Radek Čech, Veronika Vincze, Gabriel Altmann

On Motifs and Verb Valency

Abstract: The present study mainly scrutinizes the question of motif-like characters of verb valency. Do they behave as other linguistic units? Tests are performed using the Czech text Šlépěj (Footprint) written by Karel Čapek and the Hungarian translation of G. Orwell´s *1984*. The rank-frequency, the spectrum of motifs and the relation between length and frequency are examined. The Zipf-Mandelbrot distribution is used for the rank-frequency and spectrum as a model; the relationship between length and frequency is modelled by the Lorentzian function. For a determination of verb valency, a full valency approach is used. Results show that valency motifs are regular language entities.

Keywords: motifs, verb valency, distribution

1 Introduction

It is well known that any names, concepts, definitions and criteria in science are conventions. They are set up in order to identify and determine a "real" entity which should be concerned or analysed. They may refer to things, actions, properties, time, place, relations, etc. One can coin them freely but their classes must be useful for some scientific purpose, e.g. description, classification, comparison, modelling, search for links to other properties, search for laws, etc. Some linguistic entities have been known already in old Greece and India and new ones have been coined especially in the last two centuries. In recent decades, a new unit called motif has been used in linguistics. The aim of this study is to apply this unit in a syntactic analysis. Specifically, we focus on a verb valency and we scrutinize whether so called verb valency motifs behave as other linguistic units.

The inventor of motifs was the musicologist M.G. Boroda who introduced a

Radek Čech: Department of Czech Language, University of Ostrava, Ostrava, Czech Republic, cechradek@gmail.com
Veronika Vincze: Research Group on Artificial Intelligence, Hungarian Academy of Sciences, Szeged, Hungary, vinczev@inf.u-szeged.hu
Gabriel Altmann: Lüdenscheid, ram-verlag@t-online.de

formally unequivocally identifiable musical motif (Boroda 1973, 1982, 1988), not identical with motifs defined by classical musicologists. The idea has been transferred into linguistics by Köhler (2005, 2006, 2008a,b), and today it enjoys an ever more increasing interest (cf. Köhler, Naumann 2008, 2009, 2010; Mačutek 2009; Sanada 2010; Beliankou, Köhler, Naumann 2013; Köhler 2015; Milička 2015; Mačutek, Mikros 2015).

In the article, we first introduce the linguistic background of the analysis and its methodology (Section 2 and 3), then the frequency distribution, frequency spectrum and relationship between the frequency and the length of verb valency motifs is modelled (Section 4 and 5). Finally in Section 6 we present further possibilities of the research.

2 Verb Valency

As a matter of fact, valency of the verb is a subset of its polytexty: it is measured on the basis of the environment of the verb. Since the size of the subset can easily be stated, sentences of a text can be written as a numerical (or a symbolic) sequence of verb valencies. One must, however, differentiate between the possible subset of valencies which can be found in dictionaries and the topical subset observed in the given sentence which may be different for each text.

After the sequence of valencies is stated, valency motifs can be constructed which in this case are the subsequences of non-decreasing numbers. The better specified the text, i.e. the greater the valency of the verbs, the greater the mean of the sequence. Hence, even the individual motifs can be characterized, e.g. by their average, standard deviation, range, etc. Here we shall concentrate on the usual quantitative motifs representing valency.

3 Method

For a determination of verb valency, a full valency approach is used (cf., Čech et al. 2010). Contrary to an original conception of valency (Allerton 2005), there is no differentiation between obligatory complements and non-obligatory adjuncts in this approach; all sentence elements which are dependent on a verb constitute the (full) verb valency. The approach was proposed as a reaction to the lack of clear criteria for distinguishing complements and adjuncts. Present results (Čech et al. 2010; Vincze 2014) seem to justify this concept and, moreover,

the approach seems to be applicable for the analysis of syntactic dependency relationships in general.

Specifically, for each verb in a text a full valency frame is determined as follows; let us start with a sentence (1):

(1) *My father gave four books to Mary yesterday evening in Berlin*

In accordance with syntactic dependency formalism (Meľčuk 1988), it is possible to express syntactic relationships in a form of graph (see Fig. 1). There are four sentence elements (*father, books, to, yesterday, in*) directly dependent on the verb *gave*, consequently, all of them constitute the full valency frame of the verb. Another way form presentation is linear, simply joining *gave* by means of an edge with *father, books, to, yesterday*, and *in*.

In the present study, only the number of elements constituting the full valency frame is used for the analysis; for each predicative verb in a text the number is calculated. Thus, a sequence of numbers is obtained expressing the size of full valency frames and, then, it is possible to determine the motifs based on the size of full valency frame; we call them FVS-motifs.

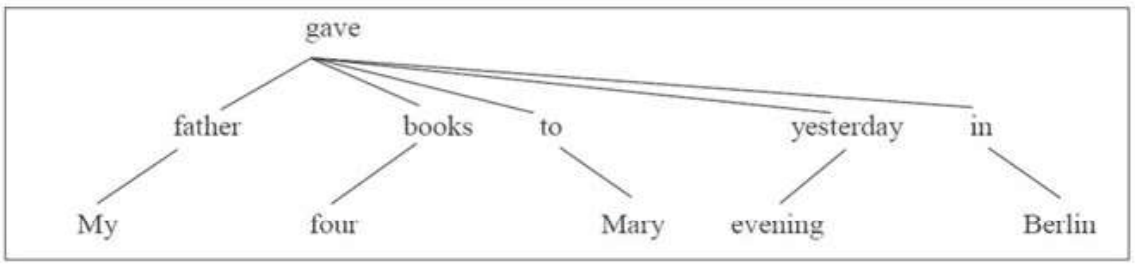

Fig. 1: Syntactic tree of the sentence (1) based on dependency syntax formalism

The minimum size of full valency frame equals zero, for instance it occurs in imperative sentences, cf. the English sentence (2)

(2) *Run*!

or in pronoun-dropping (or null-subject) realization of sentences, cf. Russian (3)

(3) *Пришёл*
[prišjol; he came]

or in sentences where the subject is not realized at all, cf. Czech (4)

(4) *Prší.*
[it rains].

Theoretically, the maximum size of full valency frame is unlimited, however, our empirical findings reveal that it does not exceed 6 (for Czech) and 10 (for Hungarian), as shown in Tab. 1 and 2 (See Appendix).

The short story Šlépěj (Footprint) written by the Czech writer Karel Čapek (1917) is used for the present analysis. The text contains 301 predicative verbs; to each verb, the size of full valency frame was annotated manually. A sequence of the sizes from the text is as follows:

[4, 3, 3, 4, 2, 3, 2, 2, 2, 1, 4, 3, 3, 3, 2, 0, 2, 1, 3, 3, 2, 1, 1, 1, 2, 2, 2, 2, 0, 2, 2, 2, 2, 0, 3, 3, 3, 1, 3, 1, 0, 1, 0, 2, 2, 1, 2, 1, 2, 2, 1, 3, 3, 2, 2, 2, 1, 1, 1, 2, 2, 2, 1, 4, 2, 3, 2, 2, 2, 1, 2, 2, 4, 0, 2, 2, 2, 2, 1, 2, 2, 2, 2, 5, 2, 0, 3, 4, 1, 3, 1, 3, 3, 3, 4, 1, 3, 2, 3, 2, 3, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2, 3, 3, 3, 2, 0, 2, 1, 2, 2, 2, 4, 2, 3, 5, 4, 4, 3, 2, 2, 2, 2, 1, 3, 2, 2, 2, 4, 3, 2, 1, 2, 2, 2, 1, 4, 4, 2, 1, 2, 2, 2, 2, 1, 2, 4, 3, 2, 2, 3, 3, 4, 0, 3, 2, 2, 0, 2, 2, 2, 3, 2, 1, 1, 2, 1, 2, 3, 2, 2, 1, 4, 1, 3, 2, 3, 2, 1, 2, 3, 3, 2, 3, 3, 1, 4, 1, 1, 3, 2, 4, 4, 2, 1, 1, 2, 2, 3, 1, 1, 2, 2, 3, 2, 2, 1, 2, 2, 2, 2, 1, 2, 2, 1, 1, 2, 2, 3, 2, 2, 2, 4, 1, 3, 3, 2, 2, 4, 1, 3, 3, 1, 3, 1, 3, 2, 2, 1, 2, 1, 1, 2, 2, 5, 3, 1, 3, 3, 1, 3, 3, 1, 3, 1, 3, 2, 2, 2, 2, 1, 4, 2, 2, 6, 2, 3, 2, 2, 2, 1, 3, 1, 3, 3, 2, 3, 1, 2, 2, 2, 4, 1, 1, 1, 2, 2, 2, 2, 1, 3, 2]

Analogically to L-motifs which are defined as a sequence of equal or increasing length values (cf. Köhler 2015), we define FVS-motifs as a sequence of equal or increasing numbers of full valency frames. For illustration, see the first 10 FVS-motifs in the text:

4
3-3-4
2-3
2-2-2
1-4
3-3-3
2
0-2
1-3-3
2

For the complete text we may present the frequency of individual motif-types in form of a rank-frequency distribution. The data are presented in Tab. 1.

Since the FVS-motif is a new linguistic unit, one may conjecture that it abides by a regular distribution. If it does not abide by a model which has been used for well-established linguistic units (such as words, morphemes, syllables

etc.), it should be either rejected as meaningless or operationalized in a different way. As for the model, we apply the well-proven Zipf-Mandelbrot distribution which was used, besides other things, with an excellent result for modelling the L-motif distribution (Köhler 2015).

Fitting the model to the data as presented in the last column of Tab. 1 yields results as follows: the parameters are estimated as $a = 0.9859$, $b = 1.9063$; the $\chi 2 = 2.7839$ with 36 degrees of freedom; the probability is $P = 1.0$, as shown in the last row of Tab. 1. The result represents an excellent fit (cf. Fig. 2), consequently, the FVS-motif can be placed, at least with regard to the distribution, in the list of other basic linguistic units. It must be remarked that the Zipf-Mandelbrot distribution is only one of several ones capturing the given data. Some of them are used frequently in linguistics (e.g. negative hypergeometric, Pólya, zeta, right truncated Zipf-Alekseev, etc.)

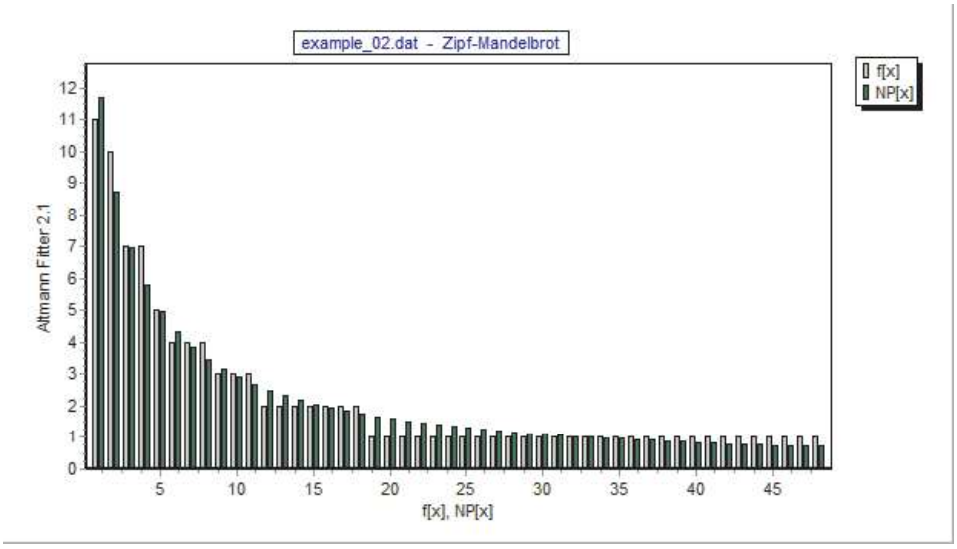

Fig. 2: Graph of the Zipf-Mandelbrot distribution as fitted to the data from Tab. 1

4 The Spectrum

Further, a transformation of the data (from Tab. 1) to a frequency spectrum, which expresses the number of FVS-motifs occurring in the text exactly x times, allows another evaluation of the behavior of the unit. Applying the transformation $x \leq f_x < x+1$ (cf. e.g. Wimmer et al. 2003: 119) to the Zipf-Mandelbrot distribution the resulting spectrum should be

$$P_x = C\left[\frac{1}{x^a} - \frac{1}{(x+1)^a}\right] \quad (1)$$

Since modelling by means of distributions or alternatively, by functions, does not distort the "reality", we apply the above function and consider C (the normalizing constant) simply as the sum of cases.

In Tab. 1, one can see that there are 30 motifs occurring exactly once; 7 motifs occurring each twice, etc. In this way we obtain the observed and computed frequency spectrum of FVS-motifs in the short story Šlépěj (Footprint) as presented in Tab. 3 (See Appendix).

The model is very satisfactory and can be preliminarily accepted. Its advantage is the existence of only one parameter. However, fitting the Zipf-Mandelbrot distribution, we would also obtain acceptable results (a = 2.314, b = 0.5333, $\chi 2$ = 2.33 with 3 degrees of freedom and the probability P = 0.721); even the simple zeta-distribution is sufficient (a = 2.047, $\chi 2$ = 1.9303 with 6 DF and P = 0.926).

Since the software yields also figures of results, below, in Fig. 3 and 4, we show the fitting of the Zipf-Mandelbrot and the zeta distributions.

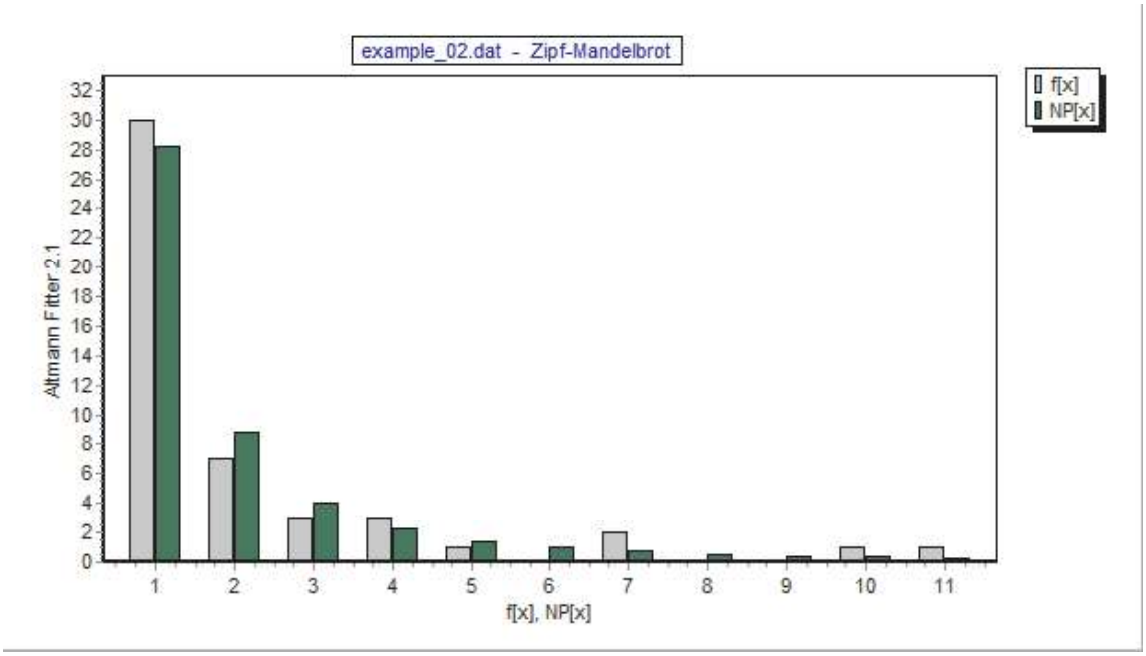

Fig. 3: Graph of the Zipf-Mandelbrot distribution as fitted to the data from Tab. 3

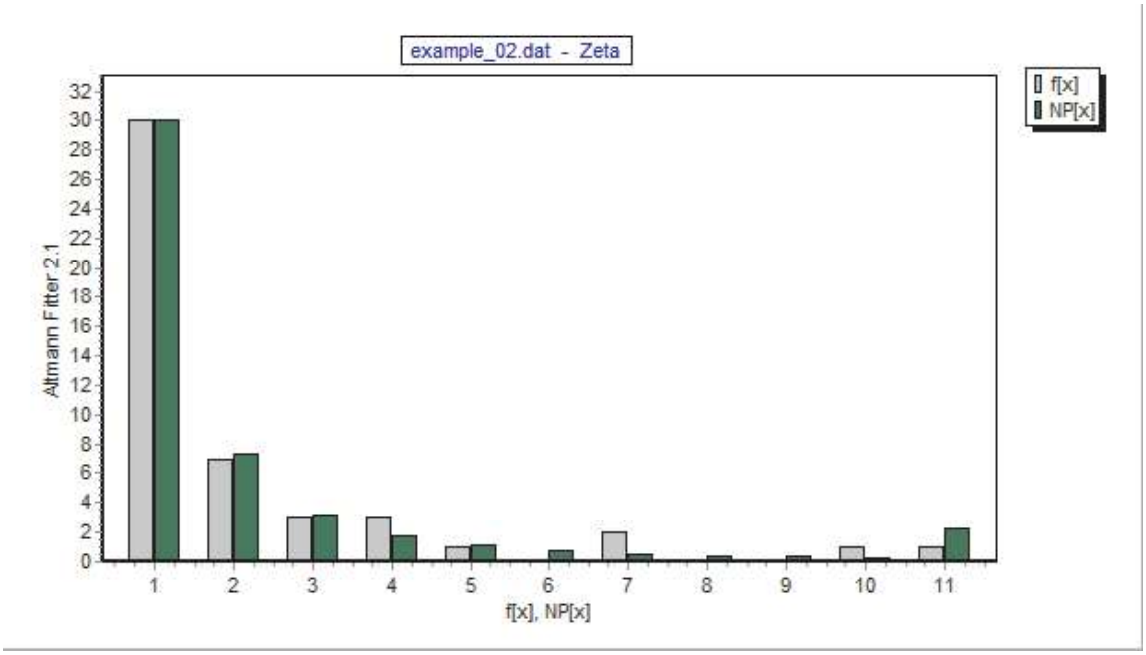

Fig. 4: Graph of the zeta distribution as fitted to the data in Tab. 3

The spectrum of the Hungarian data is presented in Tab. 4 (See Appendix).

Evidently, formula (1) is appropriate, even if here we used it as a simple function. The results testify to the fact that motifs behave like all other linguistic units.

5 The Link between Length and Frequency

Both the rank-frequency distribution and the frequency spectrum of FVS-motifs corroborate the linguistic status of this unit. Therefore, it seems reasonable to expect that the unit also meets some of the requirements fulfilled by basic linguistic ones. The hypothesis predicting a relationship between the frequency and length of the unit is ranked among the best corroborated ones. Thus, we assume that more frequent FVS-motifs must be shorter (on the average) than less frequent ones. Despite the fact that our data (see Tab. 3) did not allow to test the hypothesis properly – there are not enough instances in each frequency class, cf. Tab. 1 – it is possible, just preliminarily, to observe a tendency, if it exists at all. The length of the FVS-motif is counted in the number of words it consists of. In order to express the existence of a link between these two properties, we use the Lorentzian function

$$y = a + \frac{b}{1+\left(\frac{x-c}{d}\right)^2} \tag{2}$$

used several times in linguistics (Popescu et al. 2009, 2011, 2015). This function can easily be derived using the differential equation approach containing the classical components: the language constant, the effort of the speaker and the equilibrating force of the hearer (cf. Wimmer, Altmann 2005). The result of the computation is displayed in Tab. 5 (See Appendix) and Fig. 5.

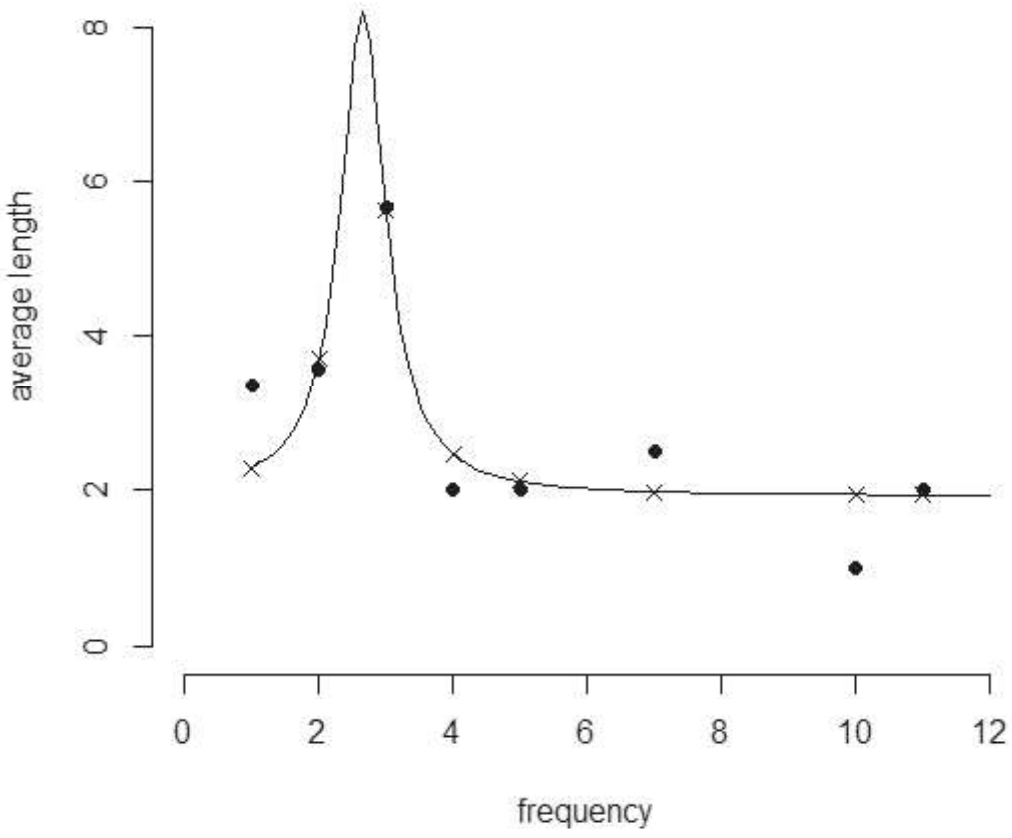

Fig. 5: Graph of the relationship between the frequency and the average length of FSV-motifs from Tab. 5

For the Hungarian data we obtain the results presented in Tab. 6 and Fig. 6. In this data we have a simple decreasing phenomenon but the oscillation is evident. Hence we try to capture it applying a simple power function

$$y = ax^b \tag{3}$$

and obtain the results presented in Tab. 6 (See Appendix).

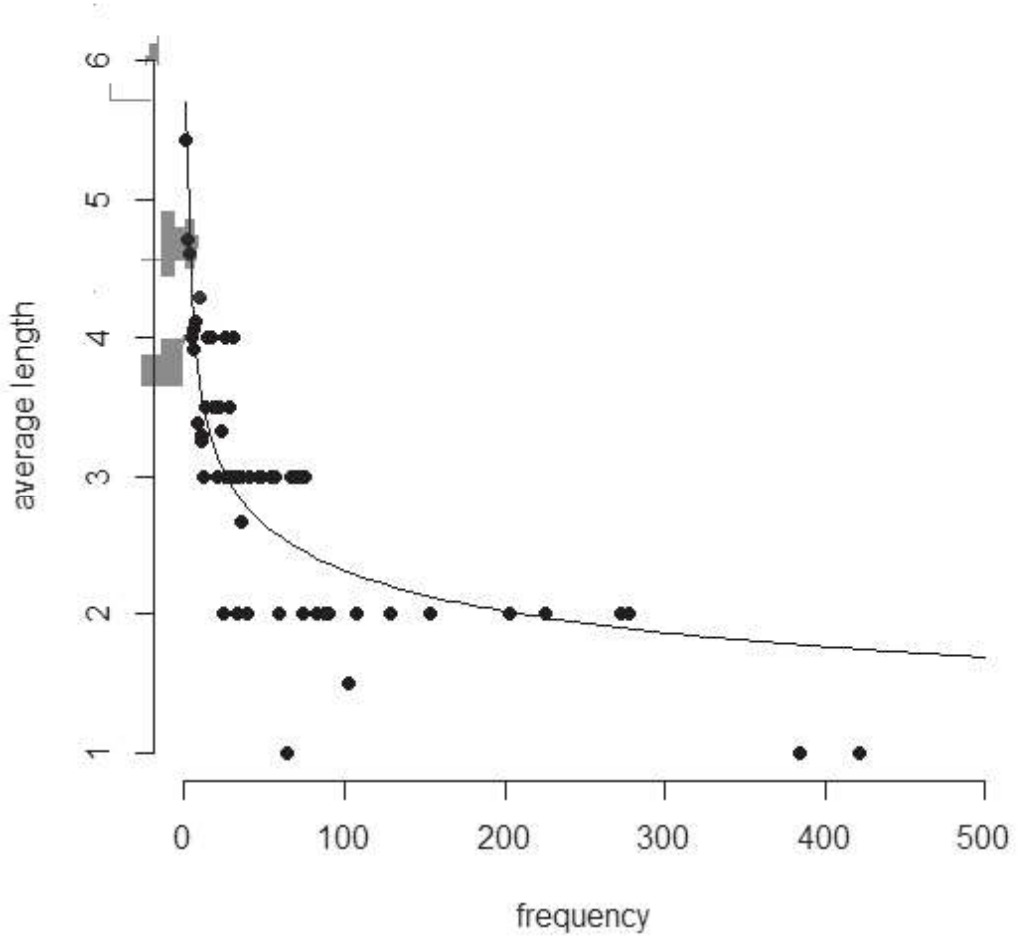

Fig. 6: Graph of the relationship between the frequency and the average length of FSV-motifs from Tab. 6

In accordance with the hypothesis, there is a tendency to emphasize the central FVS-motif length and shorten less frequent and more frequent FSV-motifs,

hence, the curve should rather bell-shaped. But the Hungarian data which are very extensive show that the dependence need not be bell-shaped. Perhaps the parameters of the functions are associated with the kind of linguistic entity. However, this is, so to say, not the last word. Analyses of many texts will, perhaps, show that the link is very complex.

Here we considered frequency as the independent variable but the other way round is possible, too.

6 Conclusion

The study reveals that verb valency motifs can be considered to be the linguistic unit which share the same characteristics as the majority of well-established traditional units. Specifically, both the rank frequency and spectrum abide by the Zipf-Mandelbrot distribution; further, there is the relationship between the frequency and the length of the motif.

Valency is a property of the verb stated on the basis of its associated environment. Verbs themselves express a number of various activities which need not be restricted to Man. Capturing them, e.g. according to Ballmer (1982), cf. also Köhler, Altmann (2014: 66 f.) in form of B-motifs, one can study the relation of the text to the reality. This aspect has nothing to do with grammar but rather with the semantics of verbs. A further possibility is to give names to the members of the valency, e.g. N, V, Adv, Prep, Pron, ... or to the phrases in which they occur, and then study the form of the valencies, i.e. consider the frequencies and similarities of valency frames.

References

Allerton, D. J. (2005). Valency Grammar. In K. Brown (Ed.), *The Encyclopedia of Language and Linguistics* (pp. 4878-4886). Elsevier Science Ltd.

Ballmer, Th.T. (1982). *Biological foundations of linguistic communication*. Amsterdam/Philadelphia: Benjamins.

Beliankou, A., Köhler, R., & Naumann, S. (2013). Quantitative properties of argu mentation motifs. In I. Obradović, E. Kelih, & R. Köhler (Eds.), *Methods and Applications of Quantitative Linguistics* (pp. 33-43). *Selected papers of the VIIIth Inter national Conference on Quantitative Linguistics (QUALICO) in Belgrade, Serbia*.

Boroda, M.G. (1973). K voprosu o metroritmičeski elementarnoj edinice v muzyke. *Bulletin of the Academy of Sciences of the Georgian SSR*, 71(3), 745–748.

Boroda, M.G. (1982). Die melodische Elementareinheit. In J.K. Orlov, M.G. Boroda, & I.Š. Nadarejšvili (Eds.), *Sprache, Text, Kunst. Quantitative Analysen* (205–222). Bochum: Brockmeyer.

Boroda, M.G. (1988). Towards a problem of basic structural units of musical texts. *Musikometrika 1* (pp. 11-69), Bochum: Brockmeyer.

Čech, R., Pajas, P., & Mačutek, J. (2010). Full valency. Verb valency without distinguishing complements and adjuncts. *Journal of Quantitative Linguistics*, 17, 291–302.

Köhler, R. (2005). Synergetic linguistics. In R. Köhler, G. Altmann, & R.G. Piotrowski (Eds.), *Quantitative Linguistics: An International Handbook* (pp. 760–774). Berlin/New York: de Gruyter.

Köhler, R. (2006). The frequency distribution of the length of length sequences. In J. Genzor, M. Bucková (Eds.), *Favete linguis. Studies in honour of Viktor Krupa* (pp. 145-152). Bratislava: Slovak Academy Press.

Köhler, R. (2008a). Word length in text. A study in the syntagmatic dimension. In S. Mislovičová (Ed.), *Jazyk a jazykoveda v pohybe* (pp. 421–426). Bratislava: Veda.

Köhler, R. (2008b). Sequences of linguistic quantities. Report on a new unit of in vestigation. *Glottotheory 1(1), 115–119.*

Köhler, R. (2015). Linguistic motifs. In G.K. Mikros, & J. Mačutek (Eds.), *Sequences in Language and Text* (pp. 89-108). Berlin/Boston: de Gryuter.

Köhler, R., & Altmann, G. (2014). *Problems in Quantitative Linguistics Vol. 4.* Lüdenscheid: RAM-Verlag

Köhler, R., & Naumann, S. (2008). Quantitative text analysis using L-, F- and T-segments. In C. Preisach, H. Burkhardt, L. Schmidt-Thieme, R. Becker, R. (Eds.), *Data Analysis. Machine Learning and Applications* (pp. 635-646). Berlin/ Heidelberg: Springer.

Köhler, R., & Naumann, S. (2009). A contribution to quantitative studies on the sentence level. In R. Köhler (Ed.), *Issues in Quantitative Linguistics* (pp. 34-57). Lüdenscheid: RAM-Verlag.

Köhler, R., & Naumann, S. (2010). A syntagmatic approach to automatic text clas sification. Statistical properties of F- and L-motifs as text characteristics. In P. Grzybek, E. Kelih, J. Mačutek (Eds.), *Text and Language. Structures, functions, interrelations, quantitative perspectives* (pp. 81-89). Wien: Praesens.

Mačutek, J. (2009). Motif richness. In R. Köhler (Ed.), *Issues in Quantitative Lin guistics* (pp. 51-60). Lüdenscheid: RAM-Verlag.

Mačutek, J., & Mikros, G. (2015). Menzerath-Altmann law for word length motifs. In G.K. Mikros, & J. Mačutek (Eds.), *Sequences in Language and Text* (pp. 125-131). Berlin/Boston: de Gryuter.

Meľčuk, I. (1988). *Dependency Syntax: Theory and Practice*. Albany: State Univers ity of New York Press.

Milička, J. (2015). Is the distribution of L-motifs inherited from the word length distribution In G.K. Mikros, & J. Mačutek (Eds.), *Sequences in Language and Text* (pp. 133-145). Berlin/Boston: de Gryuter.

Popescu, I.-I., Altmann, G., Grzybek, Jayaram, B.D., Köhler, R., Krupa, V., P., Mačutek, J., Pustet, R., Uhlířová, L., & Vidya, M.N., (2009). *Word frequency studies*. Berlin/ New York: Mouton de Gruyter.

Popescu, I. I., Čech, R., & Altmann, G. (2011). Vocabulary richness in Slovak poetry. *Glottometrics*, 22, 62–72.

Popescu, I.I., Lupea, M., Tatar, D., & Altmann, G. (2015). *Quantitative Analysis of Poetic Texts*. Berlin: de Gruyter.

Sanada, H. (2010). Distribution of motifs in Japanese texts. In P. Grzybek, E. Kelih, & J. Mačutek, (Eds.), *Text and Language. Structures, functions, interrelations, quantitative perspectives* (pp. 183-194) Wien: Praesens.

Vincze, V. (2014). Valency frames in a Hungarian corpus. *Journal of Quantitative Linguistics* 21(2), 153–176.

Wimmer, G., Altmann, G., Hřebíček, L., Ondrejovič, S., & Wimmerová, S. (2003) *Úvod do analýzy textov* [Introduction to text analysis]. Bratislava: VEDA.

Wimmer, G., & Altmann, G. (2005). Unified theory of some linguistics laws. In Köhler, R., Altmann, G., Piotrowski, R., (Eds.) *Quantitative Linguistics. An International Handbook* (pp. 791-807). Berlin, New York: Walter de Gruyter.

Appendix

Tab. 1: Rank-frequency distribution of the FSV motifs in the short story Šlépěj (Footprint)

Rank	Motif	Fr.	ZI-MA	Rank	Motif	Fr.	ZI-MA
1	1-3	11	11.68	25	2-3-3	1	1.30
2	2	10	8.73	26	3-3-3	1	1.26
3	1-3-3	7	6.97	27	1-2-4	1	1.21
4	2-3	7	5.81	28	2-2-4	1	1.17
5	1-4	5	4.98	29	3-3-4	1	1.14
6	3	4	4.35	30	1-4-4	1	1.10
7	2-2-2	4	3.87	31	2-4-4	1	1.07
8	2-2	4	3.49	32	2-3-5	1	1.04
9	1-1-1-2-2-2-2	3	3.17	33	2-2-6	1	1.01
10	1-1-2-2-3	3	2.91	34	4-4	1	0.98
11	1-2-2-2-2	3	2.69	35	0-1	1	0.95
12	1-2-2	2	2.50	36	0-2-2	1	0.93
13	1-2	2	2.33	37	0-2-2-2-3	1	0.91
14	0-2	2	2.19	38	0-3	1	0.88
15	0-2-2-2-2	2	2.06	39	0-3-3-3	1	0.86
16	1-2-2-2-4	2	1.95	40	0-3-4	1	0.84
17	2-2-2-2	2	1.84	41	1-1-2-2-5	1	0.82
18	2-2-2-4	2	1.75	42	1-2-2-2	1	0.80
19	1	1	1.67	43	1-2-2-2-3-3-3	1	0.79

Rank	Motif	Fr.	ZI-MA	Rank	Motif	Fr.	ZI-MA
20	4	1	1.59	44	1-2-2-4	1	0.77
21	5	1	1.53	45	1-2-3-3	1	0.75
22	1-1-2	1	1.46	46	1-3-3-3-4	1	0.74
23	1-1-3	1	1.40	47	2-2-2-2-2-2	1	0.72
24	1-2-3	1	1.35	48	2-2-3-3-4	1	0.71
a = 0.9859, b = 1.9063, n = 48, DF = 36, X^2 = 2.7839, P = 1.00							

Tab.2: Rank-frequency distribution of the FSV motifs in the Hungarian translation of Orwel's *1984*

Rank	Motif	Fr.	ZI-MA	Rank	Motif	Fr.	ZI-MA
1	2	422	514.92	244	1-1-2-2-2-4	2	2.22
2	3	384	391.78	245	0-0-3-3	2	2.21
3	2-3	278	312.09	246	1-1-1-1-2-3	2	2.20
4	1-3	273	258.87	247	1-1-1-2-2-2	2	2.19
5	1-4	226	216.65	248	0-1-2-5	2	2.18
6	2-4	203	186.24	249	2-2-2-2-5	2	2.16
7	1-2	153	162.54	250	2-2-3-6	2	2.15
8	2-2	129	143.63	251	0-3-3-3-4	1	2.14
9	0-3	107	128.24	252	2-4-4-6	1	2.13
10	2-5	103	115.53	253	0-0-3-4	1	2.12
11	4	103	104.81	254	2-3-3-3-3-3-3-4	1	2.11
12	3-3	90	95.73	255	1-4-7	1	2.10
13	3-4	88	87.93	256	4-4-6	1	2.08
14	0-4	82	81.18	257	2-2-3-3-3-4	1	2.07
15	2-3-3	75	75.27	258	1-2-2-5-6	1	2.06
16	1-5	74	70.08	259	0-1-1-2-6	1	2.05
17	1-3-3	73	65.47	260	2-3-8	1	2.04

Rank	Motif	Fr.	ZI-MA	Rank	Motif	Fr.	ZI-MA
18	1-2-3	71	61.37	261	4-5-6	1	2.03
19	2-2-3	69	57.70	262	0-1-2-3-3-4	1	2.02
20	1-2-4	68	54.39	263	3-4-7	1	2.01
21	2-2-4	66	51.39	264	0-3-5-9	1	2.00
22	1	64	48.67	265	3-3-3-3-4-4	1	1.99
23	0-2	59	46.20	266	2-2-3-3-3-3-3-3-3-4	1	1.98
24	1-3-4	57	43.93	267	0-0-1-2-3	1	1.98
25	2-3-4	54	41.85	268	2-4-4-4-4	1	1.96
26	1-1-3	48	39.93	269	0-1-4-4-4	1	1.95
27	1-2-2	47	38.17	270	1-1-1-1-1-4	1	1.94
28	2-2-2	40	36.53	271	2-2-2-4-5	1	1.93
29	3-5	39	35.01	272	1-1-1-1-3-3-4	1	1.92
30	0-2-3	36	33.60	273	0-1-3-3-6	1	1.91
31	2-2-5	35	32.28	274	2-3-3-3-5-5	1	1.90
32	0-5	35	31.05	275	1-1-1-3-7	1	1.89
33	1-1-4	35	29.90	276	5-5	1	1.88
34	1-1-2	34	28.83	277	0-1-2-2-2-3-3-4-4	1	1.88
35	1-6	33	27.81	278	1-1-2-3-4	1	1.87
36	0-3-3	31	26.86	279	0-3-4-4-4	1	1.86
37	2-2-2-3	30	25.97	280	1-3-3-3-5-5-6	1	1.85
38	2-4-4	28	25.12	281	2-2-2-2-4-4	1	1.84
39	2-2-3-3	28	24.32	282	1-4-6	1	1.83
40	2-3-5	27	23.56	283	0-0-2-4	1	1.82
41	1-2-2-3	26	22.85	284	1-2-7	1	1.81
42	1-4-4	25	22.17	285	0-1-3-3-3	1	1.81

Rank	Motif	Fr.	ZI-MA	Rank	Motif	Fr.	ZI-MA
43	0-1-3	25	21.52	286	3-3-4-4-4	1	1.80
44	1-2-5	25	20.91	287	0-3-3-6	1	1.79
45	4-4	24	20.33	288	0-4-6	1	1.78
46	3-3-4	23	19.77	289	2-3-4-5	1	1.77
47	1-1-5	23	19.24	290	0-2-2-3-3-4	1	1.76
48	1-2-3-3	23	18.74	291	1-3-6-6	1	1.76
49	0-2-4	22	18.25	292	2-2-2-2-2-3-4	1	1.75
50	2-2-2-4	22	17.79	293	1-1-2-5	1	1.74
51	1-2-2-4	21	17.35	294	2-2-2-3-3-5	1	1.73
52	2-6	21	16.93	295	2-2-2-3-3-4	1	1.72
53	3-3-3	20	16.52	296	0-1-2-4-6	1	1.72
54	1-3-3-3	19	16.13	297	1-2-2-2-2-3-4	1	1.71
55	0-3-4	19	15.76	298	1-2-2-2-2-2-3-3	1	1.70
56	2-3-3-3	18	15.40	299	0-0-2-3-6-6	1	1.69
57	0-2-2	18	15.10	300	1-2-5-5	1	1.69
58	1-3-3-4	17	14.72	301	4-4-4-6	1	1.68
59	2-2-3-4	17	14.41	302	1-3-5-5	1	1.67
60	1-1-3-3	15	14.10	303	1-1-3-3-3-5	1	1.66
61	3-4-4	15	13.80	304	2-3-3-3-5	1	1.66
62	1-1-2-3	15	13.52	305	1-1-2-2-6	1	1.65
63	0-3-3-3	15	13.24	306	0-2-3-3-3-3	1	1.64
64	1-1-2-2	15	12.97	307	1-1-1-2-4-5	1	1.63
65	2-2-2-2-3	15	12.72	308	1-1-3-3-3-3	1	1.63
66	1-1-2-4	15	12.47	309	2-2-2-2-3-3-3-5	1	1.62
67	2-3-3-4	14	12.23	310	1-1-3-3-3-4-4	1	1.61
68	1-4-5	13	12.00	311	2-3-3-3-3-4	1	1.61

Rank	Motif	Fr.	ZI-MA	Rank	Motif	Fr.	ZI-MA
69	1-2-2-2	13	11.78	312	4-8	1	1.60
70	0-2-5	12	11.56	313	0-8	1	1.60
71	0-6	12	11.35	314	4-4-4	1	1.59
72	3-3-5	12	11.15	315	0-2-2-2-5	1	1.58
73	1-3-5	12	10.95	316	1-2-2-2-2-3-3-3	1	1.57
74	0-2-2-3	12	10.76	317	0-0-1-4	1	1.57
75	3-6	11	10.57	318	1-3-3-6	1	1.56
76	2-3-3-5	11	10.39	319	0-1-3-5	1	1.55
77	1-1-3-4	11	10.22	320	2-2-2-2-2-2-2	1	1.55
78	2-2-2-2	11	10.05	321	1-1-1-2-3-3-3	1	1.54
79	2-2-4-5	11	9.88	322	2-2-2-2-2-5	1	1.53
80	1-7	11	9.72	323	0-4-4-6	1	1.53
81	0-1-4	11	9.56	324	1-1-1-2-2-3-3	1	1.52
82	5	10	9.41	325	2-2-3-3-4-5	1	1.51
83	1-2-3-4	10	9.27	326	0-2-2-4	1	1.51
84	2-3-4-4	10	9.12	327	1-2-2-2-3-3-4-4	1	1.50
85	1-1-1-4	10	8.98	328	1-1-3-3-3-3-3	1	1.50
86	1-2-3-5	9	8.85	329	1-3-4-8	1	1.49
87	2-2-2-3-3	9	8.71	330	2-2-5-6	1	1.48
88	2-3-6	9	8.58	331	1-8	1	1.48
89	0-3-3-5	9	8.46	332	0-2-2-6	1	1.47
90	1-2-3-3-4	9	8.34	333	0-5-6	1	1.47
91	1-2-2-2-3	9	8.22	334	0-1-1-2-2-5	1	1.46
92	0-1-2-2	9	8.10	335	0-1-1-1-3	1	1.45
93	0-1-2	8	7.99	336	0-1-1-1-2	1	1.45
94	3-3-3-4	8	7.88	337	0-4-5-6	1	1.44
95	2-2-6	8	7.77	338	1-1-1-2-2-5	1	1.44
96	0-2-2-2	8	7.66	339	0-0-2-3-3-3-6	1	1.43

Rank	Motif	Fr.	ZI-MA	Rank	Motif	Fr.	ZI-MA
97	2-2-3-5	8	7.56	340	0-1-1-1-3-3	1	1.42
98	1-1	8	7.46	341	1-2-2-3-3-3-4	1	1.42
99	2-2-2-2-4	8	7.36	342	0-3-6	1	1.41
100	2-7	8	7.26	343	0-1-1-3-3-3-3-4	1	1.41
101	2-4-5	7	7.17	344	2-2-2-8	1	1.40
102	1-2-2-3-3	7	7.08	345	0-2-2-2-2-4-4-4	1	1.40
103	2-2-2-3-4	7	6.99	346	2-2-3-5-5	1	1.39
104	3-3-6	7	6.90	347	2-2-8	1	1.39
105	1-2-6	7	6.81	348	0-1-2-2-2-4	1	1.38
106	1-1-1-3	7	6.73	349	0-2-3-6	1	1.38
107	2-2-4-4	7	6.64	350	1-3-3-3-3-3	1	1.37
108	1-2-2-2-4	7	6.56	351	0-1-1-4-5	1	1.37
109	1-3-3-3-3	7	6.48	352	1-2-2-3-3-5	1	1.36
110	1-3-3-5	7	6.41	353	1-2-2-2-5-5	1	1.35
111	1-2-4-4	7	6.33	354	1-3-4-6-6	1	1.35
112	2-2-3-3-4	7	6.26	355	1-1-2-2-2-2-3-3-3	1	1.34
113	3-4-5	6	6.18	356	1-2-2-2-2-2-2-4	1	1.34
114	2-3-3-3-4	6	6.11	357	2-3-3-4-4-5	1	1.33
115	0-4-4	6	6.04	358	2-2-2-2-3-4-4	1	1.33
116	0-3-5	6	5.97	359	0-1-2-4-5	1	1.32
117	2-4-6	6	5.91	360	1-1-2-2-2-2-2	1	1.32
118	1-3-4-4	6	5.84	361	1-1-1-1-2-4	1	1.31
119	1-3-6	6	5.78	362	2-2-3-3-3-4-4	1	1.31
120	0-2-6	6	5.71	363	0-2-4-8	1	1.30
121	2-2-3-3-3	6	5.65	364	1-2-2-3-3-4-4-4	1	1.30
122	2-3-3-3-3	6	5.59	365	0-1-1-2-5	1	1.30
123	1-2-2-2-2-3	6	5.53	366	0-0-6	1	1.29
124	1-2-2-5	6	5.47	367	0-1-1-3-4	1	1.29
125	4-5	5	5.41	368	1-2-2-3-3-3	1	1.28
126	3-3-3-3	5	5.36	369	0-0-3-3-3-3	1	1.28

Rank	Motif	Fr.	ZI-MA	Rank	Motif	Fr.	ZI-MA
127	2-8	5	5.30	370	1-1-3-4-4	1	1.27
128	0-3-3-4	5	5.25	371	1-2-3-3-4-5	1	1.27
129	0-1-6	5	5.19	372	1-1-1-2-2	1	1.26
130	2-5-5	5	5.14	373	0-2-2-2-4-4	1	1.26
131	0-4-5	5	5.09	374	0-3-5-6	1	1.25
132	1-3-3-3-4	5	5.04	375	1-2-2-2-2-2-3-3-3-3-3	1	1.25
133	1-1-2-2-4	5	4.99	376	2-2-2-3-6	1	1.24
134	1-5-5	5	4.94	377	2-2-2-3-4-4	1	1.24
135	1-2-2-3-4	5	4.89	378	1-2-2-2-3-4-4	1	1.24
136	1-1-2-2-3	5	4.84	379	2-2-2-2-6	1	1.23
137	2-2-2-2-2	5	4.80	380	1-2-2-2-4-5	1	1.23
138	1-1-1-1-3	5	4.75	381	0-1-1-1-3-5	1	1.22
139	1-2-3-3-3	5	4.70	382	0-0-3-5	1	1.22
140	3-3-4-4	5	4.66	383	1-1-3-3-4	1	1.21
141	1-1-1-2-3	5	4.62	384	1-1-1-4-4	1	1.21
142	2-3-3-4-4	5	4.57	385	1-2-2-2-2-2-2	1	1.21
143	2-2-7	4	4.53	386	1-1-1-1-2-2-2-2-2-3	1	1.20
144	2-4-4-4	4	4.49	387	1-1-2-2-2-3-4	1	1.20
145	0-1-2-4	4	4.45	388	2-2-2-2-2-2-2-3	1	1.19
146	2-3-3-6	4	4.41	389	1-1-1-1-2-2-3	1	1.19
147	1-1-4-5	4	4.37	390	0-1-3-4-5	1	1.18
148	2-2-2-3-5	4	4.33	391	3-5-6	1	1.18
149	1-4-4-4	4	4.29	392	2-2-3-4-4-4	1	1.18
150	0-1-3-3	4	4.25	393	1-2-3-3-3-4	1	1.17
151	0-1-5	4	4.22	394	4-4-5	1	1.17
152	0-2-3-3	4	4.18	395	3-3-3-3-4-6	1	1.16
153	1-1-3-5	4	4.14	396	1-3-3-4-5	1	1.16
154	1-1-1-3-4	4	4.11	397	0-2-2-2-7	1	1.16
155	2-2-2-2-3-3	4	4.07	398	1-1-3-6	1	1.15
156	2-2-2-4-4	4	4.04	399	1-3-3-3-3-3-4	1	1.15

Rank	Motif	Fr.	ZI-MA	Rank	Motif	Fr.	ZI-MA
157	0-0-4	4	4.00	400	3-3-3-5-6-6	1	1.15
158	2-4-4-5	4	3.97	401	0-3-3-4-4	1	1.14
159	0-1-1-2	4	3.94	402	4-4-4-5	1	1.14
160	1-1-2-2-2	4	3.90	403	1-3-3-7	1	1.13
161	0-1	4	3.87	404	0-1-2-2-5	1	1.13
162	1-3-7	4	3.84	405	2-2-2-5-6	1	1.13
163	1-5-6	4	3.81	406	2-5-6	1	1.12
164	1-1-2-3-3	4	3.78	407	1-1-1-2-4-4	1	1.12
165	2-2-2-5	4	3.75	408	1-4-5-6	1	1.11
166	1-1-1-2	4	3.72	409	-3-10	1	1.11
167	0-7	3	3.69	410	1-3-4-6	1	1.11
168	3-3-3-3-3	3	3.66	411	1-5-5-6	1	1.10
169	1-1-4-4	3	3.63	412	0-2-3-4-4	1	1.10
170	1-1-1-2-5	3	3.60	413	1-1-3-3-4-5	1	1.10
171	1-2-2-4-4	3	3.57	414	2-2-2-3-3-6	1	1.09
172	1-1-6	3	3.55	415	1-2-3-4-4-4-4	1	1.09
173	0-1-2-3	3	3.52	416	2-4-4-4-5	1	1.09
174	0-0-3	3	3.49	417	3-3-3-4-4	1	1.08
175	1-1-1-1-3-3	3	3.46	418	1-3-3-5-5	1	1.08
176	0-1-2-2-4	3	3.44	419	3-4-6	1	1.08
177	2-2-2-2-2-2	3	3.41	420	1-2-2-3-6	1	1.07
178	3-7	3	3.39	421	1-2-4-4-4-4	1	1.07
179	2-2-2-3-3-3	3	3.36	422	1-1-2-3-3-4-4	1	1.07
180	1-1-2-2-2-2-3	3	3.34	423	1-1-1-1	1	1.06
181	0-1-1-3	3	3.31	424	0-4-4-5	1	1.06
182	0-2-2-2-3	3	3.29	425	2-2-2-2-3-5	1	1.06
183	1-2-2-2-2-4	3	3.26	426	1-2-2-2-2-2-2-3	1	1.05
184	1-2-2-6	3	3.24	427	2-2-3-3-3-3	1	1.05
185	1-3-4-5	3	3.22	428	0-1-1-1-4	1	1.05
186	1-3-3-4-4	3	3.19	429	0-2-3-3-3-3-3-3-3-3	1	1.04

Rank	Motif	Fr.	ZI-MA	Rank	Motif	Fr.	ZI-MA
187	1-1-2-3-5	3	3.17	430	0-1-1-1-2-3	1	1.04
188	1-1-1-5	3	3.15	431	0-1-2-3-3-3	1	1.04
189	1-1-2-2-2-2	3	3.13	432	0-2-2-2-3-3	1	1.03
190	2-6-6	2	3.11	433	0-0-5	1	1.03
191	2-4-5-5	2	3.08	434	2-2-2-2-3-4	1	1.03
192	0-3-5-5	2	3.06	435	1-2-3-3-3-3-3	1	1.02
193	1-2-3-4-4	2	3.04	436	2-2-3-4-7	1	1.02
194	3-3-3-5	2	3.02	437	2-2-2-2-3-3-4	1	1.02
195	0-1-1-3-3	2	3.00	438	1-1-2-2-3-5	1	1.01
196	2-2-2-2-2-3	2	2.98	439	1-1-1-1-4	1	1.01
197	4-4-4-4	2	2.96	440	0-2-3-3-3-3-3-4	1	1.01
198	4-6	2	2.94	441	1-1-3-5-5	1	1.00
199	1-2-2-2-3-4	2	2.92	442	1-1-1-2-2-2-3	1	1.00
200	1-2-4-6	2	2.90	443	1-2-4-4-4	1	1.00
201	1-4-4-5	2	2.88	444	1-1-3-3-3	1	0.99
202	2-3-4-4-4	2	2.86	445	2-2-2-2-2-2-3	1	0.99
203	0-2-3-5	2	2.84	446	0-1-1-2-3	1	0.99
204	0-2-5-5	2	2.82	447	1-1-1-1-1-2	1	0.99
205	0-0-2	2	2.81	448	0-1-7	1	0.98
206	2-2-2-2-2-3-3	2	2.79	449	0-1-1-4	1	0.98
207	2-3-4-6	2	2.77	450	2-2-2-6	1	0.98
208	2-2-4-6	2	2.75	451	0-1-2-3-3	1	0.97
209	0-1-1-5	2	2.73	452	0-1-3-6	1	0.97
210	1-1-1-1-2-2-2	2	2.72	453	1-1-1-2-2-3-3-3-4	1	0.97
211	0-1-3-4	2	2.70	454	1-1-1-1-1-1-5	1	0.97
212	2-2-3-3-5	2	2.68	455	1-1-1-2-4	1	0.96
213	0-2-3-3-5	2	2.67	456	1-1-1-3-3-4	1	0.96
214	0-0-2-2-3	2	2.65	457	1-2-2-3-4-4	1	0.96
215	0-2-2-2-6	2	2.63	458	1-1-2-2-2-3	1	0-95
216	1-2-2-2-5	2	2.62	459	1-1-2-2-2-2-2-2-	1	0.95

Rank	Motif	Fr.	ZI-MA	Rank	Motif	Fr.	ZI-MA
					2-2-2		
217	3-4-4-5	2	2.60	460	0-2-2-2-3-3-4	1	0.95
218	0-2-3-4	2	2.58	461	0-1-2-2-2-2-3	1	0.95
219	1-2-3-3-5	2	2.57	462	2-2-2-7	1	0.94
220	0-2-2-3-4	2	2.55	463	2-2-2-2-3-3-3	1	0.94
221	1-2-2-2-2-3-3	2	2.54	464	0-1-2-2-2	1	0.94
222	1-2-2-2-2	2	2.52	465	1-2-5-6	1	0.93
223	1-2-3-6	2	2.51	466	0-2-2-2-2	1	0.93
224	0-5-5	2	2.49	467	0-2-3-3-4	1	0.93
225	0-4-5-5	2	2.48	468	0-1-2-4-4-5	1	0.93
226	1-1-3-7	2	2.46	469	0-0-1-1-1-1	1	0.92
227	1-1-1-3-3	2	2.45	470	1-2-2-2-3-3-4	1	0.92
228	0-2-2-3-3	2	2.43	471	1-1-2-2-3-3	1	0.92
229	1-2-2-3-3-4	2	2.42	472	0-0-2-3	1	0.92
230	1-2-2-2-3-3	2	2.41	473	2-3-3-3-3-3-3	1	0.91
231	0-1-4-4	2	2.39	474	0-0-1-3-3	1	0.91
232	1-1-2-3-3-3	2	2.38	475	1-2-2-2-2-2	1	0.91
233	0-2-2-5	2	2.36	476	0-2-2-2-4	1	0.91
234	0-1-2-4-4	2	2.35	477	0-3-4-4	1	0.90
235	0-4-4-4	2	2.34	478	1-1-2-7	1	0.90
236	1-2-2-4-5	2	2.32	479	1-4-4-4-4	1	0.90
237	1-1-2-2-3-4	2	2.31	480	0-3-3-3-3	1	0.90
238	1-1-2-2-2-5	2	2.30	481	1-1-4-4-5	1	0.89
239	0-1-1-2-2	2	2.29	482	0-2-3-3-7	1	0.89
240	1-1-1	2	2.27	483	1-2-3-4-4-4-5	1	0.89
241	1-2-4-5	2	2.26	484	3-3-3-7	1	0.89
242	1-1-1-1-2	2	2.25	485	0-2-2-2-3-6	1	0.88
243	1-2-3-4-5	2	2.24	486	1-2-3-3-4-4	1	0.88

Tab. 3: The frequency spectrum of the FSV-motifs in the short story Šlépěj (Footprint)

Frequency	No. of FVS-motifs	Theoretical (5)
1	30	29.90
2	7	7.87
3	3	3.41
4	3	1.84
5	1	1.13
7	2	0.53
10	1	0.24
11	1	0.19
$a = 1.4071$, $R^2 = 0.9917$		

Tab. 4: The frequency spectrum of the FSV-motifs in the Hungarian translation of Orwel's *1984*

Frequency	No. of FVS-motifs	Theoretical (5)	Frequency	No. of FVS- motifs	Theoretical (5)
1	237	235.33	34	1	0.17
2	60	67.26	35	3	0.16
3	23	30.61	36	1	0.15
4	24	17.15	39	1	0.12
5	18	10.83	40	1	0.12
6	12	7.39	47	1	0.08
7	12	5.34	48	1	0.08
8	8	4.01	54	1	0.06
9	7	3.12	57	1	0.05
10	4	2.48	59	1	0.05
11	7	2.02	64	1	0.04
12	5	1.67	66	1	0.04
13	2	1.40	68	1	0.04
14	1	1.19	69	1	0.03
15	7	1.03	71	1	0.03
17	2	0.78	73	1	0.03
18	2	0.69	74	1	0.03
19	2	0.61	75	1	0.03
20	1	0.54	82	1	0.02
21	2	0.49	88	1	0.02

Frequency	No. of FVS-motifs	Theoretical (5)	Frequency	No. of FVS- motifs	Theoretical (5)
22	2	0.44	90	1	0.02
23	3	0.40	103	2	0.01
24	1	0.36	107	1	0.01
25	3	0.33	129	1	0.01
26	1	0.30	153	1	0.01
27	1	0.28	203	1	0.003
28	2	0.26	226	1	0.002
30	1	0.22	273	1	0.002
31	1	0.21	278	1	0.001
33	1	0.18	384	1	0.0007
			422	1	0.0006
a = 1.2657, c = 402.8904, R^2 = 0.9922					

Tab. 5: The frequency and average length of the FSV-motifs in the short story Šlépěj (Footprint) by K. Čapek

Frequency	Average length	Lorentzian
1	3.36	2.29
2	3.57	3.71
3	5.66	5.63
4	2	2.47
5	2	2.12
7	2.5	1.98
10	1	1.95
11	2	1.94
a = 1.9279, b = 6.2626, c = 2.6563, d = 0.4132, R^2 = 0.82		

Tab. 6: The frequency and average length of the FSV-motifs in the Hungarian translation of Orwel's 1984

Fr	L	Power	Fr.	L	Power	Fr.	L	Power
1	5.43	5.70	22	3.50	3.11	64	1.00	2.52
2	4.70	4.97	23	3.33	3.08	66	3.00	2.51
3	4.61	4.59	24	2.00	3.06	68	3.00	2.49
4	4.00	4.34	25	3.00	3.03	69	3.00	2.49
5	4.06	4.16	26	4.00	3.01	71	3.00	2.47
6	3.92	4.01	27	3.00	2.99	73	3.00	2.46
7	4.12	3.89	28	3.50	2.97	74	2.00	2.45
8	3.38	3.79	30	4.00	2.93	75	3.00	2.45
9	4.29	3.71	31	3.00	2.91	82	2.00	2.40
10	3.25	3.63	33	2.00	2.87	88	2.00	2.37
11	3.29	3.56	34	3.00	2.86	90	2.00	2.36
12	3.00	3.50	35	2.67	2.84	103	1.50	2.30
13	3.50	3.45	36	3.00	2.82	107	2.00	2.28
14	4.00	3.40	39	2.00	2.78	129	2.00	2.20
15	4.00	3.35	40	3.00	2.77	153	2.00	2.13
17	4.00	3.27	47	3.00	2.68	203	2.00	2.01
18	3.50	3.24	48	3.00	2.67	226	2.00	1.97
19	3.50	3.20	54	3.00	2.61	273	2.00	1.90
20	3.00	3.17	57	3.00	2.58	278	2.00	1.89
21	3.00	3.14	59	2.00	2.56	384	1.00	1.78
						422	1.00	1.74

$a = 5.6962$, $b = -0.1957$, $R^2 = 0.7082$