On the dependency of word length on text length. Empirical results from Russian and Bulgarian parallel texts

Emmerich Kelih

1 Introduction

This paper tackles two basic problems of quantitative linguistics: firstly the "word length" and secondly the text length in terms of type and token numbers. It has to be shown that these two basic properties of a text are directly related. The interrelation between word length and text length can be captured by an appropriate mathematical model; hence a law-like status of the interrelation between word length and text length can be stated. Up to now, no special attention has been paid to this particular kind of self regulation (increase of the word length with increasing text length) of the text structure, which can be explained by a general control of the information flow on the lexical level. After a theoretical explanation of the interrelation of word and text length, some empirical results are presented on the basis of Russian and Bulgarian parallel texts *The Master and Margarita*).

2 Word length research and text length

Recent word length research is dedicated mainly to problems of theoretical modelling (cf. Wimmer et al. (1994) and Best (2001) with an extensive bibliography) and to the general dependency of word length on the analysed text types (cf. Antić et al. 2006; Kelih et al. 2005). Moreover, word length plays a crucial role in Menzerath's law (cf. Altmann 1980, Altmann and Schwibbe 1989) and is a central property of synergetic linguistics – cf. Köhler (1986), or Weber (2005) who gives a good overview.

Within word length studies, especially in works on modelling of word length frequencies, different aspects have been discussed: the problem of an appropriate word definition, the question of relevant units of measurement (phoneme, syllable, morpheme, etc.), the impact of text types on word length, the sampling problem, etc. – for details see Grotjahn and Altmann 1993. In other words, there are many boundary conditions which have to be taken into consideration in word length studies.

According to our knowledge, the question of a "minimal", "maximal" or "optimal" text length for word length studies has never been discussed in detail,

because it is generally accepted in quantitative linguistics that only "complete" texts should be analysed and not fragments or random samples of a text. As Orlov (1982) has shown, this is the case especially for Zipf's law, which holds true for "complete" texts more than for "mixed" texts (= corpora). The mixing of different texts or the analysis of text fragments increases the heterogeneity of the data, which can lead to further complications in theoretical modelling.

Special attention is paid in word length research to the text type "letter". It has been assumed that letters are in some way an "optimal" text sort, because in general they are written in one act of creation and are thus considered to be a good example of a spontaneous "natural" short text (Best and Zinenko 1998). In the context of text length, one general methodological problem has to be mentioned: in word length studies the χ^2 test is usually used as a measure of how good the fit is, i.e. the discrepancy between observed and expected values. However, the minimal requirement for using the χ^2 test is that the minimal frequency within one cell is greater than five. Thus, applying this methodological requirement in an orthodox way, only texts where all word length frequencies are greater than five can be analysed. A requirement that in fact is not fulfilled in all "typical" short private letters, and hence some pooling of frequency classes has to be performed.

In other words, even though much has been achieved in word length research in the past, many theoretical, methodological and empirical problems still remain. In this paper, only one particular question, which has so far not been examined in word length research, will be discussed: is there a dependency of word length on text length?

2.1 Text length, measured in the number of types and tokens

In general, text length can be measured in the number of graphemes, sounds, bigrams, syllables, morphemes, word forms, lemmas, sentences, etc. The choice of a particular linguistic unit is determined by the linguistic hypothesis analysed. A frequently used measurement unit of text length is the number of word-form types and word-form tokens. In quantitative linguistics, there are different approaches and interpretations of the number of word form types (V) and word form tokens (N) in a text.

Generally, the so-called type-token ratio (TTR) is calculated and interpreted as an indicator of vocabulary richness and stylistic diversification of a text. Nevertheless, one should keep in mind that the TTR directly depends on text length and thus the TTR cannot be understood as a reliable indicator of the stylistic richness of a text. Altmann and Altmann (2008: 107) offer a plausible alternative interpretation of the TTR: According to them, the TTR is a possibility for the measurement of the information flow in a text, i.e. the introduction of new word form types into a text equals the introduction of new information

and new semantic content – cf. Kelih (2010) for some empirical results of the TTR in a Slavic parallel corpus.

Recently, special attention has been paid to the modelling of the number of types, in dependency on the number of tokens. In the past, several models have been proposed (an extensive overview on models used and discussed in the past can be found in Altmann and Altmann (2008). To give an illustration here of this kind of Type-Token curve, the increase of types and tokens in a Russian novel (cf. Kelih 2010 for details) is captured in Figure 1. For the analysis, ten chapters of the novel *Kak zakaljalas' stal'* [How the steel was tempered] were cumulated successively and the number of word-form types and word-form tokens were counted. The definition of word-form types and tokens is based exclusively on orthographical criteria (cf. Kelih 2007).

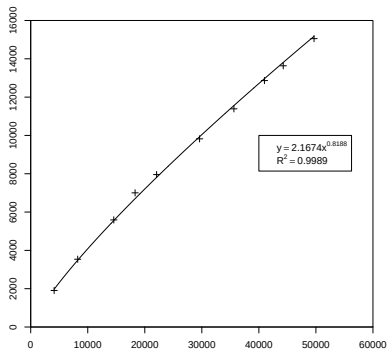


Figure 1: Increase of types and tokens: Russian novel *How the steel was tempered*

From a linguistic point of view, the increase of types and tokens can be interpreted in different ways (as already mentioned above, no information about the stylistic structure or the vocabulary richness is included), mainly in a language typological manner:[1]

1. The increase (or in other words, the steepness) of the curve gives information about the morphological richness of the language under examination. The increase in the number of types and tokens is determined by the morphological structure of the language and the richness of the inflexion system. Moreover, the degree of analytism and synthetism of the language has to be taken into consideration.

1. It has to be noted that all stated hypotheses about the increase of types hold true only for inflective languages (e.g. Slavic languages). For isolating or agglutinative languages it can be supposed that the behaviour of word-form types and tokens differs from that of inflective languages.

2. The increase of word-form types with the successive accumulation of chapters (= increasing text length) leads to the introduction of "new" word-form types (in the sense of new lexemes). Here the successive increase of types in fact represents the information structure and in some way the "semantic dynamics" of a text. In every text some theme or topic is described and there must therefore be some "semantic" development of the text with increasing text length. For instance, in a novel a particular content (or plot) is usually presented and described. The author is more or less forced to continually give new information in the text, because overly frequent repetitions of the same word-form types and word-form tokens would in some way lead to a semantic "stagnation" in the text. Concretely, this means that in the beginning of a text some "known" vocabulary is used and introduced, which is afterwards accompanied by the introduction of new and innovative lexical material. And furthermore, innovation in a text can only be given by the introduction of "infrequent" and "rare" words. Since "infrequent" words – these words are sometimes hapax legomena – are generally longer than frequent ones, it can be assumed that with increasing text length the word length of word-form types increases, too. Moreover, innovation and the provision of new information in a text can only be accomplished by the introduction of autosemantic words, whereas synsemantic words (which are very frequent in texts) are responsible for the grammatical and syntactical regulation of the text. It is well known that autosemantic words are generally longer than the frequently occurring synsemantic words in a text.
3. Furthermore, it has to be taken into consideration that with increasing text length Zipf's law comes into play. Zipf's law is responsible for a particular regulation of the lexical material in such a way that the front part of the rank frequency distribution is filled up with a few word form types that occur in a text with a very high frequency. Again, the front part is filled up with synsemantic words, which are – as a tendency – much shorter than autosemantic words.

In the case that our linguistic interpretation and assumptions about the special behaviour of number of types and tokens with increasing word length hold true, the following linguistic hypotheses can be established: with increasing text length (number of types), a successive increase of the word length, measured in terms of grapheme/syllable numbers, is observable. In other words, the longer the text in terms of the number of types, the longer the types (measured in the number of graphemes/syllables). This hypothesis will be tested, based on text material from Russian and Bulgarian, in the next chapter.

2.2 Empirical evidence from Russian

The above-mentioned hypothesis concerning the interrelation between "word length" (types length) and text length is tested by means of the Russian novel *The Master and Margarita*, written by Mikhail Bulgakov between 1928 and 1940. The novel consists of 33 chapters and, in addition to the Russian original, the Bulgarian translation will be analysed. Thus, some cross-linguistic comparisons between Russian (as a strongly synthetic language) and Bulgarian (usually considered a more analytic language) can be made. Furthermore, from a morphological point of view, Russian is of particular interest because it has a rich inflectional system, whereas Bulgarian has already largely abandoned the inflectional case system. Thus, the hypothesis can be tested in two typologically different languages.

The analysis procedure is as follows: in the 33 chapters the number of types (V) and tokens (N) is counted, after which the single chapters are cumulated successively, starting with chapter 1, followed by chapters $1+2$, chapters $1+2+3, \ldots, 1+2+3+\ldots+33$, and finishing with text 33, which is in fact the whole novel. The average token and type length is measured in terms of both the number of syllables and of graphemes. All raw data counted can be found in Tables 2 and 3 in the Appendix (p. 53f.).

First of all it has to be tested whether there is a systematic increase of types with increasing text length. As shown in Kelih (2010), the power model $V = a \cdot N^b$ is an adequate model for the type-token relation in the analysed Russian text. As can be seen in Figure 2, the model $V = 3.574 \cdot N^{0.584}$ ($R^2 = 0.9983$) does indeed seem to be adequate.

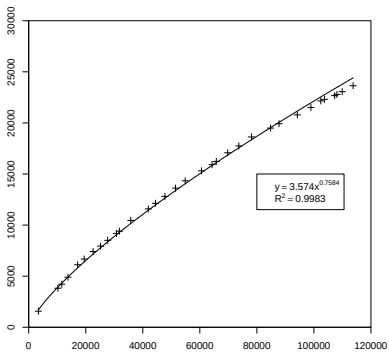


Figure 2: Relation between number of tokens and types (Russian prose text)

Based on the corroborated fact of a systematic interrelation between the number of types and tokens, in the next step the hypothesis concerning word

and text length can be analysed. As already mentioned above, a reasonable conjecture for the increase of the type length with increasing text length is the necessity of providing new information in a text. This can be easily demonstrated by one particular text property: by analysing the relative frequency of *hapax legomena* in the cumulated chapters (= increasing text length), it is possible to obtain successive and systematic growth. Table 1 represents the corresponding data.

Table 1: Frequencies (f) and relative frequencies (f_{rel}) of *hapax legomena* in *The Master and Margarita* (Russian)

f	f_{rel}	f	f_{rel}	f	f_{rel}
1170	0.049492386	7018	0.296869712	11739	0.496573604
2706	0.114467005	7653	0.323730964	12173	0.514932318
2977	0.125930626	8001	0.338451777	12366	0.523096447
3458	0.146277496	8394	0.355076142	12767	0.540059222
4349	0.183967851	8879	0.375592217	13195	0.558164129
4709	0.199196277	9277	0.392428088	13556	0.573434856
5133	0.217131980	9785	0.413917090	13618	0.576057530
5443	0.230245347	10113	0.427791878	13783	0.583037225
5789	0.244881557	10310	0.436125212	13841	0.585490694
6217	0.262986464	10819	0.457656514	13985	0.591582064
6341	0.268231810	11217	0.474492386	14308	0.605245347

A graphical representation of this increase is presented in Figure 3.

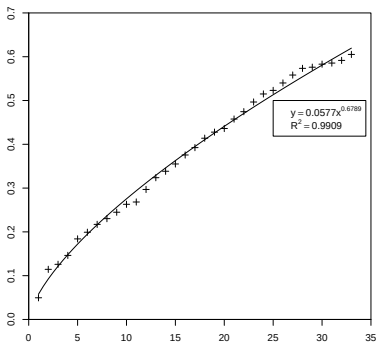
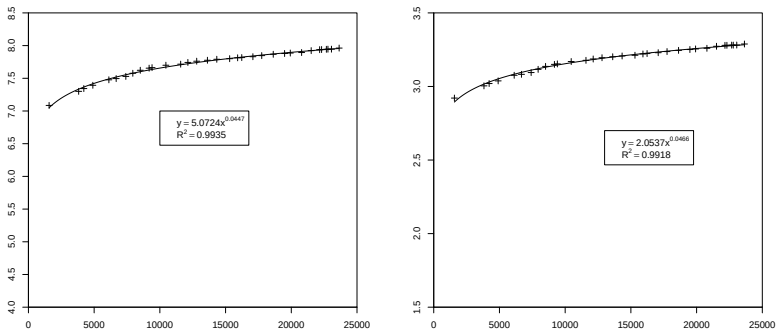


Figure 3: Relative frequency of *hapax legomena* vs. number of types in cumulated chapters

Hence, a further plausible explanation for the relation between type length and text length is to be found – *hapax legomena* are undoubtedly longer than

the frequently occurring word-form tokens.

Moving on to the empirical results of the test of the hypothesis, 'The longer the text in the number of types, the longer the types (in the number of graphemes/syllables)", as can be seen from Figures 4a and 4b, the assumed interrelation can be confirmed in both respects: the token length, measured in the number of graphemes (Figure 4a, and the token length, measured in the number of syllables, increase systematically with increasing text length. For the measurement of the word-form types in the number of graphemes, the model $WL(Gra) = 5.0724 \cdot V^{0.0447}$ ($R^2 = 0.9935$) is adequate, and for the word form types, measured in the number of syllables, the model $WL(Syl) = 2.0537 \cdot V^{0.0466}$ ($R^2 = 0.9918$). In both cases, $R^2 > 0.99$ can be obtained, and thus our hypothesis is confirmed almost perfectly.



(a) Word length in graphemes *(b)* Word length in syllables

Figure 4: Word length vs. text length (types)

Therefore, it can be stated, at least for the analysed Russian text, that the length of types ("word length") is directly related to the text length. Hence, the linguistic substantiation presented above seems to be quite adequate and reliable. Further empirical results from Bulgarian texts are presented in the next chapter.

2.3 Empirical evidence from Bulgarian

In addition to the Russian original text, the Bulgarian translation of *The Master and Margarita* was analysed. The relevance and problems of the linguistic analysis of Slavic parallel corpora are discussed in Kelih (2009a,b; 2010), where it was shown that the quantitative analysis of translations is a possibility of the minimisation of the heterogeneity of the analysed linguistic data.

The analysis of the Bulgarian translation is carried out in the same way as the analysis of the Russian text, so the procedure need not be explained in detail again. Let us start with the analysis of the relation between the number of types (V) and tokens (N). As expected, the Bulgarian translation can be captured by the power model $V = 4.5168 \cdot N^{0.7141}$ ($R^2 = 0.9974$). Figure 5 demonstrates this relation for the Bulgarian text.

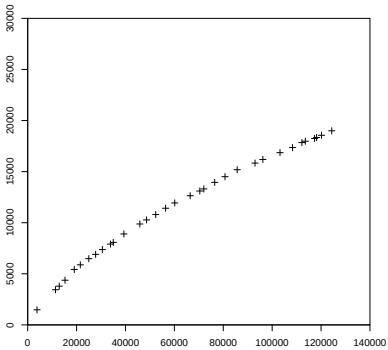
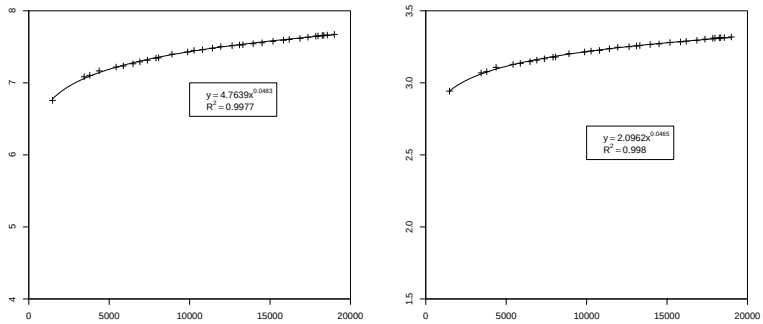


Figure 5: Increase of types and tokens: Bulgarian translation (*The Master and Margarita*)

The test of the hypothesis concerning increasing types length with increasing text length shows the following results: the relation between token length in the number of graphemes and text length can be satisfactorily modelled by $WoL(gra) = 4.7639 \cdot V^{0.0483}$, where $R^2 = 0.9977$ again indicates a good congruence between the empirical and the theoretical values. Quite a similarly good result is obtained for the tokens length, measured in the number of syllables: with $WoL(syl) = 2.0962 \cdot V^{0.0465}$ and $R^2 = 0.998$ again an almost perfect result is obtained. Both models are presented in Figures 6a and 6b.

Based on these findings for the Bulgarian text, further evidence for the validity of the systematic interrelation between token length and text length was found. It can be assumed that, based on the same analysis procedure and the same boundary conditions, similarly good results can also be obtained for other languages and texts.

Finally, two aspects have to be discussed: a language typological comparison of Bulgarian and Russian will be made, and the linguistic behaviour of the token length with increasing text length has to be tentatively analysed.



(a) Word length in graphemes *(b)* Word length in syllables

Figure 6: Word length vs. text length (types)

2.4 Comparison of word length: Russian vs. Bulgarian

The length of a linguistic form is without doubt a typological characteristic. For crosslinguistic comparisons, it has to be decided which linguistic material can be compared. Furthermore, it is well known that "word length" (type length, token length) is determined by the text type and the functional style, and thus even within one language a broad spectrum of word lengths can be obtained.

Therefore, in order to be methodologically correct, only similar text types of different languages should be compared. From this perspective, the use of parallel texts seems to be a good starting point for a systematic crosslinguistic comparison of the word length. To provide an insight into the behaviour of the word length in the Russian source text and the Bulgarian translation, the average type length in the cumulative chapters of both languages is presented in Figure 7. As can be seen, there is a parallel "course" of both average token lengths in the cumulated chapters, and no overlapping of the languages can be observed. Interestingly, in this case Russian has a slightly lower word length than Bulgarian, which is normally classified as an analytic language. It can be assumed that the postpositive article in Bulgarian leads to a "natural" lengthening of the linguistic forms. However, more systematic studies have to be conducted in order to obtain a more detailed picture of the word length in these two languages. Furthermore, it has to be noted that the frequency of the word forms in the analysis performed above is not taken into account. A statistical test for differences (power models are transformed to linear models by taking logarithms, after which the regression coefficients are tested; other tests are possible, too) between the two mean type lengths (Russian and Bulgarian) shows no significant results.

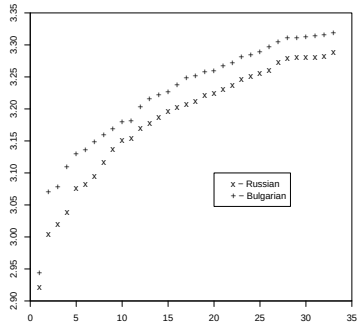


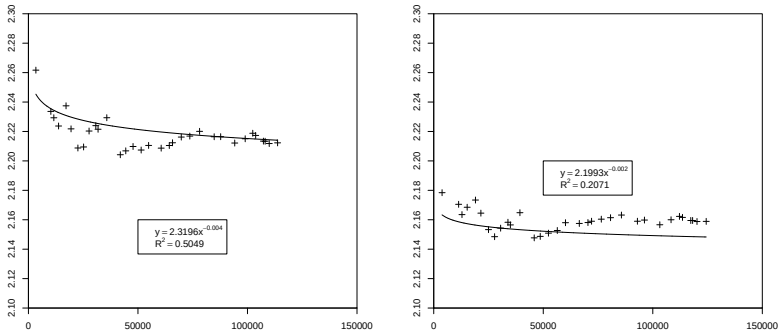
Figure 7: Word length (types) in cumulative chapters: Bulgarian vs. Russian (*The Master and Margarita*)

2.5 Token length vs. text length?

Finally, a theoretical aspect has to be briefly discussed: taking into account the systematic interrelation between type length and text length, in the next step the token length has to be tentatively analysed. From a linguistic viewpoint, it is clear that the token length must be shorter than the type length. This can be explained by the fact that tokens occur more frequently and in general are much shorter. Furthermore, as already discussed above, the front part of the rank frequency distribution of tokens is filled with synsemantic words, which are very frequent in texts. Thus, the token length can be interpreted as an indicator of the morphosyntactical text structure.

With regard to the empirical behaviour of token length with increasing text length, for the time being only some tentative and preliminary observations can be presented: in Figures 8a and 8b the relation between token length and text length (number of tokens) for Russian and Bulgarian is given.

As can be seen from the graphical representations, no entirely systematic relation is observable, but rather a decreasing trend of token length (measured by the number of syllables) with increasing text length. At the end of the curve, a kind of stabilisation of the token length can be seen. As already mentioned, this interrelation is a kind of tendency that is interrupted at certain points. In other words, the length of tokens with increasing text length is not as strongly controlled as the type length. Thus, it must be concluded here that the relation found between type length and text length represents a kind of paradigmatic self-regulation, whereas the token length is regulated by more complex mechanisms, which have to be explored in future research.



(a) Russian *(b)* Bulgarian

Figure 8: Token length (in syllables) vs. text length (tokens)

3 Summary

The results of the present paper can be summarised as follows. The length of types is in a systematic interrelation with text length: the longer the text, the longer the types. This hypothesis has been empirically tested in Russian and Bulgarian texts. Furthermore, it has been shown that this kind of regulation does not hold true at the token level, which must be analysed in more detail in the future. Moreover, the correlations that were found have to be tested in the future for many more languages with different morphological structures.

References

Altmann, Gabriel

1980 Prolegomena to Menzerath's law. In: Grotjahn, Rüdiger (ed.), *Glottometrika 2*. Bochum: Brockmeyer; 1–10.

Altmann, Vivien; Altmann, Gabriel

2008 *Anleitungen zu quantitativen Textanalysen. Methoden und Anwendungen.* Lüdenscheid: RAM.

Altmann, Gabriel; Schwibbe, Manfred

1989 *Das Menzerathsche Gesetz in informationsverarbeitenden Systemen.* Zürich, New York, Hildesheim: Olms.

Antić, Gordana; Stadlober, Ernst; Grzybek, Peter; Kelih, Emmerich

2006 Word Length and Frequency Distributions in Different Text Genres. In: Spiliopoulou, Myra; Kruse, Rudolf; Nürnberger, Andreas; Borgelt, Christian; Gaul, Wolfgang (eds.), *From Data and Information Analysis to Knowledge Engineering*. Heidelberg, Berlin: Springer; 310–317.

Best, Karl-Heinz
2001 Kommentierte Bibliographie zum Göttinger Projekt. In: Best, Karl-Heinz (ed.), *Häufigkeitsverteilungen in Texten.* Göttingen: Pest & Gutschmidt; 248–310.

Best, Karl-Heinz; Zinenko, Svetlana
1998 Wortlängenverteilungen in Briefen A.T. Twardowskis. In: *Göttinger Beiträge zur Sprachwissenschaft*, 1; 7–19.

Kelih, Emmerich
2010 The type-token relationship in Slavic parallel texts. In: *Glottometrics*, 20; 1–11.
2009a Preliminary analysis of a Slavic parallel corpus. In: Levická, Jana; Garabík, Radovan (eds.), *NLP, Corpus Linguistics, Corpus Based Grammar Research. Fifth International Conference Smolenice, Slovakia, 25-27 November 2009. Proceedings.* Bratislava: Tribun; 175–183.
2009b Slawisches Parallel-Textkorpus: Projektvorstellung von «Kak zakaljalas' stal' (KZS)». In: Kelih, Emmerich; Levickij, Viktor V.; Altmann, Gabriel (eds.), *Methods of Text Analysis. Metody analizu tekstu.* Černivci: ČNU; 106–124.
2007 Zur Frage der Wortdefinitionen in Wortlängenuntersuchungen. In: Kaliuščenko, Volodymir; Köhler, Reinhard; Levickij, Viktor V. (eds.), *Problems of Typological and Quantitative Lexicology*. Chernivtsi: Ruta; 91–105.

Kelih, Emmerich; Antić, Gordana; Grzybek, Peter; Stadlober, Ernst
2005 Classification of Author and/or Genre? The Impact of Word Length. In: Weihs, Claus; Gaul, Wolfgang (eds.), *Classification. The Ubiquitous Challenge.* Heidelberg, New York: Springer; 498–505.

Köhler, Reinhard
1986 *Zur linguistischen Synergetik: Struktur und Dynamik der Lexik*. Bochum: Brockmeyer [= Quantitative Linguistics; 31]

Orlov, Jurij K.
1982 Ein Modell der Häufigkeitsstruktur des Vokabulars. In: Guiter, Henri; Arapov, Michail V. (eds.), *Studies on Zipf's law*. Bochum: Brockmeyer; 154–233.

Weber, Sabine
2005 Zusammenhänge. In: Köhler, Reinhard; Altmann, Gabriel; Piotrowski, Rajmund G. (eds.), *Quantitative Linguistics · An International Handbook. Quantitative Linguistik · Ein Internationales Handbuch.* Berlin, New York: de Gruyter; 214–226. [= Handbücher zur Sprach- und Kommunikationswissenschaft; 27]

Wimmer, Gejza; Köhler, Reinhard; Grotjahn, Rüdiger; Altmann, Gabriel
1994 Towards a Theory of Word Length Distribution. In: *Journal of Quantitative Linguistics*, 1; 98–106.

Table 2: Data used: Russian (*The Master and Margarita*)

	Tokens					Types				
cum.	Frequencies			WoL		Frequencies			WoL	
chaps.	*Wf*	*Syl*	*Gra*	(gra)	(syl)	*Wf*	*Syl*	*Gra*	(gra)	(syl)
1	3405	7701	18332	5.3838	2.2617	1573	4595	11145	7.0852	2.2120
2	10234	2285	54381	5.3138	2.2336	3822	11482	27902	7.3004	3.0420
3	11600	25860	61547	5.3058	2.2293	4215	12729	30951	7.3431	3.0199
4	13763	30605	72854	5.2935	2.2237	4901	14889	36234	7.3932	3.0380
5	17171	38419	91504	5.3290	2.2374	6121	18829	45766	7.4769	3.0761
6	19452	43218	102925	5.2912	2.2218	6675	20576	50033	7.4956	3.0825
7	22584	49883	118789	5.2599	2.2088	7411	22934	55797	7.5289	3.0946
8	25187	55650	132448	5.2586	2.2095	7943	24754	60182	7.5767	3.1165
9	27685	61469	146260	5.2830	2.2203	8509	26688	64861	7.6226	3.1364
10	30759	68407	162837	5.2940	2.2240	9183	28928	70295	7.6549	3.1502
11	31790	70622	168063	5.2867	2.2215	9405	29658	72071	7.6631	3.1534
12	35780	79766	189830	5.3055	2.2293	10459	33149	80532	7.6998	3.1694
13	41946	92457	219666	5.2369	2.2042	11583	36806	89373	7.7159	3.1776
14	44438	98067	233192	5.2476	2.2068	12132	38661	93933	7.7426	3.1867
15	47762	105544	250894	5.2530	2.2098	12804	40913	99375	7.7612	3.1953
16	51460	113592	269835	5.2436	2.2074	13619	43606	105898	7.7758	3.2019
17	54870	121291	288268	5.2537	2.2105	14332	45966	111656	7.7907	3.2072
18	60596	133837	318516	5.2564	2.2087	15311	49180	119427	7.8001	3.2121
19	64303	142140	337822	5.2536	2.2105	15918	51270	124371	7.8132	3.2209
20	65779	145524	345707	5.2556	2.2123	16232	52328	126882	7.8168	3.2238
21	69777	154640	367279	5.2636	2.2162	17079	55157	133732	7.8302	3.2295
22	73678	163334	387808	5.2636	2.2169	17751	57463	139289	7.8468	3.2372
23	78133	173462	411760	5.2700	2.2201	18627	60451	146544	7.8673	3.2453
24	84799	187950	445961	5.2590	2.2164	19491	63367	153578	7.8794	3.2511
25	87765	194527	461597	5.2595	2.2165	19923	64851	157115	7.8861	3.2551
26	94181	208342	494155	5.2469	2.2121	20786	67752	164085	7.8940	3.2595
27	98956	219192	519989	5.2547	2.215	21514	70417	170455	7.9230	3.2731
28	102414	227235	539226	5.2652	2.2188	22157	72659	175877	7.9378	3.2793
29	103733	230008	545849	5.2621	2.2173	22306	73157	177102	7.9397	3.2797
30	107275	237448	563362	5.2516	2.2135	22669	74377	180071	7.9435	3.2810
31	108058	239171	567575	5.2525	2.2134	22789	74771	181068	7.9454	3.2810
32	109926	243135	576965	5.2487	2.2118	23056	75659	183195	7.9457	3.2815
33	113748	251641	596976	5.2482	2.2123	23640	77741	188223	7.9621	3.2885

Table 3: Data used: Bulgarian (*The Master and Margarita*)

	Tokens					Types				
cum.	Frequencies			WoL		Frequencies			WoL	
chaps.	*Wf*	*Syl*	*Gra*	(gra)	(syl)	*Wf*	*Syl*	*Gra*	(gra)	(syl)
1	3805	8289	18533	4.8707	2.1784	1476	4343	9972	6.7561	2.9424
2	11366	24670	55215	4.8579	2.1705	3442	10563	24382	7.0837	3.0689
3	12894	27896	62421	4.8411	2.1635	3792	11670	26937	7.1036	3.0775
4	15262	33096	73999	4.8486	2.1685	4372	13586	31328	7.1656	3.1075
5	19050	41404	92761	4.8693	2.1734	5429	16982	39190	7.2186	3.1280
6	21523	46586	104309	4.8464	2.1645	5877	18425	42516	7.2343	3.1351
7	24997	53825	120457	4.8189	2.1533	6482	20398	47088	7.2644	3.1469
8	27788	59702	133701	4.8115	2.1485	6905	21803	50347	7.2914	3.1576
9	30539	65790	147273	4.8225	2.1543	7373	23350	53927	7.3141	3.1670
10	33879	73120	163693	4.8317	2.1583	7910	25138	58086	7.3434	3.1780
11	34994	75466	168939	4.8277	2.1565	8072	25665	59304	7.3469	3.1795
12	39344	85173	190893	4.8519	2.1648	8902	28506	65858	7.3981	3.2022
13	45870	98514	220874	4.8152	2.1477	9877	31749	73371	7.4285	3.2144
14	48623	104478	234411	4.8210	2.1487	10276	33093	76552	7.4496	3.2204
15	52371	112644	252660	4.8244	2.1509	10799	34825	80554	7.4594	3.2248
16	56406	121428	272174	4.8253	2.1527	11418	36958	85409	7.4802	3.2368
17	60143	129789	290914	4.8370	2.1580	11935	38751	89560	7.5040	3.2468
18	66454	143375	321530	4.8384	2.1575	12641	41095	94997	7.5150	3.2509
19	70402	151937	340629	4.8383	2.1581	13102	42668	98603	7.5258	3.2566
20	72047	155547	348711	4.8400	2.1590	13309	43368	100192	7.5281	3.2585
21	76495	165260	370610	4.8449	2.1604	13950	45563	105217	7.5424	3.2662
22	80714	174455	391201	4.8468	2.1614	14503	47428	109577	7.5555	3.2702
23	85683	185346	415659	4.8511	2.1632	15184	49798	115036	7.5761	3.2796
24	92981	200743	450098	4.8408	2.1590	15836	52004	120198	7.5902	3.2839
25	96193	207754	465819	4.8425	2.1598	16191	53244	123082	7.6019	3.2885
26	103208	222577	498882	4.8338	2.1566	16857	55548	128382	7.6159	3.2952
27	108338	234015	524561	4.8419	2.1600	17356	57335	132492	7.6338	3.3035
28	112170	242546	543863	4.8486	2.1623	17844	59051	136471	7.6480	3.3093
29	113560	245467	550465	4.8473	2.1616	17979	59511	137556	7.6509	3.3100
30	117345	253410	568125	4.8415	2.1595	18248	60439	139712	7.6563	3.3121
31	118181	255206	572207	4.8418	2.1595	18340	60764	140475	7.6595	3.3132
32	120187	259455	581748	4.8404	2.1588	18559	61514	142222	7.6632	3.3145
33	124380	268523	602073	4.8406	2.1589	18996	63034	145745	7.6724	3.3183