THESES

András Rung

Some Phenomena of Hungarian Nominal Morphology in an Analogical Approach

doctoral dissertation

ELTE

2011

1. Purposes

If we want to write the analogical grammar of any language, one of the basic prerequisites is to know what phonemes/sounds/word forms (Bybee 2001), compound constituents (Krott 2009: 132), constructions (Fillmore és Kay 1987, Goldberg 1995, 2006) are similar to each other in the given language, **how closely they are similar, what is the basis of their similarity**. For this, however, we need to determine along what parameters we can measure this similarity, and in what features, relationships in the operation of the language similarity matters at all. Without these points of departure, analogical examinations can easily become hazy speculation (Bybee 2010: 62). For such an examination, the **similarity of relationships** is just as important as the similarity of the constituents themselves; but we could not possibly determine these relations without knowing the exact degree of the similarity between constituents. Furthermore, analogical grammars assume that similar elements generate similar relationships more frequently, and it is both more probable and easier to recognize these relationships between these elements.

In my dissertation I examined **how similarity (alone and in interaction with the frequency of occurrances) influences linguistic change, analogical leveling and extension,** and production. In the course of the analysis of formal similarities and hesitation, my aim was to gain an insight into their nature, and contribute with my description to the refinement of the tools of the analogical approach, to increase its accuracy and exactness. The examination of formal similarities **focused on mapping out the relationships between similar words and modeling the ways of their comparison**. With this, partly novel method I wished to offer a model for approaching morphophonological phenomena not explicitly treated in my dissertation for

both Hungarian and for other, morphologically diverse languages. It was also my purpose to show that a holistically conceived word structure and word endings containing a unit larger than a single phoneme can play a significant role in determining how words behave, also influenced by the circumstances of their use and their meanings.

2. Theoretical background

Rule-based grammars (either conventional or generative) can often give a fair approximate description of morphological behavior, but **cannot offer satisfying explanations for a number of linguistic phenomena and processes**. Thus they do not answer questions like why **transitions** between separate linguistic categories tend to be **gradual** (Chandler 2002: 57, Lakoff 1987, Taylor 1995), what the role of **probability** is in language use, what reasons cause linguistic hesitation, and how **frequency** influences linguistic change (Skousen 1989). Generative linguistics treats these problems by separating the concepts of performance and competence. Representations belonging to competence are taken as redundancy-free and categorical (e.g. binary features), linguistic elements are assigned uniform behavior, while numerous problematic phenomena (e.g. speech errors (Frisch 1996: 109)) are treated as belonging to performance, leaving their operation and nature hazy. It appears that it is rather practical reasons that are behind the seemingly conceptual separation of performance and competence (Bybee 2001, Bybee 2010, Skousen et al, 2002, Blevins and Blevins 2009).

Due to its flexibility, **analogical approach can offer a more satisfying solution to these problems,** and can handle cases where linguistic data are apparently unique, badly formed, or the speakers are limited in their interpretation or production due to noise, forgetting, or other reasons. At the same time, it can also encompass regular phenomena as well (Eddington 2003, Skousen 1989: 54–60). In such uncertain cases, rule systems become 'inoperational' if they do not contain one or more rules applicable to the given

linguistic elements. In rule-based approaches, therefore, we have to mark these numerous elements arbitrarily as 'irregular', even though in some other, more flexible framework they could be explained easily.

The fact that generative theories have **completely excluded extralinguistic factors (like the frequency of occurrances) from linguistic descriptions**, even though in a number of cases these can have an important influence on the formation of the framework itself (Ullman 1999, Pinker 1999, Kraska-Szlenk 2007, Rung 2008, Rung 2009) generates further problems. Usage-based grammars (Halliday 1961, Bybee 2010), contrary to this, focus on the efficiency of communication, its modes and other social, psychological functions, and thus **approach psychological reality** (that is, real linguistic operation) **better**. One result of such efforts is including into theories the concept of frequency, a concept whose importance the psycholinguists have demonstrated decades ago (MacDonald 1994, Hare et al 2001). In my dissertation I have also advocated the view that any **examination of language must work with a great amount of data** (Sinclair 1991, Jurafsky et al 2001), and that models describing real processes must be able to pass computerized modeling (Skousen et al 2002).

In my dissertation I **give a detailed account of research that (either due to their theoretical conceptions or to the technical solutions they propose) have a bearing on my examinations**. I survey the history of the analogical approach, its principles and the debates it raised. I discuss the role of similarity and frequency (two concepts especially relevant to analogy) in language, and some views concerning the construction and change of paradigms. I also give an overview of analogical research by Hungarian linguists, since both in my data and my conceptions these served as the closest points of departure. As my survey shows, linguistic change continually reshapes the structure of paradigms. I also discuss analogical modeling: the most important programs, AM (Analogical Modeling, Skousen 1989) and TiMBL (Tilburg Memory Based Learner, Daelemans and van den Bosch 2005) receive separate treatment, but I also give a short description of

modeling alternatives that are less-known but connected to my research. A short treatment of how I see the function of similarity influences and frequency in an analogical grammar follows; the role of prototypes, how the similarity to words that are special from some aspect, and their frequency can determine and influence the behavior of other words will be treated separately.

3. Methods and material

I examined the morphological behavior of a medium-size group of Hungarian nouns. **Epenthetic nouns** are interesting enough in themselves to merit a comprehensive examination, but their description was here merely a means to an end: to clarify the role of similarity and frequency, two features of great significance in ordering linguistic behavior.

In addition to studying internal structure, my purpose was to obtain a more thorough understanding of **the relationships between epenthetic nouns and of their differences from other nouns**. The internal relationships within the set of epenthetic nouns, and the external relationships between them and other nouns are well described by such factors as the individual words' degree of epentheticity and the distribution of these degrees in the paradigms of the applicable suffixes. If we assume that the epenthetic behavior of the words is not the result of accident, then these words must share some common features.

I have chosen **1211 epenthetic nouns from the BME MOKK *morphdb.hu* dictionary**, which is currently the largest freely available linguistic database with 130.000 words (Trón et al 2006). This corpus of 1211 words consists of 229 base words and compounds generated from them. I have also extended my examination beyond the pattern $\sim VC_{\alpha}(o/e/\ddot{o})C_{\beta}\#$, defined by Rebrus and Törkenczy (2008) to words that contain *–a* or *–u* as a final vowel (e.g. *ajak* 'lips', *bajusz* 'moustache', 17 words). During my work, I

relied on **frequency data** retrieved from the ***Szószablya Corpus*** (Halácsy et al 2003).

In looking at the similarities between words, I based my examination on their surface forms (Kálmán 2008, Bybee 2001, Fűköh and Rung 2005, Rebrus and Törkenczy 2008). My hypothesis was that the application of analogy is also triggered by other factors (mode of usage, meaning, etc.), but the most important of all were phonetic/phonological similarity and frequency (Lukács 2002).

I approached this material with three different methodologies, to be discussed in brief shortly. I characterized epenthetic nouns more accurately and thoroughly, both on a general and an atomic level, than in previous research, based on their relationships determined by their endings. I supported my findings with statistical calculations. I also studied the analysed corpus in graph structures, and I compared their two states, with support from the corpus data. After this, I modeled the behavior of the examined material with the help of the usual methods of analogical modeling. I finally tested whether or not Hungarian native speakers use suffixes on pseudowords in a way that supports my theory.

The **similarity of words**, needed for the three different examinations, **was determined by algorithms I have developed**, enabling more refined comparisons than the algorithms currently used for measuring similarity (Skousen et al 2002). The algorithms named *complex feature measure* and *complex tier measure* calculate the similarity of words starting from the right, assigning less and less weight to correspondances and similarities as they advance to the left side of the word. Thus both algorithms determine *vas* 'iron' and *sas* 'eagle' as 'more similar' than *vas* 'iron' and *vaj* 'butter'. They perform the comparison based on the feature values of the individual phonemes, but while *complex feature measure* (Rung 2008, Rung 2009) compares phonemes, *complex tier measure* determines the degree of similarity based on the similarity of the tiers of individual features. The values for similarity are given on a scale of 0 to 1: 0 means words which are not at all similar, while 1 is the

degree of similarity a word has to itself.

4. Summary of results

- Epenthetic nouns do not behave uniformly. Their similarity to each other and their frequency play a significant role in their behavior.
- How analogical sources are chosen can be formalized and accurately determined.
- Individual paradigm cells take a different part in linguistic change.
- The behavior of Hungarian nouns can be modeled analogically, in which I managed to get results equally or more accurate than current learning algorithms.
- Speakers feel, when having to choose suffixes, that two words are more similar if the phonemes closer to the end of the words are similar than if the similarities or identical elements occur within the words or at the beginning.
- The behavior of words is influenced not only by the word most similar to them, but also by their nearest neighbors.

5. An analysis of the behavior of epenthetic nouns

I **analysed** the **behavior of Hungarian epenthetic nouns** based on the principles of analogy. As the first step, I examined the reasons of their behavior based on their end sequences.

last phoneme	frequency	ratio based on frequency	type frequency	how many times more frequent than epenthetic	degree of epentheticity (based on type)	degree of epentheticity (based on occurrances)	epenthesis in all forms
m	1412628	65.39%	558	21.25	99.6%	99.9%	52%
g	378878	17.54%	55	1.97	99.7%	99.9%	52%
k	183660	8.50%	170	2.47	98.1%	99.5%	52%
r	105764	4.90%	186	0.49	97.3%	98.2%	46%

ɲ	23048	1.07%	36	0.30	99.5%	99.7%	49%
l	22492	1.04%	26	0.23	74.7%	84.1%	33%
n	15644	0.72%	20	0.27	98.1%	99.6%	37%
j	12011	0.56%	21	0.27	96.7%	98.5%	39%
ʧ	4704	0.22%	13	0.63	99.6%	99.9%	69%
s	1136	0.05%	5	0.02	32.4%	36.2%	48%
z	471	0.02%	2	0.02	43.7%	33.6%	17%

Groups of epenthetic nouns on the basis of their last phonemes

In later examinations, I showed that **among epenthetic nouns, the degree of epentheticity with suffixes that originally require epenthesis is lower in the case of those words whose total number of forms** (including derivatives) **show a low ratio of epenthetic forms.** This connection is important because the lower ratio of epenthetic forms (as projected on the set of all forms) cannot be caused solely by the hesitation that words generate. Those words in whose case the ratio of epenthetic forms is lower in the set of all their forms are more likely to be unstable in their epenthetic behavior and take part in analogical leveling. I could also support this connection between the degree of epentheticity (measured on the basis of all forms) and the degree of epentheticity in suffixed forms originally requiring epenthesis with similarity groups based on the root ends (last and last two phonemes, last two consonants).

I also surveyed the organization of epenthetic nouns into similarity groups in graph structures. I showed that change is **conditioned not only by the similarity to certain groups, but also by the difference from these groups.** Unique words of also unique structure are more distanced from the behavior prescribed by the epenthetic scheme than those epenthetic nouns that are organized into similarity group with other words of similar epenthetic patterns. This suggests that the weakening of the connections with the group of the word's own behavior pattern (formal or meaning-based autonomy) might have a primary influence on analogical regularization, and that this influence can easily be stronger than the attraction that non-epenthetic words exert on these words.

Contrary to this, **the words ending in *–alom, -elem*, forming a closed**

set among morphologically heterogenous epentheic words, **are consistently epenthetic, which might be connected with their strong similarity relations and high frequency too.** Their nearly identical behavior cannot be attributed to the morphemes *–alom, -elem*, since in many cases these cannot be clearly segmented or cannot be segmented at all (*cimbalom* 'cimbalom', *malom* 'mill', *halom* 'heap' etc.), and on a morpheme-level basis we could not even explain why a number of somewhat similar, but somewhat differently ending words (*álom* 'dream', *ólom* 'lead') behave in a way nearly identical with theirs. I also showed the significance of the factors similarity/difference by proving that the number of closest connections (as determined by *complex feature measure* and *complex tier measure*) shows a significant correlation with the degree of epentheticity. I successfully showed frequency influences in the case of compound words too: here I saw that words with a degree of epentheticity different from the head of the compound can be found in compound groups where type and token frequency are lower than the average.

I examined the exact nature of changes and the specific behavior of the various suffixed forms by comparing two large text corpora (*Szószablya Corpus* and another corpus of my own collection, made in 2010 with the help of Google: *Google Frequency Collection*). I made my *Google Frequency Collection* in the spring of 2010, and this contained the frequency numbers of the most frequently used epenthetic suffixed forms of epenthetic nouns appearing in my examination. In the comparison of *Szószablya Frequency Dictionary* and *Google Frequency Collection* my most important insight proved to be that **paradigm cells take part in changes in quite different ways.** Rebrus and Törkenczy (2008) had already brought this up (if only in passing): the behavior of individual paradigm cells is not wholly autonomous, since general tendencies can be observed in them, but the change does not happen in a way that would place words evenly or even suddenly into a different paradigm.

Some paradigm cells show lower, while others show higher rates of epentheticity, and these can be validated by statistical methods. Rates of

epentheticity and the rate of their changes are not the same with the different suffixes (going back to phonotactic, systematic or usage-based causes), and these show no obvious correlation, since the superessive shows the lowest rate of epentheticity, but only **the change of the form implying a 3rd person singular possessor or its stronger presence in the informal registers can be proven in a significant way.** The hierarchy between the suffixes, however, shows differences in the rate of epentheticity in numerous individual cases, and partial patterns can also be observed in the changes (e.g. quickly changing words which are least epenthetic in their accusative).

	Szószablya 2003 rate of epentheticity	Google 2010 rate of epentheticity	rate of change	dynamics of change
accusative	96.82%	96.36%	0.46%	1.14
superessive	95.31%	93.11%	2.2%	1.47
plural	98.62%	98.54%	0.08%	1.06
Sg.1 possessor	97.83%	97.68%	0.15%	1.07
Sg. 3 possessor	97.56%	96.66%	0.9%	1.37
Pl.3 possessor	98.71%	98.55%	0.16%	1.12
Sg.3 possessor w/ more than one possession	98.99%	98.71%	0.28%	1.28
all suffices	97.57%	96.97%	0.6%	1.25

Rate of epentheticity of epenthetic nouns with suffices, in *Szószablya Corpus* and *Google Frequency Collection*

6. Modeling similarity influences

Counterarguments against an analogical approach generally point out that **the selection of analogical sources tends to be arbitrary and not clear enough**; therefore **I have tested the validity of my suppositions about word similarity with the help of my algorithms** on tasks that could easily frustrate

rule-based theories. It is a part of our knowledge of language that we categorise new words (or at any rate, words new to us) on the basis of their similarities to known ones, and place them in an already known paradigm. It was this linguistic skill that I tried to grasp with my algorithms in several modeling tests.

My first test examined whether the different algorithms select the appropriate analogical source for a particular group of epenthetic nouns **on the basis of samples of different size.** In its methodology, this test followed one of my earlier studies on the locatives of place names (Rung 2008). **My second test** focused on how well the algorithms could **select an analogical source for all epenthetic nouns, considering the whole of the lexicon** (and supposing the available dictionary corpora to represent more or less fairly the mental lexicon of a Hungarian adult speaker). I ran **my third test** on the same vocabulary material, but this time I **compared complex feature measure with other machine learning algorithms** with tenfold cross-validation. **Finally**, based on the measure of similarity, **I selected prototypes** for the epenthetic nouns instead of closest sources, with an algorithm that synthesizes some of the conclusions of all of my examinations in its operation. With the help of these prototypes, I was seeking to find out the reasons for the differences in the degree of epentheticity of different epenthetic words.

The **algorithms** based on the phenomena observed during the study of these words' behavior were proven **to be able to grasp the formal characteristics of epenthetic words fairly well**, and managed to categorize the words better than several learning algorithms currently in use (only the Maximum Entropy Model showed similarly good results). In binary decision situations the best results were supplied by the comparison mode named *complex feature measure*, which shows that speakers primarily rely on the similarity (and not the identity) of endings if they have to make categorical decisions about the behavior of words. The differences in the individual words' rate of epentheticity, however, were also very well grasped by *complex tier measure*, an algorithm focusing more on the structural similarity of words

and therefore more holistically oriented. This comparison method produced better results than any other if the task was to model the rate of epentheticity of words based only on a few prototypes. **In the selection of prototypes** which play a role in determining the behavior of words **it was the frequency of occurrances that proved to be most important.** In this case too, it was the similarity of the root ends that mattered most.

	F score other nouns	**F score epenthetic**	**Number of errors: other noun -> epenthetic**	**Number of errors: epenthetic -> other noun**
Decision tree (J48)	0.999	0.955	39	58
Maximum entropy	0.999	0.977	27	23
Complex feature measure	0.999	0.979	31	14
TiMBL (Tilburg Memory Based Learner)	0.999	0.955	60	37

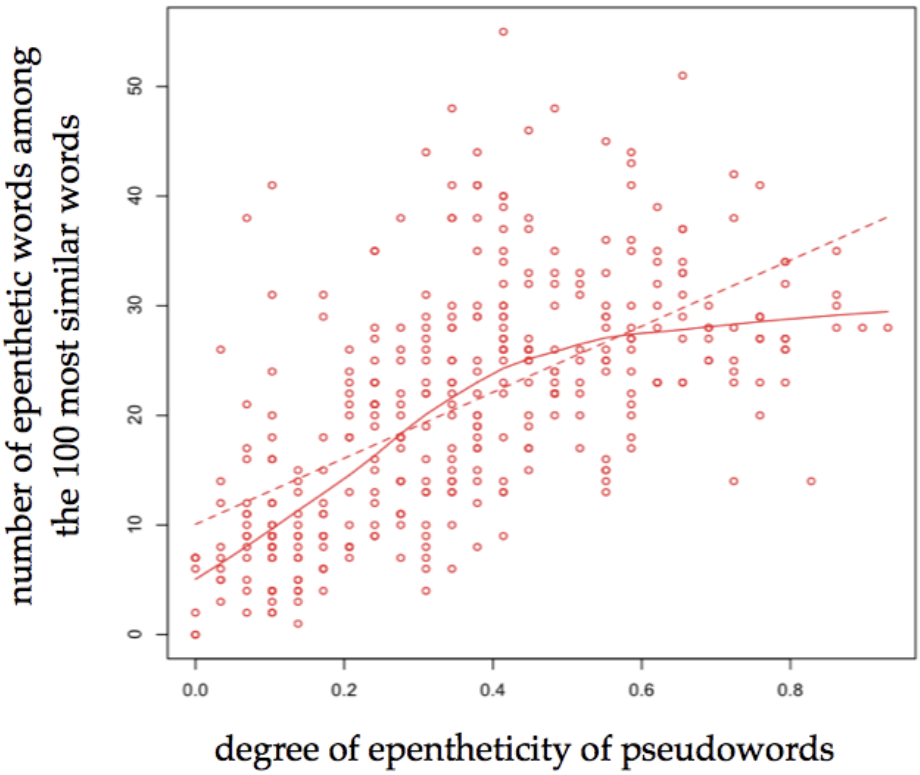
Results of testing algorithms, based on the categorization of 1.078 epenthetic and 48.389 non-epenthetic nouns, with tenfold cross-validation

7. Measuring the factors of selecting analogical sources by pseudowords of *CVCVC* structure

In this linguistic test I studied **to what degree phonemes in different positions contribute to the categorization of two words as similar**. My hypothesis was that the more a pseudoword is similar to a real word in its behavior, the closer it will be judged to it on the basis of their formal similarity. Beyond this, the test also aimed at determining what factors, taken together, influence the behavior of a new linguistic element: the most significant of these proved to be the analogical influence of structurally similar words. I executed this test with 116 Hungarian native speakers, using 91 epenthetic nouns of *CVCVC* structure.

The test confirmed several hypotheses that I had already used earlier in my analyses, and which Lukács (2002) had also outlined in her work. The

most important of these is that **the similarities and differences observable in different phoneme positions bear different degrees of significance in the comparison of words.** This is evident from the fact that the more the changed phoneme in a *CVCVC* pseudoword was positioned to the left, the more the word's behavior (degree of epentheticity) assimilated to that of the original words (from which the pseudoword was created). Independent of the influence of the phoneme's position, it was also observable that the way the new word's degree of epentheticity changes mostly depends on the word's unique system of relationships (beyond the influence of the original word): the speakers did not necessarily use the original word as an analogical source when selecting between the forms of the pseudoword, but they also evaluated the pseudoword on its own, and determined how it should behave on the basis of that. The greatest role here is played by words that are also structurally similar, complemented by the influence of the word whose ending is most closely similar.



Correspondance of the number of most similar epenthetic words (as determined by *complex tier measure*) and the degree of epentheticity of pseudowords

changed phoneme	degree of epentheticity	significant differences	variance	min. epenthesis	max. epenthesis	most epenthetic words
1st	44.2%	> 3rd phoneme ** > 5th phoneme ***	21.9	0%	93.1%	lücsök, rucsok, böcök, pürök, dücsök
2nd	41.5%	> 5th phoneme ***	20.1	0%	86.2%	pücök, vücsök, vocok
3rd	39.2%	> 5th phoneme ***	22.1	0%	79.3%	vöcök, rücök, tücök, tüszök, surom
5th	22.5%		13.5	0%	58.6%	sulyog, hurocs, bögöl

The influence of phoneme position on the degree of epentheticity

** = $p < 0,01$

*** = $p < 0,001$

8. András Rung's most relevant publications

With Borbála Fűköh: Az –esz és az –er végű becézett szóalakokról (2005). [On diminutive word forms ending in –esz and –er] *Nyelvtudomány* I. 115–130.

Determining word similarity in the Hungarian language (2008). In: Kálmán László (ed.): *Papers from the Mókus Conference.* Tinta Kiadó. Budapest.

112–118.

Szóhasonlóság mérése analógiás megközelítésben (2009). [Measuring word similarity in an analogical approach] In: Tanács Attila – Szauter Dóra – Vincze Veronika (eds.): *VI. Magyar Számítógépes nyelvészeti konferencia. MSZNY 2009.* Szegedi Tudományegyetem. Szeged. 104–113.

With László Kálmán: Klaszterek helyett prototípusok (2010). [Prototypes instead of Clusters] In: Tanács Attila – Csendes Dóra (eds.): *Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2010).* 325–332.

9. Reference List

Blevins James P.–Blevins, Juliette (eds.) 2009. *Analogy in Grammar: Form and Acquisition.* Oxford University Press. Oxford.

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

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

Chandler, Steve 2002. Skousen's analogical approach as an exemplar-based model of categorization. In: Skousen, Royal–Lonsdale, Deryle–Parkinson, Dilworth B. (eds.): *Analogical Modeling. An exemplar-based approach to language.* John Benjamins. Amsterdam. 51–105.

Daelemans, Walter–van den Bosch, Antal 2005. *Memory-Based Language Processing*. Cambridge University Press. Cambridge.

Eddington, David 2003. Issues in modeling language processing analogically. *Lingua. 114.* 849–871.

Fillmore, Charles J.–Kay, Paul. 1987. *The goals of Construction Grammar. Berkeley Cognitive Science Program Working Paper 50*. University of California at Berkeley. Berkeley, CA.

Frisch, Stefan A. 1996. *Similarity and Frequency in Phonology.* (PhD

dissertation) http://www.cas.usf.edu/~frisch/Frisch96.pdf (date of access 2010.07.01.)

Fűköh Borbála–Rung András 2005. Az –esz és az –er végű becézett szóalakokról. [On diminutive word forms ending in –esz and –er] *Nyelvtudomány* I. 115–130.

Goldberg, Adele 1995. *Constructions. A Construction Grammar approach to argument structure*. University of Chicago Press. Chicago.

Goldberg, Adele 2006. *Constructions at Work. The Nature of Generalization in Language.* Oxford University Press. Oxford.

Halácsy Péter – Kornai András – Németh László – Rung András – Szakadát István – Trón Viktor 2003. A Szószablya projekt. [Projekt Szószablya] In: Alexin Zoltán – Csendes Dóra (eds.): *Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2003).* 299.

Halliday, Michael A. K. 1961. Categories of the theory of grammar. *Word 17.* 241–292.

Hare, Mary L.–Ford, Michael–Marslen-Wilson, William D. 2001. Ambiguity and frequency effects in regular verb inflection. In: Bybee, Joan–Hopper, Paul (eds.): *Frequency and the emergence of linguistic structure.* John Benjamins. Amsterdam. 181–200.

Jurafsky, Daniel–Bell, Alan–Gregory, Michelle–Raymond, William D. 2001. Probabilistic Relations between Words: Evidence from Reduction in Lexical Production. In: Bybee, Joan–Hopper, Paul (eds.): *Frequency and the emergence of linguistic structure.* John Benjamins. Amsterdam. 229–254.

Kálmán László 2008. *A mögöttes és ami mögötte van.* [The underlying and what lies beyond] http://www.szv.hu/cikkek/a-mogottes-es-ami-mogotte-van (date of access 2010.07.01.)

Kraska-Szlenk, Iwona 2007. *Analogy. The Relation between Lexicon and Grammar.* Lincom. München.

Krott, Andrea 2009. The role of analogy for compound words. In: Blevins, James–Blevins, Juliette (eds.): *Analogy in Grammar: Form and Acquisition.*

Oxford University Press. Oxford. 118–136.

Lakoff, George 1987. *Women, fire, and dangerous things.* University of Chicago Press. Chicago.

Lukács Ágnes 2002. *Alaktanilag kivételes tövek vizsgálata a magyarban. A leíró általánosítások mentális realitása.* [A study of morphologically unique roots in Hungarian] (MA thesis)

MacDonald, M. C. 1994. Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes* 9: 157–201.

Pinker, Stephen 1999. *Words and Rules.* New York. Basic Books.

Rebrus Péter–Törkenczy Miklós 2008. *Morfofonológia és a lexikon.* [Morphology and the lexicon] In: Kiefer Ferenc (ed.): *Strukturális Magyar Nyelvtan 4. A szótár szerkezete.* [Hungarian Structural Grammar 4. The Structure of the Lexicon] Akadémiai kiadó. Budapest. 683–786.

Rung, András 2008. Determining word similarity in the Hungarian language. In: Kálmán László (ed.): *Papers from the Mókus Conference.* Tinta Kiadó. Budapest. 112–118.

Rung András 2009. Szóhasonlóság mérése analógiás megközelítésben. [Measuring word similarity in an analogical approach] In: Tanács Attila–Szauter Dóra–Vincze Veronika (eds.*): VI. magyar számítógépes nyelvészeti konferencia. MSZNY 2009.* Szegedi Tudományegyetem. Szeged. 104–113.

Sinclair, John 1991. *Corpus, concordance, collocation.* Oxford University Press. Oxford.

Skousen, Royal 1989. *Analogical Modeling of Language.* Kluwer Academic Publisher. Dordrecht.

Skousen, Royal–Lonsdale, Deryle–Parkinson, Dilworth B. (eds.) 2002. *Analogical Modeling.* John Benjamin. Amsterdam.

Trón, Viktor–Halácsy, Péter–Rebrus, Péter–Rung, András–Vajda, Péter–Simon, Eszter 2006. Morphdb.hu: Hungarian lexical database and morphological grammar. In: *Proceedings of 5th International Conference on Language Resources and Evaluation. ELRA*.1670–1673.

Taylor, John R. 1995. *Linguistic categorization prototypes in linguistic theory* (2nd ed.). Clarendon Press. Oxford.

Ullman, M. T. 1999. Acceptability ratings of regular and irregular past tense forms: Evidence for a dual system model of language. from word frequency and phonological neighborhood effects. *Language and cognitive processes* 14: 47–67.