Relation between morphological and associative structure of Hungarian words

Dániel Czégel[1], Zsolt Lengyel[2], Csaba Pléh[3]

[1] *Central European University, Budapest; Institute for Cross-Disciplinary Physics and Complex Systems, Palma de Mallorca, Spain* <czegel_d@yahoo.com>
[2] *Pannon University, Veszprém*
[3] *Central European University, Budapest*

This paper aims to relate two types of presently available large scale databases corresponding to the organization of the mental lexicon, namely, text corpora and association dictionaries, from the point of view of morphological and associative variability. We used a paper-and-pencil based Hungarian verbal association dictionary containing 188 stimulus words and about 400 responses for each stimulus, collected from children (age 10-14) and young adults (age 18-24), to compute the associative entropies of the stimulus words. Associative entropies are then compared to the morphological entropy of the lexemes corresponding to stimulus words, determined from a large web-based Hungarian corpus. We found that for both age groups, more variability in morphological forms corresponds to higher associative variability in the case of nouns (r=0.20 and 0.18). For verbs, however, the opposite relation holds: morphological and associative entropies are negatively correlated (r= -0.19 in both age groups). In order to investigate this phenomenon in more depth, corpus frequency and the fraction of paradigmatic associations are also taken into consideration, showing that the relation between associative entropy and the fraction of paradigmatic associations is highly responsible for the different behaviour of nouns and verbs.*

KEYWORDS: morphology, associative entropy, mental lexicon

1. Introduction

The issue of grammatical complexity and its role in processing has become a central topic in contemporary psycholinguistics. Along this line, several new efforts attempt to relate morphological decomposition to the information based characterization of complex words. Reading time, naming, and lexical decision studies have shown

* The authors would like to acknowledge the comments of two anonymous reviewers, especially regarding the interpretation of noun-verb differences. Discussions with Jim Blevins, Alexandr Kostic, and Petar Milin helped to clarify our use of information theoretic notions. The authors also thank Gábor Kiss (Tinta Publisher) for providing searchable versions of the associative dictionaries.

the role of morphological variability in processing complex words (Moscoso del Prado Martín *et al*. 2004). However, at the same time relatively little attention has been paid to the possible relations between corpus based morphological variability of words and measures of associative diversity, where the latter mainly reflects the organization of semantic memory. In our study, an attempt is made to relate morphological entropy in Hungarian to the distribution of associative responses in young child and adult populations.

Hungarian is an interesting example to study the relevance of information theoretical notions over word processing due to its extremely rich morphological paradigms. A Hungarian noun has over 200 different suffixed forms. However, the actual paradigms are rather unsaturated with many empty cells. As Kornai (1992, 1994) pointed out, even if we take relatively frequent words such as *asztal* 'table' or *szék* 'chair' out of the roughly two hundred possible word forms starting with these stems, only a couple of dozens show up in the frequency dictionary (37 and 23 in the given cases). Thus, to come through even once with all the forms of even a frequent word, a much bigger corpus is needed than what we would experience in our entire life span. Kornai uses this example to suggest morphemic or analytic representation, arguing for rules. In actual language performance, this is also an argument for the use of information values for segmentation purposes.

Earlier studies using an informational theoretical processing frame for Hungarian mainly worked on the level of lexical words. Pléh *et al.* (2013) have shown, for the processing of individual Hungarian nouns, that the entropy decrease near the end of stem of nouns was a better predictor of recognition than the traditionally used uniqueness points. The low entropy connected to early uniqueness points shows that the uniqueness points used in traditional word recognition studies are in fact related to a more objective and statistically more sensitive measure, the entropy at the given point. Uniqueness points are only sensitive to absolute values, while entropy and its change are sensitive to frequency distribution. In a nutshell, entropy change is important in explaining neighborhood effects in Hungarian word recognition.

In the present study, relationships between two types of entropy measures were compared. The first one is the corpus based morphological entropy of a given lexeme. This reflects how many different forms of a word are used in actual texts, and how varied is this use. The other type is of a lexico-semantic nature. Entropy values were computed over the same lexemes, but this time taking the associative diversity that was observed over a population of subjects. Starting from the original insights of Kostic (1991, 1995) regarding the use of information metrics in studying the underlying mechanisms of both lexical access and morphological decomposition, we extended this to the study of associative hierarchies. Kostic has considered the different case forms of Serbian nouns as having a certain number of functions, and he had shown that these functions had an explanatory power of the processing difficulty of a given ending. Noun forms with fewer functions were easier to process.

We extended this approach to a broader consideration of two types of entropies computed. We assumed that the more forms a given word has in actual texts the more varied sentential and perceptual contexts the word appears in. On the other hand, words that appear in many contexts would tend to evoke these contexts in a word association manner. Thus one could compute entropies over responses in a free association task. The basic aim of our study was to see the relationships of those two context based entropies.

Figure 1 shows how the relationships could be conceived, with Figure 1a showing a high/high, and Figure 1b a low/low entropy case.

The underlying idea is simple and straightforward. One possibility shown on the figure is that the two kinds of uncertainty are indeed related. The more morphological forms a word appears in, the higher the likelihood that it will appear in different syntactic and semantic contexts. Take a simple count noun, like *asztal* \`table\`. Its different case marked forms would correspond to different possible sentential contexts and different neighborhood lexical items. The different cases would activate different classes of other items or different lexical items like in the model proposed by Kintsch & Mangalath (2011). To continue with the example, the suffixed form *asztal-on* [table-ON] would correspond to neighborhood lexical items like *bread, dish* etc., while *asztal-hoz* [table-TO] would go together with items like *sit*, *chair* etc. The variability of the two classes – morphological variants of the

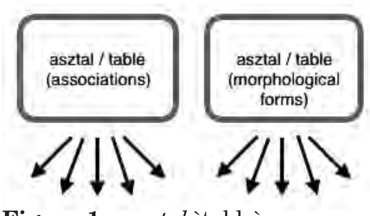

Figure 1a. *asztal* \`table\`.

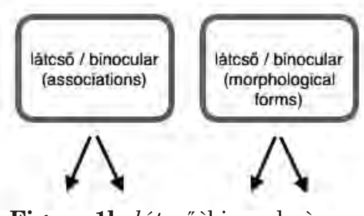

Figure 1b. *látcső* \`binocular\`.

Figure 1. Theoretical image of a high/high (A) and a low/low (B) entropy word.

same lexical entry and evoked verbal associates – would be related if the associative tasks would also evoke sentential contexts.

The idea relies on two grounds. The lexical proposal goes together with the contextual theory of meaning and lexical representation originally promoted by Titchener (1909) and later taken up by Noble (1952). In this vision, the meaning of a word is represented by the contexts it can mentally evoke. Noble operationalized this idea with a measure of meaningfulness that would correspond to the number of continuous associates a word can evoke. Prior and Bentin (2008) have experimentally shown that words coappearing in a given sentence are later bases for associative reactions. In this regard, we suggest that along with a contextual theory of meaning and lexical representation, meaningfulness can also be conceptualized as the *associative diversity* found in associative dictionaries. One can propose entropy measurements over the hierarchies found in these dictionaries. This would be a population-based characterization of the contexts some given lexical items can mobilize.

The morphological proposal is based on the idea that lexemes in a language with high morphological density can be characterized as being morphologically more or less saturated. The issue of empty cells raised by Kornai (1992, 1994) can be given a more elaborate treatment by characterizing the entire distribution of forms over a given corpus through 'the morphological entropy observed in the corpus' (Kostic 1995, Blevins 2015). As a matter of fact, the forms may be treated as showing a Zipfian distribution. With many typical nouns, for example, only a few forms appear but they do appear with high frequency, resulting in low entropy, while with some passe-partout nouns like *dolog* \`thing\` many different morphological form would appear resulting in high entropy. The psycholinguistic literature is well aware of the relationships between this morphological entropy and lexical processes, but mainly in the study of word recognition and lexical access. The textbook chapter by McQueen and Cutler (1998) provides a detailed summary, and the edited book of Baayen and Schreuder (2003) gives an insight into the intricacies of determining at what point morphological analysis is activated in the recognition of words. At the same time, the possible relationship between morphological complexity and associative complexity has not been raised as an empirical issue. We propose to relate the two data-driven measurements of uncertainty with each other: the uncertainty of forms in a large corpus, and the uncertainty in meaning based associative networks. The two entropies are not supposed to be related via some mysterious correspondence, but via the representation of these structures in the mind of individual speakers. We conjecture that in the actual associative task, the stimulus words usually presented in their citation form activate the morphological family of actual word forms, while the morphological family, via the activation of sentential contexts, facilitates and filters the actual associations recalled.

We thus propose that a higher morphological variety would correspond to more varied contexts of use and more varied associations. Tables 1a-b illustrate the entropies for a high and low entropy noun, respectively.

Table 1. A low entropy (a) and a high entropy noun (b) with their ten most frequent association responses and morphological forms.

Table 1a. A low entropy noun.

ország (country), $H_A = 4.39$, $H_M = 3.43$			
RESPONSE WORD	ASSOCIATION RESPONSE FREQUENCY	MORPHOLOGICAL FORM	FREQUENCY OF MORPHOLOGICAL FORM
Magyarország	142	ország	227352
Haza	32	országban	78778
Világ	28	országok	59066
Város	16	országot	32943
Nép	11	országokban	25988
Határ	9	országnak	19105
Térkép	8	országból	12322
Kicsi	7	országba	9646
Nagy	7	országa	8102
Nemzet	7	országában	6847

Table 1b. A high entropy noun.

barát (friend), $H_A = 6.46$, $H_M = 5.04$			
RESPONSE WORD	ASSOCIATION RESPONSE FREQUENCY	MORPHOLOGICAL FORM	FREQUENCY OF MORPHOLOGICAL FORM
Jó	36	barátja	13540
Társ	23	barátom	11446
Barátnő	22	barátok	6592
ellenség	16	barátai	6586
Fiú	15	barát	5457
Kedves	10	barátaim	4223
Haver	8	barátját	2579
bizalom	7	barátunk	2360
barátság	6	barátjával	2255
megértés	6	barátjának	2214

2. Methods

No new experiment has been designed or run particularly for this study. Rather, we reanalyzed existing data of Hungarian verbal associations with contemporary corpus linguistics tools and compared them to statistical properties of a large corpus.

Two associative data dictionaries (Lengyel, 2008, 2010) based on data collected in the 1980s have been used in this study, corresponding to children (age 10-14) and young adults (age 18-24). Both dictionaries contain about 400 free association responses for each of the 188 stimulus words. The selection of these age groups, as well as the selection of the actual stimulus words, was motivated by considerations of language teaching.

In order to investigate the relation between associative and morphological structure of words, also taking into account the effect of corpus frequency, four measures have been computed for each of the stimulus words, for both ages:

- Associative entropy: $H_A = -\sum_i f_i \log_2 f_i$, where f_i is the relative frequency of association *i*. Deese (1966) has proposed a systematization of several associative measures. However, an entropy-based measure was first introduced by Osgood, May and Miron (1975) to characterize the diversity of different qualifiers used in judging nominal concepts. The difference is that here associative entropy is defined over the distribution for a given stimulus word. Thus, while Osgood and colleagues characterized the entropy of the associative reaction words, we characterized the entropy of the stimulus words. Pléh and Czigler (1979) in a similar effort also characterized the entropy values for the constrained associations for the given nouns. They had shown by means of partial correlations that, out of the different indicators of associative variability (maximal frequency, idiosyncratic reactions, number of different associative responses), associative entropy was the most independent one, though the number of different responses had a high correlation (0.83) with entropy. A measure of stimulus word entropy was first proposed by Laffal (1955) in a clinical context. Laffal showed with 100 nouns that the entropy of associative responses to a stimulus word correlated with reaction times, and with reaction errors.
- Morphological entropy $H_M = -\sum_i F_i \log_2 F_i$, where F_i is the relative corpus frequency of the morphological form *i* of the stimulus word, computed from the Hungarian MOKK corpus (Halácsy et al, 2004) containing about 600 million words from web-based texts.
- Corpus frequency (freq): Frequency of the stimulus word as stem in the same MOKK corpus.

- Fraction of paradigmatic associations (par): fraction of associations belonging to the same part of speech as the stimulus word. This is a rough quantitative measure of paradigmatic over syntagmatic associations (Nelson, 1977). In general, most of the paradigmatic associations of a word belong to the same part of speech, and they usually represent hierarchical relations (dog vs. animal), opposites (heavy vs. light), synonyms (man vs. guy). In our case the par was defined as the ratio between number of same part-of-speech answers and number of all answers.

The Appendix shows the actual values for all stimulus words used in this study.

3. Results

The relation between morphological and associative entropies are shown in Figures 2a-d and Table 2. In both age groups, the more varied the morphology is, the more varied associative structure the word has in case of nouns. This 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. By contrast, an inverse relation is to be observed for verbs.

Correlations over the two entropy measures were computed in both age groups, taking the words as units of correlation. As shown in Table 2, the values are rather modest, but still significantly positive in nouns and negative in verbs.

In order to investigate this phenomenon in more depth, we computed the pairwise partial correlations between associative entropy, morphological entropy, corpus frequency and the fraction of paradigmatic associations. The results are shown in Tables 3a-b and visualized in Figures 3a-d.

One important factor is that, for nouns, frequency correlates with morphological entropy. The more frequent a noun, the more variable its morphological role. This is not true for verbs. At the same time, overall frequency is basically unrelated to associative entropy.

Interestingly, in both age groups, the main difference between nouns and verbs lies in the relationship between associative entropy and the fraction of paradigmatic associations. In nouns, no significant correlation is present between the two; in verbs, however, they are strongly negatively correlated (once the effect of the other factors has been factored out).

Thus, in the case of verbs, if everybody tends to associate them with the same few words from a given stimulus (i.e., the stimulus

Figure 2a-d. Relationships between the two entropies in the two age groups for different parts of speech.mulus words (nouns, age 10-14).

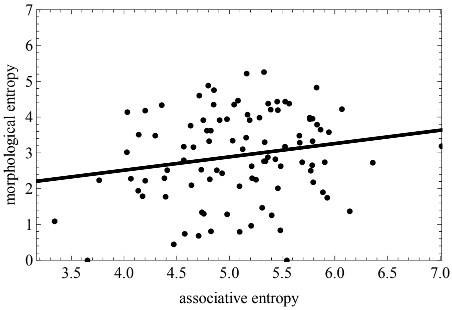

Figure 2a. The relationship between morphological and associative entropies of stimulus words (nouns, age 10-14).

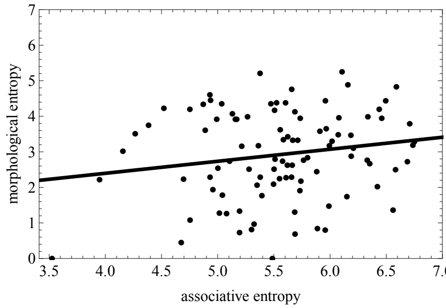

Figure 2b. The relationship between morphological and associative entropies of stimulus words (nouns, age 18-24).

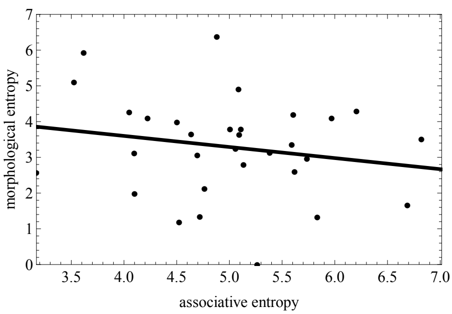

Figure 2c. The relationship between morphological and associative entropies of stimulus words (verbs, age 10-14).

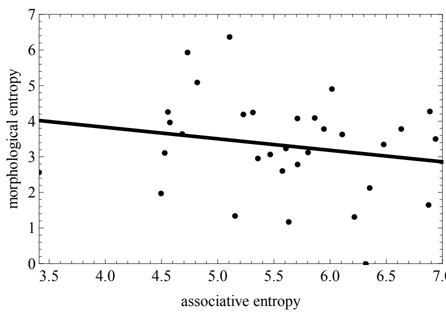

Figure 2d. The relationship between morphological and associative entropy of stimulus words (verbs, age 18-24).

Table 2. The Pearson's correlation coefficients *r* between associative and morphological entropy. For all cases, the *p*-value *p* of the null hypothesis $r=0$ is $p<0.001$.

	AGE 10-14	AGE 18-24
Nouns	0.20	0.18
Verbs	-0.19	-0.19

Table 3a-b. Partial correlations r and the p-values p of the null hypothesis $r = 0$ in the form $r(p)$ in case of age 10-14 (a) and 18-24 (b). Upper right triangles of the tables correspond to nouns, while lower left triangles correspond to verbs.

Table 3a. Verbs and nouns, age 10-14.

V\N			FREQ	PAR
		0.21* (0.05)	-0.07 (0.49)	0.06 (0.54)
	-0.12 (0.53)		0.35** (<0.01)	-0.03 (0.78)
Freq	-0.04 (0.83)	0.31 (0.10)		0.00 (0.97)
Par	-0.37* (0.05)	-0.02 (0.90)	0.32 (0.09)	

Table 3b. Verbs and nouns, age 18-24.

V\N			FREQ	PAR
		0.13 (0.22)	0.11 (0.30)	0.07 (0.51)
	-0.11 (0.56)		0.33** (<0.01)	0.01 (0.92)
freq	-0.20 (0.29)	0.32 (0.08)		-0.05 (0.65)
Par	-0.42* (0.02)	-0.01 (0.97)	0.07 (0.72)	

Figure 3a-d. Partial correlations (r) between different properties of the stimulus words. Positive r values correspond to solid lines, while negative r values to dashed lines. The thickness of the lines is proportional to r. Significance levels are marked by asterisk, *: p<0.05, **: p<0.01.

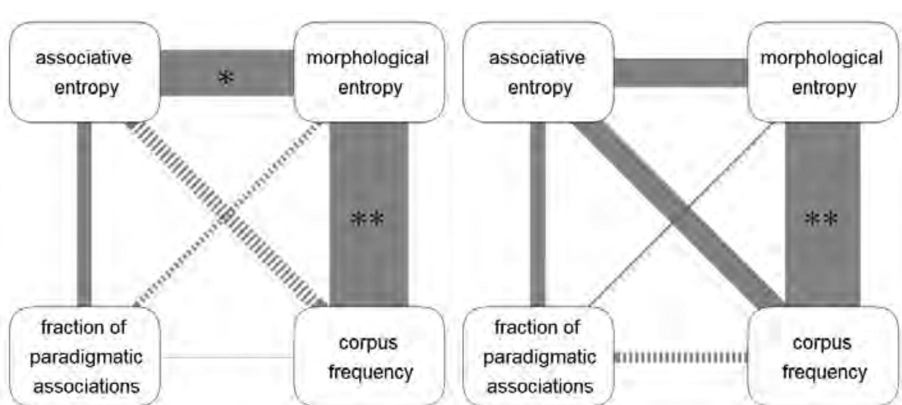

Figure 3a. Nouns, age 10-14. **Figure 3b**. Nouns, age 18-24.

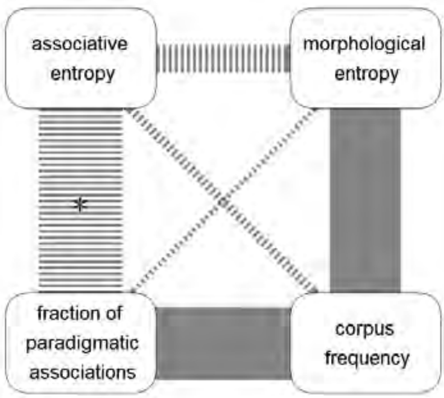

Figure 3c. Verbs, age 10-14.

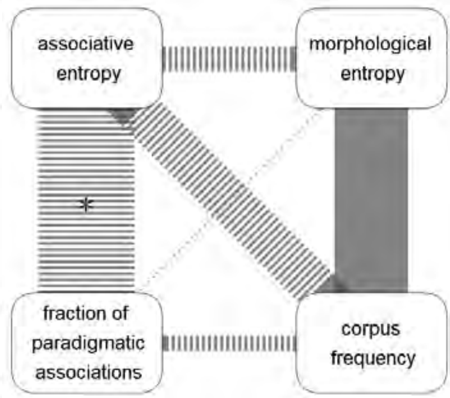

Figure 3d. Verbs, age 18-24.

word has small associative entropy), these associations are probably paradigmatic ones, like *walk-run, eat-drink* (again, for fixed values of the other parameters). By contrast, if the associations are highly varied, they turn out to be more syntagmatic than paradigmatic (*break-window, throw-ball*).

4. Discussion and conclusions

With nouns, the assumed relationship between morphological entropy and associative entropy has been proven. The more varied the morphological form of a given noun stem is, the larger variety of associations is recalled. With verbs, on the other hand, the more varied grammatical form a verb has, the less varied its associative network. The main factor underlying this difference is that, for nouns, the partial correlation between associative entropy and the fraction of paradigmatic associations is close to zero. By contrast, in the case of verbs, they are strongly negatively correlated.

This might be related to the fact that there is an interesting shift towards a reemergence of paradigmatic associations with age, a point that was already made by Kiss (2013) in connection with the same material. As Nelson (1977) also emphasized in her theoretical analysis, the nature of the syntagmatic/paradigmatic relationship seems to be influenced by several cognitive factors, beside the usually assumed linguistic ones. Further studies using a more careful sampling of words might shed more light on these issues.

Appendix

Here follows the list (in Hungarian orthography) of the words used in the study with their entropy values. The few cases of zero MorphE are due to allomoprhy relations that were treated here as a mere noise.

Word	Word	AsE 14	AsE 18	MorphE
ablak	window	4.9155	5.279438	2.525293
ad	gives	4.826541	5.106899	6.378842
ágy	bed	5.296641	5.266528	3.993813
alak	shape	5.43882	5.605361	4.385936
állni	to stand	4.990477	5.871013	2.166411
álom	dream	5.252131	6.189447	2.884661
aludni	to sleep	4.937962	5.435424	0.384858
anya	mother	4.953828	5.158617	3.920339
apa	father	4.870668	5.47456	4.355673
asszony	woman	4.44236	5.658173	2.29467
asztal	table	5.129131	5.131991	4.081509
baba	baby/doll	4.739069	5.68991	0.6926
baj	problem	5.099237	6.208051	3.116183
barát	friend	5.594353	6.460638	3.950066
bemegy	enters	4.42036	5.632428	1.18433
benyomás	impression	4.917537	6.150741	1.750539
beszél	speaks	5.083646	6.633429	3.78961
betegség	illness	5.770108	5.741048	2.184931
csinál	makes	5.049046	5.607355	3.240408
csoport	group	5.519071	6.750611	3.287541
dolgozik	works	5.280219	6.319775	0
dolog	matter	3.478641	4.755057	1.094527
drága	dear/expensive	5.469209	6.121799	0.788302
édes	sweet	3.728306	4.709695	2.931621
egészség	health	5.197063	5.764188	2.775164
egyszerű	simple	5.024032	5.703059	1.869356
éhes	hungry	5.546827	6.260959	1.040814
élet	life	5.134143	5.524645	4.383304
elmegy	goes away	5.674244	6.217442	1.31911
élni	to live	4.904646	5.999003	2.278068
ember	man	5.442928	5.994405	3.176496
emlékezni	to remember	5.819683	6.999417	0.75373
enni	to eat	5.422772	6.123211	2.85697

Word	Word	AsE 14	AsE 18	MorphE
érdekes	interesting	5.623262	5.966003	1.711808
erdő	forest	4.133265	4.932181	2.290617
eredmény	result	5.712088	5.910423	3.581982
erő	power	4.649718	4.930784	4.60847
erős	strong	5.534406	5.550962	3.726924
érteni	to understand	5.569932	5.759445	1.196841
év	year	4.866154	5.659449	4.764969
falu	village	5.416213	6.422537	2.024484
fehér	white	4.753038	4.517933	2.31322
fej	head/milks	5.500896	5.730048	5.126144
fekete	black	4.583373	5.09472	0.87085
felel	answers	5.599218	5.576057	2.604287
fény	light	4.874982	5.55704	3.635302
férfi	male/man	4.387914	5.393788	1.780727
fiatal	young	4.819408	5.680735	2.217892
fiú	boy	4.531239	5.498013	2.098582
folyó	river	4.467224	5.502629	2.52054
föld	earth	5.822635	6.709232	3.803014
fut	runs	5.049001	5.945946	3.786365
gondol	thinks	5.720434	6.88926	4.287505
gyerek	child	5.611363	6.583481	8.331527
gyomor	stomach	4.99448	5.015393	1.285583
gyors	quick	4.976591	5.574375	2.366462
gyümölcs	fruit	4.759802	5.002767	2.543294
háború	war	5.245209	5.37878	2.296711
hallani	to hear	5.39061	5.985165	0.163215
hallgatni	to be silent/to listen to	5.784715	5.040453	0.75176
hang	sound/voice	5.362891	6.496184	4.437388
hangos	loud	5.177204	5.873536	1.543793
harag	wrath	5.595401	6.353387	2.667989
ház	house	4.84839	6.159562	4.894526
hegy	mountain	4.850259	5.668652	3.334631
hely	place	5.473604	6.59043	4.837196
hideg	cold	4.412608	5.004667	1.485437
hold	moon	4.973812	5.881633	2.444915
hosszú	long	4.629996	5.247308	0.873191
idő	time/weather	4.11521	5.68726	4.140552
igazság	truth	5.140495	6.331358	2.770462

Word	Word	AsE 14	AsE 18	MorphE
ígér	promises	5.061032	6.478387	3.351213
írni	to write	5.454992	5.452219	1.780711
iskola	school	5.279414	5.551292	2.256081
ismerős	acquaintance/known/ familiar	3.837082	5.671868	2.460389
iszik	drinks	4.770598	5.155238	1.341395
jegy	mark/ticket	5.12322	4.937727	4.463256
jó	good	4.650865	5.949284	2.754155
jog	law	5.172171	4.753558	4.210928
jön	comes	4.174995	4.529641	3.111619
kalapács	hammer	4.850112	5.301547	0.821026
kályha	stove	3.697759	3.526276	0
katona	soldier	6.065934	6.561427	1.371696
kedves	dear/kind	5.687623	6.207597	1.494506
kék	blue	4.403972	4.590125	2.097858
kell	have to/must/need	5.74974	5.857883	0.942959
kemény	hard	4.632236	5.012495	1.673301
kenyér	bread	5.139727	5.957547	0.802939
kép	picture	5.302416	6.107541	5.260553
kérdezni	to ask	5.789426	5.618871	0.686838
keres	earns/searches	4.131398	4.557946	4.265013
keserű	bitter	4.642948	4.842267	1.263884
kéz	hand	5.262503	6.019076	3.302432
kicsi	little	4.433924	4.57009	1.486489
kíván	wishes	5.267662	5.805634	3.124439
könyv	book	5.780946	6.078689	3.970029
küld	sends	5.504961	5.229355	4.199469
láb	foot/leg	5.890184	5.578545	2.742017
lágy	soft	4.154132	4.390804	1.548019
lámpa	lamp	4.531337	4.676282	0.448928
lánya	daughter	4.52936	5.195837	0.743003
lassú	slow	3.925953	4.995031	0.712691
lát	sees	3.646115	4.818393	5.097701
leány	girl	4.779325	5.609261	2.274696
leül	sits down	4.172339	4.497359	1.976189
magas	high/tall	5.273596	5.490725	2.230345
magyar	Hungarian	4.95772	5.086864	8.331527
megbocsát	excuses	6.06686	6.877454	1.659885

Word	Word	AsE 14	AsE 18	MorphE
meglát	catches sight of	5.621268	5.359119	2.960959
megy	goes	5.157285	5.712294	2.796304
mély	deep	4.936625	4.564012	3.620145
mond	says	4.22056	5.709392	4.090033
mozi	cinema	3.797289	4.696913	2.242077
munka	job	5.446399	6.438172	4.20715
nagy	big/great	4.513818	5.509883	2.647765
nagymama	grandmother	5.501302	5.486805	0
nap	day/sun	5.481098	5.958993	4.439025
négyzet	square	4.215146	4.957468	1.94355
nehéz	difficult/heavy	4.218537	5.401949	1.647657
nép	people	5.225564	5.170624	3.918835
név	name	6.803331	6.737967	3.200141
nézni	to look at	4.708359	4.822271	1.740096
nyugodt	tranquil	5.337815	6.200579	1.270241
óceán	ocean	4.242104	5.043279	1.793328
oldal	page/side	4.782049	4.889433	3.622626
óra	clock/lesson	4.620126	5.362303	3.178518
orosz	Russian	5.27454	5.420779	2.477704
oroszlán	lion	4.757382	5.685611	1.307261
ország	country	4.684556	4.38516	3.761768
orvos	physician	4.697944	5.214778	3.170817
öreg	old	4.826218	4.939348	2.018295
öröm	joy	5.511874	5.106369	2.741905
papír	paper	5.21885	5.663239	2.633767
pénz	money	5.770376	5.712218	3.330047
piros	red	4.982147	5.528003	1.03944
pont	dot/point/score/full stop	5.045972	5.036801	4.357113
reggel	morning	4.840032	5.369081	1.855907
rend	order	5.029493	5.378097	5.217286
rész	part	5.466736	4.522056	4.233557
rossz	bad	4.885587	5.735306	1.866646
rövid	short	4.783621	4.528935	1.841187
sárga	yellow	5.661937	5.771227	1.050244
sarok	corner	4.730046	5.194977	1.346585
savanyú	sour	4.080681	3.737866	0
segít	helps	6.292576	6.939757	3.511355
sétál	walks	4.803186	6.353349	2.128352

Word	Word	AsE 14	AsE 18	MorphE
sima	smooth	5.923921	5.709582	0.132254
só	salt	5.140999	5.623225	3.429225
sötét	dark	4.853876	4.795839	2.007539
szabad	free	6.092063	5.803584	3.276948
szár	stem	4.194795	4.267313	3.52081
szék	chair	4.805879	4.991341	3.920809
szem	eye	4.4116	4.872152	4.345248
szép	beautiful	5.062901	5.734699	1.948281
szeret	likes	5.104512	6.108525	3.631316
szín	colour	4.111256	4.157611	4.51768
szó	word	4.275663	5.506288	4.181348
szoba	room	5.267442	5.989807	1.474996
szombat	Saturday	4.257615	3.949144	2.227968
szomjas	thirsty	4.052848	4.161769	0.579689
szőnyeg	carpet	5.284323	5.079363	1.268248
talál	finds	4.52116	4.575943	3.977038
tanít	teaches	4.687055	4.685361	3.652528
tanul	learns	5.823026	5.864897	4.101972
tart	holds	5.040251	6.016964	4.911292
tele	full	5.388435	5.758677	5.421339
tér	space	5.549828	6.079847	5.091431
termelés	production	5.766414	5.732607	1.909946
tesz	does	3.3185	3.410561	2.567909
tiszta	clean	5.027637	5.78609	1.332964
tolvaj	thief	5.235873	5.324017	0.970044
törvény	law	5.420683	5.621517	2.629709
tud	can/knows	5.032153	5.313387	4.254411
új	new	4.763863	5.141507	3.409078
újság	newspaper	4.622685	5.510926	2.799816
út	road/way	5.653682	6.337032	3.996892
utca	street	5.44995	5.887981	0.847692
ül	sits	3.731002	4.732514	5.93176
ünnep	holiday	5.043219	5.585659	3.352379
vaj	butter	5.125466	5.351236	2.073382
város	town	4.35935	6.073613	3.490038
vásárol	purchases	4.750039	5.467805	3.067549
vendég	guest	5.750395	6.689512	2.736738
világ	world	4.953456	5.735177	3.948671

Word	Word	AsE 14	AsE 18	MorphE
virág	flower	5.46717	5.801112	2.837879
víz	water	5.883976	5.966949	3.654442
zene	music	5.690638	6.18606	3.481934
zöld	green	4.431868	5.177033	2.158057

Bibliographical References

Baayen, Harald & Schreuder, Robert, (eds.) 2003. *Morphological Structure in Language Processing*. Berlin: Mouton de Gruyter.

Blevins, James P. 2015. *Word & Paradigm Morphology*. Oxford: Oxford University Press.

Deese, James 1966. *The structure of associations in language & thought*. Johns Hopkins University Press.

Halácsy, Peter; Kornai, András; Németh, Laszló; Rung, András; Szakadát, Istvan & Trón, Viktor 2004. Creating open language resources for Hungarian. In LREC, 2004.

Jurafsky, Daniel & Martin, James H. 2009. *Speech & Language Processing* Prentice Hall Series in Artificial Intelligence (2nd edition). Upper Saddle River, New Jersey: Prentice Hall.

Kintsch, Walter & Mangalath, Praful. 2011. The construction of meaning. *Topics in Cognitive Science* 3. 346-370.

Kiss, G. 2013. *Types of Hungarian word association responses* (in Hungarian). Psycholingustic Summer School, Balatonalmádi, 2013.

Kornai, András 1992. Frequency in morphology. In Kenesei, István. & Pléh, Csaba (eds.), *Approaches to Hungarian. Vol. 4.* Szeged: JATE, 247-268.

Kornai, András 1994. *On Hungarian morphology.* Budapest: Institute of Linguistics, HAS.

Kornai, András; Halácsy, Péter; Nagy, Viktor; Oravecz, Csaba; Trón, Viktor & Varga, Dániel 2006. Web-based frequency dictionaries for medium density languages. In Kilgarriff, Adam & Baroni, Marco (ed.), *Proceedings of the 2nd International Workshop on Web as Corpus,* ACL-06. 1-9.

Kostić, Aleksandar 1991. Informational approach to processing inflectional morphology: Standard data reconsidered. *Psychological Research* 53. 62-70.

Kostić, Aleksandar 1995. Information load constraints on processing inflected morphology. In Feldman Laurie B. (ed.), *Morphological aspects of language processing.* Mahlah, NJ: Erlbaum. 317-344.

Laffal, Julius 1955. Response faults in word association as a function of response entropy. *The Journal of Abnormal & Social Psychology*50.2. 265-270.

Lengyel, Zsolt 2008-2010. *Magyar asszociációs normák enciklopédiája I-II.* `Hungarian associative norms`. In Hungarian Budapest: TINTA.

McQueen, James M. & Cutler, Anne 1998. Morphology in Word Recognition. In Spencer Andrew & Zwicky Arnold M. (ed.), *The Handbook of Morphology.* Oxford: Blackwell Publishers. 405-427.

Moscoso del Prado, Martín Fermin; Kostić, Aleksandar & Baayen, Harald 2004. Putting the bits together: an information theoretical perspective on morphological processing. *Cognition* 94. 1. 1-18.

Nelson, Katherine 1977. The syntagmatic-paradigmatic shift revisited: a review of research and theory. *Psychological Bulletin* 84. 93-116.

Németh, Kornél; Varga, Dániel; Fazekas, Judit; Várhelyi, Klára & Pléh, Csaba 2013. Entropy measures & predictive recognition as mirrored in gating & lexical decision over multimorphemic Hungarian noun forms. *Psihologija* 46. 4. 397-420.

Noble, Clyde E. 1952. An analysis of meaning. *Psychological Review* 59. 421-430.

Osgood, Charles E.; May, William H. & Miron, Murray S. 1975. *Cross-Cultural Universals of Affective Meaning.* Urbana, IL: University of Illinois Press.

Pléh, Csaba & Czigler, István 1979. Kísérlet politikai kifejezések megítélésére alkalmas szemantikus differenciálskálák kialakítására és alkalmazására. (Elaboration and application of semantic differential scales for political expressions). *Pszichológiai Tanulmányok. XV.* 479-533.

Prior, Anat & Bentin, Shlomo 2008. Word associations are formed incidentally during sentential semantic integration. *Acta Psychologica* 127. 57-71.

Titchener, Edward B. 1909. *Lectures on the experimental psychology of thought processes.* New York: Macmillan.