A study of relations between associative structure and morphological structure of Hungarian words

Dániel Czégel
Budapest University of Technology
and Economics
czegel_d@yahoo.com

Zsolt Lengyel
Pannon University, Veszprém

Csaba Pléh
Central European University, Budapest
vispleh@ceu.hu

The paper mainly aims to reanalyze data with the presently available corpus linguistics tools from a relatively large scale paper-and-pencil based Hungarian verbal association dictionary with regard to two aspects. *i)* The mental lexicon issue. How are associative overlaps representing structural relations in the mental lexicon? *ii)* The systemic variability of the associative fields mobilized by the stimulus words: how variable the responses are, and how these associative entropies are related to morphological entropies of the same words.

Figure 1: Associative field of children (age 10-14)

1   Methods and materials

For the associative corpora, two dictionaries of Lengyel [3] were used. They are based on the responses of 2000 students between $10 - 14$ and $18 - 24$ to about 200 stimulus words. Digitized responses from this dictionary were related to the frequency distribution of 800 million web-based Hungarian words from the MOKK corpus [2].

2   Results

2.1   Associative overlaps and lexical fields

Based on the associative overlap measure introduced by Deese [1], a multidimensional scaling method was used to obtain associative fields depicting the pairwise associative distance of stimulus words in a two-dimensional figure. The results indicate that young adults have a more dense structure, their associative clusters are more tight compared to those of children of age 10-14, as illustrated in figure (1) and (2) and shown quantitatively in figure (3) and (4).
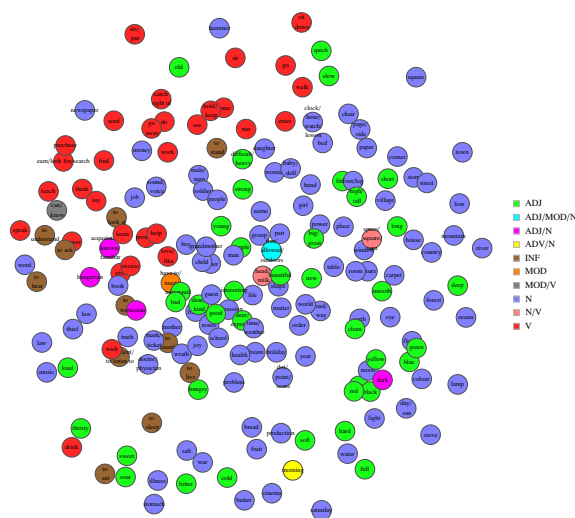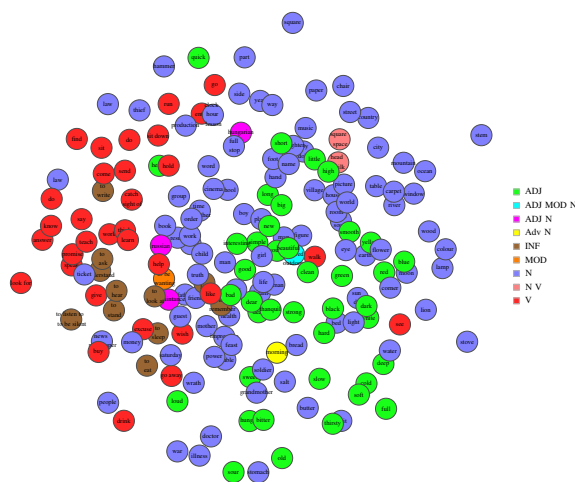

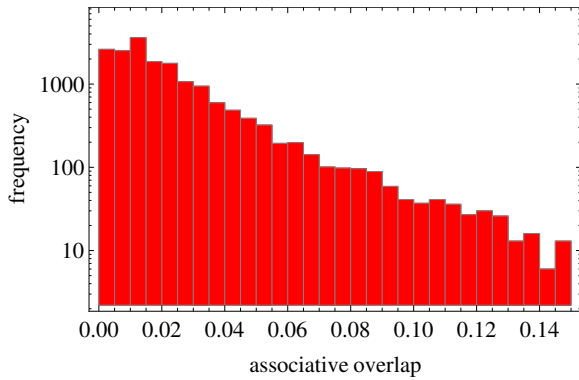Figure 2: Associative field of young adults (age 18-24)

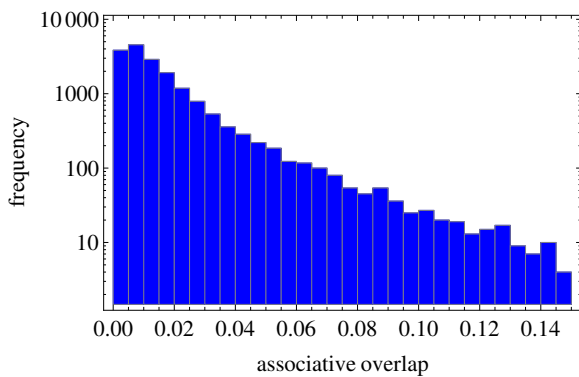Figure 3: Histogram of pairwise associative overlaps (age 10-14)

Figure 4: Histogram of pairwise associative overlaps (age 18-24)

2.2 Relationship between associative entropy, morphological entropy and frequency

We followed the methods introduced by Osgood [5] for analyzing the variability of morphological and associative structure of words. As shown in figure (5), there is an interesting difference of the relations between associative entropy (defined as $H_A = -\sum_i p_i \log_2 p_i$, where p_i is the relative frequency of the ith associated word) and corpus morphological entropy (defined as $H_M = -\sum_i q_i \log_2 q_i$, where q_i is the relative frequency of form i in the MOKK corpus) between nouns and verbs. In nouns, the more varied the morphology of the noun is in the corpus, the more variable the associative field is ($r = 0.202, 0.175$ in the two ages). That can be interpreted as implying that the more varied the suffixation of a noun is, the more variable associative relations it enters with other words. In verbs, however, if the verb has a more varied morphology, it has less associations ($r = -0.194$, both groups). As figure (6) shows, a similar relationship has been obtained between as-

sociative entropy and the logarithm of corpus frequency: In the case of nouns the correlation is positive ($r = 0.134$ and 0.367 in the two ages) while the correlation is negative ($r = -0.281, -0.222$) for verbs. This peculiar relation would be further studied with considering morphological entropy in light of the argument frames of the verbs on the one hand, and the role of syntagmatic associations in the associative fields of verbs on the other [4].
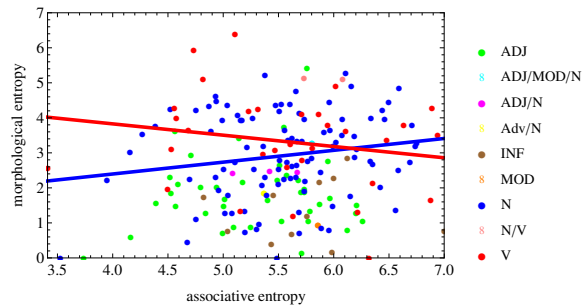

Figure 5: Relation between associative and morphological entropy (age 18-24). The blue regression line corresponds to nouns, while the red line corresponds to verbs.
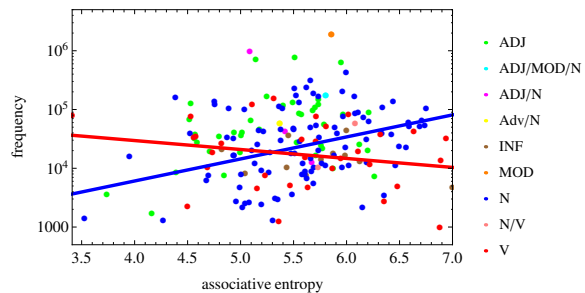

Figure 6: Relation between associative entropy and corpus frequency (age 18-24). The blue regression line corresponds to nouns, while the red line corresponds to verbs.

References

[1] James Deese. *The structure of associations in language and thought*. Johns Hopkins University Press, 1966.

[2] Péter Halácsy, András Kornai, László Németh, András Rung, István Szakadát, and Viktor Trón. Creating open language resources for hungarian. In *LREC*, 2004.

[3] Zsolt Lengyel. *Magyar asszociációs normák enciklopédiája I-II*. TINTA könyvkiadó, Budapest, 2008-2010.

[4] Katherine Nelson. The syntagmatic-paradigmatic shift revisited: a review of research and theory. *Psychological Bulletin*, 84(1):93, 1977.

[5] Charles Egerton Osgood. *Cross-cultural universals of affective meaning*. University of Illinois Press, 1975.