~ zero alternating derivational forms

Systematised in the following way:

- Base forms (uninflected forms, not necessarily analogical bases) (e.g. *irodalom* 'literature')
- Inflectional suffixed forms (e.g. *irodalma* 'his literature'),
- Derivational forms (e.g. *irodalmár* 'literary person')
- Compounds (e.g. *irodalomszociológia* 'sociology of literature')

5. Analysis

- we established the token frequency of each group (the number of base forms, inflectional and derivational forms – in the case of simple words and compounds)
- we found the most frequent suffixed form in the paradigm (both inflectional and derivational) of each stem
- we counted the number of vowelless and vowelised variants of the most frequent suffixed form in each paradigm
- examined the relationship between the most frequent (vowelised or vowelless) variant of the most frequent suffixed form and the overall token frequency of vowelised and vowelless suffixed forms in each inflectional and derivational paradigm

6. Results

- (a)** The structure (vowelless or vowelised) of the most frequent suffixed form in each inflectional and derivational paradigm seems to determine the distribution of vowelised and vowelless forms in the same paradigm

izom 'muscle'

Inflectional paradigm

- the most frequent suffixed form of *izom* 'muscle' is *izmok* 'muscle-pl', which only occurs in its vowelless variant (**izomok*)
- the distribution of vowelless variants in the paradigm is 98, 78%

The most frequent suffixed form in the inflectional paradigm	With vowel
<i>izmok</i>	
%	

Derivational paradigm

- the most frequent suffixed form of *izom* 'muscle' is *izmos* 'muscular', which only occurs in its vowelless variant (**izomos*)
- the distribution of vowelless variants in the paradigm is 99,86%

The most frequent suffixed form in the derivational paradigm	With vowel
<i>izmos</i>	
%	

alkalom 'occasion'

Inflectional paradigm

The most frequent suffixed form in the inflectional paradigm	With vowel
<i>alkalommal</i>	
%	

Derivational paradigm

The most frequent suffixed form in the derivational paradigm	With vowel
<i>alkalmas</i>	
%	

kereskedelem 'trade'

Inflectional paradigm

The most frequent suffixed form in the inflectional paradigm	With vowel
<i>kereskedelemben</i>	
%	

Derivational paradigm

The most frequent suffixed form in the derivational paradigm	With vowel
<i>kereskedelmi</i>	
%	

This tendency was observed in the case of all the examined stems. However, there seems to be no correlation between the amount of variation in the case of the most frequent suffix in the paradigm and in the case of the suffixes in general in the given paradigm. For example, *átok* 'curse' exhibits a high amount of variation (54%-46%) as far as all inflectional suffixes in general, but its most frequent suffixed form in the same paradigm does not vacillate (100% vowelless).

- (b)** At the same time, in the case of *bajusz* 'moustache', there seems to be such a correlation. Its most frequent inflectional suffixed form occurs vowelless in 35% and vowelless in 65% of the cases. The distribution of vowelless and vowelless inflectional forms in the paradigm is the same.
- (c)** There is almost always a correlation between vowelless and vowelless forms (in the case of the most frequent suffixed form and suffixed forms in general) in the derivational paradigms, as the distribution of vowelless and vowelless suffixed forms is uneven (usually around 20-80% or 10-90%.
- (d)** If one accepts McCarthy's (2002) suggestion then one would expect derivational paradigms to behave differently from inflectional ones: derivational paradigms would be expected to behave analogously to the base form, while in the case of inflectional paradigms one would expect the behaviour of the different forms to influence one another in one way or another (if anything at all).

Also, Bybee and Brewer's (1980) and Benua's (1997) claim also suggests that derivationally suffixed forms are expected to follow the behaviour of the base while inflected forms are not.

However, our data suggest that derivational forms do not necessarily copy the behaviour of the base: the base almost always only appears with the unstable stem vowel while derivationally suffixed forms are mostly vowelless. (Percentage values in proportion to all occurrences of the base)

(3)

	w
<i>akol</i>	
<i>álm</i>	
<i>ártalm</i>	
<i>átok</i>	
<i>bajusz</i>	
<i>késedelem</i>	
<i>korom</i>	
<i>kebel</i>	

In the case of inflectional paradigms Bybee and Brewer's (1980) and Benua's (1997) claims are borne out in the sense that inflected forms do not pattern with the base form as is clear from the data below:

(4)

	w
<i>kegyelem</i>	
<i>kebel</i>	
<i>kazal</i>	
<i>karom</i>	
<i>kapor</i>	
<i>irgalom</i>	
<i>izgalom</i>	

kölyök	
köpedelem	
kötelem	
küzdelem	
lakodalom	
lepel	
lucsok	
mentelem	
meder	

7. Conclusions

- Neither members of inflectional, nor those of derivational paradigms seem to follow the behaviour of the base.
- Inflectional and derivational paradigms are similar in that they seem to choose an analogical source on the basis of frequency (token frequency)
- Each inflectional and derivational paradigm have a different base (the most frequent suffixed form in the paradigm and its quality to be vowelless or vowelised) but the base is selected in the same way: on the basis of token frequency
- All the above indicate that McCarthy (2002), Bybee and Brewer (1980) and Benua (1997) are probably not entirely right in their claims and also that Albright (2002, 2004) seems to be at least partially right. However, to be able to clearly justify or falsify these claims we will have to further investigate both the above claims and their predictions on the rest of our data as well as the influence of type frequency.

8. Bibliography

Albright, Adam (2002) *The Identification of Bases in Morphological Paradigms*. PhD dissertation, UCLA, Los Angeles.

Albright, Adam (2004) „Inflectional paradigms have bases too: Arguments from Yiddish”. Paper presented at the MIT (Non)identity within Paradigms Workshop.

Benua, Laura (1997) *Transderivational Identity: Phonological Relations Between Words*. PhD dissertation, University of Maryland, College Park.

Bybee, Joan and Mary Brewer (1980) „Explanation in morphophonemics: Changes in Provençal and Spanish preterit forms”. *Lingua* 52. 201-42.

Bybee, Joan (2001) *Phonology and Language Use*. Cambridge University Press.

Bybee, Joan (2010) *Language, Usage and Cognition*. Cambridge University Press.

Halácsy, Péter, András Kornai, László Németh, András Rung, István Szakadát and Viktor Trón (2004) „Creating open language resources for Hungarian”. In: *Proceedings of the 4th international conference on Language Resources and Evaluation (LREC2004)*.

McCarthy, John (2002) „Optimal Paradigms”. ROA 485-1201.

Rung, András (2010) *Magyar főnévi alaktani jelenségek analógiás megközelítésben*. PhD dissertation. ELTE University, Budapest, Hungary.