DOI 10.2478/jazcas-2021-0037

THE MENZERATH-ALTMANN LAW AS THE RELATION BETWEEN LENGTHS OF WORDS AND MORPHEMES IN CZECH

KATEŘINA PELEGRINOVÁ[1] – JÁN MAČUTEK[2,3] – RADEK ČECH[1]

[1] Department of Czech Language, University of Ostrava, Ostrava, Czech Republic
[2] Mathematical Institute, Slovak Academy of Sciences, Bratislava, Slovakia
[3] Department of Mathematics, Constantine the Philosopher University, Nitra, Slovakia

PELEGRINOVÁ, Kateřina – MAČUTEK, Ján – ČECH, Radek: The Menzerath-Altmann law as the relation between lengths of words and morphemes in Czech. Journal of Linguistics, 2021, Vol. 72, No 2, pp. 405 – 414.

Abstract: It is shown that the mean morpheme length (measured in phonemes) decreases with the increasing length of word types (in morphemes) in Czech texts, i.e., these language units behave according to the Menzerath-Altmann law. The law is not valid in general for word tokens. Some hints towards an interpretation of parameters are presented.

Keywords: Menzerath-Altmann law, word, morpheme, phoneme, Czech

1 INTRODUCTION

The Menzerath-Altmann law [1] (henceforward MAL) is, together with the Zipf law [2] and the law of brevity [3] (which was, in fact, formulated also by Zipf), one of the best known laws in quantitative linguistics. Its special case was first articulated by Paul Menzerath [4] who studied the German vocabulary and observed that longer words consist, on average, of shorter syllables. The law was later substantially generalized by Gabriel Altmann [5]. The current version of the MAL claims that greater constructs consist on average of smaller constituents, with constructs and constituents being neighbours in the language unit hierarchy (such as, e.g., words and syllables, sentences and clauses, etc.). Sometimes even a more general form is used, namely, the size of the construct is a function of the mean size of its constituents. The function does not necessarily have to be strictly decreasing. It can increase to its peak (achieved usually for constructs of size two) and decrease from the peak to the right, see theoretical considerations in [6, p. 7] and examples in [5, p. 8; Table 4] or [6, p. 54; Table 5.7]. The general mathematical expression for the MAL is

(1) $y(x)=a x^{b} \mathrm{e}^{-cx}$,

with $y(x)$ denoting the mean size of constituents in constructs of size *x*; *a*, *b*, and *c* are parameters of the model. A special case of this general formula, namely

(2) $y(x)=ax^b$,

is very often sufficient to fit data. This simpler version is appropriate only if the mean constituent size achieves its maximum for constructs of length 1, and the mean constituent size then decreases with the increasing construct size.

The validity of the MAL was corroborated for several language levels and in many languages. As examples, we mention the MAL as the model for the relation between lengths of words (in syllables) and syllables (in graphemes) [7], canonical word forms (in syllables) and syllables (in phonemes) [8], word length motifs (in words) and words (in syllables) [9], sentences (in clauses) and clauses (in words) [10]. We note that while there seems to be a consensus on the hierarchy of low-level language units (speech sound/phoneme – syllable/morpheme – word)[1], the situation from word higher on is not so clear. Until relatively recently, clause was considered the "upper neighbour" of word (see e.g. [10]), but a rapid development of computational methods in syntax made it possible to take into account also other, intermediate level units (e.g., the MAL was shown to be valid for the relation between the lengths of clauses and syntactic phrases which are directly dependent on the clause predicate, see [11]). In addition, new, "non-traditional" units (different kinds of motifs, see an overview in [12]) were defined, which follow the MAL as well.

The first attempt to study the MAL specifically as a model for the relation between lengths of words (in morphemes) and morphemes (in phonemes) can be found in [13], where 15,011 German words' lemmas from a dictionary are analysed. A similar approach was chosen in [14] – the author uses databases of morphologically segmented word lemmas from Dutch, English, and German, which consist of 124,136, 52,447, and 50,708 word lemmas, respectively. Here, morpheme length is measured both in phonemes and graphemes. Both papers (and both choices of units in which morpheme length was measured in the latter one) result in data which can be modelled by the MAL in form (2), i.e., it holds that the longer the word, the shorter the mean length of its morphemes. The same is true for a short Turkish text [15, p.20], with morpheme length measured in the number of phonemes.[2] Both word types and tokens from a Czech novella were taken into account in [17]. A decreasing trend of morpheme lengths can be observed, but the two curves differ (the one for types is steeper).

[1] Needless to say, there are some (mainly methodological) problems related also to the analysis of these low-level units, such as e.g., their different definitions, the status of zero-syllable prepositions, etc.

[2] In [15], it is not specified whether word types or tokens are analysed. In addition, word length in syllables is studied in the same text in [16] but, curiously enough, the total numbers of words differ in the two works (559 words in Text 7 from [15], and 587 words in Text 2 from [16]). The difference between these two numbers is too small to be explained by different approaches, i.e., as the number of types in [15] and the number of tokens in [16]. In such case, the type-token ratio would achieve an extremely implausible high value of 0.95.

This paper presents the first systematic study of the relation between word length and morpheme length, performed on the sample of 15 Czech texts which are morphologically segmented by the same method. Function (1) is a good model for 14 of the texts, the only one for which the fit of the model is not sufficiently good is the shortest text in the sample. Some hints towards an interpretation of the parameters can be found in the conclusion.

2 METHODOLOGY AND DATA

The morphological segmentation applied in this paper is based on the retrograde morphemic dictionary of Czech [18]. However, [18] contains only dictionary forms of words (i.e. lemmas), and since Czech is an inflectional language, also rules for inflected word forms are needed. Moreover, there are several groups of words (proper names, pronouns, particles, interjections) which require a special approach.

The segmentation of inflected word forms of nouns, adjectives, numerals, and verbs mostly follows the Czech grammar [19]. The segmentation rules from [19] were modified or specified more in detail as follows.

1. For nouns and adjectives, markers of grammatical cases were reconsidered in order to avoid a high degree of allomorphy. For example, the morpheme *ch* serves as the marker of the locative plural in this paper, while according to [18], the locative plural is marked by six allomorphs: *-ech*, *ích*, *-ách*, *-ich*, *ěch*, *-ch*.[3] The vowel which precedes the morpheme *-ch* in the locative plural (if there is any) is considered a separate morpheme (in [20], it is called a connecteme). Other grammatical cases are treated analogously.
2. Czech pronouns can be inflected and, at the same time, they constitute a closed class containing a limited number of words. Therefore, there are less options for inter-paradigmatic comparisons (within this class as well as with other nominal parts of speech, i.e., nouns and adjectives). Consequently, the segmentation rules for pronouns found in grammars are often more ambiguous than for other parts of speech. Led again by the motivation to reduce allomorphy, we decided to apply a deeper (i.e., more detailed) morphological segmentation. We demonstrate our approach on the example of the instrumental plural form *našimi* of the first person plural possessive pronoun *náš* ‘our’. First, the inflectional morpheme is separated: *naš-imi*. Then, applying the above mentioned rules, *-mi* is considered the marker of the instrumental plural, and *-i-* which precedes it a connecteme: *naš-i-mi.* But in Czech there is also the instrumental plural affix *-ma*[4], contrasting

[3] These allomorphs occur in nouns, adjectives, and some pronouns.

[4] The affix *-ma* originally marked the dual number. In contemporary Czech it is used with nouns referring to paired body parts (and in forms agreeing with such nouns).

with *-mi* (*našim-a* vs. *našim-i*). This contrast leads to the segmentation into *našim-i.* Next, the inter-paradigmatic comparison with the second person plural possessive pronoun *váš* 'your' results in the segmentation *n-aš-i-m-i*, as these two pronouns contrast only in their consonant roots *n-* vs. *v-*. The final step is carried out on the basis of inter- and intra-paradigmatic comparisons of inflected forms of the possessive pronouns *náš* 'our', *váš* 'your' with the corresponding (non-possessive) personal pronouns *my* 'we', *vy* 'you'. Pronouns *my* and *vy* have stems *ná-* and *vá-*, respectively, in all cases with the exception of the nominative form. Therefore the final segmentation is *na-š-i-m-i.*

3. In proper names, only inflectional affixes and productive derivational suffixes are segmented. The remaining stems are not analysed morphologically here (e.g., nominative singular *Prah-a* 'Prague', genitive singular *Prah-y*; *Pelegrin-ov-á* – the female version of a surname with the suffix *-ov-*; *Fin-sk-o* 'Finland' etc).
4. Particles and interjections in Czech are uninflected parts of speech with an ambiguous delimitation and no consensus concerning their morphological segmentation. We only segment those particles and interjections which are morphologically transparent.

The texts under analysis were processed semiautomatically. Word forms were morphologically segmented manually at their first occurrence. Thus, a dictionary of morphologically segmented word forms was created. Then, a computer script[5] was written, which found the segmentation in the dictionary at further occurrences of the word forms.[6]

The texts[7] which serve as our language material were taken from a corpus of works by Czech writer Karel Čapek (the corpus described in [25]). In particular, we took five short stories (denoted as S1 – S5 below), five personal letters (L1 – L5), and five studies on philosophy (P1 – P5).[8] The texts were transcribed in such a way that the number of letters is equal to the number of phonemes the word consists of (e.g. *hoch* 'boy' is transcribed as *hox*).

[5] The script is written in Python. It is available upon request.

[6] The dictionary was enlarged every time the program encountered a hitherto unsegmented word form.

[7] Within our research framework, we prefer to work with texts rather than with corpora. At the same time, we are convinced that both of these approaches are reasonable, and both have their own advantages as well as limitations. See e.g. [21], [22], [23], and [24] for text vs. corpus discussions and related topics.

[8] The following texts were chosen: the first five short stories from the collection *Povídky z druhé kapsy*, personal letters with numbers 749, 753, 755, 756, and 761 (as they are numbered in https://search.mlp.cz/cz/titul/korespondence/43837/#book-content), and chapters one, two, three, four, and nine from the study *Pragmatismus*.

3 RESULTS

The mean morpheme lengths (expressed in the number of phonemes) in word types of particular lengths (counted in morphemes) which occur in the texts under analysis can be found in Tables 1, 2, and 3. As the mean is strongly affected by extreme values if the sample size is small, we decided to pool the data so that the minimum frequency in each category is ten.[9] In the pooled categories, word length is represented by the weighted arithmetic mean of the pooled words, with length frequencies serving as the weights.[10] In the tables, *WL* denotes word length, *MML* the mean morpheme length in words consisting of a particular number of morphemes, and *fr* the frequency with which particular word lengths occur in the texts.

The data were fitted to the MAL in form of (2), see Section 1. The fitness of the model was evaluated in terms of the determination coefficient R^2. The fit is usually considered satisfactory if $R^2>0.9$, see [27]. As in (2) it holds $y(1)=a$, the value of the parameter *a* was set as the mean length (in phonemes) of monomorphemic words in particular texts. We thus follow the approach applied to the relation between word length and syllable length from [7]. The value of the parameter *b* maximizes the determination coefficient.[11] The values of the two parameters and R^2 are also presented in the tables.

	S1		S2		S3		S4		S5	
WL	*MML*	*fr*	*MML*	*fr*	*MML*	*fr*	*MML*	*fr*	*MML*	*fr*
1	4.30	145	3.82	117	3.61	115	3.76	159	3.65	127
2	2.39	261	2.41	213	2.36	200	2.46	342	2.33	260
3	1.92	259	1.89	210	1.92	244	1.94	400	1.88	270
4	1.80	199	1.76	155	1.76	131	1.75	265	1.78	171
5	1.73	66	1.71	57	1.74	47	1.64	98	1.70	56
6					1.53	13				
6.19									1.50	16
6.21							1.63	28		
6.35			1.60	20						
6.42	1.61	26								
a	4.30		3.82		3.61		3.76		3.65	
b	-0.64		-0.55		-0.51		-0.54		-0.53	
R^2	0.941		0.959		0.973		0.973		0.968	

Tab. 1. Word length (in morphemes) and morpheme length (in phonemes): short stories, word types

[9] The minimum frequency of ten is only a rule o thumb, see e.g., [11] and [26]. Another possibility is to neglect the lengths with too low frequencies, see, e.g., [8].

[10] We demonstrate the pooling on the example of the first short story, see Table 1. There are 19 words of length six, three words of length seven, and four words of length eight. Therefore, these word lengths are pooled into one category, with the weighted mean being $(19\times6+3\times7+4\times8)/(19+3+4)=6.42$. The mean morpheme length is evaluated as the mean length of morphemes in all words from this category.

[11] The values of the parameter *b* were determined using NLREG software (www.nlreg.com).

	L1		L2		L3		L4		L5	
WL	*MML*	*fr*	*MML*	*fr*	*MML*	*fr*	*MML*	*fr*	*MML*	*fr*
1	3.05	81	2.98	49	3.17	63	2.64	47	3.12	42
2	2.18	132	2.32	79	2.07	107	2.13	82	2.03	76
3	1.85	158	1.83	65	1.81	101	1.85	82	1.76	69
4	1.77	108	1.86	45	1.82	70	1.74	37	1.62	36
5	1.62	36							1.73	25
5.18							1.58	17		
5.23					1.74	40				
5.24			1.61	25						
6.17	1.73	12								
6.20									1.76	10
a	3.05		2.98		3.17		2.64		3.12	
b	-0.39		-0.38		-0.44		-0.31		-0.42	
R^2	0.934		0.976		0.899		0.998		0.833	

Tab. 2. Word length (in morphemes) and morpheme length (in phonemes): letters, word, types

	P1		P2		P3		P4		P5	
WL	*MML*	*fr*	*MML*	*fr*	*MML*	*fr*	*MML*	*fr*	*MML*	*fr*
1	3.25	60	3.07	46	4.17	143	3.68	90	3.00	46
2	2.47	111	2.25	62	2.70	184	235	205	2.28	66
3	2.07	121	2.03	62	2.18	165	1.98	232	2.00	82
4	1.93	111	1.94	51	1.99	143	1.89	220	1.86	49
5	1.84	79	1.79	41	1.85	90	1.71	142	1.71	23
6	1.85	21					1.73	61		
6.36					1.85	39				
6.52			1.75	27						
7									1.53	11
7.14	1.74	14					1.57	28		
a	3.25		3.07		4.17		3.68		3.00	
b	-0.35		-0.34		-0.52		-0.48		-0.35	
R^2	0.967		0.955		0.965		0.947		0.995	

Tab. 3. Word length (in morphemes) and morpheme length (in phonemes):
studies on philosophy, word types

The MAL expressed by formula (2) fits the relation between word length in morphemes and the mean morpheme length in phonemes very well for 13 out of 15 texts under analysis. The fit for text L3 is practically on the threshold of 0.9. The only exception with a low value of R^2 is text L5. However, this text contains only 258 word types (it is the shortest in our sample), and, moreover, there are exactly ten words (i.e., the minimum which is accepted) in the pooled category with the longest words. If this criterion is made stricter and words with length five or more are pooled into one category, the fit becomes acceptable, with $R^2=0.91$. The relation between word length and morpheme length is demonstrated also in Figure 1, where data for text P1 (see Table 3) are used.

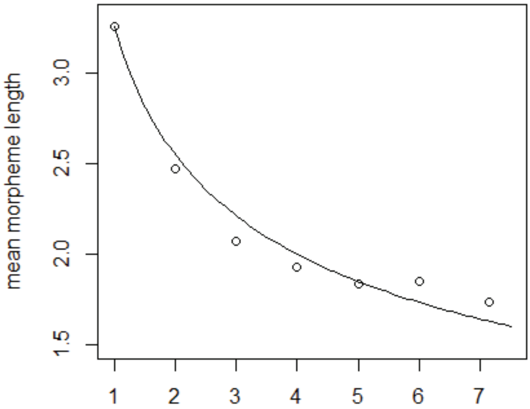

Fig. 1. The MAL in text P1

One can see a different pattern of behaviour if word tokens (as opposed to word types) are analysed (see data for short stories in Table 4). In some texts, the relation between the mean morpheme length fluctuates, and one cannot speak about a decreasing trend in general. In general, the MAL in the form given by (2) fails to achieve an acceptable fit.[12]

	S1		S2		S3		S4		S5	
WL	*MML*	*fr*	*MML*	*fr*	*MML*	*fr*	*MML*	*fr*	*MML*	*fr*
1	2.44	312	2.33	109	2.74	235	2.27	122	2.31	118
2	1.90	273	2.11	118	1.83	184	1.94	143	1.81	132
3	1.72	229	1.78	80	1.77	138	1.77	105	1.62	108
4	1.74	121	1.84	50	1.80	74	1.74	37	1.59	42
5	1.57	40								
5.20					1.72	45				
5.22							1.61	18		
5.23			1.61	26						
6.15	1.72	13								
6.17									1.80	12

Tab. 4. Word length (in morphemes) and morpheme length (in phonemes): short stories, word tokens

[12] The MAL in the form given by (1) fits the data for word tokens very well. However, using this formula would mean fitting roughly five or six data points with a function with three uninterpreted parameters.

This behaviour can be explained – admittedly, only speculatively for the time being – as a display of a competition between two "language forces" represented by the MAL on the one hand, and by the Zipf law of brevity on the other. According to the latter, shorter units are preferred. If the law of brevity is valid also within words of particular lengths (e.g., if words consisting of three shorter morphemes occur more often than words with three longer morphemes), the MAL may (see, e.g., [17]) or may not (see, e.g., text S1 in Table 4) hold for word tokens, depending on how strongly the law of brevity prefers shorter morphemes.

4 CONCLUSION

The presented results corroborate the validity of the MAL on new language material. In addition, we are able to provide some hints towards an interpretation of the parameters of the model.[13]

First, the parameter *a* is the mean number of phonemes in monomorphemic words in a text. The parameter *b* determines the steepness of the curve, which corresponds to the rate of the shortening of the mean length of morphemes when words get longer (the smaller the value of *b*, the steeper the curve). Moreover, the values of *a* and *b* strongly correlate with the number of types in a text (the values of the Pearson correlation coefficient are 0.81 and -0.83, respectively), as well as with each other (0.93). These findings are consistent with the behaviour of the length of word types measured in syllables reported in [28] – the length of word types increases with the increasing text length. The positive correlation between the values of the parameter *a* and text length in word types indicates that the length of word types in morphemes behaves analogously. On the other hand, the negative correlation between the number of word types in a text and the value of parameter *b* suggests that the mean morpheme length decreases (with the increasing length of word types) more steeply in longer texts. The same interpretation of the parameters of the MAL at the level word – syllable – phoneme in Czech prosaic texts (and much weaker correlations in poems) can be found in [29].

A more precise and empirically based characterization of the interaction of the MAL with other language laws (such as, e.g., with the law of brevity, as discussed in Section 3) remains an open question for future research.

[13] The results, of course, depend on the segmentation rules we applied. However, they mostly follow the commonly accepted rules for the Czech language from [18] and [19], with the most important modification being a deeper segmentation for pronouns. As pronouns are a closed class of words, and we consider word types (and not word tokens), the change in their segmentation should not influence the results too much. Nevertheless, an analysis of the impact of segmentation rules (the choice of which is always at least partly subjective) can be an interesting topic for a future study.

ACKNOWLEDGEMENTS

The authors are thankful to Viktor Elšík (Department of Linguistics, Faculty of Arts, Charles University in Prague) for his advices on the morphological segmentation.

The research has been supported by grant "Podpora talentovaných studentů doktorského studia na Ostravské univerzitě III (ev. č. smlouvy 07359/2019/RRC)" (K. Pelegrinová) and by VEGA grant 2/0096/21 (J. Mačutek).

References

[1] Cramer, I. M. (2005). Das Menzerathsche Gesetz. In R. Köhler, G. Altmann and R. G. Piotrowski (eds.), Quantitative Linguistics. An International Handbook. Berlin/New York: de Gruyter, pages 659–688.

[2] Popescu, I.-I., Altmann, G., Grzybek, P., Jayaram, B. D., Köhler, R., Krupa, V., Mačutek, J., Pustet, R., Uhlířová, L., and Vidya, M. N. (2009). Word Frequency Studies. Berlin/New York: de Gruyter.

[3] Bentz, C., and Ferrer-i-Cancho, R. (2016). Zipf's law of abbreviation as a language universal. In C. Bentz, G. Jäger and I. Yanovich (eds.), Proceedings of the Leiden Workshop on Capturing Phylogenetic Algorithms for Linguistics. University of Tübingen, online publication system. Available at: https://publikationen.uni-tuebingen.de/xmlui/handle/10900/68558.

[4] Menzerath, P. (1954). Die Architektonik des deutschen Wortschatzes. Bonn: Dümmler.

[5] Altmann, G. (1980). Prolegomena to Menzerath's law. In R. Grothjahn (ed.), Glottometrika 2. Bochum: Brockmeyer, pages 1–10.

[6] Altmann, G., and Schwibbe, M. H. (1989). Das Menzerathsche Gesetz in informationsverarbeitenden Systemen. Hildesheim: Georg Olms Verlag.

[7] Kelih, E. (2010). Parameter interpretation of Menzerath law: Evidence from Serbian. In P. Grzybek, E. Kelih and J. Mačutek (eds.), Text and Language. Structures, Functions, Interrelations, Quantitative Perspectives. Wien: Praesens, pages 71–79.

[8] Mačutek, J., and Rovenchak, A. (2011). Canonical word forms: Menzerath-Altmann law, phonemic length and syllabic length. In E. Kelih, V. Levickij and V. Matskulyak (eds.), Issues in Quantitative Linguistics 2. Lüdenscheid: RAM-Verlag, pages 136–147.

[9] Mačutek, J., and Mikros, G. K. (2015). Menzerath-Altmann law for word length motifs. In G. K. Mikros and J. Mačutek (eds.), Sequences in Language and Text. Berlin/Boston: de Gruyter, pages 125–131.

[10] Teupenhayn, R., and Altmann, G. (1984). Clause length and Menzerath's law. In J. Boy and R. Köhler (eds.), Glottometrika 6. Bochum, Brockmeyer, pages 127–138.

[11] Mačutek, J., Čech, R., and Milička, J. (2017). Menzerath-Altmann law in syntactic dependency structure. In S. Montemagni and J. Nivre (eds.), Proceedings of the Fourth International Conference on Dependency Linguistics (Depling 2017). Linköping: Linköping University Press, pages 100–107.

[12] Köhler, R. (2015). Linguistic motifs. In G. K. Mikros and J. Mačutek (eds.), Sequences in Language and Text. Berlin/Boston: de Gruyter, pages 89–108.

[13] Gerlach, R. (1982). Zur Überprüfung des Menzerath'schen Gesetzes im Bereich der Morphologie. In W. Lehfeldt and U. Strauss (eds.), Glottometrika 4. Bochum: Brockmeyer, pages 95–102.

[14] Krott, A. (1996). Some remarks on the relation between word length and morpheme length. Journal of Quantitative Linguistics, 3(1), pages 29–37.

[15] Hřebíček, L. (1995). Text Levels. Constructs, Constituents and the Menzerath-Altmann Law. Trier: WVT.

[16] Altmann, G., Erat, E., and Hřebíček, L. (1996). Word length distribution in Turkish texts. In P. Schmidt (ed.), Glottometrika 15. Trier: WVT, pages 195–204.

[17] Milička, J. (2014). Menzerath's law: The whole is greater than the sum of its parts. Journal of Quantitative Linguistics, 21(2), pages 85–99.

[18] Slavíčková, E. (1975). Retrográdní morfematický slovník češtiny s připojenými inventárními slovníky českých morfémů kořenových, prefixálních a sufixálních. Praha: Academia.

[19] Komárek, M., Kořenský, J., Petr, J., and Veselková, J. (1986). Mluvnice češtiny. Svazek 2. Tvarosloví. Praha: Academia.

[20] Komárek, M. (2016). Příspěvky k české morfologii. 2nd ed. Olomouc: Periplum.

[21] Altmann, G. (1992). Das Problem der Datenhomogenität. In B. Rieger (ed.), Glottometrika 13. Bochum: Brockmeyer, pages 287–298.

[22] Grzybek, P. (2013). Homogeneity and heterogeneity within language(s) and text(s): Theory and practice of word length modeling. In R. Köhler and G. Altmann (eds.), Issues in Quantitative Linguistics 3. Lüdenscheid: RAM-Verlag, pages 66–99.

[23] Williams, J. R., Bagrow, J. P., Danforth, C. M., and Dodds, P. S. (2015). Text mixing shapes the anatomy of rank-frequency distributions. Physical Review E, 91, 052811.

[24] Čech, R., Kosek, P., Mačutek, J., and Navrátilová, O. (2020). Proč (někdy) nemíchat texty aneb Text jako možná výchozí jednotka lingvistické analýzy. Naše řeč, 103(1–2), pages 24–36.

[25] Kubát, M. (2016). Kvantitativní analýza žánrů. Ostrava: Ostravská univerzita.

[26] Mačutek, J., Chromý, J., and Koščová, M. (2019). A data-based classification of Slavic languages: Indices of qualitative variation applied to grapheme frequencies. Journal of Quantitative Linguistics, 26(1), pages 66–80.

[27] Mačutek, J., and Wimmer, G. (2013). Evaluating goodness-of-fit of discrete distribution models in quantitative linguistics. Journal of Quantitative Linguistics, 20(3), pages 227–240.

[28] Kelih, E. (2012). On the dependency of word length on text length. Empirical results form Russian and Bulgarian parallel texts. In S. Naumann, P. Grzybek, R. Vulanović and G. Altmann (eds.), Synergetic Linguistics. Text and Language as Dynamic Systems. Wien: Praesens, pages 67–80.

[29] Čech, R., and Mačutek, J. (2021). The Menzerath-Altmann law in Czech poems by K. J. Erben. Proceedings of the Conference Plotting Poetry 4 (in print).