

Vowel ~ zero alternations in Hungarian nominal inflectional and derivational paradigms: An analogy-based statistical approach

Lilla Magyar (ELTE / HAS)

Hungarian vowel ~ zero stem alternations (e.g. *torony-tornyok* 'tower' nom.sg.-nom.pl., 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. Such alternations cannot be analysed either as cases of insertion or deletion for widely discussed reasons. Furthermore, these stem alternations exhibit some inter-speaker or even intra-speaker vacillation: some of the stems may or may not undergo vowel ~ zero alternation (e.g. *bajuszt* 'moustache' acc.sg. vs. *bajszot*).

Gradient phenomena like this raise another important question concerning the nature of different morphological paradigms. According to the more traditional view (advocated by McCarthy (2002) amongst many others) only inflectional paradigms are regarded as true paradigms, derivational paradigms are not. Only derivational paradigms have bases (as derivational forms are derived from the base form), whereas inflectional paradigms do not have an analogical source: there is a correspondence relation from every paradigm member to every other, i.e. members of the paradigm equally influence each other. Bybee and Brewer (1980) and Benua (1997) also argue that morphologically derived surface forms are in an Output-Output relation with their common base forms, which are morphologically simpler (less marked) than their derivations (which are more marked). Inflectional forms, however, are of equal complexity, the base form is not immediately apparent. Albright (2002) and (2004), however, argues that inflectional paradigms do have bases, too. The difference between the selection of bases in the case of inflectional and derivational paradigms lies in that bases of inflectional paradigms are determined not on the basis of morphological-morphophonological markedness but on the informativeness of the relevant lexical properties. Our goal is to explore which of these assumptions hold in the case of Hungarian vowel ~ zero nominal stem alternations. Our other aim is to find out whether there is a significant difference between the selection of bases in the case of inflectional and derivational paradigms, as inflectional and derivational paradigms are gradient (as any other morphological category) (Bybee 2010), not strict, unseparable categories.

Our data, 240 stems and their different forms, were collected from the Hungarian Webcorpus (Halácsy et al., 2004). The samples contain all the possible forms of the nouns which are potentially capable of vowel ~ zero alternation – including the inflectional suffixes which trigger vowel ~ zero alternation (Rung 2010), but we also look at vowel ~ zero alternating derivational forms as well as inflectional and derivational forms which do not trigger a vowel ~ zero alternation. The data are systematised the following way: base forms (uninflected forms, not necessarily analogical bases, e.g. *irodalom* 'literature' nom.sg.), inflectional forms (e.g. *irodalma* 'his / her literature'), derivational forms (e.g. *irodalmár* 'literary person') and compounds (e.g. *irodalomszociológia* 'sociology of literature'). We use all the forms which contain the given (potentially) vowel ~ zero alternating stems so that we could determine the overall frequency of stems. We also determine the frequency of the forms without inflection, inflectional and derived forms and compounds for each stem and for each inflectional form in the case of each stem. We take type frequency as well as token frequency into account since both have an effect on the formation of paradigmatic patterns (Bybee 2001).

Our further goal is to explore what the effect of frequency is in both the inflectional and derivational paradigms and how it influences the propensity to exhibit vowel ~ zero alternations. We will also determine whether there is a paradigm uniformity effect which results in paradigm levelling and whether it is due to frequency effects (including type and token frequency), the informativeness of relevant lexical properties or an equal influence of paradigm members on each other. We also determine how the different morphological classes of suffixes – e.g. lowering stems, classes of suffixes – influence the likeliness of the alternation in suffixed forms. We present the results in a statistical analysis of the data.