Glottometrics 22, 2011, 44-56

Word length in Slovak poetry

Radek Čech[1]
Ioan-Iovitz Popescu
Gabriel Altmann

Abstract. In the present article the place of word-length in Slovak poetry in the framework of the general theory is sought. Different models are presented. The possibility of applying Menzerath's law in this domain is scrutinized.

Keywords: word length, verse length, Slovak, binomial distribution, Poisson distribution, Ord's scheme, Menzerath's law

1. Introduction

Word-length is a problem which has occupied generations of both linguists and mathematicians. The literature is enormous (cf. Best 2001, 2006; Grzybek 2006) but the number of problems associated with word length seems to increase, e.g., the distribution of word length in different languages, word properties associated with word length, etc. The latter problem belongs to the domain of synergetic linguistics where word length plays a central role already since G.K. Zipf (1935/1968) (cf. Köhler 2005).

The first problem has two aspects: (1) Word length is a general phenomenon obeying some laws (except for monosyllabic languages where it is not a variable), and (2) even within one and the same language it is characteristic for text sorts, style, author, etc. (cf. Best 2001, 2006; Grzybek 2006) In both cases boundary conditions play an enormous role, but if a model functions in about 90% of cases, one usually does not care any more for subsuming the rest of the cases under the given law or searching for the causes of deviation. Usually, 5% of bad fits is no reason for a rejection. However, some authors do it, introduce modifications of the model or derive the distribution of word length in a quite different way (e.g. Uhlířová 1995; Wimmer, Witkovský, Altmann 1999). Though the history abounds in models (cf. Grzybek 2006a), a theory be cannot easily constructed. Language must fulfil a number of requirements. These are not the general requirements listed by Köhler (2005) but specific ones, so to speak local ones enabling the author/writer to write a text for exactly the given occasion. He must express something, express himself and give the text an adequate form. Seen from this perspective, every text is a unique creation, hence mixing texts, e.g. taking a corpus as a whole and scrutinise in it the word-length distribution, may lead to distortions.

A well known problem connected with testing a model is the character of the sample. Sometimes the number of different length classes is too small for testing a model, e.g. the words are not longer than four syllables and the model has three para-

[1] Radek Čech, Department of Czech Language, University of Ostrava, Reální 5, Ostrava, 701 03, Czech Republic, e-mail: radek.cech@osu.cz,

meters. The chi-square test for goodness-of-fit cannot be applied because there are no degrees of freedom. Some other tests for fitting are obsolete and should not be used. The chi-square itself is not quite safe because the result depends on the sample size (cf. Rietveld et al. 2004). Thus the creation of a theory has both theoretical and empirical hindrances.

In our analysis we shall try to characterize Bachletová's poetry as a solid block of length distributions placed in a restricted area (Chapter 2). The distributions are always binomial being a further characteristic of the author (Chapter 3). We further try to find a distribution of verse length in terms of word numbers (Chapter 4) and lastly the relation of mean word length to the verse length, a view that has been neglected so far.

2. Methodology

In the present article we restrict ourselves to one author, Eva Bachletová, and her poetic texts written in Slovak. The poems by Bachletová are rhymeless, not following a metric prescription, and the individual verses are rather short, many times consisting only of one word.

We count for each poem separately the number of words having length x = 1,2,3,… measured in terms of syllable numbers. The length zero which is frequent in Slavic languages has been omitted because the pertinent words are consonantal prepositions (e.g. Slovak *k, s, v, z*) which can be considered proclitics of the next word. Their separate counting leads to a modification of every model (cf. Uhlířová 1995; Wilson 2006). Hence, the numbers in the second column of Table 1 are to be read as follows: in the poem *Aby spriesvitnela* there are 14 words of length 1; 26 words of length 2; 15 words of length 3; 5 words of length 4, and 4 words of length 5. All operations are performed upon these empirical data, e.g. the average length of verses is [1(14) + 2(26) + 3(15) + 4(5) + 5(4)]/64 = 151/64 = 2,3594 words.

An empirical distribution can be characterized by many means, e.g. moments and their functions (asymmetry and excess), entropy, repeat rate, etc. Here we shall use two functions of moments proposed by J.K. Ord (1972) used frequently in text analysis, namely

$$(1) \qquad I = \frac{m_2}{m_1'}$$

and

$$(2) \qquad S = \frac{m_3}{m_2},$$

where $m_1' = \bar{x}$, the mean of the distribution, and m_2 and m_3 are the second and third central moments defined as

$$m_r = \frac{1}{N}\sum_{x=1}^{k}(x - m_1')^r f(x) \quad (3)$$

where N is the sum of the frequencies $f(x)$ and k is the greatest length. The second central moment ($r = 2$) is the variance (i.e. a measure of dispersion) and the third moment ($r = 3$) is the indicator of skewness of the distribution. Hence, I and S are functions of some properties of the distribution.

The results of counting are presented in Table 1. The frequencies in the column "Distribution" represent those of lengths $x = 1,2,3,\ldots$

Table 1
Word-length distributions and Ord's criterion <*I,S*> in poems by E. Bachletová

Poem	Distribution	I	S
Aby spriesvitela	14,26,15,5,4	0,5082	0,8249
Bez rozlúčky	15,13,5	0,3030	0,3739
Čakáme šťastie	8,19,9,6,2	0,4747	0,6625
Čakanie na boží jas	30,24,18,2	0,3938	0,3986
Čas pre nádych vône	23,36,16,4,0,1	0,4284	1,0712
Dielo Stvoriteľa	41,53,23,13,1	0,4545	0,6582
Dnešný luxus	14,9,8,3,1	0,5855	0,7951
Do večnosti beží čas	15,22,11,1	0,3116	0,2563
Hľadanie odpovedí	22,20,18,4	0,4223	0,3145
Iba neha	51,53,16,5,0,2	0,4925	1,5436
Iba život	10,18,10,5	0,3924	0,3497
Idem za Tebou	26,28,10,5	0,4203	0,6943
Ihly na nebi	26,18,7,1	0,3614	0,6887
Keď dohorí deň	22,16,11,2	0,4215	0,5399
Kým ich máme	16,20,5,1,1	0,4145	1,1458
Len áno	7,17,5,1	0,2867	0,3750
Malé modlitby	13,25,8,2	0,3050	0,4421
Malý ošiaľ	32,22,12,1	0,3721	0,5581
Mladé oči	7,8,3,1	0,3830	0,6075
Moje určenie	55,54,25,6,2	0,4448	0,8493
Neopusť ma	6,17,7,1,0,1	0,4432	1,6529
Náš chrám	28,26,19,6,1,1	0,5509	0,9998
Naše dejiny	6,7,6,5,1	0,5435	0,2941
Naše mamy	22,19,11,4	0,4481	0,5909
Naše svetlo	16,23,11,7	0,4356	0,4740
Neha domova	10,11,3,1	0,3556	0,6750
Nepoznateľné	39,33,12,4,1,1	0,5381	1,4643
Podobnosť bytia	23,32,22,5,1,1	0,4726	0,9170

Prvotný sen	23,30,13,9,2	0,5206	0,7757
Rozdelená bytosť	26,31,16,3	0,3634	0,4184
Rozťatá prítomnosť	30,34,9,3	0,3507	0,6667
Som iná	20,24,6,4,1,1	0,5928	1,5737
Spájania	18,14,8,2	0,4249	0,6133
Stály smútok pre šesť písmen	54,64,19,4	0,3287	0,5505
Tak málo úsmevu	19,25,11,5,1	0,4614	0,7605
Tiché verše	8,10,11	0,3064	-0,1515
To všetko je dar	16,18,12,1	0,3469	0,2531
Večerná ruža	13,18,9,2,1,1	0,5672	1,4309
Večerné ticho	25,27,9,5	0,4242	0,7226
Vo večnosti slobodná	38,71,44,5,0,1	0,3417	0,5724
Vrátili sa	17,18,9,4	0,4375	0,5714
Vyznania	17,25,9,3	0,3578	0,5331
Z neba do neba	13,27,18,5,0,1	0,4174	0,8585
Zasľúbenie jasu	12,23,10,3	0,3367	0,4026
Zbytočné srdce	12,15,5,4	0,4517	0,6749

If we plot <I,S> in a Cartesian coordinate system, we obtain the results presented in Figure 1. One can see that the points are placed on a straight line with relative great dispersion. The trend can be expressed by $S = -0{,}5842 + 3{,}0397I$ yielding an $R^2 = 0.41$ which is small, but in poetry of this kind – without any binding – it is sufficient. As a matter of fact, continuing to evaluate more poems of the author one would obtain an ellipse, but preliminarily we only want to show the unity of the author. The ellipse can be constructed using our results: the slope of the longer axis is given by the regression coefficient of the straight line and the shorter axis is given as $1 - R^2$ placed orthogonally to the mean of the long axis. The straight line $S = 2I - 1$ represents the upper boundary of the beta-binomial (negative hypergeometric) distribution and serves here for orientation.

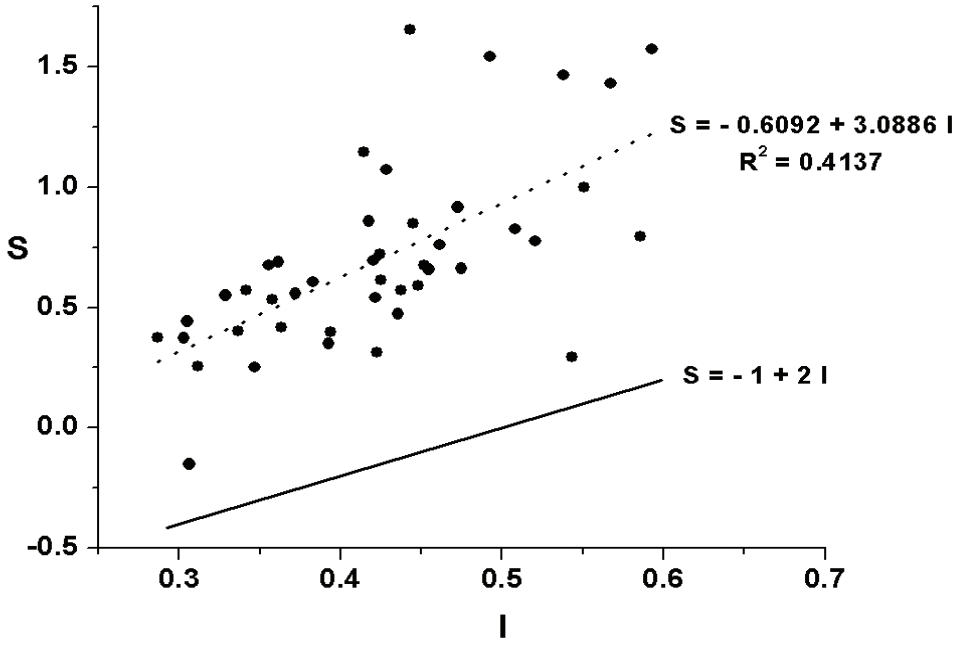

Figure 1. The <I,S> domain of word-length distributions with Bachletová

The model of the individual distributions is not easy to set up. First, the support of the data is very short. Though all data lie in the domain of Ord's hypergeometric distribution ($1 > S > 2I - 1$, $I < 1$) one needs at least five well represented length classes in the empirical distribution in order to have at least 1 degree of freedom. Evidently we must test models with smaller number of parameters. As candidates there are the Poisson and the binomial distributions which are limiting cases of the hypergeometric distribution. (Poisson when $N \to \infty$, $M \to \infty$, $n \to \infty$, $nM/N \to a$; binomial when $N \to \infty$, $M \to \infty$, $M/N \to p$). As a matter of fact, the great majority of data abide by these models. The points above $S > 1$ belong to the domain of the beta-Pascal distribution which has, unfortunately, three parameters and is not useful here. In Table 2, these distributions are fitted to the individual poems and tested using the chi-square test for goodness-of-fit. Of course, both distributions are displaced 1 step to the right because the support of the data is always x = 1,2,…,*n*. The formulas are

Binomial distribution: $P_x = \binom{n}{x-1} p^{x-1} q^{n-x}, \; x = 1,2,...n+1$

Poisson distribution: $P_x = \frac{a^{x-1} e^{-a}}{(x-1)!}, \quad x = 1,2,...$

The right truncated Poisson distribution would be more realistic because length data can never be infinite, but we did not obtain good results using it. Besides, it has one parameter more. This caused the impossibility to apply several times the binomial distribution, too.

Table 2
Fitting the Poisson and the binomial distributions to word-length data
in poems by E. Bachletová

Poem	**Distribution**	**Parameters**	**X^2**	**DF**	**P**
	Poisson	a = 1,3871	1,67	3	0,64
Aby spriesvitela	Binomial	n = 1374, p = 0,0010	1,67	2	0,43
Bez rozlúčky	Poisson	a = 0.7408	0,26	1	0,61
	Binomial	n = 7, p = 0,3182	2,60	3	0,46
Čakáme šťastie	Poisson	a = 2,2538	5,14	4	0,27
	Poisson	a = 0,9441	4,48	2	0,11
Čakanie na boží jas	Binomial	n = 6; p = 0,1529	3,70	1	0,05
	Binomial	n = 5; p = 0,2091	0,56	1	0,45
Čas pre nádych vône	Poison	a = 1,0588	3,09	3	0,38
	Poisson	a = 1,1161	3,90	3	0,27
Dielo Stvoriteľa	Binomial	n = 9; p = 0,1230	3,33	2	0,19
	Poisson	a = 1,0914	1,81	2	0,40
Dnešný luxus	Binomial	n = 1088; p = 0,0010	1,82	1	0,18
	Binomial	n = 3; p = 0,3216	0,30	1	0,59
Do večnosti beží čas	Poisson	a = 1,0206	4,00	2	0,14
	Poisson	a = 1,1189	3,43	2	0,18
Hľadanie odpovedí	Binomial	n = 6; p = 0,1799	2,82	1	0,09
	Poisson	a = 0,8616	2,17	3	0,54
Iba neha	Binomial	n = 864, p = 0,0010	2,16	2	0,34
	Poisson	a = 1,2761	0,95	2	0,62
Iba život	Binomial	n = 4; p = 0,3130	0,58	1	0,44
	Poisson	a = 0,9361	0,63	2	0,73
Idem za Tebou	Binomial	n = 925; p = 0,0010	0,62	1	0,43
	Poisson	a = 0,6852	0,40	2	0,82
Ihly na nebi	Binomial	n = 7; p = 0,0969	0,24	1	0,63
	Poisson	a = 0,8926	1,71	2	0,42
Keď dohorí deň	Binomial	n = 10; p = 0,0083	1,67	1	0,20
	Poisson	a = 0,8562	1,99	2	0,37
Kým ich máme	Binomial	n = 6; p = 0,1421	1,22	1	0,27
	Binomial	n = 3; p = 0,3337	1,84	1	0,18
Len áno	Poisson	a = 1,0219	5,56	2	0,06
	Binomial	n = 3; p = 0,3281	1,41	1	0,23
Malé modlitby	Poisson	a = 1,0018	5,25	2	0,07

	Poisson	a = 0,7653	2,27	2	0,32
Malý ošiaľ	Binomial	n = 5; p = 0,1493	2,01	1	0,16
	Poisson	a = 0,9120	0,26	2	0,88
Mladé oči	Binomial	n = 904; p = 0,0010	0,26	1	0,61
	Poisson	a = 0,9174	0,41	3	0,94
Moje určenie	Binomial	n = 13, p = ß,0707	0,28	2	0,87
	Binomial	n = 5; p = 0,2310	2,33	1	0,13
Neopusť ma	Poisson	a = 1,1701	4,80	2	0,09
	Poisson	a = 1,1173	0,91	3	0,82
Náš chrám	Binomial	n = 1123; p = 0,0010	0,91	2	0,63
	Poisson	a = 1,5794	1,53	3	0,68
Naše dejiny	Binomial	n = 1548; p = 0,0010	1,52	2	0,47
	Poisson	a = 0,9660	0,26	2	0,88
Naše mamy	Binomial	n = 958; p = 0,0010	0,26	1	0,61
	Poisson	a = 1,2015	0,51	2	0,77
Naše svetlo	Binomial	n = 1181; p = 0,0010	0,51	1	0,48
	Poisson	a = 0,8167	0,71	2	0,70
Neha domova	Binomial	n = 808; p = 0,0010	0,70	1	0,40
	Poisson	a = 0,8493	0,50	2	0,78
Nepoznateľné	Binomial	n = 859; p = 0,0010	0,50	1	0,48
	Binomial	n = 7; p = 0,1701	0,81	2	0,67
Podobnosť bytia	Poisson	a = 1,2003	2,05	3	0,56
	Poisson	a = 1,2049	1,93	3	0,59
Prvotný sen	Binomial	n = 1194; p = 0,0010	1,93	2	0,38
	Binomial	n= 4; p = 0,2375	0,14	1	0,70
Rozdelená bytosť	Poisson	a = 0,9810	2,31	2	0,31
	Poisson	a = 0,8171	2,55	2	0,28
Rozťatá prítomnosť	Binomial	n = 4, p = 0,2040	1,31	1	0,25
	Poisson	a = 0,8812	2,94	2	0,23
Som iná	Binomial	n = 1000; p = 0,0010	2,94	1	0,09
	Poisson	a = 0,8757	0,47	2	0,79
Spájania	Binomial	n = 868; p = 0,0010	0,47	1	0,47
Stály smútok pre	Binomial	n = 3; p = 0,2716	1,26	1	0,26
šesť písmen	Poisson	a = 0,8273	5,93	2	0,05
	Poisson	a = 1,0914	0,79	3	0,85
Tak málo úsmevu	Binomial	n = 1089, p = 0,0010	0,79	2	0,67
Tiché verše	Poisson	a = 1,3057	0,01	1	0,92
	Binomial	n = 3; p = 0,3221	1,17	1	0,28
To všetko je dar	Poisson	a = 1,0330	3,43	2	0,18
	Poisson	a = 1,0576	0,57	2	0,75
Večerná ruža	Binomial	n = 9; p = 0,1234	0,11	1	0,74
	Poisson	a = 0,9329	0,87	2	0,65
Večerné ticho	Binomial	n = 922; p = 0,0010	0,87	1	0,35
Vo večnosti slobodná	Binomial	n = 5; p = 0,2287	6,12	2	0,05

	Poisson	a = 1,0244	0,01	2	0,99
Vrátili sa	Binomial	n = 1015; p = 0,0010	0,01	1	0,92
	Binomial	n = 4; p = 0,2432	0,69	1	0,41
Vyznania	Poisson	a = 0,9809	2,22	2	0,33
	Binomial	n = 5; p = 0,2555	0,58	2	0,78
Z neba do neba	Poisson	a = 1,3094	4,10	3	0,25
	Binomial	n = 3; p = 0,3568	0,69	1	0,41
Zasľúbenie jasu	Poisson	a = 1,0857	2,82	2	0,24
	Poisson	a = 1,0746	0,97	2	0,61
Zbytočné srdce	Binomial	n = 1051; p = 0,0010	0,98	1	0,32

The results of fitting are very persuading. There is no exception; all fittings are significant. In some cases only the Poisson distribution was applicable, because the number of classes was too small for the binomial (*Tiché verše; Bez rozlúčky*); in one case only the binomial was applicable (*Vo večnosti slobodná*). In many cases one can see that the binomial distribution converges towards the Poisson: this is evident in cases where the parameter n is very great and p is very small (usually 0,0010 because of computing restriction). The product np is almost identical with the parameter a of the Poisson distribution. In Table 2 we wrote for every poem first the distribution whose P was greater.

Thus the only model expressing the word-length behaviour of Bachletová's poetry is the binomial distribution with its limiting case, the Poisson distribution ($n \to \infty$, $p \to 0$, $np \to a$). The result shows that the author has a certain "casting-mould" represented by a restricted <*I,S*> domain. It should be mentioned that for other texts, other variables and other languages, the <*I,S*> criterion yields very different results (c.f. Popescu et al. 2009) which may turn out to be characteristic of the author, style or language, etc.

3. Verse length

In some poetry the length of the verse measured in terms of word numbers is a constant. In such cases the bounding is too strong and verse length is not a variable. The same may hold for the number of feet (e.g. hexameter), number of syllables (e.g. thirteen), etc. However, in the poems by Bachletová which are free of any binding, verse length is a variable, most probably applied according to an internally pre-formed pattern and uttered spontaneously. Though the poems are short, the <*I,S*> characterization is always possible and if there are also longer verses, a probability distribution may be found.

Not all of the poems could be analyzed. Some of them were very short and the representation of individual frequency classes was far from being reliable. We selected poems having at least 15 verses and at least 4 frequency classes, and obtained the results presented in Table 3. The distribution is slightly more complex than the Poisson. Starting from Wimmer-Altmann's general theory (2005), the Poisson distribution follows from the simple difference equation

$$(4) \qquad P_x = \left(1 + a_0 + \frac{a_1}{x}\right) P_{x-1}$$

where a_0 = -1 and $a_1 = a$, yielding the formula presented above. For verse lengths in non-metric poetry this approach is not sufficient and a further modifying parameter must be added. We conjecture

.

$$(5) \qquad P_x = \left(1 + a_0 + \frac{a_1}{x + b - 1}\right) P_{x-1}$$

whose solution (setting again a_0 = -1, $a_1 = a$) yields

$$(6) \qquad P_x = \frac{a^x}{b^{(x)}\,{}_1F_1(1;b;a)}, \quad x = 0,1,2,\ldots,$$

where $b^{(x)} = b(b+1)\ldots(b+x-1)$ is the ascending factorial function, and ${}_1F_1(1;a;b)$ is the confluent hypergeometric function yielding the normalizing sum. Of course, since the distributions do not have $x = 0$, the expression (6) must be displaced one step to the right, i.e.

$$(7) \qquad P_x = \frac{a^{x-1}}{b^{(x-1)}\,{}_1F_1(1;b;a)}, \; x = 1,2,3,...$$

an operation made by the software (Fitter) automatically. This is the one-displaced Hyperpoisson distribution. In one case, namely with the poem *Čas pre nádych vône* which does not have verses consisting of only one word, the distribution is displaced two step to the right.

In some cases the Poisson distribution which is a special case of Hyperpoisson (when $b = 1$) would be sufficient and in one case we were forced to use the limiting case of the Hyperpoisson, namely the geometric distribution, following for $a \to \infty$, $b \to \infty$, $a/b \to q$, where q is the parameter of the geometric distribution $P_x = pq^x$, $x = 0,1,2\ldots$

Table 3
Verse lengths in terms of word numbers

Poem	**Frequencies**	**Parameter** **a**	**b**	**X^2**	**DF**	**P**	**I**	**S**
Aby spriesvitela	4,15,4,3,1	0,6658;	0,1776	2,39	1	0,12	0,41	0,92
Bez rozlúčky	4,6,5,1	0,8294;	0,4813	0,73	1	0,39	0,36	0,15
Čakanie na boží jas	7,6,10,4,1,0,1	1,9925;	1,5222	2,39	2	0,30	0,71	1,31
Čas pre nádych vône	0,1,3,4,2,5,2	2,6402;	0,8863	1,85	3	0,60	0,58	-0,21

Hľadanie odpovedí	5,4,6,9	21,9759;27,4687	1,64	1	0,20	0,48	-0,46
Iba neha	13,15,13,9,3,1	1,9655; 1,4877	0,78	3	0,85	0,63	0,65
Ihly na nebi	2,12,3,1,3	0,7178; 0,1196	2,80	1	0,09	0,54	1,24
Malý ošiaľ	2,14,6,5	0,7569; 0,1081	1,22	1	0,27	0,31	0,38
Moje určenie	10,15,12,9,2,4	2,4553; 1,6677	1,93	3	0,59	0,73	0,94
Nepoznateľné	26,16,2,6,1	3,9217; 6,6769	6,96	2	0,03	0,64	1,39
Podobnosť bytia	4,6,10,7,1,1	1,8377; 0,8675	2,65	3	0,45	0,49	0,30
Rozdelená bytosť	1,9,8,5,2,1	1,2280; 0,1534	0,17	2	0,92	0,44	0,77
Som iná	1,7,10,2,1	0,9604; 0,1372	1,96	2	0,37	0,27	0,42
Stály smútok pre šesť písmen	4,13,13,15,2,1	1,6148; 0,4969	5,41	3	0,14	0,42	0,15
Tak málo úsmevu	1,3,10,4,2	1,6701; 0,5567	4,53	2	0,10	0,29	0,03
Vo večnosti slobodná	8,22,15,6,5,2,2	2,1271; 1,3296	4,20	3	0,24	0,75	1,50
Vyznania	7,13,4,0,2	0,5561; 0,2994	0,51	1	0,48	0,52	1,43
Z neba do neba	22,12,5,1	0,8735; 1,5399	0,20	1	0,65	0,39	0,85

The poem *Nepoznateľné* can be captured rather using the geometric distribution with parameter $p = 0{,}5517$ ($X^2 = 1{,}33$, $DF = 1$, $P = 0{,}27$) or using the Poisson distribution with parameter $a = 0{,}7134$, ($X^2 = 0{,}31$, $DF = 1$, $P = 0{,}57$). All the other data can be well fitted using the Hyperpoisson. Only in one case (*Hľadanie odpovedí*) the parameters seem to increase but there is no necessity to use a different distribution.

Looking at the *<I,S>* domain we see that verse-lengths are placed in the same domain as word-lengths. However, here the dispersion is still greater than with words, hence everything points to the existence of an ellipsis

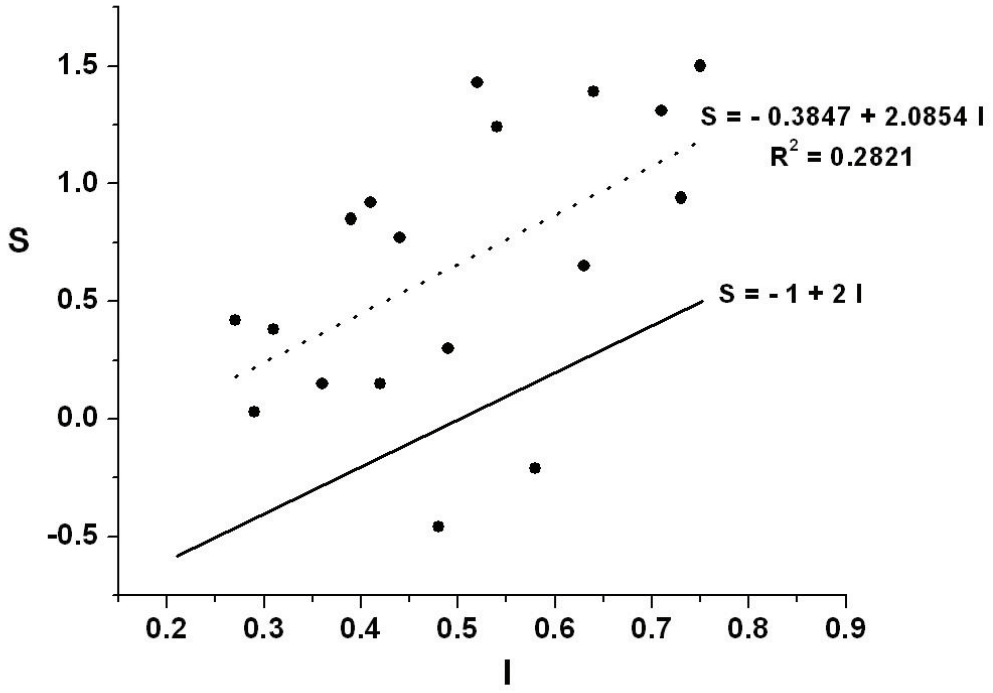

Figure 2. The <I,S> domain of verse length distributions

4. Verse length and Menzerath's law

According to Menzerath's law "the greater is a construct, the smaller are its components". The components are always immediate constituents. The law is stochastic and there is no transitivity, i.e. it does not hold necessarily that the greater the construct, the greater the components of the components. But just this is the contents of the so called Arens' law concerning sentence length and word length. This relationship evokes many problems and Grzybek, Stadlober and Kelih (Grzybek, Stadlober 2007; Grzybek, Stadlober, Kelih 2006, 2008) have shown that the result depends on text sort.

Since Bachletová poetry has its singular character (no rhyme, no fixed verse length, no meter) expressed especially by the shortness of verses, we can consider verse as a poetic construct whose immediate components are words. Bachletová's verse is a poetic substitute for the linguistic clause. If this conjecture is correct, then it must hold that the longer the verse, the shorter are its words on the average. The relationship follows from a simple logic: if the poet puts only one word in the verse, then it is most probably an autosemantic, and autosemantics are usually longer than synsemantics; but if he prolongs the verse, he has the chance to insert short synsemantics between autosemantics whereby the mean word length decreases. If our hypothesis is correct, we

have two problems: to test the hypothesis on individual poems and to obtain the parameters of this relationship characteristic for Bachletová.

This hypothesis is easily testable but the result strongly depends on the representativeness of verse numbers of a certain length. It is not testable using short poems. In its present form the hypothesis is an analogue to Menzerath's hypothesis but does not concern directly linguistic constructs but rather poetic, textual ones.

To test the above hypothesis we have chosen 14 longer poems. The results are presented in Table 4. Though not all length classes were representative, the general power trend representing Menzerath's law, $y = ax^{-b}$, could be shown. The parameters and the determination coefficients are shown in the last three columns of Table 4.

Table 4
Menzerath's law for verses

Poem	**Verse length (Number of words)**								**a**	**b**	**R^2**
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**			
Aby spriesvitnela	3,5	2,63	1,92	1,83	1,4				3,5534	-0,5215	0,98
Iba neha	2,54	2,22	1,58	1,64	1,6	1,5			2,5586	-0,3147	0,91
Čakanie na boží jas	2,38	1,8	1,97	1,75	1,6	1,5			2,3303	-0,2281	0,86
Hladanie odpovedí	3	2,38	2,04	1,79					2,4823	-0,0174	0,80
Idem za tebou	---	2,5	2,5	1,88	1,9	1,89	1,71	1,5	3,3009	-0,3446	0,86
Moje určenie	3,3	2,17	1,93	1,68	1,7	1,5			3,2094	-0,4466	0,97
Rozťatá pritomnosť	2,67	1,83	1,71	1,25					2,6658	-0,4920	0,96
Tak málo úsmevu	5	1,83	2,1	2	1,9				4,6872	-0,7409	0,81
Môj ošiaľ	2,5	1,89	1,67	1,5					2,4883	-0,3703	~1.0
Podobnosť bytia	2,75	2,57	2,19	2	2,2	1,5			2,8539	-0,2487	0,78
Nepoznateľné	2,89	1,53	1,67	1,38	1,2				2,7772	-0.5479	0,90
Dielo Stvoriteľa	3,5	2,54	2,06	1,97	1,67	1,33			3,4948	-0,4542	0,99
Z neba do neba	2,73	2,13	1,78						2,7380	-0,3813	~1.0
Rozdelená bytosť	4,00	2,59	1,53	1,80	1,60	1,33			3,9520	-0,6328	0,95

The result has some textological consequences. We see that in text still other constructs than the usual linguistic ones abide by some regularities. In our case it was the verse but it can be asked whether there are still other entities, which underlie stochastic regularities, e.g. the strophe, the rhymed pair of verses, semantically or associatively constructed images of reality, etc. The setting up of such units and finding the relevant regularities holding for them are very demanding tasks whose solution must be delayed to future research.

References

Best, K.-H. (ed.) (2001). *Häufigkeitsverteilungen in Texten*. Göttingen: Peust & Gutschmidt.

Best, K.-H. (2006). *Quantitative Linguistik. Eine Annäherung*. Göttingen: Peust & Gutschmidt.

Grzybek, P. (ed.) (2006). *Contributions to the science of text and language. Word length studies and related issues*. Dordrecht: Springer.

Grzybek, P. (2006a). History and methodology of word length studies. In: Grzybek (2006): *15-90.*

Grzybek, P., Stadlober, E. (2007). Do we have problems with Arens' law? A new look at the sentence-word relation. In: Grzybek, P., Köhler, R. (eds.), *Exact methods in the study of language and text: 205-217.* Berlin-New York: Mouton de Gruyter.

Grzybek, P., Stadlober, E., Kelih, E. (2006) The relationship of word length and sentence length. The inter-textual perspective. In: Decker, R., Lenz, H.-J. (eds.), *Advances in Data Analysis. Proceedings of the 30th Annual Conference of the Gesellschaft für Klassifikation e.V., Freie Universität Berlin, March 8-10, 2006: 611-618*. Berlin, Heidelberg: Springer

Grzybek, P., Stadlober, E., Kelih, E. (2008). The relation between word Length and sentence length: an intra-systemic perspective in the core data structure. *Glottometrics 16, 111-121.*

Köhler, R. (2005). Language synergetics. In: Köhler, R., Altmann, G., Piotrowski, R.G. (eds.), *Quantitative Linguistics. An International Handbook: 760-774*. Berlin-New York: de Gruyter.

Popescu, I.-I. et al. (2009). *Word frequency studies*. Berlin-New York: Mouton de Gruyter.

Rietveld, T., Hout, R.v., Ernestus, M. (2004). Pitfalls in corpus research. *Computers and the Humanities 38(4), 343-362.*

Uhlířová, L. (1995). On the generality of statistical laws and individuality of texts. A case of syllables, word forms, their length and frequencies. *Journal of Quantitative Linguistics 2, 238-247.*

Wilson, A. (2006). Word-length distribution in present-day Lower Sorbian newspaper texts. In: Grzybek (2006): *319-327.*

Wimmer, G., Altmann, G. (2005). Towards a unified derivation of some linguistic laws. In: Köhler, R., Altmann, G., Piotrowski, R.G. (eds.), *Quantitative Linguistics. An International Handbook: 307-316*. Berlin-New York: de Gruyter.

Wimmer, G., Witkovský, V., Altmann, G. (1999). Modification of probability distributions applied to word length research. *Journal of Quantitative Linguistics 6, 257-268.*

Zipf, G.K. (1935/1968). *The psycho-biology of language: an introduction to dynamic philology.* Cambridge, Mass.: The M.I.T. Press.