Sonnets: Quantitative Inquiries

by

Sergey Andreev
Michal Místecký
Gabriel Altmann

2018
RAM-Verlag

Studies in Quantitative Linguistics

Editors

Fengxiang Fan	(fanfengxiang@yahoo.com)
Emmerich Kelih	(emmerich.kelih@univie.ac.at)
Reinhard Köhler	(koehler@uni-trier.de)
Haitao Liu	(htliu@163.com)
Ján Mačutek	(jmacutek@yahoo.com)
Eric S. Wheeler	(wheeler@ericwheeler.ca)

1. U. Strauss, F. Fan, G. Altmann, *Problems in quantitative linguistics 1*. 2008, VIII + 134 pp.
2. V. Altmann, G. Altmann, *Anleitung zu quantitativen Textanalysen. Methoden und Anwendungen.* 2008, IV+193 pp.
3. I.-I. Popescu, J. Mačutek, G. Altmann, *Aspects of word frequencies*. 2009, IV +198 pp.
4. R. Köhler, G. Altmann, *Problems in quantitative linguistics 2.* 2009, VII + 142 pp.
5. R. Köhler (ed.), *Issues in Quantitative Linguistics.* 2009, VI + 205 pp.
6. A. Tuzzi, I.-I. Popescu, G.Altmann, *Quantitative aspects of Italian texts*. 2010, IV+161 pp.
7. F. Fan, Y. Deng, *Quantitative linguistic computing with Perl.* 2010, VIII + 205 pp.
8. I.-I. Popescu et al., *Vectors and codes of text*. 2010, III + 162 pp.
9. F. Fan, *Data processing and management for quantitative linguistics with Foxpro***.** 2010, V + 233 pp.
10. I.-I. Popescu, R. Čech, G. Altmann, *The lambda-structure of texts*. 2011, II + 181 pp
11. E. Kelih et al. (eds.), *Issues in Quantitative Linguistics Vol. 2*. 2011, IV + 188 pp.
12. R. Čech, G. Altmann, *Problems in quantitative linguistics 3*. 2011, VI + 168 pp.
13. R. Köhler, G. Altmann (eds.), *Issues in Quantitative Linguistics Vol 3.* 2013, IV + 403 pp.
14. R. Köhler, G. Altmann, *Problems in Quantitative Linguistics Vol. 4.* 2014, VIII+148 pp.
15. Best, K.-H., Kelih, E. (eds.), *Entlehnungen und Fremdwörter: Quantitative Aspekte.* 2014. VI + 163 pp.
16. I.-I. Popescu, K.-H. Best, G. Altmann, G. *Unified Modeling of Length in Language.* 2014, VIII + 123 pp.
17. G. Altmann, R. Čech, J. Mačutek, L. Uhlířová (eds.), *Empirical Approaches to Text and Language Analysis dedicated to Luděk Hřebíček on the occasion of his 80th birthday.* 2014. VI + 231 pp.
18. M. Kubát, V. Matlach, R. Čech, *QUITA Quantitative Index Text Analyzer*. 2014, VII + 106 pp.

19. K.-H. Best, *Studien zur Geschichte der Quantitativen Linguistik.* 2015. III + 158 pp.
20. P. Zörnig, K. Stachowski, I.-I. Popescu, T. Mosavi Miangah, P. Mohanty, E. Kelih, R. Chen, G. Altmann, *Descriptiveness, Activity and Nominality in Formalized Text Sequences.* 2015. IV+120 pp.
21. G. Altmann, *Problems in Quantitative Linguistics Vol. 5*. 2015. III+146 pp.
22. P. Zörnig, K. Stachowski, I.-I. Popescu, T. Mosavi Miangah, R. Chen, G. Altmann, *Positional Occurrences in Texts: Weighted Consensus Strings.* 2015. II+178 pp.
23. E. Kelih, R. Knight, J. Mačutek, A.Wilson (eds.), *Issues in Quantitative Linguistics Vol. 4*. 2016. III + 231 pp.
24. J. Léon, S. Loiseau (eds.), *History of quantitative linguistics in France*. 2016. II + 232 pp.
25. K.-H. Best, O. Rottmann, *Quantitative Linguistics, an Invitation*. 2017, III+171 pp.
26. M. Lupea, M. Rukk, I.-I. Popescu, G. Altmann, *Some Properties of Rhyme,* 2017, IV + 134 pp.
27. G. Altmann, *Unified Modeling of Diversification in Language*. 2018, VIII+119 pp.
28. E. Kelih, G. Altmann, *Problems in Quantitative Linguistics, Vol. 6.* 2018, IX+118 pp.

RAM-Verlag
Stüttinghauser Ringstr. 44
D-58515 Lüdenscheid
Germany
RAM-Verlag@t-online.de
http://ram-verlag.eu

ISBN: 978-3-942303-71-2

Contents

1. Introduction 1

2. Parts-of-Speech: Complex Assessment 3
1.1. General Tendencies 3
2.2. POS Motifs 6
2.3. POS of Rhyme Words 8
2.4. Rhyme-Word POS as Motifs 14

3. Belza-Chains 22
3.1. General Measurements 22
3.2. Belza-chain Motifs 30

4. Hreb Analysis 32
4.1. Generalities 32
4.2. Ordered Positional Hrebs 36
4.3. Text Compactness 39

5. Equalities 40

6. Adnominals 41
6.1. General Measurements 41
6.2. Ornamentality 45
6.3. Attributiveness 45
6.4 Adnominal Runs 50
6.5. Modelling Tendencies 54
6.6. Valency Motifs 58

7. Consensus Strings 62

8. Measuring Activity 66
8.1. Busemann's Coefficient and Chi-Square Testing 66
8.2. Activity Development 75

9. H-point and Lambda Indicator: Generalities 79

10. Frequency Studies 83
10.1. General Measurements 83
10.2. H-point and Lambda Indicator 93
10.3. Repeat Rate and Herfindahl's Indicator of Concentration 95

11. Investigating Syllables **99**
11.1. Type Distributions 99
11.2. Comparisons 111

12. Weighting **114**

13. Nominal Valency **117**

14. Conclusions **119**

Register **120**

Author Index **127**

Subject Index **129**

1. Introduction

The birth and the history of sonnets are well-known and need not be repeated here. The fact that they consist of 4 + 4 + 3 + 3 lines, but the rhymes are not always equally positioned, is also a piece of basic literary knowledge. Literary scholars have written a lot about the contents, intentions, language, etc., of sonnets (cf. e.g., Jacob et al. 2017. Zhang et al. 2010; Delmonte 2016; Kernot et al. 2017; Pan et al. 2018). Here, we want to analyze only the textological aspects which can be partially or in various ways quantified, and make a small step into the deep. The main goal of the book is to provide a material background for multifarious detailed studies on particular problems. The individual types of rhyming (e.g., *abab*…) have been sufficiently described; one distinguishes Petrarca, Shakespeare, de Ronsard, and other types, but this is a rather superficial observation. Considering sonnets, one has the advantage of the same structure, and the disadvantage of text shortness. The numbers obtained – in whatever sense – are not quite firm, but no statistician has ever said how long a text must be in order to yield safe results.

Any text has an infinite number of properties. They are not inherent, but constructed by us conceptually. In order to find some laws, all sonnets in all languages should be analyzed – but such a task is impossible to fulfill. One would always enter new conceptual levels, and the work would thus be endless. Nevertheless, in the case of sonnets, one has a great advantage, compared to other text types: to an extent, all sonnets are written in the same way. But the shortness causes that some of the well-known textual properties cannot be measured; or if measured, they give very unreliable results.

The following (preliminary) questions may be scrutinized:

(1) Are there some tendencies in using special parts-of-speech in the rhyme words? And what about the distribution of parts-of-speech in general?

(2) What are the Belza-chains of the poem like? They cannot be longer than 14, but even this number is quite seldom. Nevertheless, they have lengths, one can construct their motifs and study their number.

(3) What is the hreb-organization of a sonnet? That means – what is the semantic organization of the sonnet?

(4) What are the distances between equal entities? The entities can be chosen freely, not all display a regularity, but one can surely find some special ones, even if only for individual sonnets. Equality must be strictly defined.

(5) Which adnominals can be found in the sonnet and what is their distribution?

(6) Are there expressed any consensus strings introduced by Zörnig (2016), and for which kinds of entities do they hold, and in what form?

(7) What ways of activity measurement can be used in investigating sonnets?
(8) In what manners can h-point, concentration, and lambda indicator contribute to the knowledge of the sonnet structure?
(9) What type of type-token mathematics is useful for investigating sonnets?
(10) What can be researched about syllabic structures of sonnets?
(11) What about the weighting of individual sonnet features?
(12) What are the features of nominal valency in sonnets?

One can (and will) ask much more questions – the problems are discussed in many publications on text analysis –, but we shall restrict ourselves to some few ones.

As it is not possible to analyze all sonnets in all languages, we must restrict ourselves to some chosen ones. Some of them represent collections, i.e. consist of several sonnets, have the same theme, but develop it. In the present book, we investigate Czech, Slovak, German, Russian, English, French, and Hungarian texts.

Some questions concerning texts are not very fruitful for the study of sonnets. For example, the study of vocabulary richness of individual sonnets in the usual way is a rather frustrating enterprise. There are scarcely words that are repeated. Even a complete collection of sonnets cannot comply to our understanding of vocabulary richness. The poets try to omit word repetitions. On the other hand, counting the occurrences of words belonging to some classes (e.g., POS) does not signalize the richness. It would be possible to study this problem taking into account a complete collection of sonnets, but the collections avoid repetitions in the similar way. They are written with time breaks, and even if they are presented as a collection, they tend to differentiate thematically. The only possibility is to study the spectrum of word frequencies, which abides – at least in sonnets – by the same law.

Our aims are manifold. First, we want to look at the properties that can be defined for sonnets and yield variable results. Second, we want to show some theoretical models which adequately capture the empirical results and yield a possibility to develop a specified theory. And third, we endeavour to demonstrate various ways of stylometric evaluations of poetic texts.

2. Parts-of-Speech: Complex Assessment

2.1 General Tendencies

Though the grammar of a language has various prescriptions concerning the use and sequence of parts-of-speech, in poems, many prescriptions may be ignored. However, in sonnets, one usually tends to keep basic grammatical rules. In this way, every language presents in sonnets a "normal" picture, which can be variegated according to the topic and author. Here, we shall analyze only a few sonnets in several languages in order to show the regularity of this phenomenon. Regularity means behaviour according to a general tendency which can be modelled. Consider, e.g., the Russian sonnet written by Trediakovskij (1735), yielding the POS as follows:

N,	PA,	PA,	N,	N,	V,	A,	
V,	PN,	AV,	PP,	PN,	A,	V,	
C,	PN,	L,	PP,	PN,	N,	AV,	A,
C,	AV,	N,	PN,	PA,	AV,	V,	V,
I,	PA,	N,	N,	C,	PA,	A,	
L,	V,	C,	PN,	AV,	N,	V,	
PN,	PP,	PA,	N,	PA,	ST,	A,	
PA,	C,	N,	PN,	PA,	V,	V,	
V,	L,	AV,	C,	AV,	PN,	AV,	
N,	PP,	PA,	V,	N,	V,	C,	AV,
V,	V,	AV,	N,	A,			
V,	N,	PN,	PN,	AV,	V,		
N,	PP,	N,	V,	PA,	A,		
PN,	PA,	A,	N,	AV,	V,		

Here, the abbreviations are:
N – noun, A – adjective, V – verb, AV – adverb, C – conjunction, I – interjection, L – particle, PTV – "verbal" participle, PTA – "adjectival" participle, PN – pronoun nominal, PA – pronoun adjectival, NU – numeral, ST – stative, PP – preposition.

By collecting the individual POS, we obtain the results presented in Table 2.1. The ranked frequencies can be fitted by different functions; here, we choose the exponential one (with added 1) because of its simplicity both in theory and in practice.

Table 2.1
POS in Trediakovskij's sonnet (1759)

POS type	Rank	Frequency	Exponential ft.
N	1	24	23.77
V	2	15	16.14
PN	3	12	11.07
AV	4	8	7.70
PA	5	6	5.45
A	6	5	3.96
I	7	2	2.97
L	8	1	2.31
C	9	1	1.87
ST	10	1	1.58
a = 35.2450, b = 0.4080, R^2 = 0.9855			

Since the number of sonnets is very large, we present only some of the results, namely the parameters and the determination coefficients. Following the values of R^2, the fits are usually highly successful. In Russian, Slovak, Hungarian, and German sonnets, we use the exponential function with added 1, and accept it if $R^2 > 0.8$. In Czech, English, and French sonnets, the exponential function is employed; if it does not provide the appropriate fit, the generalized form of the power-law distribution is made use of – i.e.,

$$y = ax^{-b}e^{-c} \quad . \tag{2.1}$$

The poems fitting the latter are marked with asterisks.

Table 2.2
Fitting the rank-order of POS in several Russian, Slovak, Hungarian, and German poems

Russian Sonnets	**a**	**b**	**R^2**
Tredjakovskij 1 (1735)	26.123	0.2635	0.8536
Tredjakovskij 2 (1759)	35.245	0.408	0.9855
Cheraskov: Sonet i epitafia 1755	37.355	3.2284	0.9818
Cheraskov: *Kol' budu v zhizni ja nakazan nischchetoju*	35.636	3.0055	0.9776
Rzhevskij: *Sonet, zakljuchajushchij v sebe tri mysli*	29.247	4.0564	0.9159
Slovak Sonnets	**a**	**b**	**R^2**
Hviezdoslav 1	39.585	0.2833	0.9652

Hviezdoslav 2	44.983	0.396	0.9732
Hviezdoslav 3	32.253	0.2258	0.8521
Hviezdoslav 4	37.786	0.3173	0.8684
Hviezdoslav 5	68,5568	0.6569	0.9552
Hungarian Sonnets	**a**	**b**	**R^2**
Faludi, F.: *A Pipárul*	24.88	0.3816	0.9832
Babits, M.: *A lírikus epilógja*	27.825	0.2781	0.9359
Jozsef, A.: *A kozmosz éneke 1*	29.167	0.3042	0.9775
A kozmosz éneke 2	38.817	0.4383	0.9633
A kozmosz éneke 3	28.731	0.3502	0.9182
German Sonnets	**a**	**b**	**R^2**
Goethe J. W.: *Sonett 1*	31.858	0.2545	0.8357
Goethe, J. W.: *Sonett 2*	30.33	0.2841	0.9497
Goethe, J. W.: *Freundliches Begegnen*	26.563	0.2218	0.9348
Goethe, J. W.: *Mächtiges Überraschen*	27.24	0.23	0.9037
Goethe, J. W.: *Kurz und gut*	28.224	0.2353	0.9159

Table 2.3
Fitting the rank-order of POS in several Czech, English, and French poems

Czech Sonnets	**a**	**b**	**c**	**R^2**
J. S. Machar: *E. Zolovi*	22.3087	0.7552	–	0.9226
J. S. Machar: *Matce*	26.4645	0.839	–	0.8896
J. S. Machar: *Sonet cynický**	33.2384	-0.4825	0.4114	0.964
J. S. Machar: *Sonet de vanitate*	24.7776	0.8161	–	0.9403
J. S. Machar: *Sonet elegický**	28.1431	-0.8279	0.4849	0.9435
English Sonnets	**a**	**b**	**c**	**R^2**
G. G. Byron: *Sonnet on Chillon*	31.318	0.7482	–	0.8835
E. B. Browning*: I**	29.7648	-0.2868	0.2736	0.9311
J. Keats: *On the Grasshopper and Cricket*	29.8811	0.6577	–	0.9566
P. B. Shelley: *England in 1819*	27.8681	0.5895	–	0.8573
A. Tennyson: *The Kraken*	26.5861	0.5794	–	0.9215
French Sonnets	**a**	**b**	**c**	**R^2**
Ch. Baudelaire: *Correspondances*	27.4537	0.5217	–	0.8295
Ch. Cros: *Révolte*	28.6863	0.6702	–	0.8187

A. de Musset: *Jamais**	30.381	-0.3915	0.292	0.9306
S. Prudhomme: *Aux jeunes**	34.7324	-0.1528	0.2737	0.8973
P. Verlaine: *Allégorie*	27.0142	0.6667	–	0.9109

2.2 POS Motifs

Motifs have been discovered by R. Köhler (2008, 2015) (cf. also Liu, Liang 2017). As soon as one has a sequence of numbers or symbols, one can study the motifs which have, in turn, their own properties, e.g., length. Besides, quantitative motifs have a range and a mean; qualitative ones are poorer, but they have a ranked distribution of types. In a qualitative motif, the same symbol must not occur in the same motif; in the next motif, only one symbol of the previous motif may occur. In sonnets, the examination will be somewhat restricted because they are short texts. Nevertheless, one can study the motifs of rhyme-word sequences, those of Belza-chains, those of word frequencies, etc.

Above, we have presented the POS of the Russian sonnet written by Trediakovskij (1735). Below, the POS motifs of the same poem are listed, accompanied with the motifs in G.G. Byron's *Sonnet on Chillon*. Since the frequencies of individual motifs are very low, the modelling will be carried out on the basis of motifs' lengths. The fit will be provided by the Lorentzian function, as the data manifest a bell-shaped distribution – i.e.,

$$y = \frac{a}{1 + \left(\frac{x - b}{c}\right)^2}. \tag{2.2}$$

Table 2.4
POS motifs in Tredjakovskij's (1735) sonnet

Motif	Fr.	Motif	Fr.	Motif	Fr.
[N,PA]	1	[N,C]	1	[AV,N,PP,PA,V]	1
[PA,N]	2	[PA,A,L,V,C,PN,AV]	1	[N,V,C]	1
[N,V,A]	1	[N,V]	1	[AV,V]	1
[V,PN,AV,PP]	1	[PN,PP,PA,N]	1	[AV,N,A,V]	1
[PN,A]	1	[PA,ST,A]	1	[N,PN]	1
[V,C,PN,L,PP]	1	[PA,C,N,PN]	1	[PN,AV,V]	1
[PN,N,AV,A]	1	[PA,V]	1	[N,PP]	1
[C,AV]	1	[V]	2	[N,V,PA,A,PN]	1
[N,PN,PA,AV,V]	1	[V,L,AV,C]	1	[PA]	1
[V,I]	1	[AV,PN]	1	[A,N,AV,V]	1

Table 2.5
Lengths of POS motifs in the sonnet by Tredjakovskij (1735)

Length	**Frequency**	**Frequency (Lorentzian)**
1	3	4.66
2	13	12.32
3	4	6.36
4	7	2.26
5	4	1.07
6	1	1.79
7	1	1.79
Lorentzian fit	a = 12.6228, b = 2.1370, c = 0.8700 R^2 = 0.6168	

Table 2.6
POS motifs in Byron's *Sonnet on Chillon*

Motif	**Fr.**
[A,C,PN]	1
[A,N,DET]	1
[A,N,PP,DET]	1
[A,N,V,DET]	1
[A,N,V]	1
[A,N]	2
[A,PP]	1
[A]	1
[C,PN,N,PP]	1
[C,PN]	1
[N,A,C]	1
[N,C,DET]	1

[N,C,PN]	1
[N,C]	1
[N,DET]	1
[N,PN,V,PP,AV]	1
[N,PN]	2
[N,PP,PN,A,V]	1
[N,PP,PN,V]	1
[N,PP,PN]	1
[N,PP]	1
[N,V,A]	1
[N,V,DET,A]	1
[N,V,PN]	1
[N,V,PP,PN]	1

[N,V,PP]	1
[N,V]	1
[N]	4
[PN,N,C]	1
[PN,N,V,DET]	1
[PN]	1
[PP,N]	1
[PP]	2
[V,C]	1
[V,DET]	1
[V]	1

Table 2.7
Lengths of POS motifs in Byron's *Sonnet on Chillon*

Length	**Frequency**	**Frequency (Lorentzian)**
1	5	5.35
2	11	10.65
3	11	11.44
4	7	5.97
5	2	2.99
Lorentzian fit	a = 12.5633, b = 2.5742, c = 1.3563, R^2 = 0.9592	

As can be seen, the Lorentzian fit is not always ideal for modelling the distribution of the data; this may be caused by discrepancies in motif lengths, and can thus be attributable to stylistic peculiarities of the given work.

2.3 POS of Rhyme Words

A longer study concerning the membership of rhyme words in the classes of POS has already been published (cf. Lupea, Rukk, Popescu, Altmann 2018); here, we shall consider merely sonnets. The parts-of-speech are those that were proposed already in the Antiquity, and hold true still today. They are:

N = nouns, V = verbs, A = adjectives, Av = adverbs, Pn = pronouns, Nu = numerals, I = interjections, C = conjunctions.

In some non-European languages, one finds still other ones, but in case of necessity, one can insert them in the above collection. In some languages, one can find participles (active or passive) which have the form of an adjective or an adverb; in that case, they will be considered as such (A or Av). In English, it depends on the environment whether forms with *-ing* and *-ed* can be considered verbs or adjectives. One can perform even sub-classifications and order the words in as many classes as appropriate. The data itself can sometimes require a more detailed classification, especially if such an operation leads to the capturing of boundary conditions. One can adhere to "official" grammars, but one can define the classes in a different way, too.

In Table 2.8, the sequences of rhyme-word POS are presented. If one ranks them according to decreasing frequency, one obtains a sequence which can be satisfactorily captured by the Zipf-Alekseev function, defined as

$$y = 1 + c * e^{a+b*\ln(x)}, \tag{2.3}$$

which can be derived from the unified theory of language laws (cf. Wimmer, Altmann 2005) by means of a differential equation. Nevertheless, we use also other functions. For example, for German, we fitted the Zipf-Mandelbrot function, defined as

$$y = 1 + \frac{a}{(b+x)^c}, \tag{2.4}$$

and used in many domains of ranking. There are a number of other possibilities; one should care merely for simplicity and unity.

In Table 2.8, the individual rows concern individual sonnets. There are always 14 symbols in each row, as the sonnet has 14 lines.

Table 2.8
POS of rhyme words in some sonnets

Author	Sonnet	Parts of speech of rhyme words
P.O. Hviezdoslav *Sonety* (Slovak)	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21	[N,N,N,N,N,Av,N,N,V,N,Pn,N,N,N; N,N,N,N,N,N,N,N,N,N,N,N,V,N; N,N,V,N,A,V,N,N,A,V,N,V,N,N; N,N,N,N,N,V,N,N,N,N,N,N,N,V; N,N,N,V,N,N,N,V,V,N,N,N,A,V; N,N,N,N,N,V,N,A,N,N,N,V,N,N; Av,N,V,N,N,N,N,A,A,N,V,N,N,V; N,A,V,A,A,N,V,A,N,N,V,V,N,N; N,N,N,V,N,V,V,N,Av,V,N,N,N,V; N,V,Av,V,V,N,N,N,N,N,N,N,N,N; N,N,N,V,V,Nu,N,N,N,A,V,N,N,N; N,N,V,N,V,N,V,A,N,N,N,N,N,N; N,I,N,N,N,N,N,V,N,N,N,A,N,N; N,N,N,N,V,N,N,N,N,N,N,N,N,A; N,V,N,N,N,N,V,V,Av,Pn,N,N,N,N; V,V,N,N,N,N,N,N,N,V,V,A,N,V; N,Av,N,N,N,N,N,N,Pn,A,N,N,N,N; N,N,N,V,N,N,V,Pn,N,N,N,N,N,Av; N,N,Av,A,N,N,V,N,N,N,N,N,N,V; Av,N,N,N,Av,N,N,N,N,N,N,Nu,V,N; A,N,V,N,N,N,A,V,V,N,N,N,V,A]
	Frequencies Classes Zipf-Alek-seev ft. + 1	203 54 20 10 4 2 N V A Av Pn Nu 202.99 54.15 19.81 9.02 4.87 3.05 a = -1.5244, b = -0.5794, c = 201.9893, $R^2 = 0.9999$
P.O. Hviezdoslav *Krvavé sonety* (Slovak)		[N,V,N,V,N,V,N,V,A,N,N,V,N,N; N,N,Av,N,N,N,N,Pn,V,N,V,N,V,N; N,N,A,N,N,A,N,N,N,N,A,V,N,N; N,N,N,N,N,A,N,N,N,V,N,N,N,N; N,N,N,N,N,V,N,N,N,N,A,N,N,V; N,N,N,N,N,V,N,N,N,N,N,A,N,N; V,V,V,C,V,N,C,N,V,V,N,A,V,V; A,N,Av,Av,N,A,V,N,N,V,N,N,Av,N; N,V,V,N,N,V,V,Av,N,N,V,N,N,N; N,Pn,A,V,N,A,A,V,N,N,N,N,N,N; N,N,N,N,N,A,N,N,N,N,V,N,A,N; V,N,N,V,V,A,A,N,N,N,N,N,N,N; N,N,N,N,N,A,N,N,V,A,A,Pn,N,Pn; A,N,V,N,N,N,V,V,N,V,N,A,N,N;

		N,A,N,V,A,N,N,V,N,A,N,Av,N,V; N,Av,N,N,N,V,N,A,N,A,A,N,N,Pn; N,N,N,Av,N,A,N,N,N,A,V,N,N,N; V,N,Pn,N,A,N,Av,N,N,N,N,N,A,N; N,N,V,V,N,N,N,N,N,A,V,N,V,N; V,N,V,N,N,N,N,V,N,N,N,N,Nu,Av; V,V,V,N,V,V,N,Pn,V,V,N,Av,N,V; Pn,N,N,N,Av,Av,N,N,N,N,N,N,A,N; N,N,N,N,A,A,N,N,N,N,V,N,N,N; N,N,N,V,N,V,N,N,A,V,N,N,Av,V; V,V,N,Av,V,N,N,V,N,N,N,N,N,V; N,N,A,N,N,N,N,N,N,N,N,V,N,V; V,Pn,N,I,V,N,N,N,N,N,N,N,N,A; A,A,N,N,N,V,N,A,N,N,N,A,N,Pn; N,N,Av,Pn,N,N,A,N,Pn,N,N,V,N,N; A,N,V,V,N,N,A,V,V,N,V,N,N,N; A,A,N,N,N,N,A,A,N,N,I,N,N,N; A,A,A,N,Av,N,A,A,Av,A,Av,A,N,N]
	Frequencies Classes Zipf-Alek-seev ft. +1	274 81 57 19 12 2 2 1 N V A Av Pn I C Nu 273.48 88.19 40.20 21.94 13.45 8.96 6.38 4.78 a = -1.4372, b = -0.2983, c = 272.4829, $R^2 = 0.9929$

One can compare the two collections in at least two different ways: (1) One compares ranked frequencies; or (2), one can compare the frequencies in individual POS-classes. Both tests can be performed applying the chi-square test. Solving the first problem, we obtain – using the data in Table 2.8 – the numbers in Table 2.9.

Table 2.9
Comparison of ranked frequencies of POS
in Hviezdoslav's *Sonety* and *Krvavé sonety*

Rank	***Sonety***	***Krvavé sonety***	**Sums**
1	203	274	477
2	54	81	135
3	20	57	77
4	10	19	29
5	4	12	16
6	2	2	7
7		2	
8		1	
	293	448	741

Since the last three numbers in *Krvavé sonety* are too small, we added them obtaining 5. Performing the test according to the well-known formula, we obtain $\chi^2 = 9.83$, which is – with 5 degrees of freedom – not significant. That means, in this sense, that Hviezdoslav had a fixed tendency of placing POS in rhymes. Of course, one should compare all his sonnets with all the other ones.

In the German sonnets by Goethe, we find the situation presented in Table 2.10. Here, the number of verbs is greater than that of nouns, and the ranking can be better captured by the Zipf-Mandelbrot function.

Table 2.10
POS in the rhyme words of German sonnets

Author	Sonnet	Part of speech of rhyme words
J.W.v. Goethe	*Das Sonett 1*	[V,V,V,V,V,V,V,V,N,N,V,V,N,V;
	2	V,V,V,V,N,N,V,V,V,N,V,V,N,V;
	Mächtiges Überraschen	N,V,N,N,N,N,V,N,V,V,N,V,N,N;
	Freundliches Begegnen	V,V,N,V,V,V,N,V,V,N,V,V,V,N;
	Kurz und gut	V,N,N,N,V,V,N,N,A,N,V,V,N,V;
	Das Mädchen spricht	N,V,N,A,N,V,V,N,V,V,V,V,V,V;
	Wachstum	N,N,N,V,V,V,V,V,V,N,V,V,Av,A;
	Reisezehrung	N,V,N,Av,N,Pn,V,N,N,N,V,V,V,N;
	Abschied	N,V,N,V,N,N,N,N,V,V,Av,V,V,V;
	Die Liebende schreibt	Pn,N,N,V,Pn,N,N,V,Av,N,V,N,N,N;
	Die Liebende abermals	V,V,V,N,V,V,N,N,V,N,V,V,V,V;
	Sie kann nicht enden	V,V,N,N,V,N,V,V,N,V,V,V,V,V;
	Nemesis	V,V,N,V,V,V,N,V,N,N,V,N,N,N;
	Christgeschenk	N,N,N,V,N,V,N,V,N,N,V,N,N,V;
	Warnung	V,N,V,V,Pn,N,V,V,N,V,V,V,V,N;
	Die Zweifelnden	N,V,V,N,N,V,V,N,N,V,V,N,V,V;
	Mädchen	N,N,V,V,V,V,V,V,N,V,N,A,Av,N;
	Epoche	V,N,V,N,V,V,V,V,A,A,N,A,A,N;
	Charade	V,V,V,V,N,V,V,N,V,V,V,V,V,V]
	Frequencies	148 101 8 5 4
	Classes	V N A Av Pn
	Zipf-Mandelbrot ft. +1	148.00 100.98 8.46 1.41 1.02
		a = 3.1323, b = -5.3214, c = 147.0029
		$R^2 = 0.9988$

The POS-inertia or monotony of rhyme POS of individual sonnets can be expressed in various ways. One can apply the usual repeat rate, the entropy, one can compare the similarity of the vectors, etc. (see below). The sonnets can be ordered according to the year of creation, too.

The differences between individual sonnet collections or authors or languages can be performed by direct comparison of frequencies applying, e.g., the

chi-square test, or by ordering individual POS by ranks and comparing them by a nonparametric test. If one compares two languages, one can take into consideration all sonnets analyzed.

Let us compare Goethe's *Sonnets*, with the sum of both Slovak collections by Hviezdoslav. We obtain the results presented in Table 2.11.

Table 2.11
Comparison of Goethe's and Hviezdoslav's sonnets

Rank	***Hviezdoslav***	***Goethe***	**Sums**
1	477	148	625
2	135	101	236
3	77	8	85
4	29	5	34
5	16	4	20
6	7	0	7
	741	266	1007

For the numbers in Table 1.3 we obtain $\chi^2 = 52.96$, which is, with 5 DF, highly significant. Most probably, in both languages one thus follows a subconscious ranking law.

The second kind of testing is the comparison of the frequencies of the same POS. For the two collections by Hviezdoslav, we obtain the same results because the proportional correspondence of POS is the same as the ranking correspondence. But comparing Hviezdoslav with Goethe yields a very significant result because the two languages are quite different and the authors have their own customs. In Table 2.12, one can see that Hviezdoslav uses rather nouns and adjectives in the rhymes, while Goethe prefers verbs.

Table 2.12
Frequencies of equal POS in collections by Hviezdoslav and Goethe

POS	***Hviezdoslav***	***Goethe***	**Sum**
N	477	101	578
V	135	148	283
A	77	8	85
Av	29	5	34
Pn	16	4	20
Nu	3	0	7
I	2	0	
C	2	0	
	741	266	1007

In this case, we obtain a very significant difference, namely $\chi^2 = 139.28$ (with 5 DF).

Now, let us consider Czech sonnets by J. S. Machar, who wrote 4 volumes of them. The results of the first volume, *Letní sonety*, are displayed in Table 2.13. The POS are ordered according to their frequency. Here, "Ab" means abbreviations, such as "etc.", which, as a matter of fact, do not represent a POS, but must be taken into account. The Zipf-Alekseev function with added 1 is satisfactory; one could simplify the fit by using the exponential function, too.

Table 2.13
Frequency of POS in rhyme words in *Letní sonety* by J.S.Machar (Czech)

POS	**Rank**	**Frequ.**	**ZA+1**
N	1	272	271.06
V	2	186	190.97
A	3	105	99.92
Av	4	63	51.97
Pn	5	25	28.18
Nu	6	4	16.10
I	7	2	9.72
C	8	1	6.22
Ab	9	1	4.22
a = 0.1878, b = -1.0031, c = 270.0607, R^2 = 0.9944			

Comparing *Letní sonety* by Machar with the Slovak sonnets by Hviezdoslav, we obtain χ^2 = 79.43 for the ranking of POS (with 5 DF), and χ^2 = 131.28 for the POS alone. In both cases, the difference is evident.

For the comparison of Machar with Goethe, we obtain χ^2 = 61.93 for the ranking, and χ^2 = 87.48 for the POS alone.

As can be seen, languages differ in the construction of sonnets, even if they are very similar (Czech and Slovak). This may be attributed to the different schools the poets belong to (Hviezdoslav taking part in the Slovak National Revival, Machar forming the backbone of the 1890s generation). Linguistically, Hviezdoslav's production is noun-based, as the substantives in his sonnets are three times more frequent than verbs; on the other hand, Machar's sonnets seem more balanced. This may have been caused by the narrative tendency of some of the latter, and the abstract, static character of the former.

Next, in Czech and Slovak sonnets there are more nominal rhyme POS than in German, where the verbal ones prevail.

In the following tables, results for English and French sonnets are presented; the first one covers the POS distribution in five poems by Charles Baudelaire ("Bohémiens en voyage", "Correspondances", "L'Ennemi", "L'Idéal", and "La beauté"), and the second one the same in the first five sonnets of the collection "Modern Love" by George Meredith. The latter are special, as they contain 16 lines.

Table 2.14
Frequency of POS in rhyme words in five selected sonnets by Baudelaire

POS	Rank	Frequ.	ZA+1
N	1	45	45.60
A	2	19	14.45
V	3	5	7.67
Pn	4	1	5.06
a = 3.3835, b = -1.7293, c = 1.5134, R^2 = 0.9624			

Table 2.15
Frequency of POS in rhyme words in five selected sonnets by Meredith

POS	Rank	Frequ.	ZA+1
N	1	41	41.24
V	2	16	15.87
A	3	12	9.31
AV	4	7	6.50
PN	5	3	4.99
PP	6	1	4.07
a = 2.7715, b = -1.4358, c = 2.5177, R^2 = 0.9805			

2.4 Rhyme-Word POS as Motifs

The same as POS in general, the POS of rhyme words may be modelled as motifs. Consider, for example, the sequence of rhyme words classified as POS in Table 2.16 (Goethe).

Table 2.16
Motifs of rhyme word POS in the sonnets by Goethe

Motifs
[V],[V],[V],[V],[V],[V],[V],[V,N],[N],[V],[V,N],[V],
[V],[V],[V],[V,N],[N],[V],[V[,[V,N],[V],[V,N],[V],
[N,V],[N],[N],[N],[N,V],[N],[V],[V,N],[V],[N],[N],
[V],[V,N],[V],[V],[V,N],[V],[V,N],[V],[V],[V,N],
[V,N],[N],[N,V],[V],[N],[N,A],[N,V],[V],[N,V],
[N,V],[N,A],[N,V],[V],[N,V],[V],[V],[V],[V],[V],
[N],[N],[N,V],[V],[V],[V],[V],[V,N],[V],[V,Av,A],
[N,V],[N,Av],[N,Pn,V],[N],[N],[N,V],[V],[V,N],
[N,V],[N],[V,N],[N],[N],[N,V],[V,Av],[V],[V],[V],
[Pn,N],[N,V],[Pn,N],[N,V,Av],[N],[V,N],[N],[N],
[V],[V],[V,N],[V],[V,N],[N],[V,N],[V],[V],[V],[V],

[V],[V,N],[N],[V,N],[V],[V,N],[V],[V],[V],[V],[V],
[V],[V,N],[V],[V],[V,N],[V],[N],[N,V],[N],]N],[N],
[N],[N],[N,V],[N],[V,N],[V],[N],[N,V],[N],[N,V],
[V,N],[V],[V,Pn,N],[V],[V,N],[V],[V],[V],[V,N],
[N,V],[V],[N],[N,V],[V],[N],[N,V],[V],[N,V],[V],
[N].[N,V],[V],[V],[V],[V],[V,N],[V],[N,A,Av],[N],
[V,N],[V],[N,V],[V],[V],[V,A],[A,N],[A],[A,N],
[V],[V],[V],[V,N],[V],[V,N],[V],[V],[V],[V],[V],[V]

The motifs of rhyme-word POS can be classified according to the type of the motif, e.g. [N], [N,V], etc., or according to its length – i.e., the number of POS in it. Considering the first problem, we obtain the ranked types as presented in Table 2.17. Though the numbers look somewhat strange, the exponential function with added 1 expresses the ranking trend satisfactorily.

Table 2.17
Ranked types of POS-motifs in the sonnets by Goethe

Motif	**Number**	**Exponential ft.**
[V]	87	84.16
[N]	35	45.43
[V,N]	31	24.74
[N,V]	24	13.69
[A,N]	2	7.78
[N,A]	2	4.62
[Pn,N]	2	2.94
[A]	1	2.03
[V,A]	1	1.55
[V,Av]	1	1.30
[N,Av]	1	1.16
[N,A,Av]	1	1.08
[N,Pn,V]	1	1.05
[N,V,Av]	1	1.02
[V,Av,A]	1	1.01
[V,Pn,N]	1	1.01
a = 155.6378, b = 0.6268, R^2 = 0.9621		

From Table 2.17, one can easily compute the frequencies of motifs of the given length. We obtain the results presented in Table 2.18. One can fit the given data by a simple linear function, $y = a + bx$. The perfect agreement is, perhaps, a property of German, but it can be tested for other languages, too.

Table 2.18
Lengths of rhyme POS motifs in the sonnets by Goethe

Length	Number	Linear ft.
1	123	123.00
2	64	64.00
3	5	5.00
a =182.0000, b = -59.0000, R^2 = 1.0000		

For the sonnets by Hviezdoslav, we obtain the results concerning the POS rhyme-word motifs presented in Table 2.19.

Table 2.19
Motifs of rhyme POS in *Sonety* by Hviezdoslav

Rank/Type	Number	Exponential + 1
N	147	146.90
N,V	25	26.03
V	7	5.30
N,A,V	6	1.74
N,A	4	1.13
V,N	4	1.02
A	3	1.00
A,N,V	2	1.00
V,A	2	1.00
N,Av,A,	1	1.00
N,I	1	1.00
N,Nu,V	1	1.00
N,Pn	1	1.00
N,Pn,A	1	1.00
N,V,A,	1	1.00
N,V,Av	1	1.00
N,V,Pn	1	1.00
V,Av,Pn	1	1.00
V,N,Av	1	1.00
V,Nu	1	1.00
Av,N	1	1.00
Av,N,V	1	1.00
a = 850.2695, b = 1.7627, R^2 = 0.9978		

For the lengths of individual types, we obtain the results in Table 2.20. Again, the linear function yields a satisfactory fit.

Table 2.20
Lengths of motifs of rhyme POS in *Sonety* by Hviezdoslav

Length	Number	Linear ft.
1	157	156.59
2	39	42.06
3	17	11.30
a = 582.9372, b = 0.2686, R^2 = 0.9963		

For the work *Krvave sonety* by Hviezdoslav, we obtain the following two tables.

Table 2.21
Motifs of rhyme POS in *Krvavé sonety* by Hviezdoslav

Rank/Type	Number	Zipf-Alekseev
N	171	170.12
N,V	27	37.85
V	24	18.01
N,A	24	11.30
A	10	8.15
V,N	9	6.38
N,Av,Pn	8	5.27
A,N	5	4.52
N,A,V	5	3.99
N,Pn	4	3.59
A,N,Av	2	3.28
A,N,V	2	3.04
Av	2	2.84
N,V,A	2	2.68
V,N,Av	2	2.54
V,N,Pn	2	2.43
Av,A,N	1	2.33
C,N	1	2.25
N,Av,Pn	1	2.18
N,I	1	2.12
N,Nu,Av	1	2.06
N,Pn,A,V	1	2.02
N,Pn,V	1	1.97
Pn,N	1	1.94
V,A	1	1.90
V,Av	1	1.87
V,C	1	1.85
V,N,A	1	1.82
V,Pn,N,I	1	1.80

A,Av	1	1.78
A,Pn	1	1.77
A,V,N	1	1.75
a = -2.3800, b = 0.3058, c = 170.1168, R^2 = 0.9875		

For lengths, the linear function is not adequate – it yields negative values. Hence, we fitted the exponential function again and obtained good results (cf. Table 2.22).

Table 2.22
Lengths of motifs of rhyme-POS in *Krvavé sonety* by Hviezdoslav

Length	**Number**	**Exponential ft.**
1	207	207.52
2	77	75.18
3	29	22.24
4	2	9.87
a = 572.8012, b = 0.9849, R^2 = 0.9972		

The data for Machar are presented in Table 2.23.

Table 2.23
Motifs of rhyme POS in *Letní sonety* by Machar

Rank/Type	**Number**	**Exponential + 1**
N	121	116.66
V	67	70.44
N,V	31	42.69
A	26	26.03
V,N	22	16.02
N,A	16	10.02
Av	14	6.42
V,A	9	4.25
A,N	8	2.95
N,Av	8	2.17
Av,N	7	1.70
N,Pn	7	1.42
A,N,V	5	1.25
V,Av,A	5	1.15
A,V,N	4	1.09
N,V,Av	4	1.05
Pn,V,N	4	1.03
Av,A,V	3	1.02

N,Av,A	3	1.01
V,A,N	3	1.01
V,N,A	3	1.00
A,Av,N	2	1.00
A,V	2	1.00
Av,V	2	1.00
Av,V,A	2	1.00
Av,V,N	2	1.00
N,A,V,Av	2	1.00
N,Av,V	2	1.00
N,V,Pn	2	1.00
V,A,Av,N	2	1.00
A,N,Av	1	1.00
A,N,Pn,V	1	1.00
A,Nu,N	1	1.00
A,V,Pn	1	1.00
Av,N,V	1	1.00
Av,V,A,Pn,N	1	1.00
C,N,A	1	1.00
N,A,Pn	1	1.00
N,A,V	1	1.00
N,Ab	1	1.00
N,Av,V,Nu	1	1.00
N,C	1	1.00
N,Pn,Av	1	1.00
N,Pn,V,A	1	1.00
N,V,A	1	1.00
N,V,Nu	1	1.00
Pn,Av,A,N	1	1.00
Pn,V,A	1	1.00
V,N,A,Pn	1	1.00
V,Pn,A	1	1.00
V,Pn,N	1	1.00
V,Pn,Nu	1	1.00
a = 192.6631, b = 0.5103 R^2 = 0.9727		

Table 2.24
Lengths of motifs of rhyme POS in *Letní sonety* by Machar

Length	Number	Exponential ft.+1
1	228	230.75
2	114	107.41
3	58	50.28
4	9	23.82
5	1	11.57
a = 496.0617, b = 0.7697, R^2 = 0.9873		

Some more data for French and English sonnets are presented in the tables below.

Table 2.25
Motifs of rhyme POS in the five selected sonnets by Baudelaire

Rank/Type	Number	ZA + 1
[N]	27	27.10
[N,A]	9	8.77
[A]	6	4.82
[N,V]	3	3.31
[A,N]	2	2.56
[N,A,V]	1	2.14
[N,Pn]	1	1.87
a = 2.8429, b = -1.7488, c = 1.5208, R^2 = 0.9924		

Table 2.26
Lengths of motifs of rhyme-POS in the five selected sonnets by Baudelaire

Length	Number	ZA + 1
1	33	33.45
2	15	11.00
3	1	6.02
a = 2.2185, b = -1.6983, c = 3.5300, R^2 = 0.9195		

Table 2.27
Motifs of rhyme POS in the five selected sonnets by Meredith

Rank/Type	Number	ZA + 1
[N]	19	19.00
[V,N]	5	5.10
[N,V]	3	2.72

[N,A]	2	1.93
[A,V,N]	2	1.58
[N,V,Av]	1	1.39
[N,A,Pp,Pn]	1	1.28
[N,Av]	1	1.21
[N,Av,V]	1	1.17
[V]	1	1.13
[V,Pn]	1	1.11
[V,A]	1	1.09
[A,Av,N]	1	1.08
[Av]	1	1.06
[Av,A,N]	1	1.06
[Av,N,A]	1	1.05
[N,Pn,A]	1	1.04
a = 1.4010, b = -2.1348, c = 4.4346, R^2 = 0.9980		

Table 2.28
Lengths of motifs of rhyme-POS in the selected sonnets by Meredith

Length	**Number**	**Exponential ft.+1**
1	21	21.67
2	13	10.33
3	8	6.86
4	1	5.21
a = 4.2504, b = -1.1474, c = 0.2947, R^2 = 0.8749		

3. Belza-Chains

3.1 General Measurements

Belza-chains are usually considered as (uninterrupted) sequences of sentences containing the same concept. This is easy to find in strongly analytic languages, but it requires a decision in strongly synthetic ones. Can one consider grammatical affixes as identical with the given concept? For example, in German, *ich schreibe*, *du schreibst*, *er schreibt*; do the personal endings of the verb represent the same concept as the pronouns? From the linguistic point of view, this is correct, but there are cases where, e.g., the gender does not help to identify the identity of concepts because the language does not have gender. In German, we have *schöner Garten*, *schöne Frau, schönes Kind*, but in Hungarian, one finds *szép kert*, *szép asszony*, *szép gyerek*. Thus even the same (translated) text may contain quite different Belza-chains. Further, a morpheme may refer to two or more other concepts, e.g., in German *ich sah* ("I saw"), *er sah* ("he saw"), the zero ending of verbs can refer to the first or to the third person singular. The same holds for the first and third person plural (*wir sahen*, *sie sahen*). In English, there is no person distinction in the past at all. In Hungarian, there is a special ending for each person. The researcher must decide himself how to solve the polyvalency.

The situation is even more complex in translation – translating prose, one can translate the text exactly; in poetry, one presents the idea, not the text. Besides these issues, it also needs to be said that finding links in a text is always at least partly a subjective matter.

Nevertheless, one must introduce the concept of Belza-chains also in poetry and study its existence. But here, the unit of the sequence is not the sentence, but the line (verse). Since the line is short, it may happen that the same concept is not present in the neighbouring line, but in the following; in that case, we have no chain. But it may happen that the same concept occurs in almost all lines – e.g., in Shakespeare's *Sonnet Nr. 66*:

Tir'd with all these, for restful death I cry,
As, to behold desert a beggar born,
And needy nothing trimm'd in jollity,
And purest faith unhappily forsworn,
And guilded honour shamefully misplaced,
And maiden virtue rudely strumpeted,
And right perfection wrongfully disgraced,
And strength by limping sway disabled,
And art made tongue-tied by authority,
And folly (doctor-like) controlling skill,
And simple truth miscall'd simplicity,
And captive good attending captain ill:

Tired with all these, from these would I be gone,
Save that, to die, I leave my love alone.

Here, one finds, e.g., the word "and" in 10 subsequent lines. Further, in a chain, another chain may be placed, or begin, or be finished. A line, not having any concept common with the subsequent (or the preceding) line, has the Belza-length of 1.

One should take into account grammatical morphemes only if they directly refer to a noun, verb, or adjective, etc.

Since sonnets have the same length (14 lines), the sum of Belza-chain lengths, or their averages, or parameters of the fitted functions are measures of ***association***. As an indicator of this property, we shall use the sum of the chain-length divided by their number, i.e.

$$A_s = \frac{\sum CL}{n_{ch}}, \tag{3.1}$$

where n_{ch} is the number of chains, and CL the length of chains. The number of chains can be minimally 1, but the maximum cannot be stated because there can be several chains of different lengths. In general, the greater A_s is, the stronger the semantic associativity of the text is. As a matter of fact, A_s is the mean of lengths.

Consider, for example, *Krvavé sonety* by Hviezdoslav, in which we find the Belza chains as given in Table 3.1.

Table 3.1
Lengths of Belza-chains in *Krvavé sonety* by Hviezdoslav

Sonnet	Chain Length	Association
1	2,1,2,1,1,1,1,5,1	1.67
2	1,4,1,1,7	2.80
3	2,2,2,1,3,1,1,1,1,1,1	1.45
4	1,1,1,1,1,1,1,1,1,1,1,1,1,1	1.00
5	1,1,1,1,1,1,1,1,2,3,2	1.36
6	1,1,1,1,2,1,1,1,1,1,1,1,1	1.08
7	2,1,1,1,2,1,1,2,1,1,1	1.27
8	1,1,1,1,1,1,1,1,5,1	1.40
9	3,1,5,1,4,1	2.50
10	3,2,1,2,1,2,1,1,1	1.56
11	7,1,1,1,1,2,2	2.14
12	1,1,1,2,1,1,2,2,2,1,1	1.36
13	1,1,3,1,1,2,1,2	1.50
14	3,1,2,1,1,2,1,1,1,2	1.50
15	5,1,1,1,2,2,2	2.00

16	2,2,3,2,1,1,1,1,2	1.67
17	1,1,2,1,2,1,2,1,1,1,1	1.27
18	1,1,1,1,1,2,1,1,2,1,1,1	1.17
19	1,1,1,1,1,1,1,1,1,1,1,2,1	1.08
20	2,2,1,1,1,1,1,1,1,1,2,1	1.25
21	3,1,1,1,1,3,1,1,2	1.56
22	1,1,1,1,1,1,1,1,1,2,1,1,1	1.08
23	1,1,1,1,1,1,3,3,1,1	1.40
24	1,3,1,1,2,1,1,2,2,3	1.70
25	1,1,1,1,1,1,1,1,1,2,2,2	1.25
26	1,1,1,1,1,2,1,1,1,1,3	1.27
27	2,1,1,2,1,1,4,2,	1.75
28	2,2,1,3,3,1,2	2.00
29	5,2,1,2,3,3	2.67
30	3,1,3,1,4,1,1	2.00
31	3,1,1,1,2,1,1,1,1,1,1	1.27
32	2,1,1,1,1,2,3,1,2	1.56

The collection may be characterized in many various ways. One can compute the mean of associativities, one can find a function expressing the sequence of associativities, and consider one of its parameters, etc. We shall characterize it by a function, but first, we must subdivide the associations in intervals. We observe numbers from 1 to 3. Having subdivided it into intervals <1.0, 1.5), <1.5, 2.0), <3.0, 2.5), and <2.5, 3.0>, and taken the mean of each interval into account, we fit the exponential function, as shown in Table 3.2. Again, we obtained a simple function. Unfortunately, we have merely 32 sonnets, hence a more detailed subdivision into intervals would not be recommended.

Table 3.2
Associations in *Krvavé sonety* by Hviezdoslav

Interval	**Interval Mean**	**Frequency**	**Exponential ft.**
<1, 1.5)	1.25	16	16.05
<1.5, 2.0)	1.75	9	8.67
<2.0, 2.5)	2.25	4	4.68
<2.5, 3.0>	2.75	3	2.53
$a = 74.8433, b = 0.8120, R^2 = 0.9925$			

For *Sonety* by Hviezdoslav, we obtain the results displayed in Tables 3.3 and 3.4. As can be seen, there is an anomaly in *Sonnet Nr. 8*, where there is a very long chain consisting of repeated occurrences of the first person. However, the fit with added 1 to the exponential function is satisfactory.

Table 3.3
Length of Belza-chains in *Sonety* by Hviezdoslav

Sonnet	Chain Lengths	Association
1	4,1,1,2,2,1,1,6,2	2.22
2	1,1,1,1,3,3,3,2	1.88
3	2,1,1,1,1,1,1,1,2,3	1.40
4	2,1,2,3,1,1,1,1,1,1	1.40
5	1,1,2,4,6	2.80
6	1,1,1,1,2,1,1,3,1,2	1.40
7	4,3,2,3,1,1,2,1,2,3	2.20
8	10,4,1,2	4.25
9	1,1,1,2,1,1,1,1,2,1,3,2	1.42
10	1,3,3,1,1,2,3	2.00
11	1,1,2,1,3,1,1,1,1,2	1.40
12	2,1,1,1,1,1,1,2,3,1,1	1.36
13	2,1,1,1,1,1,1,2,1,2	1.30
14	2,1,1,1,1,1,1,1,1,1,2,1	1.17
15	1,1,1,1,2,1,1,1,2,2,1,1	1.25
16	2,1,1,5,1,1,1,1,1	1.56
17	4,2,3,3,1,2	2.50
18	1,1,2,2,1,1,2,1,1,2	1.40
19	2,3,2,2,1,1,4,2	2.12
20	2,2,2,2,2,1,1,1,1,2	1.60
21	2,7,1,1,1,2,1	2.14

Table 3.4
Belza-chain associations in S*onety* by Hviezdoslav

Interval	Interval Mean	Frequency	Exponential ft. + 1
<1.0, 1.5)	1.25	10	9.52
<1.5, 2.0)	1.75	3	5.03
<2.0, 2.5)	2.25	5	2.91
<2.5, 3.0)	2.75	2	1.90
<3.0, 3.5)	3.25	–	–
<3.5, 4.0)	3.75	–	–
<4.0, 4.5)	4.25	1	1.10
a = 55.2333, b = 1.4957, R^2 = 0.83			

Table 3.5
Lengths of Belza-chains in the sonnets by Goethe

Sonnet	Chain Lengths	Association
Das Sonett	1,3,2,1,2,1,1,1,1,3	1.60
	2,2,1,1,1,1,1,2,1,1,1	1.27
Mächtiges Überraschen	2,2,3,1,1,2,1,1,1,1	1.50
Freundliches Begegnen	1,1,1,1,1,1,1,7	1.75
Kurz und gut	6,3,1,1,1,1,1,2	2.00
Das Mädchen spricht	4,2,1,1,2,2,3,1,2,2	2.00
Wachstum	1,1,1,1,2,2,1,5,3	1.89
Reisezehrung	2,2,1,5,2,1,1,4	2.25
Abschied	2,1,1,1,1,1,1,1,2,1,2	1.27
Die Liebende schreibt	2,3,2,2,2,1,1,1,3,2	1.90
Die Liebende abermals	1,3,2,1,1,1,1,2,3	1.67
Sie kann nicht enden	2,2,2,1,1,2,6,6	2.75
Nemesis	1,1,2,3,2,1,1,1,2,1	1.50
Christgeschenk	1,1,1,1,1,1,1,1,2,4	1.40
Warnung	1,1,2,3,4,1,2,1	1.88
Die Zweifelnden	1,2,1,4,1,1,1,1,1,1	1.40
Mädchen	4,3,1,1,1,1,1,1,1	1.56
Epoche	2,4,1,1,2,1,3	2.00
Charade	2,2,1,1,1,1,3,3	1.75

Table 3.6
Belza-chain associations in the sonnets by Goethe

Interval	Interval Mean	Frequency	Zipf-Alekseev ft.
<1.0, 1.5)	1.25	4	4.02
<1.5, 2.0)	1.75	10	9.97
<2.0, 2.5)	2.25	4	4.10
<2.5, 3.0)	2.75	1	0.77
a = 10.9949, b = -10.5989, c = 0.5860 R^2 = 0.9985			

As can be seen, the frequencies in Goethe are bell-shaped, hence a monotonically decreasing function would not be adequate. We tested the generalization of the exponential distribution, namely the Zipf-Alekseev function, and obtained very satisfactory results. The sonnets by Goethe express a stronger Belza-chaining than those by Hviezdoslav. It must be further tested whether this is a property of a language, or of an author. Adding 1 to the function does not give better results.

For the Czech sonnets by Machar, we obtain the results presented in Table 3.7.

Table 3.7
Lengths of Belza-chains in *Letní sonety* by Machar

Sonnet	Chain Length	Association
E. Zolovi	8,1,1,1,2,1	2.33
Matce	2,2,2,5,1,4	2.67
Sonet cynický	1,1,2,2,1,1,3,1,1,1	1.40
Sonet de vanitate	3,1,1,1,2,3,2,1,2,2	1.80
Sonet elegický	2,1,1,1,1,2,2,3,2	1.67
Sonet ironický	2,1,1,3,2,1,2,1,3	1.78
Sonet k sociální otázce	1,1,1,1,1,1,2,1,2,2,2,1	1.33
Sonet k teorii; Boj o život	5,2,2,4,2,2	2.83
Sonet materialistický	2,2,4,1,4,1	2.33
Sonet mystický	2,2,1,1,1,1,5,1	1.75
Sonet na Chopinovu melodii	2,2,2,1,1,2,3,1,2,2,2	1.82
Sonet na sentenci z Goetha	2,4,2,1,1,1,2,2,3	1.78
Sonet na sklonku století	2,2,2,3,1,1,1,1,2	1.67
Sonet nad verši z mládí	1,2,2,3,3,3,2,2,2	2.22
Sonet noční	2,1,1,2,2,1,1,1,1,2	1.40
Sonet o antice a vlasech	2,3,2,2,2,1,1,1,2,2	1.80
Sonet o bídě	1,2,5,2,1,2,2,2	2.13
Sonet o hodinách	2,2,2,1,3,2,1,5	2.25
Sonet o lásce	1,1,1,1,1,1,1,1,1,3,1,1	1.17
Sonet o minulosti	1,1,2,2,5,1,1,1,1	1.67
Sonet o Panně Marii	1,1,2,2,2,3,2,2,3	2.00
Sonet o rokoku	1,3,2,2,2,2,3,1,2	2.00
Sonet o staré metafoře	2,2,2,3,1,3,1,2	2.00
Sonet o starém líci a rubu	2,8,3,3,1,1,3,3	3.00
Sonet o třech metaforách	1,1,1,1,2,2,2,1,1,1,1	1.27
Sonet o třetí hodině v červenci	2,2,1,2,1,1,4,2,1	1.78
Sonet o vídeňských kosech	3,5,1,4,2,2,1	2.57
Sonet o západu slunce	3,2,1,1,2,1,2,2	1.75
Sonet o zlatém věku naší poezie	1,1,1,1,2,1,1,2,1,3	1.40
Sonet o životě	2,8,2,2,2	3.20
Sonet patologický	1,1,1,1,2,4,1,1,1,1,1,1	1.33
Sonet polední	1,1,1,1,1,2,1,1,2,2,1	1.27
Sonet sarkastický	3,1,3,4,2,1,2,1	2.13
Sonet svatební	2,12,3	5.67
Sonet úvodní	2,1,1,2,2,1,2,2,2,2,1	1.64
Sonet večerní	2,1,1,1,1,1,1,1,1,1,1,1,1	1.08
Sonet z dvacátého září	1,1,2,2,2,2,1,1,1,3	1.60
Sonet-apostrofa	1,3,2,1,1,1,3,2,1	1.67
Sonet-epilog čtenáři	1,3,1,1,1,1,1,2,1,2	1.40

Sonet-intermezzo$_2$	2,6,2,2,2	2.80
Sonet-intermezzo	1,1,2,11	3.75
Sonety-Causerie I.	3,1,4,2,2,2,2,1	1.88
Sonety-Causerie II.	2,1,1,2,2,9,2,2	2.63
Sonety-Causerie III.	1,4,2,2,1,1,1,1,1,1	1.50
Sonety-Causerie IV.	2,4,2,1,1,1,1,3,1,2	1.80
Sonety-Causerie V.	1,1,1,3,1,1,2,2,1,1	1.40
Své ženě s předešlým sonetem	3,6,2,3,3,1	3.00

The association may be captured by the usual Lorentzian function, as shown in Table 3.8.

Table 3.8
Associations in *Letní sonety* by Machar

Interval Mean	**Frequency**	**Lorentzian ft.**
1.25	11	11.01
1.75	19	18.99
2.25	7	7.10
2.75	5	2.82
3.25	1	1.44
3.75	1	2.82
5.25	1	0.30
a = 19.9234, b = 1.6510, c = 0.4458, R^2 = 0.9675		

Table 3.9
Belza-chains in Hungarian sonnets

Text	**Belza-Chains**	**Associativity**
Babits, M.	5,2,7,2	4.00
Faludi, F.	4,8,3,1,1,2	3.00
Jószef, A.: 1	1,1,2,3,2,5,1	2.14
Jószef, A.: 2	2,1,1,1,1,3,3,1,1,1	1.50
Jószef, A.: 3	3.1,2,1,1,3,2,1,1	1.67
Jószef, A.: 4	2,2,1,1,1,1,3,1,1,1	1.40
Jószef, A.: 5	1,1,1,1,1,2,2,1,1,1,1,1	1.16

Since we have always 14 lines, one can, of course, consider only the length of Belza chains and their distribution. For the lengths in *Sonette* by Goethe, we obtain the results presented in Table 3.10.

Table 3.10
Lengths of Belza-chains in *Sonette* by Goethe

Length	**Frequency**	**Exp. + 1**
1	101	101.49
2	44	41.90
3	16	17.64
4	7	7.77
5	2	3.76
6	3	2.12
7	1	1.46
$a = 246.9367$, $b = 0.8991$, $R^2 = 0.9985$		

One can consider a writer, a part of his production, a language, etc. For some other sonnets, we get the results presented in Table 3.11.

Table 3.11
Some length measurements of Belza chains

Length	**Hviezdoslav: *Krvavé sonety***	**Exp+1**	**Hviezdoslav: *Sonety***	**Exp+1**	**József: 1–5**	**Exp+1**
1	217	216.81	111	111.92	32	31.81
2	62	63.40	50	46.43	9	10.30
3	22	19.05	18	19.61	6	3.81
4	4	6.22	6	8.62	–	–
5	5	2.51	1	4.12	1	1.26
6	–	–	2	2.28		
7	2	1.13	1	1.52		
8			–	–		
9			–	–		
10			1	1.04		
	$a = 746.2981$, $b = 1.2407$, $R^2 = 0.9994$		$a = 270.8093$, $b = 0.8926$, $R^2 = 0.9969$		$a = 102.0863$, $b = 1.1981$, $R^2 = 0.9883$	

Since here, we have the same regularity, we can suppose that the semantic construction of sonnets is regulated in the same way as parts of speech. There may be, of course, many exceptions, but on the whole, Belza-chains are supra-line constructs, and even if the text experiences posterior changes, the whole adheres to a special structuring.

3.2 Belza-chain Motifs

Motifs show the non-decreasing chain length, and one can conjecture that the shorter the motifs, the stronger the chaining is in the given sonnet. Let us consider the motifs of Belza chains in *Krvavé sonety* by Hviezdoslav. Again, we obtain three types of distributions: the first concerns the types ranked according to their frequency; the second is the distribution of chain lengths; and finally, one could study the sum of element lengths in the individual motifs. For example, the motif [1,1,2] has the sum 4, the motif [1,2,3,5] has the sum 11, etc.

First, we look at the types and their frequencies. In Table 3.12, we can see that the number of sonnets is too small in order to yield a smooth distribution. But if we consider the lengths of individual types, we obtain the results presented in Table 3.13. Evidently, the exponential function with added 1 satisfactorily expresses the existing trend.

Table 3.12
Belza-chains motifs in *Krvave sonety* by Hviezdoslav

Type	Numb.	Type	Numb.	Type	Numb.
[1]	5	[1,1,4]	1	[1,1,1,1,1,2]	2
[2]	9	[1,1,7]	1	[1,1,1,1,2,2]	1
[3]	6	[1,3,3]	1	[1,1,1,1,2,3]	1
[5]	2	[2,2,2]	1	[1,1,1,2,2,2]	1
[7]	1	[2,2,3]	1	[1,1,1,1,1,1,1,1]	1
[1,1]	3	[1,1,1,1]	1	[1,1,1,1,1,1,3,3]	1
[1,2]	9	[1,1,1,2]	4	[1,1,1,1,1,1,1,1,2]	1
[1,3]	3	[1,2,3,3]	1	[1,1,1,1,1,1,1,1,5]	1
[1,4]	3	[1,1,1,1,2]	2	[1,1,1,1,1,1,1,1,1,2]	1
[1,5]	1	[1,1,1,1,3]	2	[1,1,1,1,1,1,1,1,2,3]	1
[2,2]	2	[1,1,1,1,5]	1	[1,1,1,1,1,1,1,1,1,1,1,2]	1
[1,1,1]	4	[1,1,2,2,2]	1	[1,1,1,1,1,1,1,1,1,2,2,2}	1
[1,1,2]	8	[1,1,2,2,3]	1	[1,1,1,1,1,1,1,1,1,1,1,1,1,1]	1
[1,1,3]	1	[1,1,1,1,1,1]	2		

Table 3.13
Lengths of Belza-chain motifs in *Krvave sonety* by Hviezdoslav

Length	Number	Exponential ft. +1
1	23	25.20
2	21	18.55
3	18	13.73
4	6	10.23
5	7	7.69
6	7	5.85

8	2	3.55
9	2	2.85
10	2	2.34
12	2	1.71
14	1	1.37
a = 33.3778, b = 0.3214, R^2 = 0.9245		

In the following tables, the same is carried out for Machar's sonnets.

Table 3.14
Belza-chains motifs in *Letní sonety* by Machar

Type	**Freq.**	**Type**	**Freq.**	**Type**	**Freq.**
[2]	18	[2,4]	2	[1,1,4]	1
[1]	13	[5]	2	[1,2,2,2,2]	1
[2,2,2]	8	[1,1,1,1,1,1,1,1,1,1,1]	1	[1,2,2,3,3,3]	1
[1,1,2,2]	6	[1,1,1,1,1,1,1,1,1,3]	1	[1,2,5]	1
[2,2]	6	[1,1,1,1,1,1,2]	1	[1,3,4]	1
[1,3]	5	[1,1,1,1,2,2,3]	1	[1,5]	1
[1,4]	5	[1,1,1,1,2,4]	1	[2,1]	1
[3]	5	[1,1,1,1,3]	1	[2,12]	1
[1,2]	4	[1,1,1,1,5]	1	[2,2,2,2]	1
[1,1,1,1,2]	3	[1,1,1,1]	1	[2,2,2,3]	1
[1,1,1,3]	3	[1,1,1,2,2,3]	1	[2,3,3]	1
[1,1,1,1,1,1]	2	[1,1,1,2,2]	1	[2,3]	1
[1,1,1,1,1,2]	2	[1,1,1,2,3]	1	[2,6]	1
[1,1,2]	2	[1,1,1,2]	1	[2,8]	1
[1,1,3]	2	[1,1,1]	1	[3,5]	1
[1,1]	2	[1,1,2,11]	1	[3,6]	1
[1,2,2,2]	2	[1,1,2,2,2,2]	1	[8]	1
[1,2,2]	2	[1,1,2,2,5]	1	[9]	1
[2,2,4]	2	[1,1,2,3]	1		

Table 3.15
Lengths of Belza-chain motifs in *Letní sonety* by Machar

Lengths of Belza chains	**Frequ.**	**Expon. ft.**		**Lengths of Belza chains**	**Frequ.**	**Expon. ft.**
1	44	46.97		6	7	8.46
2	38	32.96		7	3	6.19
3	23	23.21		10	1	2.74
4	17	16.44		12	1	1.84
5	12	11.73				
				a = 66.1347, b = 0.3636, R^2 = 0.9751		

4. Hreb Analysis

4.1 Generalities

A hreb – called after its father, L. Hřebíček (1992) – is originally the collection of all sentences containing some commonality. But since in a sentence there may occur more words/lines containing the same concept, one usually collects all such words and creates stepwise all hrebs of the text (cf. Ziegler, Altmann 2002). Hence, a hreb is a set of morphemes/words having the same meaning, the same reference, etc. Hrebs can be constructed from various aspects. Practically, the Belza-chains having a commonality, but separated by one or two lines located between them can be considered a kind of hreb. On the other hand, a hreb may contain also elements occurring in the same line. Hrebs have various properties, e.g. their number in a sonnet and their distribution in the work of an author, cardinal number (= length = a number of elements in each hreb), diffuseness measured in form of distances between the hreb elements, etc. If one applies the hreb analysis to sonnets, one should always be used.

Even if one performs a word count (tokens) in a sonnet, one must put the hrebs – consisting here of morphemes – together by individual analysis. The same as in Belza-chains, one puts together both the main word and all grammatical affixes or pronouns referring to it (cf. Ziegler, Altmann 2002). But a word whose affix relates to another word, too, may belong itself at the same time to another hreb. For example, in German, the verb *ist* ("is") belongs to the hreb whose main word is *sein* ("to be") and to the noun to which it refers. Such an analysis must be performed – as a matter of fact – word-by-word for the complete sonnet.

The study of hrebs provides a deeper insight into the semantic coherence of a sonnet: it concerns the complete sonnet and takes into account all semantic relations (identities). Needless to say, in strongly analytic languages, there will be a smaller semantic coherence, but this must be proved by treating concrete languages.

Let us illustrate the procedure using a sonnet by Goethe in German. The principle is: all morphemes must belong to a hreb. Sometimes, the set of hrebs is very extensive, but their number shows, at the same time, the semantic coherence of the poem. The smaller the number of hrebs, the stronger the coherence is. One begins with words, and searches for identities or references.

Goethe: Sonett

1 Sich in erneutem Kunstgebrauch zu üben,
2 Ist heilge Pflicht, die wir dir auferlegen.
3 Du kannst dich auch, wie wir, bestimmt bewegen
4 Nach Tritt und Schritt, wie es dir vorgeschrieben.

5 Denn eben die Beschränkung läßt sich lieben,
6 Wenn sich die Geister gar gewaltig regen;
7 Und wie sie sich denn auch gebärden mögen,
8 Das Werk zuletzt ist doch vollendet blieben.

9 So möcht ich selbst in künstlichen Sonetten,
10 In sprachgewandter Mühe kühnem Stolze,
11 Das Beste, was Gefühl mir gäbe, reimen;

12 Nur weiß ich hier mich nicht bequem zu betten.
13 Ich schneide sonst so gern aus ganzem Holze,
14 Und müßte nun doch auch mitunter leimen.

The first word, „sich“, occurs in four lines; we write the lines next to the word: {sich 1, sich 5, sich 6, sich 7}. But it is merely part of the lexeme “sich zu üben”, which has a reference in line 2 in “ist”: hence {sich zu üben 1, ist 2}; “ist” itself occurs in line 8, hence {ist 2, ist 8}. The next word is “in”, occurring in lines 1, 9 and 10, hence {in 1, in 9, in 10}; but the German “in” may evoke a dative, hence {in 1 – (erneut)-*em* 1; in 9 – (künstlich)-*en* 9; in 10 – (kühn)-*em* 10}. Continuing this way, we obtain a long list of hrebs, and can characterize the sonnet.

Table 4.1
The list of hrebs in Goethe’s *Sonett*

Hrebs
{auch 3, auch 7, auch 14}
{auferlegen 2, wir 2}
{aus 13, (ganz)-em 13, (Holz)-e 13}
{bequem 12}
{Beschränkung 5, (läß)-t 5}
{Beste 11, das 11, was 11}
{bestimmt 3}
{betten 12}
{bewegen 3 }
{blieben 8}
{denn 5, denn 7}
{die 2, die 5, die 6, das 8, das 11}
{dir 2, du 3, (kann)-st 3, dich 3, dir 4}
{doch 14}
{eben 5}

{erneut 1}
{es 4]
{gäbe 11}
{ganzem 13}
{gar 6}
{gebärden 7}
{Gefühl 11, (gäb)-e 11}
{Geister 6, (d)ie 6, (reg)-en 6, sie 7, (mög)-en 7}
{gern 13}
{gewaltig 6]
{heilig 2}
{hier 12}
{Holze 13}
{ich 9, (möcht)- 9, mir 11, ich 12, mich 12, (weiß)- 12, ich 13, (schneid)-e 13, (müßt)-e 14}
{in 1, (erneut)-em 1, in 9, (künstlich)-en 9, in 10, (kühn)-em 10}
{ist 2, ist 8}.
{kühnem 10}
{Kunstgebrauch 1, erneutem 1}
{künstlichen 9}
{läßt 5}
{leimen 14}
{lieben 5}
{mitunter 14}
{mögen 7, möcht 9}
{Mühe 10, (sprachgewandt)-er 10}
{müßte 14}
{nach 4}
{nicht 12]
{nun 14}
{nur 12}
{Pflicht 2, heilge 2, ist 2, die 2}
{reimen 11}
{schneide 13}
{Schritt 4}
{selbst 9}

{sich 1, sich 5, sich 6, sich 7}
{sich zu üben 1, ist 2}
{so 9, so 13}
{Sonetten 9, (künstlich)-en 9}
{sonst 13}
{sprachgewandter 10}
{Stolze 10, (kühn)-em 10}
{Tritt 4}
{und 4, und 7, und 14}
{vollenden 8}
{vorschreiben 4}
{weiß 12}
{wenn 6}
{Werk 8, das 8, ist 8, vollendet 8}
{wie 3, wie 4, wie 7}
{wir 2, wir 3}
{zu 12}
{zuletzt 8}

As can be seen, the analysis can be performed in many different ways. Here, for example, we also took into account zero morphemes like *möcht.* But whatever the analysis, the result has its form. First of all, we found H = 68 hrebs. In order to present an indicator, we recommend the formula

$$C = \frac{H}{14}, \tag{4.1}$$

which is unique for all languages while one considers word types or tokens would be different. Hence we obtain the semantic concentration expressed by hrebs in form

$$C = \frac{68}{14} = 4.86\,.$$

Further, the hrebs have a certain cardinal number, i.e., the number of elements in them. If we count the ones in the sonnet by Goethe, we obtain the results presented in Table 4.2. This seems to be a quite usual distribution. Ziegler and Altmann (2002) captured it either by the Waring or by the Zipf-Alekseev distribution. In order to simplify the problem, we use here the exponential function, which yields good results.

Table 4.2
Hreb sizes in the *Sonette* by Goethe

Cardinal Number of Hrebs	Frequency	Computed Exponential
1	42	41.89
2	13	13.52
3	5	4.83
4	3	2.17
5	3	1.36
6	1	1.11
9	1	1.00
a = 133.5814, b = 1.1837, R^2 = 0.9972		

The properties, or the parameters of this function could be used for comparisons. One can set up hypotheses concerning the type of language, the writer, the creation time, etc. For the sake of exemplification, we show also the first sonnet by József (Hungarian) in Table 4.3.

Table 4.3
Hreb sizes in József's *S 1*

Cardinal Number of Hrebs	Frequency	Computed Exponential
1	44	43.95
2	11	11.25
3	3	3.44
4	4	1.58
5	2	1.14
a = 180.0308, b = 1.4331, R^2 = 0.9946		

4.2 Ordered Positional Hrebs

One of the possibilities to quantifiy the properties of hrebs is their diffuseness**.** Hrebs containing only one element are not diffused, and thus need not be taken into account. For a quantification, we need to know the place of the hreb members. To this end, all the words of the whole sonnet must be numerated, and instead of morphemes/sememes, etc., one obtains an **ordered positional hreb.** For the sake of exemplification, we present the above sonnet and add to each word its position number.

Sich(1) in(2) erneutem(3) Kunstgebrauch(4) zu(5) üben(6),
Ist(7) heilge(8) Pflicht(9), die(10) wir(11) dir(12) auferlegen(13).

Du(14) kannst(15) dich(16) auch(17), wie(18) wir(19), bestimmt(20) bewegen(21)
Nach(22) Tritt(23) und(24) Schritt(25),wie(26) es(27) dir(28) vorgeschrieben(29).

Denn(30) eben(31) die(32) Beschränkung(33) läßt(34) sich(35) lieben(36),
Wenn(37) sich(38) die(39) Geister(40) gar(41) gewaltig(42) regen(43);
Und(44) wie(45) sie(46) sich(47) denn(48) auch(49) gebärden(50) mögen(51),
Das(52) Werk(53) zuletzt(54) ist(55) doch(56) vollendet(57) blieben(58).

So(59) möchte(60) ich(61) selbst(62) in(63) künstlichen(63) Sonetten(65),
In(66) sprachgewandter(67) Mühe(68) kühnem(69) Stolze(70),
Das(71) Beste(72), was(73) Gefühl(74) mir(75) gäbe(76), reimen(77);

Nur(78) weiß(79) ich(80) hier(81) mich(82) nicht(83) bequem(84) zu(85) betten(86).
Ich(87) schneide(88) sonst(89) so(90) gern(91) aus(92) ganzem(93) Holze(94),
Und(95) müßte(96) nun(97) doch(98) auch(99) mitunter(100) leimen(101).

Considering the hrebs whose cardinal number is at least 2, we obtain the sets listed in the following table.

Table 4.4
The list of ordered positional hrebs in Goethe's *Sonett*

Ordered Positional Hrebs
{auch 17, auch 49, auch 99}
{auferlegen 13, wir 11}
{aus 92, (ganz)-em 93, (Holz)-e 94}
{Beschränkung 33, (läß)-t 34}
{Beste 72, das 71, was 73}
{denn 30, denn 48}
{die 10, die 32, die 39, das 52, das 71}
{dir 12, du 14, (kann)-st 15, dich 16, dir 28}
{Gefühl 74, (gäb)-e 76}
{Geister 49, (d)ie 39, (reg)-en 43, sie 46, (mög)-en 51}
{ich 61, (möcht)-e 69, mir 75, ich 80, mich 82, (weiß)- 79, ich 80, (schneid)-e 88, (müßt)-e 96}
{in 2, (erneut)-em 3, in 63, (künstlich)-en 64, in 66, (kühn)-em 69}
{ist 7, ist 55}
{Kunstgebrauch 4 erneutem 3}
{mögen 51, möcht 60}
{Mühe 68, (sprachgewandt)-er 67}

{Pflicht 9, heilge 8, ist 7, die 10}
{sich 1, sich 35, sich 38, sich 47}
{sich zu üben 6, ist 7}
{so 59, so 90}
{Sonetten 65, (künstlich)-en 64}
{Stolze 70, (kühn)-em 69}
{und 24, und 44 und 95}
{Werk 53, das 52, ist 55 vollendet 57}
{wie 18, wie 26, wie 45}
{wir 11 wir 19}

Replacing the line number by the position and considering the first position as *inf(H_p),* the last position as *sup(H_p)* and the cardinal number of the given set as |*H*|, we obtain an indicator (cf. Ziegler, Altmann 2002: 55)

$$D_H = \frac{\sup(H_p) - \inf(H_p)}{|\mathrm{H}|}. \tag{4.2}$$

For example, for the hreb {auch} we obtain:

$$D_{\{auch\}} = \frac{99 - 17}{3} = 27.33\,.$$

For the other hrebs ordered according to increasing D_H, we obtain the results presented in Table 4.5.

Table 4.5
Diffuseness of hrebs in the *Sonett* by Goethe

Hreb	**D_H**	**Hreb**	**D_H**
Beschränkung	0.5	du	3.2
Kunstgebrauch	0.5	wir	4.0
Mühe	0.5	mögen	4.5
sich üben	0.5	den	9.0
Sonett	0.5	wie	9.0
Stolz	0.5	ich	10.0
Beste	0.67	in	11.17
aus	0.67	sich	11.50
Pflicht	0.75	der/die/das	12.2
auferlegen	1.0	so	15.5
Gefühl	1.0	und	23.67
Werk	1.25	ist	24.0
Geist	1.6	auch	27.33

As can be seen, the nouns and verbs manifest a small diffuseness (the first column), the other POS have a greater diffuseness (the second column). The only exception is the preposition *aus*. Hence, we can state that main POS (nouns, verbs, adjectives) have a smaller diffuseness than the other POS.

The total diffuseness of the sonnet can be computed as the mean of individual values, here:

$$D_M = \frac{175.01}{26} = 6.73\,.$$

Only a comparison with other sonnets can tell us whether this value is high or low. One can use the simple normal test. In any case, many investigations in many languages are necessary in order to formulate some hypotheses.

4.3 Text Compactness

The text is the more compact, the smaller the number of hrebs is. In this case, the content of it is concentrated to a few concepts. In other words, the higher the cardinal numbers of hrebs are, the more compact the text is. Since the sonnets have approximately the same length also in terms of word numbers, it is possible to compare them directly, i.e., without testing. It can be remarked that *Dada* texts, in which there is an exaggerated repetition of the same "word", have a very high compactness.

Let *n* be the number of hrebs in a sonnet, and *L* be the number of words in the given sonnet. Then Ziegler and Altmann (2002: 60) propose to compute the relative text compactness as

$$C = \frac{1 - \frac{n}{L}}{1 - \frac{1}{L}}, \tag{4.3}$$

since the minimal number of hrebs is 1. The above indicator moves in the interval <0, 1>, and can also be considered a proportion. In the above *Sonett* by Goethe, there are n = 68 hrebs and L = 101 words. Hence,

$$C = \frac{1 - \frac{68}{101}}{1 - \frac{1}{101}} = 0.33\,.$$

As to other sonnets, József's *S 1* yields, for instance, the compactness value of 0.23.

5. Equalities

Whatever property of the sonnet is considered (phonetic, grammatical, semantic, Belza-chains, motifs, hrebs, etc.), the lines may be equal or express a degree of equality. One can compare the number of syllables, number of words, number of morphemes, types of sentences; one can set up semantic classes and compare the membership of words in them, motifs of various kind, etc. One can construct measures of similarity, set up tests or set up fuzzy sets, graphs of similarity, measure the distances between similar entities, etc. The field is infinite, and taking into account all properties scholars have defined up to now would require a new book.

Since in some domains, equality is sometimes intentional – e.g., rhythm, or semantics –, its degree can be quantified and measured. Either one defines an interval within which it will be measured, or a complete sonnet will be considered. Usually, this is the normal way because sonnets are short. In other texts, an exact definition must be performed. The domain is, *expressis verbis*, infinite (cf. Altmann, Köhler 2015) because everybody can define new units which must have several properties. The sonnet is very short in order to present views about writers or languages. First, a number of investigatons must be performed with other texts, many languages, etc., before the hypotheses may be corroborated.

6. Adnominals

6.1 General Measurements

The quantitative study of adnominals has been performed especially in the recent time (cf. Pan, Liu 2014; Andreev, Popescu, Altmann 2016, 2017; Andreev, Lupea, Altmann 2017). As to sonnets, one can study them in the work of an author, classify them, consider their distributions, and compare authors or languages. Adnominals include all means which can serve as descriptions, specifications of a given noun; simply, they are all components of a noun phrase. Many languages have the same set of adnominals, but there are also some special expressions in some languages absent in other ones. Besides, languages themselves may prefer some kinds of them, and omit other ones. Their study in sonnets and modelling their properties are possible only if one investigates a collection of sonnets by the same author, as individual sonnets are very short. An adnominal may have the following properties: (1) distribution of types; (2) length in terms of syllables or words; (3) position in front of, or behind the noun; (4) being a word, or a phrase, or a clause; (5) being a dynamic indicator of a poet's change in time; (6) being an indicator of ornamentality of the text. Further properties will be found with an increasing number of investigations. Here, we want to analyze some Czech and Russian authors only.

A classification of Russian adnominals can be found in Andreev, Fan, Altmann (2017). For the sake of simplicity, we present the list below:

A – adjective
AE – adjective in elliptical construction (*I have a green and a white hat*)
AP – apposition *(a foreigner, a man with a peculiar pronunciation*)
APNA – apposition without agreement (*Hotel "Bellevue*")
APX – apposition of a proper name (*the physicist Einstein*)
AV – adverb
AY – adjectival phrase (*the cheek white of fear)*
CN – compound (*a flower-pot*)
DAT – dative case (*a letter to a friend*)
DETF – demonstrative pronoun (*this house*)
DETH – indefinite pronoun *(some houses*)
DETN – negative pronoun (*no houses*)
DETQ – qualifying pronoun (*all books*)
DETS – possessive pronoun (*his friend*)
DETV – relative pronoun (*what a book is it*)
DETW – interrogative pronouns (*which book is missing?)*
G – genitive case (*father's house, laugh of devil*)
I – infinitive (*tendency to increase*)
INSTR – instrumental case (*way through the forest*)
PR – prepositional case (*the house on the hill*)

PT – participle (*broken glass, running dog*)
PTY – participial construction (*a glass broken to pieces*)
RC – subordinate clause (*a man who does not know fear*)

As to the Czech research, the list has been simplified, as some distinctions were considered too detailed. One thus distinguishes here:

A – adjective
AP – apposition
AV – adverb
CN – compound component
DAT – dative noun
DET – determiner
GEN – genitive noun
NUM – numeral
P – participle
PR – prepositional phrase
RC – relative clause

The distributions of individual types in Czech sonneteers are presented in Table 6.1.

Table 6.1
Types of adnominals in some Czech sonnets

	Machar	**Kollár**	**Karásek**	**Kvapil**	**Nezval**
A	78	76	119	72	33
ADV	2	0	0	0	0
AP	0	0	0	2	6
CN	0	0	0	1	0
DET	24	22	11	48	33
GEN	20	20	41	34	24
NUM	4	5	2	0	3
PART	1	0	2	8	0
PR	1	3	6	2	1
RC	8	9	25	13	6

Here, we consider merely the rank-frequency of types. The results are presented in Table 6.2. As can be seen in Table 6.1, Czech authors prefer some types of adnominals. The question remains whether this is usual in Czech, or it concerns only sonnets. Though all cases can be captured by the exponential function, one may expect that some sonnets will deviate not only because of the given language, but also due to the fact that the authors may perfom additional changes. Notably, Nezval's distribution scores less in the coefficient of determination, as it

idiosyncratically shows the same number of adjectives and determiners (Ranks 1 and 2).

Table 6.2
Rank-frequencies of adnominals in Czech sonnets

Rank	Machar	Exp. + 1	Kollár	Exp. + 1	Karásek	Exp. + 1
1	78	76.51	76	74.19	119	117.61
2	24	30.99	22	30.18	41	47.29
3	20	12.91	20	12.63	25	19.37
4	8	5.73	9	5.64	11	8.29
5	4	2.88	5	2.85	6	3.89
6	2	1.75	3	1.74	2	2.15
7	1	1.30			2	1.46
8	1	1.12				
	a = 190.1012 b = 0.9233 R^2 = 0.9974		a = 183.5669 b = 0.9195 R^2 = 0.9620		a = 291.7886 b = 0.9240 R^2 = 0.9919	

Rank	Kvapil	Exp. + 1	Nezval	Exp. + 1
1	72	74.51	33	38.04
2	48	45.02	33	25.33
3	34	27.36	24	16.98
4	13	16.78	6	11.50
5	8	10.45	3	7.90
6	2	6.66	1	5.53
7	2	4.39		
8	1	3.03		
	a = 122.7633 b = 0.5128 R^2 = 0.9770		a = 56.3886 b = 0.4202 R^2 = 0.8163	

For the sonnets written in Russian, we always added the types of adnominals in the works of single authors (cf. Sources and List of Abbreviations), and obtained the results presented in Table 6.3.

Table 6.3
Adnominals in Russian sonnets

Rank	T 1–2		T 3–4		T 9–11		T 12–13	
1	9	9,28	5	5.43	13	13.82	14	14.96
2	7	6.02	5	4.06	11	8.51	7	6.68
3	3	4.05	3	3.11	4	5.40	3	3.47
4	3	2.85	2	2.46	3	3.57	2	2.07
5	3	2.12	2	2.01	2	2.51	2	1.47
6	1	1.68	2	1.70	2	1.88	1	1.20

7	1	1.41	1	1.48	2	1.52	1	1.09
8					1	1.30	1	1.03
9					1	1.18		
10					1	1.10		
	a =13.6625, b = 0.5003, R^2 = 0.9352		a = 6.4211 b = 0.3792 R^2 = 0.8926		a = 21.9037 b = 0.5354 R^2 = 0.9423		a = 30.0345 b = 0.8328 R^2 = 0.9954	
Rank	**T 14–15**		**T 16–18**		**T 19–25**		**T 27–28**	
1	20	21.65	27	27.41	49	49.41	16	15.53
2	17	11.43	15	14.81	27	25.93	6	7.47
3	3	6.27	11	8.22	14	13.84	4	3.88
4	2	3.66	3	4.78	7	7.61	3	2.28
5	1	2.35	1	2.97	3	4.41	3	1.57
6	1	1.68	1	2.03	3	2.75	3	1.25
7	1	1.34	1	1.54	2	1.90	2	1.11
8					1	1.46	1	1.05
9					1	1.24	1	1.02
10					1	1.12	1	1.01
11					1	1.06	1	1.00
	a = 40.8582 b = 0.6826 R^2 = 0.8808		a = 50.5009 b = 0.6484 R^2 = 0.9723		a = 94.0136 b = 0.6636 R^2 = 0.9983		a = 32.6121 b = 0.8088 R^2 = 0.9538	
Rank	**T 30–31**		**T 32–33**		**T 36–37**		**T 39–41**	
1	8	8.57	19	19.06	11	11.59	19	20.61
2	7	5.24	12	12.19	9	6.96	17	12.63
3	3	3.38	9	7.94	3	4.35	6	7.89
4	3	2.33	7	5.30	2	2.89	4	5.08
5	1	1.75	3	3.66	2	2.06	4	3.42
6	1	1.42	1	2.65	2	1.60	1	2.43
7	1	1.23			2	1.34	1	1.85
8	1	1.13			1	1.19	1	1.50
9	1	1.07			1	1.11	1	1.30
10	1	1.04						
	a = 13.5029 b = 0.5793 R^2 = 0.8980		a = 29.1473 b = 0.4786 R^2 = 0.9711		a = 18.8305 b = 0.5752 R^2 = 0.9279		a = 33.0938 b = 0.5231 R^2 = 0.9247	

Evidently, the simple exponential function is sufficient for capturing the rankings. The ranks do not always concern the same adnominal. A separate fitting for individual sonnets and a table of all adnominals will be shown in an article by Andreev (2018).

In the same way, one could fit the ranking of any special classification of words. All of them show that the writer abides by some subconscious regularities, and our aim is to find at least some of them.

6.2 Ornamentality

Adnominals are of a huge potential when it comes to stylometric analysis of a text – they are, for instance, able to determine the degree with which an author decorates a poem. The usage of embellishment techniques can be measured in various ways; here, the count of ornamentality will follow the formula

$$\omega = \frac{A + P + G}{\sum Ad}, \tag{6.1}$$

in which *A* stands for the sample's number of adjectives, *P* for the participles, and *G* for the genitive constructions. Their addition is divided by the sum of all the adnominals.

Table 6.4
The degrees of ornamentality in selected Czech sonneteers

Author	Ornamentality
Kollár	0.51
Čelakovský	0.56
Chmelenský	0.38
Jan z Hvězdy	0.64
Mácha	0.51
Šolc	0.55
Sládek	0.33
Machar	0.45
Bouška	0.51
Kvapil	0.57
Karásek	0.57
Hlaváček	0.43
Mařík	0.36
Nezval	0.29
Seifert	0.28

6.3 Attributiveness

Sometimes, it is needed to assess whether an author employs more attributes (i.e., adjectives modifying nouns), or genitive constructions. This may be of use when the style changes in authors, periods, and movements are being investigated. To this end, the coefficient of attributiveness (T) has been introduced, the workings of which are analogical to Busemann's one (see 8.1). Its formula reads –

$$T = \frac{A}{A + G}, \tag{6.2}$$

with *A* standing for the number of adnominal adjectives, and *G* for the number of adnominal genitives. The text of T > 0.5 is considered attributive, the one of T < 0.5 genitival, and the one with T = 0.5 balanced.

Let us have an example. Čelakovský's first sonnet manifests the adjective-genitive sequence

G – G – A – G – A – G – A – G – A – A – G – A – A – A – A – A – A – A – A,

which comprises 13 adjectives and 6 genitives. This yields the value of T –

$$T = \frac{13}{13 + 6} = 0.67\,,$$

classifying Čelakovský's first sonnet as attributive.

To make the nomenclature subtler, a chi-square test may be introduced, working on the same basis as in the activity measurements (see 8.1). In this case, it is calculated as follows:

(6.3) $$\chi^2 = \frac{(A - G)^2}{A + G}\,,$$

with the meanings of the abbreviations being the same as in Formula 6.2.

As the critical value of the test is, given the circumstances, 3.84, texts may be divided into:

(1) AS – significantly attributive (T > 0.55, $\chi^2 > 3.84$);
(2) AT – attributive (T > 0.55, $\chi^2 < 3.84$);
(3) B – balanced (0.45 < T < 0.55);
(4) GE – genitival (T < 0.45, $\chi^2 < 3.84$);
(5) GS – significantly genitival (T < 0.45, $\chi^2 > 3.84$).

In the aforementioned example of Čelakovský's first sonnet, one obtains the result

$$\chi^2 = \frac{(13 - 6)^2}{13 + 6} = 2.58\,,$$

which places the poem among the attributive ones (AT).

Some results for selected Czech and Russian sonneteers are presented in Tables 6.5 and 6.6.

Table 6.5
Coefficients of attributiveness and chi-square classifications of selected Czech sonnets

Author	Poem	A	G	T	Type
Čelakovský	1	13	6	0.67	AT
	2	11	2	0.85	AS
	3	9	3	0.75	AT
	4	4	3	0.57	AT
	5	6	2	0.75	AT
	6	8	1	0.89	AS
	7	6	4	0.60	AT
	8	10	2	0.83	AS
Mácha	1	6	2	0.75	AT
	2	5	3	0.63	AT
	3	6	4	0.60	AT
	4	14	3	0.82	AS
	5	5	2	0.71	AT
	6	14	3	0.82	AS
	7	14	3	0.82	AS
Sládek	Znělky1	3	2	0.60	AT
	Znělky2	5	1	0.83	AT
	Znělky3	5	7	0.42	GE
	Znělky4	1	0	1.00	AT
	Znělky5	9	3	0.75	AT
	Znělky6	8	4	0.67	AT
	Znělky7	5	2	0.71	AT
	Znělky8	7	1	0.88	AS
	Znělky9	9	1	0.90	AS
	Znělky10	3	4	0.43	GE
Machar	Sonet úvodní	15	2	0.88	AS
	Sonet o životě	5	1	0.83	AT
	Sonet o třetí hodině v červenci	10	3	0.77	AT
	Sonet ironický	7	2	0.78	AT
	Sonet o staré metafoře	8	1	0.89	AS
	Sonet večerní	12	4	0.75	AS
	Sonet o minulosti	2	1	0.67	AT
	Sonet o zlatém věku naší poezie	7	3	0.70	AT
	Sonet-intermezzo	5	0	1.00	AS
	Sonet o hodinách	5	3	0.63	AT
Karásek	Horečka	7	5	0.58	AT

	Japonerie	7	5	0.58	AT
	Kalný západ	26	6	0.81	AS
	Narkózy	16	4	0.80	AS
	Nemoc	12	3	0.80	AS
	Noční sonet	9	5	0.64	AT
	Pokušení	6	5	0.55	B
	Příšerná loď	19	1	0.95	AS
	Rozklad	13	7	0.65	AT
Nezval	fetišista	3	1	0.75	AT
	měšťák	5	3	0.63	AT
	svedený	0	5	0.00	GS
	circulus vitiosus	9	1	0.90	AT
	básník	5	3	0.63	AT
	literatura	4	0	1.00	AS
	upír	2	5	0.29	GE
	její moc	1	2	0.33	GE
	sonet	2	0	1.00	AT
	rytmy	2	4	0.33	GE

Table 6.6
Coefficients of attributiveness and chi-square classifications of selected Russian sonnets

Author	**Poem**	**A**	**G**	**T**	**Type**
Trediakovskij	Sonet	2	1	0.67	AT
	Sonet iz seja grecheskija rechi	1	2	0.33	GE
Cheraskov	Sonet i jepitafija	4	2	0.67	AT
	Kol' budu v zhizni ja nakazan nishhetoju (…)	3	0	1.00	AT
Rzhevskij	Sonet, zakljuchajushhij v sebe tri mysli	1	1	0.50	B
	Sonet, tri raznye sistemy zakljuchajushhij	0	1	0.00	GE
Dmitriev	Sonet	2	1	0.67	AT
Zhukovskij	Sonet	3	2	0.60	AT
Deľvig	N. M. Jazykovu	8	1	0.89	AS
	Vdohnovenie	3	1	0.75	AT
	Ja plyl odin s prekrasnoju v gondole (…)	2	1	0.67	AT
Baratynskij	My p'jom v ljubvi otravu sladkuju (…)	6	5	0.55	B
	Hotja ty malyj molodoj (…)	8	2	0.80	AT
Jazykov	K. K. Janish	12	11	0.52	B
	Na prazdnik vash prines ja dva priveta (…)	9	6	0.60	AT
Pushkin	Sonet	6	6	0.50	B
	Pojetu	11	3	0.79	AS

	Madona	10	6	0.63	AT
Benediktov	Priroda	6	4	0.60	AT
	Kometa	11	4	0.73	AT
	Vulkan	3	5	0.38	GE
	Groza	7	5	0.58	AT
	Cvetok	7	6	0.54	B
	Krasavica, kak rajskoe viden'e (…)	10	1	0.91	AS
	Kogda vdali ot suety vsemirnoj (…)	6	2	0.75	AT
Sologub	Sonet	1	1	0.50	B
Brjusov	Sonet	4	1	0.80	AT
	Egipetskij rab	12	5	0.71	AT
Blok	Ne ty l' v moih mechtah, pevuchaja, proshla (…)	9	6	0.60	AT
Ivanov	Pritcha o devah	3	3	0.50	B
	Hramina chuda	5	4	0.56	AT
Voloshkin	Venok sonetov. Sonet 1	7	9	0.44	GE
	Venok sonetov. Sonet 2	5	10	0.33	GE
Severjanin	Sonet	3	1	0.75	AT
Belyj	Prosti	0	2	0.00	GE
Gumilev	Popugaj	3	8	0.27	GE
	Roza	8	3	0.73	AT
Esenin	Moej carevne	7	6	0.54	B
Bal'mont	Mikel' Andzhelo	3	3	0.50	B
	Leonardo da Vinci	4	10	0.29	GE
	Marlo	10	6	0.63	AT
Gorodeckij	Mudrost'	4	3	0.57	AT
Sel'vinskij	Sonet	2	0	1.00	AT
Prokoshin	Vozvraschenije	7	3	0.70	AT
Averina	Stremis' k mechte	4	1	0.80	AT
Beljaeva	Kto mozhet zapretit' tebja lubit'	7	1	0.88	AS

The results of individual authors may be presented in the form of pie charts, and provide grounds for various stylistic interpretations. The distributiuons of the types of Czech and the Russian sonnets are listed in Table 6.7.

Table 6.7
Types of coefficients of attributiveness in Czech and Russian

Czech Types	**Rank**	**Frequ.**	**Expon.**	**Russian Types**	**Rank**	**Frequ.**	**Expon.**
AT	1	30	30.77	AT	1	25	24.39
AS	2	17	14.17	B	2	9	11.55

GE GS B	3 4 5	5 1 1	6.53 3.01 1.38	GE AS	3 4	8 4	5.76 3.14
		a = 66.8075 b = 0.7754 R^2 = 0.9761				a = 51.8620 b = 0.7964 R^2 = 0.9509	

A comparison of Czech and Russian sonnets could be performed either by using the chi-square test, or performing a rank test, but we postpone this problem until much more sonnets of many languages have been analyzed.

We can merely state that the ranking of attributiveness is simply exponential.

6.4 Adnominal Runs

Runs are chains of a sequence that share the same feature. For instance, if one develops the adjective-genitive sequence of the sonnet *Portrét* ("A Portrait") by Jaroslav Kvapil, one gets the vector

$$A – A – G – A – G – A – A – A,$$

which can be divided into runs as follows –

$$\{A, A\} – \{G\} – \{A\} – \{G\} – \{A, A, A\};$$

it comprises 3 adjective runs, and 2 genitive ones; the total number of runs is 5.

If the number of runs in a poem is high, it may be concluded that its interior structure is swaying from adjectives to genitives, creating thus a complicated patchwork of interchanging trends; on the other hand, if the number is low, there are stretches of monotonous adjective or genitive chains.

As the number of runs depends on the overall number of adjectives and genitives in a text, another way may be devised for the evaluation of their structure – counting their average lengths. Given the example above, one can see that the runs are of the following lengths –

$$\{2, 1, 1, 1, 3\},$$

the average length being

$$\phi_R = \frac{2 + 1 + 1 + 1 + 3}{5} = 1.6\,.$$

The interpretation of the results follows the same path as in case of the number of runs. The same calculations may be effected for a collection of poems, or for all the author's corpus. Some figures of Czech sonnets are shown in Table 6.8.

Table 6.8
Number of runs and their average lengths in selected Czech sonnets

Author	Poem	Sequence	Runs	Run Lengths	Φ_R
Mácha	1	A,G,A,A,G,A,A,A	5	{1,1,2,1,3}	1.60
	2	G,A,A,G,A,A,G,A	6	{1,2,1,2,1,1}	1.33
	3	A,A,G,A,A,A,G,G,G,A	5	{2,1,3,3,1}	2.00
	4	A,G,A,A,A,G,A,A,A,A,A,A,G,A,A,A,A,A	7	{1,1,3,1,5,1,5}	2.43
	5	G,A,G,A,A,G,A,A	6	{1,1,1,2,1,2}	1.33
	6	A,A,G,A,G,A,A,A,A,A,G,A,A,A,A,A,A,A	7	{2,1,1,1,5,1,7}	2.57
	7	A,A,G,A,A,A,A,G,A,A,A,A,A,G,A,A,A,G	8	{2,1,4,1,5,1,3,1}	2.25
Kvapil	Duše	A,A,G,A,A,A,G,G,A,A,G,A,A,G,G,A,A,A,G,A,A	11	{2,1,3,2,2,1,2,2,3,1,2}	1.91
	Iniciála	A,A,A,G,A,A,A,G,A,A,G,A,G	8	{3,1,3,1,2,1,2,1,1,1]	1.60
	Koflík	A,A,G,A,A,G,A,G,A,G,A,G,G,A,A,G,A,A,A,A,A,G	14	{2,1,2,1,1,1,1,1,1,2,2,1,5,1}	1.57
	Má poesie	A,A,A,A,A,A,A,G,A,G,A,G,A,G,A	9	{7,1,1,1,1,1,1,1,1}	1.67
	Pagoda	A,A,A,A,A,A,A,A,A,A	1	{10}	10.00
	Portrét	A,A,G,A,G,A,A,A	5	{2,1,1,1,3}	1.60
	Sonet o mém umění	A,G,G,G,A,G,A,A,A,A,A,G,A,A,A,A,G,G,G,G	8	{1,3,1,1,5,1,3,4}	2.38
Karásek	Horečka	A,A,G,A,A,A,A,A,G,G,G,G	4	{2,1,5,4}	3.00
	Japonerie	A,G,A,A,A,A,G,A,G,G,G,A,A	7	{1,1,4,1,1,3,2}	1.86
	Kalný západ	A,A,G,A,A,A,A,A,G,A,A,A,A,A,G,A,A,A,G,A,A,A,A,A,A,A,G,A,A,A,G,A	13	{2,1,5,1,5,1,3,1,7,1,3,1,1}	2.46
	Narkózy	G,A,A,A,A,A,G,A,A,A,G,A,A,A,A,A,G,	8	{1,5,1,3,1,4,1,4}	2.50

		A,A,A,A			
	Nemoc	G,A,A,A,A,A,G,A, A,A,G,A,A,A,A	6	{1,5,1,3,1,4}	2.50
	Noční sonet	A,G,A,A,A,A,A,A, G,G,G,A,A,G,A,A	7	{1,1,6,3,2,1,2}	2.29
	Pokušení	A,A,G,G,A,A,G,A, A,G,G	6	{2,2,2,1,2,2}	1.83
	Příšerná loď	A,A,A,A,A,A,A,A, A,A,A,A,A,A,A,A, A,A,A,G	2	{19,1}	10.00
	Rozklad	A,A,A,A,A,A,A,A, G,G,A,A,G,G,A,A, A,G,G,G	6	{8,2,2,2,3,3}	3.33
Nezval	fetišista	A,A,A,G	2	{3,1}	2.00
	měšťák	A,G,G,A,A,A,A,G	4	{1,2,4,1}	2.00
	svedený	G,G,G,G,G	1	{5}	5.00
	circulus vitiosus	A,A,G,A,A,A,A,A, A, A	3	{2,1,7}	10.00
	básník	A,G,G,G,A,A,A,A	3	{1,3,4}	2.67
	literatura	A,A,A,A	1	{4}	4.00
	upír	G,A,G,A,G,G,G	5	{1,1,1,1,3}	1.40
	její moc	G,A,G	3	{1,1,1}	1.00
	sonet	A,A	1	{2}	2.00
	rytmy	A,G,A,G,G,G	4	{1,1,1,3}	3.00

A complex tool to analyse whether the sequence of runs is structured is the statistical z-test (cf. Zörnig et al. 2016). The count, which may be easily transformed for the purpose of the present research, reads

$$z = \frac{n(r-1) - 2n_a n_g}{\sqrt{\frac{2n_a n_g * (2n_a n_g - 1)}{n-1}}}, \tag{6.4}$$

with n_a standing for the number of adjectives, n_g for the number of genitives, r for the total of runs, and n for the sum of adjectives and genitives.

The critical value of the test is 1.96 and -1.96 ($\alpha = 0.05$); this divides texts into three groups:

(1) $z > 1.96$ – texts with a lot of changes in the runs, which points at possible structures to be found in them (e.g., G–A–G–A–G–A–…);
(2) $-1.96 < z < 1.96$ – texts that do not show any particular structuration;
(3) $z < -1.96$ – texts in which runs tend to be very long (e.g., A–A–A–A–A–G–G–G–G–G–…).

If the aforementioned poem by Kvapil is taken as an example, one handles two genitives, six adjectives, and five runs. The calculation yields

$$z = \frac{8 * (5 - 1) - 2 * 6 * 2}{\sqrt{\frac{2 * 6 * 2 * (2 * 6 * 2 - 1)}{8 - 1}}} = 1.08\,,$$

signifying there is no structuration in the sonnet.

Some results for Czech sonnets are listed in Table 6.9.

Table 6.9
Z-test values in selected Czech sonnets

Author	Poem	A	G	z-test
Mácha	1	6	2	1.08
	2	5	3	1.03
	3	6	4	-0.56
	4	14	3	0.96
	5	5	2	1.03
	6	14	3	0.92
	7	14	3	1.87
Kvapil	Duše	14	7	0.34
	Iniciála	9	4	1.01
	Koflík	14	8	1.34
	Má poesie	11	4	1.49
	Pagoda	10	0	N/A
	Portrét	6	2	1.08
	Sonet o mém umění	8	9	-0.74
Karásek	Horečka	7	5	-1.77
	Japonerie	8	5	-0.09
	Kalný západ	26	6	1.36
	Narkózy	16	4	0.44
	Nemoc	12	3	0.18
	Noční sonet	11	5	-0.53
	Pokušení	6	5	-0.29
	Příšerná loď	19	1	-3.00*
	Rozklad	13	7	-2.08*
Nezval	fetišista	3	1	-1.00
	měšťák	5	3	-0.62
	svedený	0	5	N/A
	circulus vitiosus	9	1	0.50
	básník	5	3	-1.44
	literatura	4	0	N/A

	upír	2	5	1.22
	její moc	1	2	1.41
	sonet	2	0	N/A
	rytmy	2	4	0.35

The main drawback of the method lies in the fact that most texts do not show any structuration; this may be either due to the over-strictness of the measure significance threshold, or due to the genuine lack of organization in the studied poems. The two exceptions are Karásek's poems *Příšerná loď* ("A Dreadful Ship") and *Rozklad* ("Decay"), which both display long adjectival chains. In the former case, the show of adjectives is checked by the final genitive – the only one in the poem –, whereas the latter demonstrates a monotonous introductory chain of attributes, which is, step by step, breached by an infiltration of genitives (for the sequences, see Table 6.8).

6.5 Modelling Tendencies

The same as in activity measurements, adjective-genitive relations can be modelled both graphically and mathematically. This may be employed in micro-analyses of individual poems, but if all collections are taken as wholes, the sequences of all the poems may be put together into a single one, too. Here, the research will be restricted to a couple of examples of graphic projections.

The count is done in the following way: genitives, which are placed on the x-axis, are numbered continually; then, the number of adjectives preceding each genitive is calculated. That means that the position of G is the independent variable x, and the number of A in front of it is the dependent variable y. Let us have an example – Kvapil's sonnet *Má poesie* ("My Poetry") demonstrates the sequence

A – A – A – A – A – A – A – G – A – G – A – G – A – G – A,

which contains four genitives; the first one is preceded by seven adjectives, the second one by eight, the third one by nine, and the last one by ten. In order to include all adjectives, another, "artificial" genitive is added. The figures are listed in Table 6.10, and projected in Figure 6.10.

Table 6.10
The coordinates of genitives and adjectives in Kvapil's poem *Má poesie*

G	A
1	7
2	8
3	9
4	10
5	11

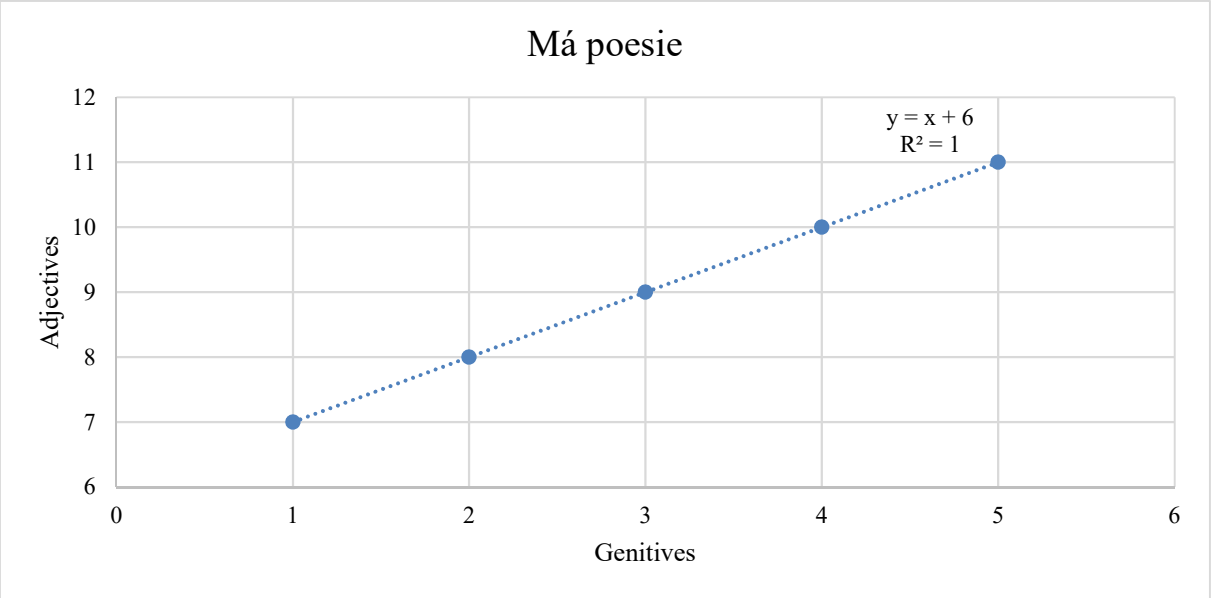

Fig. 6.1. Graphic projection of adjective and genitive adnominals in Kvapil's poem *Má poesie*

Evidently, this course can be perfectly captured by a linear function, but it does not hold for all sonnets. Hence, we must search for a general function satisfying all courses of adnominal genitive. Since the sequence represents a development, we shall use the Piotrowski-law function –

$$y = \frac{c}{1 + a * e^{-bx}} \text{ –,} \tag{6.5}$$

yielding, in the given case, parameters a = 1.9740, b = 0.2247, and c = 18.0521; as the determination coefficient R^2 equals, the function covers all the values of attributes.

Some more examples, including the figures, are presented below. The issue will be paid detailed attention in a study of its own, which is currently under preparation.

Table 6.11
Modelling figures of genitives and adjectives in the introductory sonnet of Mácha's poem *Máj*

G	**A**	**Comp.**
1	2	2.05
2	6	5.94
3	11	11.05
4	14	13.98
a = 26.3961, b = 1.4097, c = 15.2930, R^2 = 0.9999		

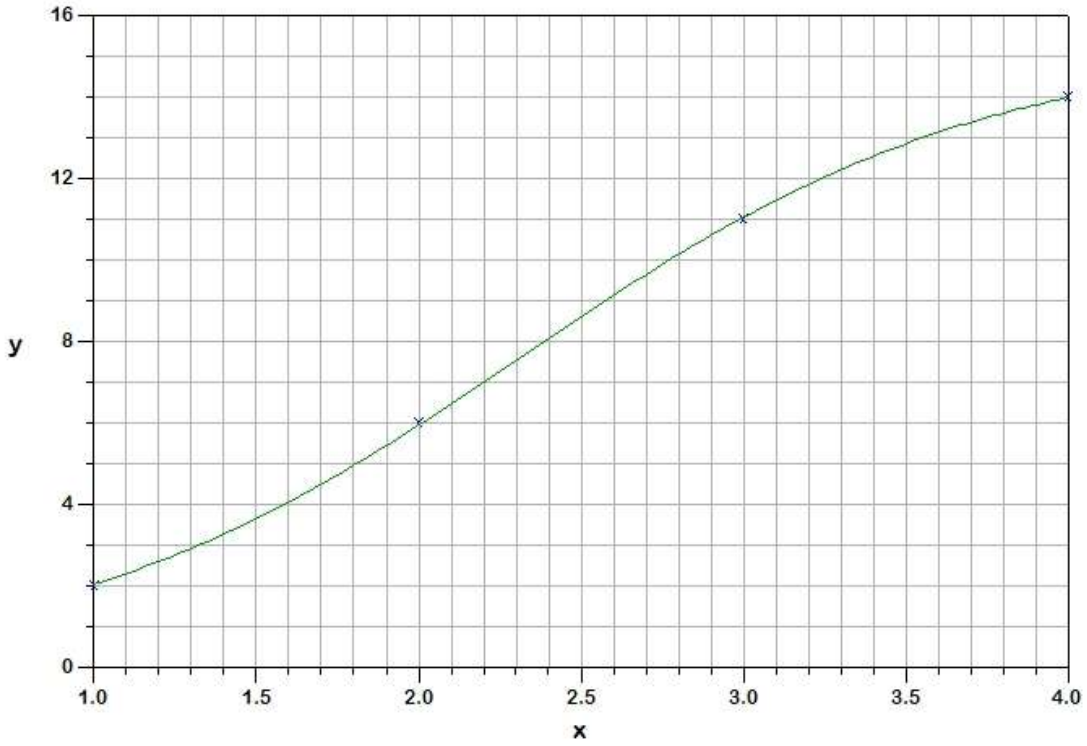

Fig 6.2. Graphic projection of adjective and genitive adnominals in the introductory sonnet of Mácha's poem *Máj*

Table 6.12
The coordinates of genitives and adjectives in Karásek's poem *Rozklad* ("Decay")

G	A	Comp.
1	8	7.48
2	8	8.66
3	10	9.82
4	10	10.91
5	13	11.89
6	13	12.75
7	13	13.48
a = 1.5608, b = 0.2919, c = 16.2025, R^2 = 0.9019		

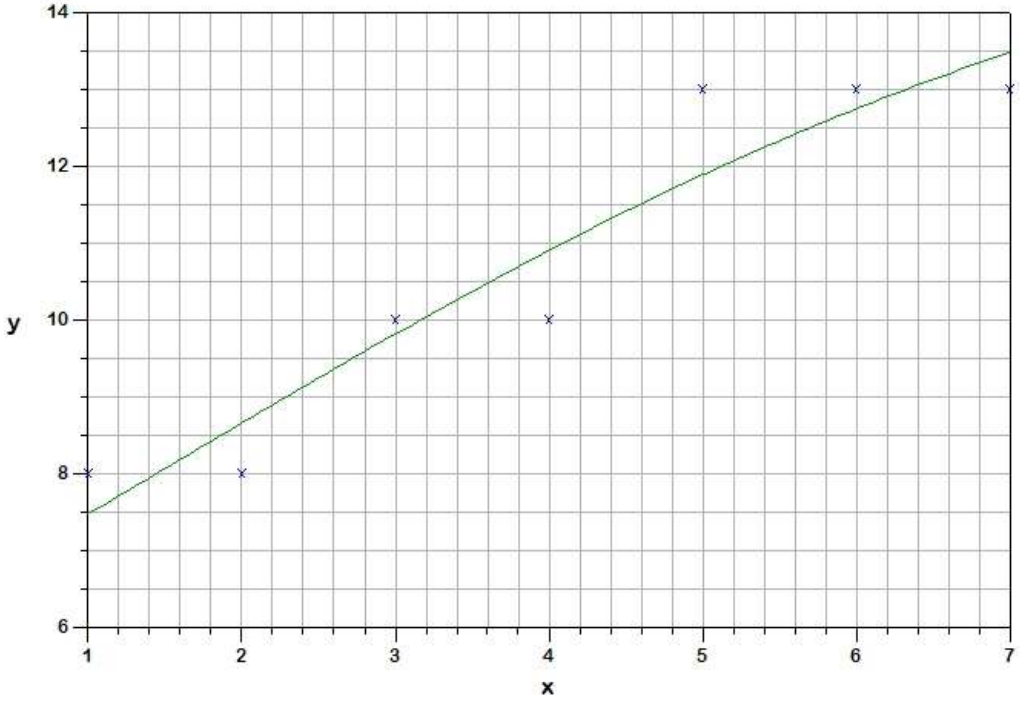

Fig 6.3. Graphic projection of adjective and genitive adnominals in Karásek's poem *Rozklad* ("Decay")

Table 6.13
The coordinates of genitives and adjectives in Kvapil's poem *Sonet o mém umění* ("A Sonnet on My Art")

G	A	Comp.
1	1	0.01
2	1	0.04
3	1	0.34
4	2	2.23
5	7	7.03
6	10	9.61
7	10	10.07
8	10	10.13
9	10	10.14
a =14596.33, b = 2.0807 c = 10.1370, R^2 = 0.9834		

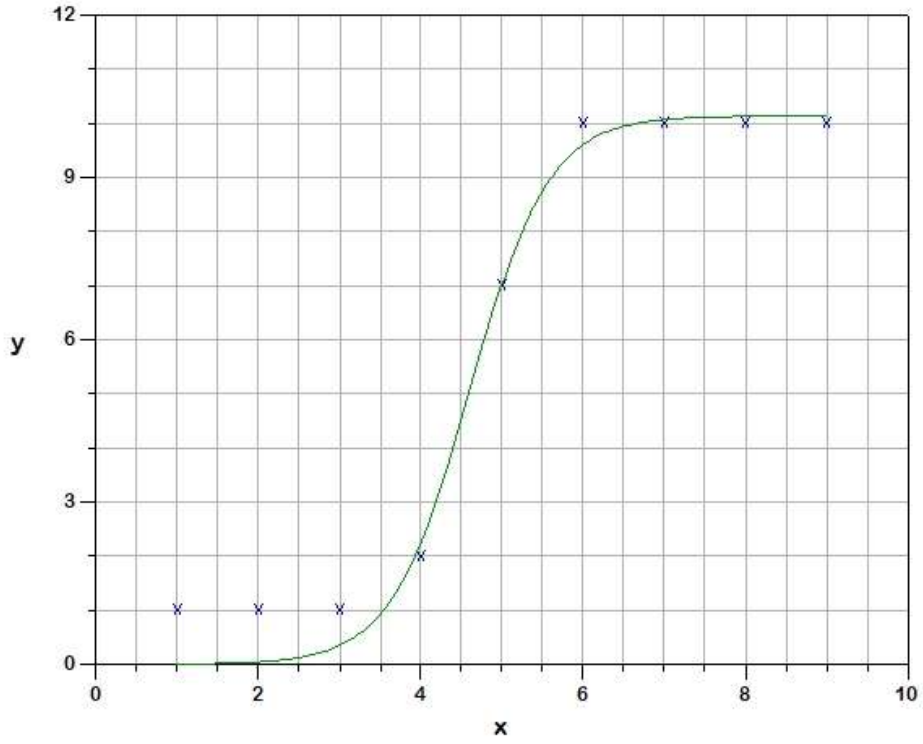

Fig. 6.4. Graphic projection of adjective and genitive adnominals in Kvapil's poem *Sonet o mém umění* ("A Sonnet on My Art")

6.6 Valency Motifs

Adnominal valencies of nouns, reflecting the number of attributes by which they are modified, can be modelled into motifs (cf. Zhang, Liu 2017). Thus, describing the adnominal structure of the sonnet *Michelangelo* by K. Bal'mont (T 39), one gets the following sequence, which takes into account the type and the order of adnominals in the texts and positions of nouns (N):

N(1), AP, N(1), AP, N(0), N(0), N(1), G, N(1), G, DETS, N(1), N(1), G, A, N(1), PT, N(1), A, N(1), N(1), PT, DETS, N(0), N(0), N(0), N(0), N(1), DETS, DETQ, N(1), N(0), N(0), A, N(1), N(0), N(0), N(0)

The types of adnominals and their abbreviations are explained in Section 6.1. The figures in brackets next to nouns show the number of adnominals linked with the given noun – i.e., its adnominal valency. These valencies may be represented in the following sequence, in which motifs are marked by square brackets:

[1,1], [0, 0, 1, 1, 1, 1, 1, 1, 1, 1], [0, 0, 0, 0, 1, 1], [0, 0, 1], [0, 0, 0].

The five motifs include one motif with 2 elements, two motifs with three elements, one 6-element one, and one 10-element motif.

Performing this analysis for the Russian sonnets and measuring the length of motifs, we obtain the results presented in Table 6.14.

Table 6.14
Lengths of adnominal motifs in Russian sonnets

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**≥ 10**
Text 1	2	–	1	1	1	–	–	–	–	–
Text 2	2	–	2	–	–	1	1	–	–	–
Text 3	2	–	2	1	1	–	–	–	1	–
Text 4	1	1	2	–	2	–	–	–	–	–
Text 5	1	–	1	–	–	–	–	–	1	–
Text 6	–	–	–	2	–	–	–	1	–	–
Text 7	1	–	1	2	1	–	1	–	–	–
Text 8	1	–	–	–	1	–	1	–	–	1
Text 9	1	3	1	2	–	1	–	–	–	–
Text 10	1	5	2	–	–	–	–	1	–	–
Text 11	1	3	2	1	–	1	–	–	–	–
Text 12	2	1	1	–	2	–	–	–	–	–
Text 13	–	–	1	–	–	–	–	–	1	–
Text 14	–	1	1	1	–	2	–	–	–	–
Text 15	1	2	2	–	1	–	1	–	–	–
Text 16	2	2	1	1	–	–	–	–	1	–
Text 17	–	–	–	–	1	–	1	–	–	1
Text 18	–	–	4	1	1	1	–	–	–	–
Text 19	–	–	2	2	–	–	–	–	1	–
Text 20	–	1	1	–	–	1	–	–	–	1
Text 21	2	1	1	2	–	–	1	–	–	–
Text 22	–	3	1	2	1	–	1	–	–	–
Text 23	–	4	1	1	–	1	1	–	–	–
Text 24	–	–	3	1	–	–	1	–	–	–
Text 25	1	1	2	2	–	1	–	–	–	–
Text 26	1	1	1	–	–	–	1	–	–	–
Text 27	–	2	–	–	–	1	1	–	–	–
Text 28	1	3	1	2	1	1	–	–	–	–
Text 29	1	4	3	1	–	–	–	–	–	–
Text 30	1	4	1	–	–	1	–	–	–	1
Text 31	2	4	1	–	2	–	–	–	–	–
Text 32	1	–	2	1	–	–	1	1	–	–
Text 33	–	2	–	1	–	–	1	–	–	1

Text 34	2	1	–	1	–	–	–	–	–	1
Text 35	1	1	–	2	–	–	–	–	–	–
Text 36	–	1	1	–	1	1	–	–	1	–
Text 37	3	2	1	1	1	–	–	–	–	–
Text 38	1	2	–	1	2	1	–	–	–	–
Text 39	–	1	2	–	–	1	–	–	–	1
Text 40	1	5	1	–	1	–	–	–	–	–
Text 41	3	4	1	–	–	2	–	–	–	–
Text 42	1	1	–	–	1	1	–	–	–	–
Text 43	–	–	1	1	1	–	–	1	–	–
Text 44	–	3	2	3	–	–	–	–	–	–
Text 45	–	–	3	2	–	1	–	–	–	–
Text 46	–	1	1	–	–	–	1	1	–	–
Total	39	70	58	38	22	19	14	5	6	7

For the total numbers of lengths of motifs, the Lorentzian function – see Sections 2.2 and 3.1 – yields a very good fit (Table 6.15).

Table 6.15
Lorentzian function of the length-motifs

Length of Motifs	**Number of Motifs**	**Lorentzian Function**
1	39	43.50
2	70	65.24
3	58	61.23
4	38	38.46
5	22	22.56
6	19	14.02
7	14	9.35
8	4	6.61
9	6	4.90
≥ 10	7	3.76
a = 68.0779, b = 2.3841, c = -1.8416, R^2 = 0.9747		

After obtaining and ordering the ranks of the motif lengths, it was found out that the exponential function with added 1 expresses the ranking trend very satisfactorily (Table 6.16).

Table 6.16
Fitting the exponential function with added 1 to motif-length ranks
in Russian sonnets

Length Ranks	**Number**	**Exponential ft. +1**
1	70	73.21
2	58	54.64
3	40	40.84
4	38	30.60
5	22	22.99
6	19	17.33
7	14	13.13
8	6	10.01
9	4	7.69
10	3	5.97
11	1	4.69
12	1	3.74
13	1	3.04
14	1	2.51
a = 5.5794, b = 0.0185, R^2 = 0.9941		

The results obtained in this section demonstrate a constant regularity, which is observed in Russian sonnets over the period of about two hundred years; it may be further tested on different samples of poetry and prose.

7. Consensus Strings

These forms have been introduced into linguistics by P. Zörnig et al. (2016). As a matter of fact, this is a vertical look at the text. One considers the individual lines, takes any property, and the sum of degrees (or elements) of the same sort – or the maximal number of some elements in the column – yields a string which has a course for each sonnet and each property. The course can be hypothesized, and the hypotheses may be tested. One can study the individual collections, the development of an author, the development in the given language, the difference of two languages, etc. Since the sonnet consists of 14 lines, the study of some indicators of each line is a good presupposition for finding hypotheses.

Since sonnets have always 14 lines, one can consider also the rhyme-word POS, construct a vector of POS – see above –, and one can set up a consensus string using all sonnets of an author. Consider, for example, the POS of rhyme words in *Sonety* by Hviezdoslav, as given in Table 2.8.

N,N,N,N,N,Av,N,N,V,N,Pn,N,N,N;
N,N,N,N,N,N,N,N,N,N,N,N,V,N;
N,N,V,N,A,V,N,N,A,V,N,V,N,N;
N,N,N,N,N,V,N,N,N,N,N,N,N,V;
N,N,N,V,N,N,N,V,V,N,N,N,A,V;
N,N,N,N,N,V,N,A,N,N,N,V,N,N;
Av,N,V,N,N,N,N,A,A,N,V,N,N,V;
N,A,V,A,A,N,V,A,N,N,V,V,N,N;
N,N,N,V,N,V,V,N,Av,V,N,N,N,V;
N,V,Av,V,V,N,N,N,N,N,N,N,N,N;
N,N,N,V,V,Nu,N,N,N,A,V,N,N,N;
N,N,V,N,V,N,V,A,N,N,N,N,N,N;
N,I,N,N,N,N,N,V,N,N,N,A,N,N;
N,N,N,N,V,N,N,N,N,N,N,N,N,A;
N,V,N,N,N,N,V,V,Av,Pn,N,N,N,N;
V,V,N,N,N,N,N,N,N,N,V,V,A,N,V;
N,Av,N,N,N,N,N,N,Pn,A,N,N,N,N;
N,N,N,V,N,N,V,Pn,N,N,N,N,N,Av;
N,N,Av,A,N,N,V,N,N,N,N,N,N,V;
Av,N,N,N,Av,N,N,N,N,N,N,Nu,V,N;
A,N,V,N,N,N,A,V,V,N,N,N,V,A.

There are 21 sonnets. In the first column, the most frequent POS is N, occurring there 17 times; in the second column, it is again N, occurring 15 times; in the third column it is N, occurring 14 times; etc. Counting the most frequent POS in the column, we obtain the sequence:

17, 15, 14, 14, 14, 15, 14, 12, 12, 15, 15, 15, 17, 12,

in which the noun is always the most productive POS.

Consider, for example, the Slovak collection S*onety* by P.O. Hviezdoslav and the columns of POS for each sonnet separately. In each column, we count the most frequent POS. The first sonnet, as given below, yields the following sequences:

Keď slnko zájde skonom a noc temné krídla
hviezd oky posiate v šír rozstrie ponad zem:
jak včely do úľa hneď každý záujem
sa vracia z poľa činov v úkryt svojho bydla.

Do seba vstúpi človek, sňatý s motovidla,
kde ako Ixion sa svíjal tam a sem,
do seba vojde zas, otrasúc šatný lem,
v sieň, ktorú vyzdobily vidín maľovidlá.

Tu doma je, kam vlastne dľa poslania patrí,
z pút smyslov unikol jak vtáča nástrahe,
a žije plným žitím pre seba a v sebe.

Len keď strasť zemskú vrhol do večernej vatry,
len vtedy octne sa zas človek na dráhe,
po nejž, i kým tu hlivie, zájsť môž' na čas v nebe.

C,N,V,N,C,N,A,N
N,N,A,Pr,N,V,Pr,N,
C,N,Pr,N,Av,Pn,N,
Pn,V,Pr,N,N,C,Pn,N,

Pr,Pn,V,N,A,Pr,N
Av,Av,N,Pn,V,Av,C,Av
Pr,Pn,V,Av,V,A,N,
Pr,N,Pn,V,N,N,

Av,Av,V,Av,Av,Pr,N,V,
Pr,N,N,V,Av,N,N,
C,V,A,N,Pr,Pn,C,Pr,Pn,

Av,Av,N,A,V,Pr,A,N,
Av,Av,V,Pn,Av,N,Pr,N,
Pr,Pn,C,Av,Av,V,V,V,Pr,N,Pr,N.

If we consider only the most frequent POS in the column and divide it by the number of elements in the column (the individual columns are not equally long), we obtain the following table.

Table 7.1
The count of the consensus-string function: POS

5/14	0.36
5/14	0.36
5/14	0.36
5/14	0.36
5/14	0.36
4/14	0.29
5/13	0.38
5/9	0.56

Here, one can speak of a horizontal curve which changes only if the number of elements gets smaller than 14. That means that the consensus string of POS in rhyme can be considered a horizontal line. Nevertheless, one can find sonnets in other languages, or in other authors displaying a different behaviour.

Let us consider the syllabic length of words in the above sonnet. We obtain

1,2,2,2,1,1,2,2.
1,2,3,1,2,2,1
1,2,1,2,1,2,3,
1,2,2,2,2,2,2,2,
1,2,2,2,2,4,
1,2,3,1,2,1,1,1
1,2,2,1,3,2,1
1,2,4,2,4,
1,2,1,1,2,1,3,2,
1,2,3,1,2,3,
1,2,2,2,1,2,1,2
1,1,1,2,2,1,3,2,
1,2,2,1,1,2,1,2,
1,1,1,1,1,2,1,1,1,1,2.

By means of the previous procedure, we get the following numbers.

Table 7.2
The count of the consensus-string function: syllables

14/14	1.0
12/14	0.86
6/14	0.43

7/14	0.5
7/14	0.5
7/13	0.54
6/11	0.55
5/7	0.71
1/1	1.0
1/1	1.0
1/1	1.0

This course represents a regular convex function. One cannot say which regularity is concealed behind individual entities. Many of them must be investigated in order to find a trace.

8. Measuring Activity

8.1 Busemann's Coefficient and Chi-Square Testing

This coefficient is usually expressed as the number of verbs divided by the sum of adjectives and verbs. Adjectives are "static", verbs mean activity. One can study the development of the coefficient in the sonnet, the development of an author, differences between authors, languages. Some results are presented in Table 8.1, 8.2, and 8.3. In some languages, one must decide whether a participle is an adjective or a verb – e.g., the English sentence "the man going in the street" can be translated into Slovak as "muž, idúc po ulici" or "muž idúci po ulici". In the first case, in can be considered a verb (a transgressive form); in the second case, it can be considered an adjective.

Busemann's coefficient is defined as

$$B = \frac{V}{A+V}, \tag{8.1}$$

but one can define it also with the adjective in the numerator.

In the first sonnet by Goethe, we obtain the following sequence:

A – V – A – V – V – V – V – V – V – A – V – A – A – A – V – V – V – V – A – V,

in which there are 7 adjectives and 13 verbs. The verbs may also be joined with a modality (*müssen*, *können*, *sollen*, *wollen*, …). Here, we obtain

$$B = \frac{13}{20} = 0.65\,.$$

In the given case, the activity is greater than the descriptivity, but one may ask whether the difference is significant. In Zörnig et al. (2015: 4–19), several tests were proposed. Here, we shall show some of them and classify the sonnets in the column *Type* of Tables 8.1, 8.2, and 8.3. The exact test considers the expected relation between verbs and adjectives as –

$$p = \frac{1}{2} = 0.5\,.$$

If the observed number of verbs is smaller than that of adjectives (i.e. smaller than the half), we compute

$$P(X \leq V) = \frac{1}{2^n} \sum_{x=0}^{V} \binom{n}{x}, \tag{8.2}$$

where n is the sum of all adjectives and verbs. If P is smaller than 0.05, we consider the sonnet strongly descriptive, for $P > 0.05$, it is merely descriptive. On the other hand, if V > A, one computes

$$P(X \geq V) = \frac{1}{2^n} \sum_{x=V}^{n} \binom{n}{x}; \tag{8.3}$$

if P is smaller than 0.05, the text is strongly active, otherwise it is merely active.

Nonetheless, one can use a simpler formula, namely (cf. Altmann 1988, Altmann, Köhler 2015) the usual chi-square defined as

$$\chi^2 = \frac{(V - A)^2}{V + A}, \tag{8.4}$$

with 1 degree of freedom. The respective probabilities can be found in the usual chi-square tables.

For example, the first sonnet by Goethe yields

$$\chi^2 = \frac{(13 - 7)^2}{13 + 7} = 1.8\,.$$

Since the given value of χ^2 with 1 DF is not significant, the poem can be assessed as active.

We interpret the results as follows:

(1) SA – significantly active (B > 0.55, $\chi^2 > 3.84$);
(2) AC – active (B > 0.55, $\chi^2 < 3.84$);
(3) N – neutral (0.45 < B < 0.55);
(4) DE – descriptive (B < 0.45, $\chi^2 < 3.84$);
(5) SD – significantly descriptive (B < 0.45, $\chi^2 > 3.84$).

Table 8.1
Busemann's coefficient in German, Slovak, Czech, and Hungarian sonnets

Sonnet	A	V	B	Type
J.W.v. Goethe				
Das Sonett 1	7	13	0.65	N
2	4	14	0.78	SA

Mächtiges Überraschen	4	19	0.83	SA
Freundliches Begegnen	10	16	0.62	N
Kurz und gut	7	19	0.73	SA
Das Mädchen spricht	3	15	0.83	SA
Wachstum	7	14	0.67	AC
Reisezehrung	3	15	0.83	SA
Abschied	6	15	0.71	SA
Die Liebende schreibt	6	13	0.68	AC
Die Liebende abermals	4	16	0.90	SA
Sie kann nicht enden	6	15	0.71	SA
Nemesis	1	11	0.92	SA
Christgeschenk	9	11	0.55	N
Warnung	3	16	0.84	SA
Die Zweifelnden	5	17	0.77	SA
Mädchen	5	16	0.76	SA
Epoche	7	10	0.53	N
Charade	5	17	0.77	SA
P.O. Hviezdoslav, *Sonety*	**A**	**V**	**B**	**Type**
1	7	16	0.70	AC
2	8	19	0.70	SA
3	7	17	0.71	SA
4	6	14	0.70	AC
5	6	17	0.74	SA
6	6	17	0.74	SA
7	7	23	0.77	SA
8	11	17	0.61	AC
9	7	19	0.73	SA
10	5	16	0.76	SA
11	10	15	0.60	AC
12	10	17	0.63	AC
13	9	10	0.53	N
14	9	12	0.57	N
15	7	18	0.72	SA
16	4	14	0.78	SA
17	12	15	0.55	N
18	5	18	0.78	SA
19	7	25	0.78	SA
20	10	17	0.63	AC
21	7	21	0.75	SA
P.O. Hviezdoslav, *Krvavé Sonety*	**A**	**V**	**B**	**Type**
1	7	11	0.61	N
2	5	17	0.77	SA
3	11	11	0.50	N
4	7	12	0.63	N

5	9	10	0.53	N
6	7	13	0.65	N
7	7	11	0.61	N
8	6	12	0.67	N
9	1	17	0.94	SA
10	5	9	0.64	N
11	5	11	0.69	N
12	7	11	0.61	N
13	4	15	0.79	SA
14	5	14	0.74	SA
15	11	12	0.52	N
16	5	12	0.71	CA
17	9	14	0.61	N
18	2	13	0.87	SA
19	2	12	0.96	SA
20	9	8	0.47	N
21	6	15	0.71	SA
22	12	10	0.45	N
23	4	9	0.69	N
24	4	20	0.83	SA
25	2	19	0.90	SA
26	4	15	0.79	SA
27	7	19	0.73	SA
28	7	14	0.67	CA
29	5	12	0.71	CA
30	6	14	0.70	CA
31	7	11	0.61	N
32	10	8	0.44	DE
Machar, *Letní sonety*				
E. Zolovi	7	8	0.47	DE
Matce	12	8	0.60	AC
Sonet cynický	12	8	0.60	AC
Sonet de vanitate	8	13	0.38	DE
Sonet elegický	9	9	0.50	N
Sonet ironický	8	11	0.42	DE
Sonet k sociální otázce	12	6	0.67	AC
Sonet k teorii: Boj o život	13	8	0.62	AC
Sonet materialistický	8	12	0.40	DE
Sonet mystický	11	10	0.52	AC
Sonet na Chopinovu melodii	11	6	0.65	AC
Sonet na sentenci z Goetha	9	6	0.60	AC
Sonet na sklonku století	7	8	0.47	DE
Sonet nad verši z mládí	8	7	0.53	AC
Sonet noční	11	7	0.61	AC

Sonet o antice a vlasech	7	16	0.30	DE
Sonet o bídě	11	5	0.69	AC
Sonet o hodinách	13	4	0.76	SA
Sonet o lásce	12	8	0.60	AC
Sonet o minulosti	10	2	0.83	SA
Sonet o Panně Marii	7	13	0.35	DE
Sonet o rokoku	10	10	0.50	N
Sonet o staré metafoře	10	9	0.53	AC
Sonet o starém líci a rubu	12	7	0.63	AC
Sonet o třech metaforách	6	13	0.32	DE
Sonet o třetí hodině v červenci	5	10	0.33	DE
Sonet o vídeňských kosech	14	6	0.70	AC
Sonet o západu slunce	15	7	0.68	AC
Sonet o zlatém věku naší poezie	8	4	0.67	AC
Sonet o životě	19	7	0.73	SA
Sonet patologický	15	5	0.75	SA
Sonet polední	10	11	0.48	DE
Sonet sarkastický	9	7	0.56	AC
Sonet svatební	16	5	0.76	SA
Sonet úvodní	8	18	0.31	SD
Sonet večerní	6	10	0.38	DE
Sonet z dvacátého září	10	10	0.50	N
Sonet-apostrofa	12	6	0.67	AC
Sonet-epilog čtenáři	12	9	0.57	AC
Sonet-intermezzo$_2$	8	9	0.47	DE
Sonet-intermezzo	8	10	0.44	DE
Sonety-Causerie I.	14	5	0.74	SA
Sonety-Causerie II.	14	5	0.74	SA
Sonety-Causerie III.	5	9	0.36	DE
Sonety-Causerie IV.	9	5	0.64	AC
Sonety-Causerie V.	11	8	0.58	AC
Své ženě s předešlým sonetem	11	9	0.55	AC
M. Babits, *A lírikus epilógja* (Hungarian)	5	20	0.80	SA
A. József, *A kozmosz éneke*				
1	10	16	0.62	AC
2	20	10	0.33	DE
3	16	11	0.41	DE
4	11	11	0.50	N
5	10	11	0.52	N
6	16	9	0.36	N
7	13	12	0.48	DE
8	12	16	0.57	N

9	14	14	0.50	N
10	8	15	0.65	AC
11	18	7	0.28	SD
12	3	12	0.80	SA
13	13	13	0.50	N
14	9	13	0.59	AC
15 (Mesterszonett)	12	11	0.48	N

Table 8.2
Busemann's coefficient in English sonnets

	A	**V**	**B**	**Type**
W. Wordsworth				
Beloved Vale!	13	17	0.57	AC
Calm Is All Nature (…)	15	14	0.48	DE
Glasmore Lake	20	8	0.29	SD
How Sweet It Is (…)	14	11	0.44	DE
London, 1802	14	8	0.36	DE
Nuns Fret Not (…)	15	8	0.35	DE
Scorn Not the Sonnet (…)	14	13	0.48	DE
The World Is Too Much with Us (…)	9	13	0.59	AC
Those Words Were Uttered (…)	17	11	0.39	DE
With How Sad Steps (…)	11	12	0.52	AC
S. T. Coleridge				
To Southey	26	11	0.30	SD
To Priestley	23	11	0.32	SD
To Pitt	22	10	0.31	SD
To Mrs Siddons	23	11	0.32	SD
To Kosciusko	21	8	0.28	SD
To Godwin	22	12	0.35	DE
To Fayette	21	13	0.38	DE
To Burke	23	11	0.32	SD
To Bowles	24	9	0.27	SD
To Erskine	16	14	0.47	DE
G. Rossetti				
A Venetian Pastoral	8	20	0.71	SA
Heart's Haven	17	9	0.35	DE
Life-in-Love	14	9	0.39	DE
Mary Magdalene	7	22	0.76	SA
Silent Noon	19	12	0.39	DE
The Heart of the Night	7	9	0.56	AC
The Choice, I	10	20	0.67	AC
The Choice, II	8	20	0.71	SA

Through Death to Love	17	10	0.37	DE
Without Her	22	3	0.12	SD

Table 8.3
Busemann's Coefficient in French sonnets

	A	**V**	**B**	**Type**
Ch. Baudelaire				
Bohémiens en voyage	15	10	0.40	DE
Correspondances	14	7	0.33	DE
L'Ennemi	10	16	0.62	AC
L'Idéal	11	8	0.42	DE
La beauté	12	14	0.54	N
La masque	11	15	0.58	AC
La Muse malade	19	7	0.27	SD
La Muse vénale	8	15	0.65	AC
La vie antérieure	16	10	0.38	DE
Le guignon	12	6	0.33	DE
Le mauvais moine	11	11	0.50	N
Parfum exotique	16	10	0.38	DE
T. Corbière				
1 Sonnet	4	9	0.69	AC
A l'éternel madame	15	19	0.56	AC
Bonsoir	10	17	0.63	AC
Déclin	9	15	0.63	AC
Duel aux camélias	15	12	0.44	DE
Féminin singulier	9	18	0.67	AC
Fleur d'art	14	5	0.26	SD
Litanie	7	10	0.59	AC
Pauvre garçon	5	22	0.81	SA
Pudentiane	8	11	0.58	AC
Sonnet à sir Bob	8	15	0.65	AC
Sonnet de nuit	7	6	0.46	N
S. Mallarmé				
Angoisse	11	16	0.59	AC
Dame sans trop d'ardeur (…)	12	15	0.56	AC
Le pitre châtié	8	12	0.60	AC
Le sonneur	11	17	0.61	AC
O si chère de loin (…)	16	9	0.36	DE
Placet futile	10	14	0.58	AC
Quand l'ombre menaça (…)	13	12	0.48	N
Remémoration d'amis belges	12	11	0.48	N
Renouveau	14	16	0.53	N

Salut	6	9	0.60	AC
Sonnet	12	17	0.59	AC
Tristesse d'été	13	13	0.50	N

Let us perform a comparison of Busemann types (preliminarily only for some writers). We obtain the results presented in Table 8.4.

Table 8.4
Types of Busemann results

Author	SA	AC	N	DE	SD	Number
Goethe	13	2	4	–	–	19
Hviezdoslav,						
Sonety	11	4	16	1	–	32
Krvavé sonety	12	6	3	–	–	21
Machar, *Letní sonety*	7	22	3	13	1	46
József, *A kozmosz éneke*	1	3	7	3	1	15
English sonnets	3	5	–	13	9	30
French sonnets	1	18	7	8	2	36

As can be seen, the differences are considerable. But if one ranks the results of individual writers, one obtains comparable results – cf. Tables 8.5 to 8.11.

Tables 8.5, 8.6, 8.7, 8.8 and 8.9
Ranking of tested Busemann's coefficients from Table 8.1

Goethe, *Sonette*			
Rank	**Type**	**Number**	**Exp ft.+1**
1	SA	13	13.00
2	N	4	3.10
3	AC	2	1.82
$a = 46.3184$, $b = 1.3517$, $R^2 = 0.9993$			

Hviezdoslav, *Sonety*			
Rank	**Type**	**Number**	**Exp + 1**
1	N	16	16.67
2	SA	11	8.76
3	AC	4	4.84
4	DE	1	2.90
$a = 31.6360$, $b = 0.7025$, $R^2 = 0.9290$			

Hviezdoslav, *Krvavé sonety*			
Rank	**Type**	**Number**	**Exp + 1**
1	SA	12	12.03
2	AC	6	5.87
3	N	3	3.14
a = 24.9854, b = 0.8178, R^2 = 0.9990			

Machar, *Letní sonety*			
Rank	**Type**	**Number**	**Exp. + 1**
1	AC	22	22.43
2	DE	13	11.95
3	SA	7	6.60
4	N	3	3.86
5	SD	1	2.46
a = 41.9121, b = 0.6721, R^2 = 0.9850			

József, *A kozmosz éneke*			
Rank	**Type**	**Number**	**Exp. + 1**
1	N	7	6.92
2	AC	3	3.50
3	DE	3	2.06
4	SA	1	1.45
5	SD	1	1.19
a = 14.0003, b = 0.8512, R^2 = 0.9424			

Table 8.10
Ranking of tested Busemann's coefficients of the English sonnets

Rank	**Type**	**Number**	**Exp. + 1**
1	DE	13	13.28
2	SD	9	8.21
3	AC	5	5.24
4	SA	3	3.49
a = 20.9052, b = 0.5320, R^2 = 0.9832			

Table 8.11
Ranking of tested Busemann's coefficients of the French sonnets

Rank	Type	Number	Exp. + 1
1	AC	18	17.84
2	DE	8	9.07
3	N	7	4.87
4	SD	2	2.85
5	SA	1	1.89
$a = 35.1412, b = 0.7357, R^2 = 0.9604$			

As can be seen, even in sonnets, activity is extremely important. The phenomenon is individualized both as to the languages, and as to the authors, with no generalizing conclusions to be formulated at the moment. It seems that declamatory sonnets, which focus on particular famous people (Coleridge's ones), tend to be very descriptive, and that there are considerable differences in individual authors' productions as well. A complex treatment of activity will be provided in a book on its own.

A step deeper would be the scaling of activity – e.g., *to run* is "more active" than *to go*. The same could be done with adjectives, but this way, the investigation became a broad discipline.

8.2 Activity Development

Another view is the development of the descriptiveness/activity in the sonnet. Let us begin with the descriptiveness evolution. One numerates the verbs (V) in the sonnet, and considers the numbers as x. Then, one counts all adjectives that occur before the given V, and considers them as y. Since these numbers increase, we can see the equilibrium of descriptiveness/activity. Neither the number of verbs, nor that of adjectives can decrease – hence the slope of the function may be used as an indicator of the development. Consider, for example, the first sonnet by Goethe (given below). We may write the sequence of adjectives and verbs and numerate them. We consider a modal verb and a "normal" verb as two verbs.

Sich in erneutem(A) Kunstgebrauch zu üben(V),
Ist heilge(A) Pflicht, die wir dir auferlegen(V).
Du kannst(V) dich auch, wie wir, bestimmt bewegen(V)
Nach Tritt und Schritt, wie es dir vorgeschrieben(V).

Denn eben die Beschränkung läßt(V) sich lieben(V),
Wenn sich die Geister gar gewaltig regen(V);
Und wie sie sich denn auch gebärden(V) mögen(V),
Das Werk zuletzt ist doch vollendet(A) blieben(V).

So möchte(V) ich selbst in künstlichen(A) Sonetten,
In sprachgewandter(A) Mühe kühnem(A) Stolze,
Das Beste, was Gefühl mir gäbe(V), reimen(V);

Nur weiß(V) ich hier mich nicht bequem zu betten(V).
Ich schneide(V) sonst so gern aus ganzem(A) Holze,
Und müßte(V) nun doch auch mitunter leimen(V).

The sequence is as follows –

A–V–A–V–V–V–V–V–V–V–V–A–V–A–A–A–V–V–V–V–V–A–V–V.

The numeration can be displayed in a table (cf. Table 8.12). The only important indicator is that of parameter *b*, which shows the increase. It must be remarked that in many languages, the identification of adjectives and verbs is not a simple task, so there will be differences even between researchers. Another problem is the fact that in some languages all words may have the "same" form – i.e., without any affixes. Here, one must decide which is which on purely semantic grounds.

Table 8.12
Increase of descriptivity in the first sonnet by Goethe

V	**A**	**Linear ft.**
1	1	0.53
2	2	0.94
3	2	1.35
4	2	1.75
5	2	2.16
6	2	2.57
7	2	2.97
8	2	3.38
9	2	3.79
10	3	4.20
11	6	4.60
12	6	5.01
13	6	5.42
14	6	5.82
15	6	6.23
15	7	6.23
17	7	7.05
a = 0.1251 , b = 0.4071, R^2 = 0.8277		

As can be seen, the number of A in front of a V is not even the half (V = 17, A = 7; SA; B = 0.71).

However, this needs not be the only possibility of modelling. We present results from several sonnets in Table 2.7. Here, we show merely the sequence of A and V, and the resulting dependence: number of A in front of the x-th V. For the sake of colourfulness, Machar's sonnets are treated in a reverted manner, taking adjectives as x and verbs as y (activity development). Usually, the linear function is sufficient to express the dependency, but here, we have to do with a development in a text. Though the development in texts differs from that in language, one can try to fit the so called Piotrowski-Law function given as –

$$y = \frac{c}{1 + a * e^{-bx}} \; ; \tag{8.5}$$

it is applied in historical linguistics, in biology, etc. Immediately, the question arises whether the law is adequate only for the development of Busemann's coefficient, or for sequences of any enitities presented in the binary form. In Busemann's coefficient, we have to do with two variables (A and V) both of which may be considered as Aristotelian noun predicates of the first order. Further, one may examine other text types and ask the same question. Thus, here a new research domain can be opened concerning textual elements, text types, languages, history of languages, etc. In Table 8.13, some results are presented.

Table 8.13
Increase of descriptiveness in various sonnets: Piotrowski-Law fits

Text	**Sequence**	**Piotrowski**
Babits, M. *A lírikus epilógja*	V, V, A, A, V, V, V, V, V, V, V, V, A, V, V, V, A, V, V, V, V, A, V, V, V	a = 19.3508 b = 0.1882 c = 6.0423, R^2 = 0.9394
Wordsworth, W. *Beloved vale*	A, V, V, A, V, V, A, A, V, V, V, V, A, A, A, V, V, V, A, A, A, A, A, A, V, V, V, V, V, V	a = 26.5687 b = 0.3411 c = 15.0893 R^2 = 0.9409
Rossetti, D. G. *A Venetian Pastoral*	V, V, V, V, V, A, V, A, V, V, A, V, V, A, V, A, V, V, V, A, A, A, V, V, V, V, V, V	a = 82.6872 b = 0.3877 c = 8.6209 R^2 = 0.9690
Coleridge, S.T. *To Southey*	V, A, A, A, A, V, A, V, A, A, V, A, A, A, A, A, V, A, V, A, V, A, V, A, A, A, V, A, A, A, A, V, A, A, V, A, A	a = 11.4076 b = 0.3368 c = 31.6485 R^2 =0.9636
Baudelaire, Ch.	A, A, V, A, V, A, A, A, V, A, A,	a = 18.4902

Bohémiens en voyage	V, A, A, A, A, V, V, V, V, V, V, A, A, A	b = 0.9183 c = 12.9440 R^2 = 0.9506
Corbière, T. *A l'éternel madame*	A, A, V, V, V, V, V, A, A, V, V, A, A, V, A, V, V, V, V, A, A, A, V, A, A, A, A, A, V, V, V, V, V, V	a = 21.8486 b = 0.2673 c = 18.4085 R^2 = 0.9438
Mallarmé, S. *Angoisse*	V, V, V, V, A, A, A, V, V, A, V, A, V, A, V, V, A, V, A, V, V, A, A, A, V, V, V	a = 31.1126 b = 0.3552 c = 12.1573 R^2 = 0.9584
Machar, J. S. *Sonet cynický*	A, A, V, V, A, V, V, V, A, A, V, A, V, V, V, V, A, V, A, A, V	a = 42.4483 b = 0.6870 c = 12.4431 R^2 = 0.9632
Karásek ze Lvovic, J. *Horečka*	A, V, A, V, V, A, A, A, A, V, V, V, V, A, V, A, V, V, V, V	a = 20857303.2 b = 0.3903 c = 6766529.28 R^2 = 0.9551
Kvapil, J. *Duše*	A, A, V, V, V, A, A, A, V, A, A, V, A, V, A, V, V, A, A, V, A, V, V, A, A, V	a = 17.6071 b = 0.2834 c = 14.9519 R^2 = 0.9675
Hviezdoslav, *Sonety 1*	V, A, A, V, V, V, A, V, V, V, A, V, V, V, V, V, A, A, V, A, V, V, V	a = 61.2533 b = 1.0198 c = 14.8579 R^2 = 0.9720
Hviezdoslav, *Sonety 2*	A, A, A, V, V, V, V, V, V, V, V, V, V, A, V, V, V, A, V, V, A, A, A, V, V, V, V	a= 94066.78 b = 2.9609 c = 15.7307 R^2 = 0.9579
Faludi F., *A Pipárul*	V, A, V, A, A, A, V, V, V, V, V, V, V, V, V, V, V, V, A, V, V, V, V	a =2008125.4 b = 3.1348 c = 18.2865 R^2 = 0.9692

9. H-point and Lambda Indicator: Generalities

The h-point has been used in linguistics since its introduction by Popescu (2007). It is appropriate for any kind of ranking in language, and divides the low-ranked entities from the higher ranked ones. A number of applications can be found, e.g., in Popescu et al. (2010), or Tuzzi, Popescu, Altmann (2010).

The h-point is defined as

$$h = \begin{cases} r, & if\ there\ is\ an\ r = f_r\ ; \\ \frac{f_i r_j - f_j r_i}{r_j - r_i + f_i - f_j} & if\ there\ is\ no\ r = f_r\,. \end{cases} \tag{9.1}$$

Consider, for example, the distribution of parts-of-speech in *Letní sonety* by Machar in Table 2.13. Here, we have the numbers presented in Table 9.1.

Table 9.1
Frequencies of POS in rhyme words of Machar's *Letní sonety*

1	272
2	186
3	105
4	63
5	25
6	4
7	2
8	1
9	1

Since is no $r = f_r$, we use the second line of the formula. Since r = 5 is smaller than $f_5 = 25$, and r = 6 is greater than $f_6 = 4$, we compute

$$h = \frac{25 * 6 - 4 * 5}{6 - 5 + 25 - 4} = 5.91\,.$$

Hence, the h-point is between the ranks 5 and 6. The concentration of the given phenomenon, here POS, can be computed by means of the formula

$$C = 2 * \sum_{r=1}^{m} \frac{|h - r'| * f(r')}{m(m-1) * f(1)}, \tag{9.2}$$

where *r'* are the ranks smaller than *h,* and *f(r')* are the frequencies of these ranks, and *m* is the number of ranks smaller or equal to *f(h).* In the above case, our *m* = 5 (smaller than 5.91), and we obtain

$$C = 2 * \frac{|5.91 - 1| * 272 + |5.91 - 2| * 105 + |5.91 - 4| * 63 + |5.91 - 5| * 25}{5 * 4 * 272} = 0.9233$$

The POS-concentration is very high; some POS occur only peripherally. This fact shows that in the rhyme of the sonnet, not all POS play the same role. Every language and author have their own styles, preferences for the rhyme words, etc.

In Table 9.2, we present concentrations of some entities discussed in this book.

Table 9.2
H-points and concentrations of entities

Author and work	Entity	h	C
Machar, *Letni sonety*	POS in rhyme	5.91	0.92
Hviezdoslav, *Sonety*	POS in rhyme	4.00	0.61
Hviezdoslav, *Krvavé sonety*	POS in rhyme	5.64	0.64
Goethe, *Sonette*	POS in rhyme	4.50	0.88
Machar, *Letni sonety*	Motifs of POS	8.17	0.49
Hviezdoslav, *Sonety*	Motifs of POS	4.67	0.71
Hviezdoslav, *Krvavé sonety*	Motifs of POS	7.25	0.40
Goethe, *Sonette*	Motifs of POS	4.87	0.99
Machar, *Letní sonety*	Busemann	3.80	1.37
Hviezdoslav, *Sonety*	Busemann	3.00	0.89
Hviezdoslav, *Krvavé sonety*	Busemann	3.00	0.83
Goethe, *Sonette*	Busemann	2.67	1.83
József, *A kozmosz ének*	Busemann	3.00	0.80

As can be seen in this short survey, the h-point lays in a relatively small interval, and its study thus does not bring new vistas. Nevertheless, for longer texts, it is a good indicator. To this aim, it is more appropriate to consider all words (types, or tokens) of a sonnet, and compute the lambda indicator. This indicator (cf. Popescu, Čech, Altmann 2011) is defined on the basis of the approximated Euclidean distance between the neighbouring frequencies (L*), and then relativized (cf. Popescu, Altmann 2015) in the form

$$\Lambda^* = \frac{L^*(\log N)}{N} = \frac{(V + f_1 - h - 1) * (\log N)}{N}, \tag{9.3}$$

where V is the number of different words (vocabulary), N is the frequency of all words, and f_1 is the frequency of the most frequent word. In order to compare texts, the variance of the indicator is given as

$$Var(\Lambda^*) = \frac{Var(f_1) * (\log N)^2}{N^2} = \frac{f_1 * (N - f_1) * (\log N)^2}{N^3}, \tag{9.4}$$

so that a normal test can be performed for two texts in the form

$$u = \frac{|\Lambda_1^* - \Lambda_2^*|}{\sqrt{Var(\Lambda_1^*) + Var(\Lambda_2^*)}}. \tag{9.5}$$

Let us consider the sonnet *To Burke* by S. T. Coleridge; it contains 118 tokens in 91 types, the most frequent word being represented six times. After calculating the h-point, which equals four, one obtains the lambda count –

$$\Lambda_{Col}^* = \frac{(91 + 6 - 4 - 1) * (\log 118)}{118} = 1.62 .$$

Now, we will focus our attention to the French sonnet *La beauté* ("The Beauty") by Charles Baudelaire. There are 123 tokens in 90 types, and the most used unit appears five times. The value of the h-point being 3.5, the lambda indicator yields –

$$\Lambda_{Bau}^* = \frac{(90 + 5 - 3.5 - 1) * (\log 123)}{123} = 1.54 .$$

As the two texts are of different lengths, the results need to be tested on statistical significance. To perform the test, variances will be counted; the calculations gave –

$$Var(\Lambda_{Col}^*) = \frac{6 * (118 - 6) * (\log 118)^2}{118^3} = 0.0018 ,$$

and

$$Var(\Lambda_{Bau}^*) = \frac{5 * (123 - 5) * (\log 123)^2}{123^3} = 0.0014 .$$

Stemming from the aforedone counts, the u-test value is –

$$u = \frac{|1.62 - 1.54|}{\sqrt{0.0018 + 0.0014}} = 1.36 ,$$

which indicates no significant difference between the two samples (u < 1.96).

Table 9.3 shows the results in more sonnets.

Table 9.3
The lambda indicator values in selected English and French poems

Poem	**f(1)**	**V**	**N**	**h-point**	**Λ***
S. T. Coleridge *To Burke*	6	91	118	4	1.62
D. G. Rossetti *Life-in-Love*	7	85	121	4	1.50
G. Meredith *I*	6	100	127	3	1.68
A. C. Swinburne *Dickens*	7	88	120	4	1.56
W. Wordsworth *London, 1802*	5	90	110	3.5	1.68
Poem	**f(1)**	**V**	**N**	**h-point**	**Λ***
Ch. Baudelaire, *La beauté*	5	90	123	3.5	1.54
S. Mallarmé, *Le sonneur*	7	89	123	4	1.55
S. Prudhomme, *Immortelle*	6	92	126	4	1.56
J.-A. Rimbaud, *Au Cabaret-Vert*	5	84	116	3.5	1.51
P. Verlaine, *Circonspection*	6	93	116	3	1.68

10. Frequency Studies

10.1 General Measurements

The number of words in a sonnet is small. The most frequent are the prepositions, conjunctions, articles, etc. Nevertheless, one can search for a distribution/ function capturing it and characterize the word stock using some indicators, such as repeat rate, entropy, lambda, h-point, arc length, Ord's criterion, etc., which can be either ordered, or compared. One can also perform tests for equality, search for the differences between languages, analyze the style of the authors, their developments, or the development of the sonnet in a language, etc. That means one can perform the usual analysis of word frequencies. Some of these will be performed in this chapter.

One knows that in a short poem, one cannot repeat the same word many times; hence, the majority of them occurs exactly once. As a matter of fact, one should count the types (not the tokens) because in synthetic languages the number of different tokens will be automatically greater. In our investigation, we do not join some tokens. We present some results without showing the complete distributions in Table 10.1. Again, it is not relevant whether we consider the frequencies as a distribution, or a sequence to which a function can be fitted. Usually, one applies the Zipf (power) function or alternatively the Zipf-Mandelbrot function, but it can be shown that the simple exponential function with added 1 fulfils this task excellently.

Table 10.1
Some properties of word frequency distributions

Text	Parameters	V, N	Ord
Goethe: *Sonette* 1	a = 3.2660 b = 0.1227 R^2 = 0.9228	V = 78 N = 101	I = 18.3666 S = 8.6431
Goethe: *Mächtiges Überraschen*	a = 7.7164 b = 0.3549 R^2 = 0.9410	V = 80 N = 97	I = 19.0684 S = 5.6185
Goethe: *Freundliches Begegnen*	a = 4.8653 b = 0.2295 R^2 = 0.9644	V = 82 N = 100	I = 19.1877 S = 6.3107
Goethe: *Kurz und gut*	a = 3.3743 b = 0.2416 R^2 = 0.9088	V = 89 N = 101	I = 18.8787 S = 3.8508
Goethe: *Das Mädchen spricht*	a = 4.9135 b = 0.1787 R^2 = 0.9450	V = 79 N = 104	I = 19.3185 S = 9.1221

Hviezdoslav: *Sonety* 1	a = 4.6882 b = 0.1879 R^2 = 0.9592	V = 90 N = 112	I = 21.3530 S = 7.8988
Hviezdoslav: *Sonety 2*	a = 12.8619 b = 0.6780 R^2 = 0.9242	V = 86 N = 98	I = 19.0172 S = 3.2150
Hviezdoslav: *Sonety 3*	a = 4.1277 b = 0.1670 R^2 = 0.9494	V = 88 N = 109	I = 20.7572 S = 7.5788
Hviezdoslav: *Sonety 4*	a = 3.1441 b = 0.1726 R^2 = 0.9122	V = 87 N = 103	I = 19.2458 S = 5.6990
Babits: *A lirikus epilógja*	a = 4.1905 b = 0.2155 R^2 = 0.9486	V = 71 N = 87	I = 16.5493 S = 5.7211
József: *A kozmosz éneke 1*	a = 3.3743 b = 0.2416 R^2 = 0.9068	V = 72 N = 84	I = 15.6883 S = 4.1501
József: *A kozmosz éneke 2*	a = 3.5195 b = 0.3672 R^2 = 0.8682	V = 69 N = 76	I = 14.0696 S = 1.9085
József A. *A kozmosz éneke 3*	a = 18113.77 b = 9.1113 R^2 = 1.000	V = 68 N = 70	I = 12.0859 S = 0.2239
Machar: *Sonet ironický*	a = 2.9175 b = 4.6150 R^2 = 0.9147	V = 73 N = 84	I = 15.6744 S = 3.7169
Machar: *Sonet úvodni*	a = 2.8044 b = 0.2526 R^2 = 0.9131	V = 75 N = 84	I = 15.5444 S = 2.8222
Machar: *Sonet o třetí hodině v červenci*	a = 3.4345 b = 0.2107 R^2 = 0.9303	V = 77 N = 91	I = 17.0707 S = 4.9165
Machar: *Sonet o životě*	a = 2.2646 b = 0.1389 R^2 = 0.8705	V = 74 N = 88	I = 16.0328 S = 5.2465
Wordsworth: *Beloved vale!*	a = 18.4257 b = 1.7005 R^2 = 0.9110	V = 85 N = 119	I = 22.7656 S = 11.6212
Wordsworth: *The world is too much with us (...)*	a = 4.3394 b = 6.6499 R^2 = 0.9520	V = 91 N = 117	I = 21.8890 S = 9.5373
Wordsworth: *Nuns fret (...)*	a = 5.8718 b = 4.7244 R^2 = 0.9127	V = 88 N = 113	I = 21.6532 S = 9.0165
Wordsworth: *Great men have been among us (...)*	a = 5.5140 b = 3.0595 R^2 = 0.9669	V = 84 N = 98	I = 18.7807 S = 4.5041

Coleridge: *To Bowles*	a = 7.1060 b = 0.3613 R^2 = 0.9554	V = 89 N = 106	I = 20.5610 S = 5.5840
Coleridge: *To Fayette*	a = 10.4288 b = 0.4458 R^2 = 0.8609	V = 90 N =111	I = 21.8420 S = 7.1269
Coleridge: *To Erskine*	a = 6.5912 b = 0.2349 R^2 = 0.9628	V = 86 N = 109	I = 21.3988 S = 8.0923
Coleridge: *To Burke*	a = 4.7759 b = 0.2119 R^2 = 0.9564	V = 88 N = 107	I = 20.5963 S = 6.6285
Rossetti: *A Venetian Pastoral*	a = 23.0647 b = 0.8906 R^2 = 0.8537	V = 87 N = 112	I = 22.0482 S = 8.7455
Rossetti: *Life-in-Love*	a = 5.6897 b = 0.1821 R^2 = 0.9224	V = 80 N = 109	I = 20.1188 S = 10.4937
Rossetti: *Heart's Heaven*	a = 7.1292 b = 0.3207 R^2 = 0.9199	V = 85 N = 106	I = 20.3864 S = 7.4091
Rossetti: *Silent Noon*	a = 78.3952 b = 2.0610 R^2 = 0.9083	V = 89 N = 110	I = 21.3591 S = 7.2015
Byron: *Sonnet on Chillon*	a = 4.9107 b = 0.1985 R^2 = 0.9666	V = 85 N = 106	I = 20.3661 S = 7.5042
Byron: *Sonnet to Genevra*	a = 2.9263 b = 0.1363 R^2 = 0.8978	V = 83 N = 101	I = 18.8311 S = 6.6536
Byron: *Sonnet on the Nuptials*	a = 8.0301 b = 0.2641 R^2 = 0.9732	V = 81 N = 106	I = 20.9001 S = 8.8668
Byron: *Sonnet to Lake Leman*	a = 13.4244 b = 0.4111 R^2 = 0.9652	V = 84 N = 114	I = 22.2269 S = 10.5809
Baudelaire: *Bohémiens en voyage*	a = 8.3272 b = 0.3482 R^2 = 0.9830	V = 83 N = 103	I = 20.2415 S = 6.8416
Baudelaire: *L'ennemi*	a = 6.7492 b = 0.2292 R^2 = 0.9518	V = 92 N = 118	I = 22.7577 S = 9.3303
Baudelaire: *La beauté*	a = 3.9897 b = 6.1334 R^2 = 0.9163	V = 87 N = 115	I = 20.7706 S = 10.2521
Baudelaire: *La masque*	a = 3.8047 b = 0.1617 R^2 = 0.9142	V = 84 N = 107	I = 20.0202 S = 7.6923

Verlaine: *À Horatio*	a = 8.2986 b = 3.8171 R^2 = 0.9470	V = 80 N = 169	I = 20.8114 S = 10.3632
Verlaine: *Le clown*	a = 7.9544 b = 2.9751 R^2 = 0.9685	V = 89 N = 109	I = 21.3900 S = 6.7587
Verlaine: *Allégorie*	a = 5.2359 b = 3.3523 R^2 = 0.9566	V = 82 N = 97	I = 18.5639 S = 5.0207
Verlaine: *Le pitre*	a = 5.6536 b = 4.5850 R^2 = 0.9364	V = 93 N = 117	I = 22.3603 S = 8.5920
Rimbaud: *La maline*	a = 6.02386 b = 6.1411 R^2 = 0.9573	V = 85 N = 118	I = 21.9896 S = 11.9242
Rimbaud: *Le mal*	a = 6.0539 b = 4.4571 R^2 = 0.9169	V = 90 N = 114	I = 22.0133 S = 8.5596
Rimbaud: *Voyelles*	a = 50.1353 b = 0.6573 R^2 = 0.9521	V = 86 N = 106	I = 21.0497 S = 6.6006
Rimbaud: *Le dormeur du val*	a = 6.0763 b = 6.0775 R^2 = 0.9621	V = 84 N = 116	I = 21.8722 S = 11.6205
Mallarmé: *Angoisse*	a = 4.2644 b = 6.0903 R^2 = 0.9491	V = 97 N = 120	I = 22.8056 S = 8.2879
Mallarmé: *Le sonneur*	a = 7.8689 b = 3.6800 R^2 = 0.9027	V = 84 N = 112	I = 21.2717 S = 10.0358
Mallarmé: *Salut*	a = 3.3743 b = 4.1383 R^2 = 0.9047	V = 60 N = 72	I = 13.3392 S = 4.3443
Mallarmé: *Le pitre châtié*	a = 6.4691 b = 0.2544 R^2 = 0.9216	V = 87 N = 109	I = 21.2082 S = 7.7431
Trediakovskij: *Sonet iz seja…*	a = 69.9991 b = 5.9751 R^2 = 0.9955	V = 75 N = 88	I = 0.3948 S = 2.6058
Cheraskov: *Kol' budu v zhizni ya nakazan nishchetoyu*	a = 65.0182 b = 2.9114 R^2 = 0.9998	V = 68 N = 93	I = 0.9100 S = 5.8762
Cheraskov: *Sonet i efitafia*	a = 54.9977 b = 2.7894 R^2 = 0.9995	V = 78 N = 102	I = 0.7936 S = 4.5552
Rzhewskij: *Sonet, zakluchajushchij v sebe tri mysli*	a = 64.0203 b = 2.7418 R^2 = 0.9993	V = 79 N = 104	I = 0.5105 S = 2.8649

In Figure 10.1, one can see the relation between I and S.

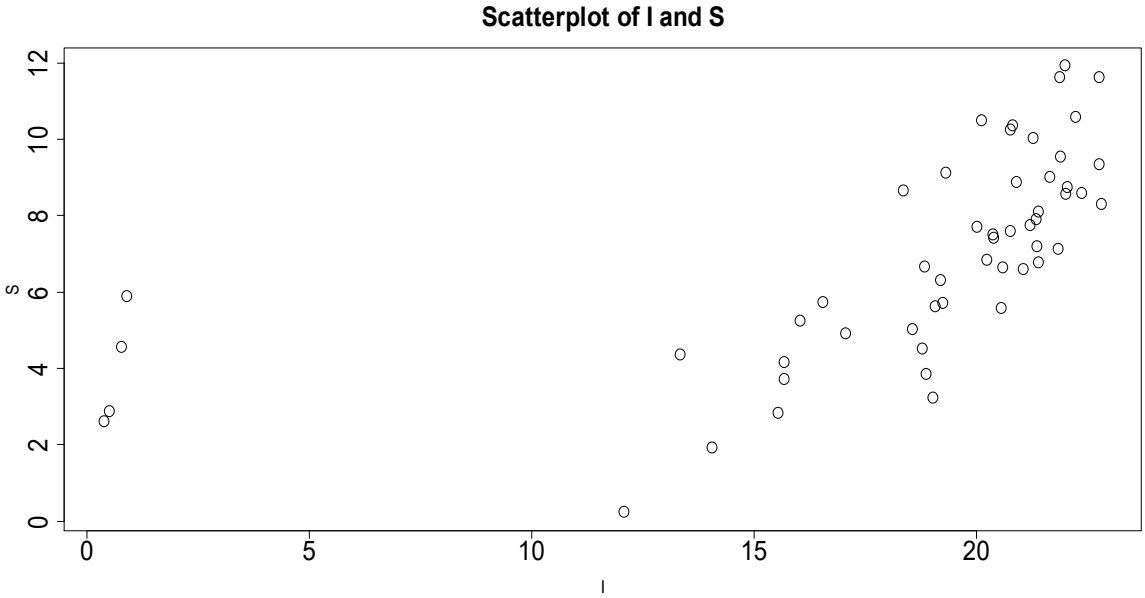

Figure 10.1
Scatterplot of I and S in some sonnets

If one replaces the ranking by the spectrum, one obtains another view of frequencies. Since the frequencies of $f_x = 1$ are the most current ones, we obtain, again, a monotonically decreasing function. In most cases, the most repeated token is not greater than 5, the majority of tokens occurring only once. In this case, the exponential function is a sufficient model, too. Some results are presented in Table 10.2.

Table 10.2
Token spectra of some sonnets (exponential function + 1)

Text	Results	Text	Results
Goethe: *Freundliches Wiedersehen*	1 72 71.99 2 5 5.21 3 3 1.25 4 1 1.01 5 1 1.00 a = 1197.9358 b = 0.3539 R^2 = 1.00	Hviezdoslav: *Sonety 1*	1 78 78.00 2 7 7.10 3 2 1.48 4 2 1.04 5 1 1.00 a = 971.8646 b = 0.3944 R^2 = 0.9997
Goethe: *Kurz und gut*	1 64 64.01 2 10 9.93 3 2 2.26 4 1 1.18	Hviezdoslav: *Sonety 4*	1 75 75.00 2 9 9.02 3 2 1.87 4 1 1.09

	5 2 1.03 a = 444.7589 b = 0.5117 R^2 = 1.00		a = 682.3599 b = 0.4504 R^2 = 1.00
Goethe: *Mächtiges Überraschen*	1 73 73.00 2 4 4.00 5 2 1.00 6 1 1.00 a = 1727.9329 b = 0.3147 R^2 = 0.9997	Hviezdoslav: *Sonety 3*	1 76 75.99 2 6 6.26 3 3 1.37 4 3 1.03 a = 1069.5565 b = 0.3763 R^2 = 0.9983
Goethe: *Sonett 1*	1 74 73.99 2 6 6.26 3 3 1.38 4 3 1.03 5 1 1.00 a = 1011.9742 b = 0.3803 R^2 = 0.9994	Hviezdoslav: *Krvavé sonety 1*	1 69 69.00 2 9 9.00 4 1 1.11 9 1 1.00 a = 578.3376 b = 0.4671 R^2 = 1.00
József: *A kozmosz éneke 1*	1 63 63.01 2 7 6.89 3 1 1.56 4 1 1.05 a = 652.5023 b = 0.4249 R^2 = 1.00	Hviezdoslav: *Krvavé sonety 2*	1 67 67.00 2 4 4.08 3 2 1.14 4 1 1.01 5 1 1.00 a = 1414.2637 b = 0.3263 R^2 = 0.9998
József: *A kozmosz éneke 2*	1 67 67.00 3 1 1.00 a = 12913.8559 b = 0.1895 R^2 = 1.00	Hviezdoslav: *Krvavé sonety 3*	1 77 77.00 3 1 1.03 4 2 1.00 5 1 1.00 a = 3852.4362 b = 0.2547 R^2 = 0.9998
József: *A kozmosz éneke 3*	1 65 65.00 2 1 1.01 3 3 1.00 a = 419870.91 b = 0.1137 R^2 = 0.9985	Hviezdoslav: *Krvavé sonety 4*	1 74 74.00 2 1 1.01 4 1 1.00 a = 440223.26 b = 0.1149 R^2 = 1.00
Machar: *Sonet úvodni*	1 68 68.00 2 6 6.00 4 1 1.03 a = 897.8841	Machar: *Sonet o životě*	1 64 64.02 2 10 9.66 3 1 2.19 4 1 1.16

	b = 0.3853 R^2 = 1.00		a = 458.4782 b = 0.5039 R^2 = 0.9994
Machar: *Sonet o třetí hodině v červenci*	1 70 69.99 2 6 6.24 3 3 1.40 a = 907.7494 b = 0.3880 R^2 = 0.9991	Machar: *Sonet ironický*	1 60 59.99 2 5 5.40 3 4 1.33 5 1 1.00 a = 751.0080 b = 0.3852 R^2 = 0.9970
Wordsworth: *Beloved vale!*	1 75 75.01 2 9 8.82 3 1 1.83 5 1 1.01 6 1 1.00 9 1 1.00 a = 699.9827 b = 0.4450 R^2 = 0.9998	Wordsworth: *Great men have been among us (...)*	1 74 74.00 2 6 5.95 3 1 1.34 4 1 1.02 5 1 1.00 a = 1075.7309 b = 0.3717 R^2 = 1.00
Wordsworth: *The world is too much with us (...)*	1 76 75.98 2 10 10.25 3 3 2.14 4 2 1.14 5 1 1.02 a = 607.7798 b = 0.4779 R^2 = 0.9996	Wordsworth: *Nuns fret (...)*	1 75 74.97 2 7 7.61 3 5 1.59 4 1 1.05 7 1 1.00 a = 727.0958 b = 0.4149 R^2 = 0.9971
Coleridge: *To Bowles*	1 80 80.00 2 6 5.96 3 1 1.31 5 1 1.00 6 1 1.00 a = 1259.0832 b = 0.3613 R^2 = 1.00	Coleridge: *To Burke*	1 70 69.99 2 4 4.27 3 4 1.15 4 1 1.01 5 1 1.00 a = 1455.8699 b = 0.3279 R^2 = 0.9978
Coleridge: *To Erskine*	1 78 78.00 2 3 3.00 3 1 1.05 4 3 1.00 5 1 1.00 6 1 1.00 a = 2862.5027 b = 0.2740 R^2 = 0.9992	Coleridge: *To Fayette*	1 81 81.00 2 3 3.22 3 5 1.06 9 1 1.00 a = 2894.6912 b = 0.2789 R^2 = 0.9966

Rossetti: *A Venetian Pastoral*	1 57 57.00 2 4 4.09 3 2 1.17 11 1 1.00 a = 1014.2782 b = 0.3452 R^2 = 0.9997	Rossetti: *Heart's Heaven*	1 73 73.00 2 8 8.06 3 2 1.69 4 1 1.07 7 1 1.00 a = 3953 b = 0.1029 R^2 = 1..00
Rossetti: *Life-in-Love*	1 63 62.96 2 10 10.53 3 4 2.46 4 2 1.23 7 1 1.00 a = 402.9378 b = 0.5341 R^2 = 0.9988	Rossetti: *Silent Noon*	1 72 72.00 2 10 10.00 11 1 1.00 a = 560.1111 b = 0.4842 R^2 = 1.00
Byron: *Sonnet on Chillon*	1 73 72.99 2 5 5.22 3 3 1.25 4 2 1.01 5 1 1.00 a = 1229.5451 b = 0.3524 R^2 = 0.9990	Byron: *Sonnet on the Nuptials*	1 71 71.00 2 3 3.00 3 1 1.06 4 3 1.00 6 1 1.00 7 1 1.00 a = 2448.0033 b = 0.2813 R^2 = 0.9990
Byron: *Sonnet to Genevra*	1 68 67.93 2 9 10.02 3 6 2.22 a = 496.6504 b = 0.4989 R^2 = 0.9937	Byron: *Sonnet to Lake Leman*	1 71 71.01 2 8 7.87 3 1 1.67 5 1 1.01 7 1 1.00 10 1 1.00 a = 713.6084 b = 0.4307 R^2 = 0.9999
Baudelaire: *Bohémiens en voyage*	1 74 74.00 2 5 4.98 3 1 1.22 4 1 1.01 5 1 1.00 7 1 1.00 a = 1340.2144 b = 0.3436 R^2 = 1.00	Baudelaire: *L'Ennemi*	1 79 78.99 2 6 6.38 3 4 1.37 4 2 1.03 7 1 1.00 a = 1131.2247 b = 0.3739 R^2 = 0.9983

Baudelaire: *La beauté*	1 68 68.04 2 14 13.55 3 2 3.35 4 2 1.44 5 1 1.08 a = 358.0924 b = 0.5968 R^2 = 0.9993	Baudelaire *La masque*	1 72 71.98 2 7 7.44 3 4 1.58 5 1 1.00 a = 782.5352 b = 0.4166 R^2 = 0.9983
Verlaine: *À Horatio*	1 66 66.00 2 9 9.00 3 2 1.99 5 1 1.01 7 2 1.00 a = 527.8820 b = 0.4774 R^2 = 0.9997	Verlaine: *Le clown*	1 80 89.00 2 4 4.15 3 3 1.13 5 1 1.00 7 1 1.00 a = 1981.4600 b = 0.3103 R^2 = 0.9993
Verlaine: *Allégorie*	1 73 73.00 2 6 5.95 3 1 1.34 4 1 1.02 5 1 1.00 a = 1046.7290 b = 0.3737 R^2 = 1.00	Verlaine: *Le pitre*	1 79 78.99 2 9 9.24 3 3 1.87 5 1 1.01 6 1 1.00 a = 738.2336 b = 0.4449 R^2 = 0.9997
Rimbaud: *La maline*	1 69 68.98 2 8 8.33 3 3 1.79 4 3 1.09 6 2 1.00 a = 630.6269 b = 0.4489 R^2 = 0.9982	Rimbaud: *Le mal*	1 78 77.99 2 5 5.43 3 5 1.26 4 1 1.01 7 1 1.00 a = 1337.6628 b = 0.3503 R^2 = 0.9969
Rimbaud: *Voyelles*	1 86 86.00 2 6 5.97 3 1 1.29 12 1 1.00 a = 1454.8985 b = 0.3521 R^2 = 1.00	Rimbaud: *Le dormeur du val*	1 75 74.99 2 5 5.45 3 5 1.27 5 3 1.00 6 1 1.00 a = 1229.8941 b = 0.3558 R^2 = 0.9956
Mallarmé: *Angoisse*	1 83 82.98 2 8 8.42 3 4 1.67 4 1 1.06	Mallarmé: *Le sonneur*	1 69 68.99 2 10 10.20 3 3 2.25 7 2 1.00

	5 1 1.01 a = 905.7828 b = 0.4163 R^2 = 0.9999		a = 502.1641 b = 0.0237 R^2 = 0.9995
Mallarmé: *Salut*	1 51 51.01 2 7 6.84 3 1 1.68 4 1 1.08 a = 428,4173 b = 0.4656 R^2 = 0.9997	Mallarmé: *Le pitre châtié*	1 77 77.00 2 3 3.23 3 5 1.07 7 1 1.00 a = 2588.7332 b = 3.5282 R^2 = 0.9962
Trediakovskij: *Sonet iz seja…*	1 70 70.00 2 1 1.01 4 4 1.00 a = 547547.68 b = 8.9791 R^2 = 9970	Cheraskov: *Sonet i efitafia*	1 65 65.00 2 9 8.99 3 2 2.00 4 1 1.12 9 1 1.00 a = 512. 3593 b = 2.0801 R^2 = 1.0
Cheraskov: *Kol' budu v zhizni ya nakazan nishchetoyu*	1 55 54.99 2 8 8.06 3 2 1.92 4 2 1.12 8 1 1.00 a = 412.6629 b = 2.0337 R^2 = 0.9996	Rzhewskij: *Sonet, zakluchajushchij v sebe tri mysli*	1 64 64.00 2 10 9.97 3 2 2.28 4 2 1.18 6 1 1.00 a = 442.4497 b = 1.9492 R^2 = 0.9997

The presentation in the previous table is not complete. We merely wanted to show that the spectrum abides by a simple exponential function, which can be considered a law of token spectra in sonnets. We considered the words in their written form. Mostly, parameter *a* is very great and should not be taken into account in comparisons. It attains very great values especially in short spectra. Parameter *b* can be considered characteristic for the given sonnet and may be used for comparisons as soon as many sonnets (in more languages) have been analyzed. One can study this parameter historically, too.

Of course, a phonetic transcription of the sonnets could yield somewhat different results. Many homophones would fall into the same class, and the spectrum could change.

In each case, the determination coefficient signalizes a very good fitting. It must be remarked that the study of sonnets is an infinite undertaking. We have touched only some problems in order to show the range of possibilities; however, comparisons with other text types, languages, historical and personal develop-

ments of an author, etc., must be omitted and will be performed in individual articles.

10.2 H-point and Lambda Indicator

Here, we will compute the h-point and the lambda indicators of all the above sonnets and order them. We shall not compare the sonnets using the normal test; we will use the spectra of the word frequencies. For example, the h-point of the German sonnet *Freundliches Wiedersehen* by Goethe yields

$$h = \frac{5 * 3 - 3 * 2}{5 - 3 + 3 - 2} = 3\,,$$

and the lambda yields

$$\Lambda = \frac{82 + 72 - 3 - 1}{100} \log(100) = 3\,.$$

The h-point and lambda values of the above analyzed sonnet spectra are presented in Table 10.3. It is to be remarked that if one finds in Table 10.2 some data in which there are a spectrum number and frequency equal to zero, for computing h-point, one must take it into account. For example, in *Mächstiges Überraschen* by Goethe, there is

1 73
2 4
5 2
6 1;

however, for the computing of the h-point one must take into account

3 0

because 3 is already greater than 0 (see Formula 9.1). Hence,

$$h = \frac{4 * 3 - 2 * 0}{4 - 0 + 3 - 2} = 2.4\,.$$

Table 10.3
Values of *h* and lambda in the spectra of the above sonnets

Text	h	Λ	Text	h	Λ
Goethe: *Freundliches Wiedersehen*	3.00	3.00	Hviezdoslav: 1	2.83	3.02
Goethe: *Kurz und gut*	2.89	2.72	Hviezdoslav: 4	2.88	3.09,
Goethe: *Mächtiges Überraschen*	2.40	3.08	Hviezdoslav: 3	3.00	3.27
Goethe: *Sonett 1*	3	2.94	Hviezdoslav: *Krvavé sonety* 1	2.70	2.89
József: *A kozmosz éneke 1*	2.71	3,10	Hviezdoslav: *Krvavé sonety 2*	2.67	3.00
József: *A kozmosz éneke 3*	1.97	3.48	Hviezdoslav: *Krvavé sonety 3*	1.97	3.27
József: *A kozmosz éneke 2*	1.98	3.17	Hviezdoslav: *Krvavé sonety 4*	1.99	3.43
Machar: *Sonet úvodni*	2.57	3.19	Machar, *Sonet ironický*	3.5	2.94
Machar: *Sonet o třetí hodině v červenci*	3	3.08	Machar: *Sonet o životě*	2.8	2.97
Wordsworth: *Beloved vale*	2.44	2.63	Wordsworth: *Great men have been among us (...)*	2.67	3.14
Wordsworth: *The world is too much with us (...)*	3	2.88	Wordsworth: *Nuns fret (...)*	3.67	2.88
Coleridge: *To Bowles*	2.67	3.15	Coleridge: *To Fayette*	4.17	3.06
Coleridge: *To Erskine*	2.33	3.00	Coleridge: *To Burke*	3.25	2.92
Rossetti: *A Venetian Pastoral*	2.67	2.57	Rossetti: *Heart's Heaven*	2.86	2.95
Rossetti: *Life-in-love*	3.33	2.59	Rossetti: *Silent noon*	2.73	2.92
Byron:*Sonnet on Chillon*	3.00	2.94	Byron: *Sonnet on the nuptials*	2.33	2.83
Byron: *Sonnet to Genevra*	3.43	2.91	Byron: *Sonnet to Lake Leman*	2.75	2.73
Baudelaire: *Bohémiens en voyage*	2.60	3.00	Baudelaire: *L`Enemie*	3.33	2.93
Baudelaire: *La beauté*	2.92	2.71	Baudelaire: *La masque*	2.75	2.89
Verlaine: *A Horatio*	2.88	1.87	Verlaine: *Le clown*	3.00	3.08
Verlaine: *Allegorie*	2.67	3.10	Verlaine: *Le pitre*	3.00	2.97
Rimbaud: *La maline*	3.00	2.63	Rimbaud: *Le mal*	3.40	2.95
Rimbaud: *Voyelles*	2.67	3.22	Rimbaud: *Le dormeur du val*	3.33	2.75
Mallarmé: *Angoisse*	3.25	3.05	Mallarmé: *Le sonneur*	3.00	2.73
Mallarmé: *Salut*	2.71	2.82	Mallarmé: *Le pitre châtié*	3.33	3.19

Trediakovskij, *Sonet iz seja…*	1.99	3.14	Cheraskov: *Sonet i efitafia*	2.88	2.74
Cheraskov: *Kol` budu v zhizni ja nakazan, nishchetoju*	2.86	2.52	Rzhewskij: *Sonet, zakluchajushchij v sebe tri mysli*	2.89	2.70

Preliminarily, no classification can be made according to the h-point. However, the h-point and the lambda indicatiors, if ordered increasingly, can be captured by the exponential function. Here, a very broad investigation is necessary in order to say something about the relation of these two indicators to the language or to the authors. To do the least, the present results can be shown graphically. Since lambda considers also the lenth of the text, we present the results in form [Λ,h], as shown in Figure 10.2.

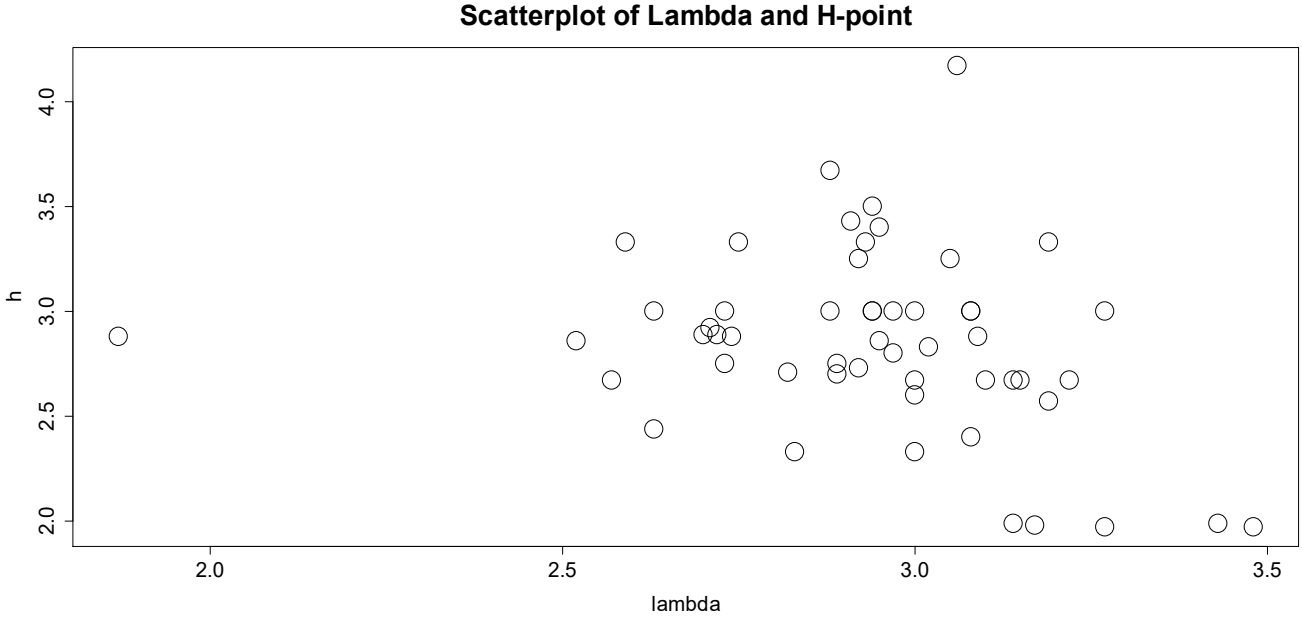

Figure 10.2
Scatterplot of lambda and h-point

10.3 Repeat Rate and Herfindahl's Indicator of Concentration

In order to characterize a text, one considers the distribution of words (types and tokens) and sets up a function expressing it. Since short poems cannot have an easily expressed distribution, one can capture it rather applying an indicator – e.g., repeat rate, or Shannon's entropy, which sufficiently describe the situation. One can do the same computation with any kind of characterization. Tests for differences are possible, too.

If one compares word frequencies, one has always a number of frequency classes, but if one compares, e.g., the types of Busemann coefficients yielding only 5 classes, one must relativize the result by a simple operation.

If one has some numbers/frequencies, one can always compute the usual repeat rate and the entropy. The repeat rate shows the concentration of entities in one direction, and the entropy does the same in another. They can be transformed into each other (cf. Altmann, Köhler 2015: 33–40). As a matter of fact, they can be interpreted also as indicators of monotony, stability, etc. In order to show some examples, we consider the occurrence of adjectives in some sonnets using the results presented in Tables 8.1, 8.2, and 8.3. We can see that individual sonnets do not contain the same number of adjectives, and want to study the stability of their use. For the three collections, *Sonety* and *Krvave sonety* by Hviezdoslav, and for sonnets by Goethe, we obtain the numbers presented in Table 10.4.

Table 10.4
Adjectives in sonnets by Hviezdoslav and Goethe

Texts	Number of adjectives in individual sonnets	k	N
Goethe	7,4,4,10,7,3,7,3,6,6,4,6,1,9,3,5,5,7,5	19	102
Hviezdoslav, *Krvavé sonety*	7,5,11,7,9,7,7,6,1,5,5,7,4,5,11,5,9,2,2,9,6,12, 4,4,2,4,7,7,5,6,7,10	32	198
Hviezdoslav, *Sonety*	7,8,7,6,6,6,7,11,7,5,10,10,9,9,7,4,12,5,7,10,7	21	160
József, *A kozmosz éneke*	10,20,16,11,10,16,13,12,14,8,18,3,13,9,12	15	185
Machar, *Letní sonety*	8,8,8,13,9,11,6,8,12,10,6,6,8,7,7,16,5,4,8,2,13,10,9, 7,13,10,6,7,4,7,5,11,7,5,18,10,10,6,9,9,10,5,5,9,5,8,9	47	389

The repeat rate, called also Herfindahl's (1950) indicator of concentration, can be computed as

$$R = \sum_{i=1}^{k} p_i^2 , \tag{10.1}$$

where k is the number adjectives, p is the relative frequency of a particular adjective –

$$p_i = \frac{x_i}{N} \text{ –}, \tag{10.2}$$

and N is the sum of all adjectives. For example, for Goethe, we obtain

$$R = \frac{(7^2 + 4^2 + 4^2 + \cdots + 5^2 + 7^2 + 5^2)}{102^2} = 0.0611\,;$$

since this number is too small, one usually relativizes it using the formula

$$R_{rel} = \frac{1 - R}{1 - \frac{1}{N}}, \tag{10.3}$$

or, according to McIntosh (1967),

$$R_{rel} = \frac{1 - \sqrt{R}}{1 - \frac{1}{\sqrt{N}}}. \tag{10.4}$$

In this way, we obtain 0.9482 for Goethe, 0.9549 for *Sonety* by Hviezdoslav, and 0.9677 for *Krvavé sonety* by Hviezdoslav. The Hungarian sonnets by József yield R_{rel} = 0.9210.

According to McIntosh, we would obtain 0.8355for Goethe, 0.8404 for *Sonety* by Hviezdoslav, and 0.8690for *Krvavé sonety* by Hviezdoslav. That means that in *Krvavé sonety* by Hviezdoslav, there is the greatest concentration of adjectives, etc. The above indicator can be transformed into several other ones (cf. Altmann, Lehfeldt 1980: 181; Altmann, Köhler 2015) well-known from statistics, hence their computations may be omitted.

Some more results are presented in Table 10.5.

Table 10.5
Herfindahl's Repeat Rate in English and Czech sonnets

Texts	**Number of Adjectives in Individual Sonnets**	**k**	**N**	**R**	**R (rel)**	**R (McIntosh)**
Wordsworth	13, 15, 20, 14, 14, 15, 14, 9, 17, 11	10	142	0.10	0.90	0.74
Coleridge	26, 23, 22, 23, 21, 22, 21, 23, 24, 16	10	221	0.10	0.90	0.73
Rossetti	8, 17, 14, 7, 19, 7, 10, 8, 17, 22	10	129	0.12	0.89	0.72
Machar, *Letní sonety*	8, 8, 8, 13, 9, 11, 6, 8, 12, 10, 6, 6, 8, 7, 7, 16, 5, 4, 8, 2, 13, 10, 9, 7, 13, 10, 6, 7, 4, 7, 5, 11, 7, 5, 18, 10, 10, 6, 9, 9, 10, 5, 5, 9, 5, 8, 9	47	389	0.02	0.98	0.89

11. Investigating Syllables

11.1 Type Distributions

In every text, one can state the types of syllables, count them, and find a model that fits them. One may conjecture that they are not uniformly distributed, as people prefer some types which are dictated by the language. Even in hexameter-poetry, only the lengths are dictated, not the syllable types. Concerning syllables, there are the following issues to be studied:

(1) What is the rank distribution of syllables?
(2) What is the distribution of syllable lengths?

In the first case, one orders the resulting frequencies simply from the greatest to the lowest one; in the second case, one sums the frequencies of syllables containing the same number of phonemes.

Let us consider the syllables in the first sonnet by J. W. Goethe. We obtain the sequence:

[CVC,VC,VC,CVC,CVC,CVCCC,CV,CCVC,CV,V,CVC,
VCC,CVC,CV,CCCVCC,CV,CVC,CVC,VC,VC,CV,CVC,
CV,CVCCC,CVC,VC,CV,CVC,CV,CCVCC,CV,CV,CVC,
CVC,CCVC,VCC,CCVC,CV,VC,CVC,CVC,CV,CCV,CVC,

CVC,V,CVC,CV,CV,CCVC,CVC,CVCC,CVC,CV,CVC,
CVC,CVC,CV,CV,CCVC,CVC,CV,CVC,CVC,CV,CVC,
VCC,CV,CV,CVC,CVC,VC,CV,CVC,CVC,CV,CVC,
CVC,CVCC,CV,CVCC,VCC,CVC,CVC,VC,CVC,CCV,CVC,

CV,CVCC,VC,CVCCCC,VC,CVCCC,CV,CVC,CV,CVC,CVC,
VC,CCCVC,CV,CVCC,CVC,CV,CV,CV,CVC,CCVC,CV,
CVC,CV,CCV,CVC,CV,CVC,CVC,CV,CV,CV,CVC,

CVC,CVC,VC,CVC,CVC,CVCC,CV,CVC, CV,CVC,CVC,
VC,CCV,CV,CVCCC,CV,CVCC,VC,CVC,CVC, CVC,CV,
VCC,CVC,CV,CVC,CVC,VC,CVC,VC,CVC,CV,CVC]

Having counted identical syllable types, we obtain the ranked results presented in Table 11.1. Since this is a kind of diversification, we can fit to the numbers with the usual exponential function with added 1, and obtain the displayed figures.

Table 11.1
Ranked distribution of syllable types in the first sonnet by Goethe

Type	Rank	Frequency	Exponential ft. +1
CVC	1	62	64.47
CV	2	44	35.90
VC	3	16	20.19
CVCC	4	7	11.55
CCVC	5	6	6.80
VCC	6	5	4.19
CCV	7	4	2.75
CVCCC	8	4	1.96
V	9	2	1.53
CCVCC	10	1	1.29
CVCCC	11	1	1.16
CVCCCC	12	1	1.08
CCCVC	13	1	1.05
	a = 115.4256, b = 0.5981, R^2 = 0.9731		

As can be seen, the result is as expected. Now, we add the types having the same length (in terms of phoneme numbers), and obtain the results presented in Table 11.2. Evidently, the distribution is bell-shaped, and the only discrete distribution giving satisfactory results is the Conway-Maxwell-Poisson one. In order to yield better results, we fit the Zipf-Alekseev function, which is a generalization of the exponential one.

Table 11.2
Lengths of syllables in the first sonnet by Goethe

Length	Frequency	Zipf-Alekseev ft. + 1
1	2	1.05
2	60	60.05
3	71	70.84
4	13	14.09
5	7	2.28
6	1	1.10
a = 16.5271, b = -8,9928, c = 0.0470, R^2 = 0.9950		

In German, the syllables may be very complex, as can be seen in Table 11.1. However, in a language like Hungarian, which has a very complex morphology, the syllables are relatively simple. Consider, e.g., the first sonnet from the book *A kozmosz éneke* by Attila József (1923), for which one obtains the ranking presented in Table 11.3. Nevertheless, the exponential function is adequate here, too.

Table 11.3
Ranking of syllable types in the Hungarian sonnet Nr. 1 by Attila József

Type	Rank	Frequency	Exponential ft + 1
CVC	1	66	68.92
CV	2	47	37.37
VC	3	14	20.47
CVCC	4	10	11.43
V	5	6	6.58
CCVC	6	2	3.99
CCVCC	7	2	2.60
CCV	8	1	1.86
a = 126.8564, b = 0.6249, R^2 = 0.9639			

The lengths of syllables follow the same model as the German sonnet, as can be seen in Table 11.4.

Table 11.4
Syllable lengths in the Hungarian sonnet Nr. 1 by Attila József

Length	Frequency	Zipf-Alekseev ft. + 1
1	6	1.04
2	61	61.02
3	67	66.97
4	12	12.08
5	2	1.96
a = 16.8657, b = -9.2828, c = 0.0435, R^2 = 0.9939		

The data for many other sonnets are presented in Tables 11.5, 11.6, and 11.7. To make sense of the abbreviations, see the Appendix to the book.

Tables 11.5, 11.6, and 11.7
Syllabic structures in selected sonnets

Types	H S 1	H S 2	H S 3	H S 4	H S 5	H S 6	H S 7
V	13	11	11	14	11	7	4
VC	1	7	3	3	1	2	6
CV	74	80	88	80	90	94	76
VCC	–	–	1	–	–	–	–
CVC	41	33	35	44	38	41	51
CCV	26	24	15	15	13	14	16
VCCC	–	–	–	–	–	–	–
CVCC	1	2	1	2	1	–	4
CCVC	8	12	13	10	13	10	10

CCCV	4	2	1	2	2	4	4
VCCCC	–	–	–	–	–	–	–
CVCCC	1	–	–	–	–	–	–
CCVCC	2	–	2	1	–	–	2
CCCVC	–	3	–	3	2	1	1
CCCCV	–	–	–	–	–	–	–
CCCCVC	–	–	1	1	–	–	–
CCCVCC	–	–	–	–	1	–	–
C	–	–	–	–	–	–	–
CC	2	–	3	1	2	2	2
CCC	1	–	–	–	2	–	2
CCCC	–	1	–	1	1	–	–

Types	**BM**	**FF**	**JS 1**	**JS 2**	**JS 3**	**JS 4**	**JS 5**
V	6	6	6	9	12	10	16
VC	12	13	14	12	8	8	14
CV	51	25	47	48	49	46	49
VCC	1	–	–	1	–	1	–
CVC	71	59	66	70	74	72	58
CCV	1	–	1	2	–	–	1
VCCC	–	–	–	–	–	–	–
CVCC	5	8	10	3	3	10	9
CCVC	1	1	2	2		–	–
VCCCC			–			–	–
CVCCC			–			–	–
CCVCC			2			1	1

Types	**M LS 1**	**M LS 2**	**M LS 3**	**M LS 4**	**M LS 5**	**M LS 6**
V	12	9	7	9	9	12
VC	1	1	2	3	1	2
CV	69	63	73	92	63	73
CVC	16	15	17	16	26	13
CCV	35	37	26	33	24	27
CVCC	2	3	3	2	2	3
CCVC	9	10	10	11	11	9
CCCV	2	–	4	2	7	1
CCVCC	1	1	–	1	–	1
CCCVC	–	–	2	1	1	–
CCCCV	–	1	1	1	1	–
CCCCVC	–	1	–	–	–	–
CCC	–	–	–	–	1	1

Syllables consisting only of consonants (CC, CCC, CCCC) are frequent in Slavic languages. However, non-syllabic prepositions, etc., are considered phonetical proclitics, and are thus joined with the next word.

The data suggests at once a number of further problems, e.g.:

(a) What is the syllabic similarity of subsequent lines like?

(b) Does the similarity change if one considers also further lines, i.e., in distances of 2, 3, 4, …?

That means, does the Skinner hypothesis hold also for syllable types?

(c) Do the sonnets of a writer display some similar behavior?

(d) Which syllable type is preferred in sonnets of a writer?

(e) Can one find a repetitive, sequential, or other structure in all sonnets of a writer? Evidently, publishing many data will lead to very extensive research.

(f) One may investigate also the number of consonants in the syllables. In the above tables, one finds syllables up to five consonants; but in Polynesian languages there is maximally one consonant in the syllable. In any case, any such investigation must consider the phonemic form of the word, not the way it is written.

Let us begin with the usual ranking of syllables according to their frequencies. Evidently, we deal here with the well-known functions, such as Zipf's power function, exponential function, Zipf-Alekseev function, Zipf-Mandelbrot function, or Lorentzian function. The choice may be decided on the basis of the determination coefficient. Taking, for example, the first sonnet by Hviezdoslav, we obtain $R^2 = 0.9983$ for the exponential function, $R^2 = 0.9533$ for the power (Zipf) function, $R^2 = 0.9964$ for the Mandelbrot function, $R^2 = 0.9964$ for the Zipf-Alekseev function, and $R^2 = 0.9950$ for the Lorentzian function. Except for the Lorentzian, we added always 1 to the function – i.e., we derived it from the differential equation

$$\frac{dy}{y-1} = f(x)\,dx\,. \tag{11.1}$$

The exponential function seems to be not only the best one, but also the simplest one from the linguistic point of view. The results are presented in the following tables.

Table 11.8
Fitting the exponential function to the rank-order of syllable frequencies: Hviezdoslav's *Sonety*

H S 1		**H S 2**		**H S 3**		**H S 4**	
Frequ.	**Comp.**	**Frequ.**	**Comp.**	**Frequ.**	**Comp.**	**Frequ.**	**Comp.**
74	74.05	80	77.55	88	86.77	80	80.27
41	41.76	33	40.29	35	38.37	44	41.31
26	23.75	24	21.17	15	17.28	15	21.50
13	13.69	12	11.35	13	8.09	14	11.42
8	8.08	11	6.31	11	4.09	10	6.30
4	4.95	7	3.73	3	2.35	3	3.69
2	3.21	3	2.40	3	1.59	3	2.37
2	2.23	2	1.72	2	1.26	2	1.70
1	1.69	2	1.37	1	1.11	2	1.35
1	1.38	1	1.19	1	1.05	1	1.18
1	1.21			1	1.02	1	1.09
1	1.12			1	1.01	1	1.05
						1	1.02
a = 130.8936, b = 0.5933, R^2 = 0.9983		a = 149.1285, b = 0.6669, R^2 = 0.9810		a = 196,8592, b = 0.8308, R^2 = 0.9867		a = 155.8876, b = 0.6763, R^2 = 0,9890	
H S 5		**H S 6**		**H S 7**		**H S 8**	
Frequ.	**Comp.**	**Frequ.**	**Comp.**	**Frequ.**	**Comp.**	**Frequ.**	**Comp.**
90	89.31	94	93.94	76	78.52	83	84.71
38	39.14	41	40.28	51	41.97	49	41.69
13	17.47	14	17.60	16	22.66	13	20.78
13	8.11	10	8.02	10	12.45	10	10.61
11	4.07	7	3.97	6	7.05	6	5.67
2	2.33	4	2.25	4	4.20	5	3.27
2	1.57	2	1.53	4	2.69	2	2.10
2	1.25	2	1.22	4	1.89	2	1.54
2	1.11	1	1.09	2	1.47	2	1.26
1	1.05			2	1.25	1	1.13
1	1.02			2	1.13	1	1.06
1	1.01			1	1.07	1	1.03
1	1.00						
a = 204.4768 b = 0.8396 R^2 = 0.9875		a = 210.8822 b = 0.8612 R^2 = 0.9959		a = 146.6574 b = 0.6376 R^2 = 0.9762		a = 172.2150 b = 0.7214 R^2 = 0.9829	

Table 11.9
Fitting the exponential function to the rank-order of syllable frequncies:
Hviezdoslav's *Krvavé sonety*

H KS 1		**H KS 2**		**H KS 3**		**H KS 4**		**H KS 5**	
Fr.	**Cp.**	**Fr.**	**Cp.**	**Fr.**	**Cp.**	**Fr.**	**Cp.**	**Fr.**	**Cp.**
68	66.01	58	59.08	59	59.73	67	65.92	68	69.50
24	30.18	38	33.66	38	34.21	30	33.32	40	34.78
14	14.09	15	19.36	14	19.78	19	17.09	15	17.66
12	6.88	11	11.32	13	11.62	8	9.01	6	9.22
11	3.64	7	6.80	8	7.00	7	4.99	4	5.05
3	2.18	4	4.26	6	4.39	5	2.99	3	3.00
3	1.53	4	2.83	3	2.92	4	1.99	3	1.99
3	1.24	3	2.03	2	2.08	3	1.49	2	1.49
3	1.11	3	1.58	1	1.61	2	1.25	2	1.24
2	1.05	2	1.33	1	1.35	1	1.12	1	1.12
2	1.02	1	1.18	1	1.20	1	1.06	1	1.06
1	1.01	1	1.10	1	1.11			1	1.03
								1	1.01
a = 144.8436 b = 0.8011 R^2 = 0.9659		a = 103.3096 b = 0.5759 R^2 = 0.9873		a = 103.8811 b = 0.5702 R^2 = 0.9849		a = 130.3887 b = 0.6973 R^2 = 0.9919		a = 138,8723 b = 9,7968 R^2 = 0.9898	

Table 11.10
Fitting the exponential function to the rank-order of syllable frequencies:
Selected Russian sonnets

	Nr 1		**Nr 2**		**Nr 3**		**Nr 4**		**Nr 5**	
Rank	**Frequ.**	**Exp+1**	**Frequ.**	**Exp+1**	**Frequ.**	**Exp+1**	**Frequ.**	**Exp+1**	**Frequ.**	**Exp+1**
1	94	93.30	69	70.96	85	85.73	97	96.25	89	87.82
2	41	43.80	49	40.99	47	42.77	37	39.68	36	40.35
3	24	20.85	16	23.86	14	21.59	17	16.70	22	18.84
4	8	10.21	12	14.06	14	11.15	10	7.38	9	9.09
5	8	5.27	9	8.47	9	6.00	8	3.59	7	4.67
6	4	2.98	8	5.27	3	3.47	2	2.05	5	2.66
7	2	1.92	6	3.44	2	2.22	1	1.43	3	1.75
8	1	1.43	4	2.39	1	1.60	1	1.17	2	1.34
9			1	1.80	1	1.30	1	1.07	1	1.15
10			1	1.46						
11			1	1.26						
	a = 199.0425, b = 0.7684, R^2 = 0.9955		a = 122.3924, b = 0.5593, R^2 = 0.9693		a = 171.8836, b = 0.7073, R^2 = 0.9855		a = 234.5943 , b = 0.9013, R^2 = 0.9956		a = 191.5495, b =0.7913, R^2 = 0.9933	

Tables 11.11, 11.12, and 11.13
Fitting the exponential function to the rank-order of syllable frequencies:
German and Hungarian sonnets

BM		**FF**		**GTA**		**GTD**		**RA**	
Fr.	**Exp+1**	**Fr.**	**Exp+1**	**Fr.**	**Exp+1**	**Fr.**	**Exp+1**	**Fr.**	**Exp+1**
71	74.74	59	58.40	51	51.79	50	51.31	46	48.22
51	38.22	25	26.98	32	28.46	39	32.95	38	30.61
12	19.78	13	12.76	12	15.85	14	21.30	15	19.57
6	10.48	8	6.32	8	9.03	13	13.89	9	12.64
5	5.79	6	3.41	6	5.34	12	9.19	8	8.30
1	3.42	1	2.09	5	3.35	7	6.20	8	5.58
1	2.22			3	2.27	7	4.30	6	3.87
1	1.62			3	1.69	2	3.10	5	5.58
				1	1.37	2	2.33	3	2.13
						1	1.85		
						1	1.54		
a = 146.0903 b = 0.6837 R^2 = 0.9479		a = 126.8020 b = 0.7926 R^2 = 0.9934		a = 93.9560 b = 0.6150 R^2 = 0.9849		a = 79.2140 b = 0.4539 R^2 = 0.9583		a = 75.3100 b = 0.4667 R^2 = 0.9460	

RATA (Ge)		**RBT (Ge)**		**RB (Ge)**		**RDS (Ge)**		**RJG**	
Fr.	**Exp+1**	**Fr.**	**Exp+1**	**Fr.**	**Exp+1**	**Fr.**	**Exp+1**	**Fr.**	**Exp+1**
54	54.79	58	55.65	55	55.82	43	44.59	43	47.29
37	33.19	28	31.75	35	30.89	36	28.72	40	31.68
15	20.27	14	18.31	12	17.29	11	18.63	22	21.34
14	12.53	12	10.74	9	9.88	10	12.21	11	14.48
7	7.90	12	6.48	8	5.84	9	8.13	10	9.93
7	5.13	7	4.08	5	3.64	8	5.54	5	6.92
5	3.47	7	2.74	4	2.44	6	3.89	4	4.93
2	2.48	5	1.98	2	1.78	5	2.83	2	3.60
2	1.89	1	1.55	1	1.43	2	2.17	1	2.72
1	1.53	1	1.31			2	1.74	1	2.14
						1	1.47		
						1	1.30		
a = 89.9682 b = 0.5133 R^2 = 0.9811		a = 97.10 b = 0.5749 R^2 = 0.9599		a = 100.5709 b = 0.6067 R^2 = 0.9793		a = 68.5404 b = 0.4526 R^2 = 0.9353		a = 69.8413 b = 0.4112 R^2 = 0.9508	

J S 1		**J S 2**		**J S 3**		**J S 4**		**J S 5**	
Fr.	**Exp+1**	**Fr.**	**Exp+1**	**Fr.**	**Exp+1**	**Fr.**	**Exp+1**	**Fr.**	**Exp+1**
66	68.92	70	73.25	74	76.87	72	72.13	58	62.15
47	37.37	49	37.79	49	38.58	46	37.35	49	37.39
14	20.47	12	19.73	12	19.61	10	19.07	16	22.66

10	11.43	9	10.54	8	10.22	10	9.99	14	13.89
6	6.58	3	5.86	3	5.57	8	5.47	9	8.67
2	3.99	2	3.47			1	3.22	1	5.56
2	2.60	2	2.26			1	2.10	1	3.72
1	1.86	1	1.64						
a = 126.8561		a = 141.8785		a = 153.1798		a = 147.0962		a = 102.7593	
b = 0.6247		b = 0.6748		b = 0.7026		b = 0.6989		b = 0.5190	
R^2 = 0.9639		R^2 = 0.9565		R^2 = 0.9514		R^2 = 0.9607		R^2 = 0.9292	

Tables 11.14
Fitting the exponential function to the rank-order of syllable frequencies: Czech sonnets

M LS 1		**M LS 2**		**M LS 3**		**M LS 4**		**M LS 5**	
Fr.	**Exp+1**	**Fr.**	**Exp+1**	**Fr.**	**Exp+1**	**Fr.**	**Exp+1**	**Fr.**	**Exp+1**
67	68.15	65	67.06	66	61.30	71	67.70	69	68.62
40	35.78	41	35.96	21	33.68	24	34.92	35	35.38
16	19.01	20	19.50	21	18.71	22	18.24	16	18.48
8	10.33	9	10.79	15	10.60	21	9.77	12	9.89
7	5.83	2	6.19	13	6.21	3	5.46	9	5.52
4	3.50	1	3.74	8	3.82	3	3.27	2	3.30
2	2.30	1	2.45	4	2.53	1	2.15	2	2.17
		1	1.77	1	1.83			1	1.59
		1	1.41	1	1.45			1	1.30
				1	1.24				
				1	1.13				
a = 129.6788		a = 124.8313		a = 111.2614		a = 131.1958		a = 132.9912	
b = 0.6581		b = 0.6364		b = 0.6125		b = 0.6764		b = 0.6764	
R^2 = 0.9899		R^2 = 0.9854		R^2 = 0.9241		R^2 = 0.9219		R^2 = 0.9938	

Now, the syllables can have different lengths. If we sum all frequencies of syllables of the same length, we obtain a function/distribution which is bell-shaped – that means the exponential function will not be adequate. However, the Zipf-Alekseev and the Lorentzian functions are appropriate, but in some cases the Zipf-Alekseev function yields quite different numbers for the length 1. Hence, we shall use the Lorentzian function and present the results in Table 11.15. The function, which has already been used in sections 2.2 and 3.1, follows the formula

$$y = \frac{a}{1 + \left(\frac{x - b}{c}\right)^2}, \tag{11.2}$$

which is a solution of the differential equation

$$\frac{y'}{y^2} + \frac{2(x-b)}{a * c^2} = 0\,, \tag{11.3}$$

where one sees that the dependent variable changes relatively to the square of its actual value and to the change of the independent variable around the parameter *b*. This is equivalent to the formula

$$\int \frac{2(x-b)}{a * c^2} dx + \int \frac{1}{y^2} dy = K\,, \tag{11.4}$$

yielding

$$\frac{a}{\frac{-b - a * c^2 * K}{c^2} + \left(\frac{x-b}{c}\right)}\,. \tag{11.5}$$

Choosing the constant

$$K = -\frac{c^2 + b^2}{a * c^2}\,, \tag{11.6}$$

we obtain (11.1).

The formula is used in various domains of science; in linguistics, it is appropriate if the bell-form of the data is not symmetrical. This is often the case when the variable is strongly language-dependent. Using this function, we look at the rather complex fact that the difference of *y* between neighbouring classes is not directly proportional to *y* (usually we have $\frac{y}{y'}$), but changes with the square of *y* – i.e., $\frac{y'}{y^2}$). As can be seen in (11.3), the expression can be presented also as *Ax* + *B*, and interpreted in the usual way: *A* represents the writer, *B* represents the language.

Table 11.15
Fitting the Lorentzian function to syllable lengths in sonnets:
Hviezdosl-av's *Sonety*

Length	**H S 1**		**H S 2**		**H S 3**		**H S 4**	
	Fr.	**Comp.**	**Fr.**	**Comp.**	**Fr.**	**Comp.**	**Fr.**	**Comp.**
1	13	12.59	11	14.18	11	13.37	14	15.35
2	77	77.03	87	86.74	90	89.84	84	83.86
3	68	67.98	57	57.60	51	51.51	55	55.36
4	12	11.82	17	11.36	15	10.04	15	11.97
5	3	4.49	3	4.47	3	3.97	4	4.80
6					1	2.10	1	2.55
	a = 188.5612 b = 2.4746 c = 0.3944 R^2 = 0.9995		a = 164.5122 b = 2.4100 c = 0.4331 R^2 = 0.9913		a = 172.2607 b = 2.3848 c = 0.4018 R^2 = 0.9947		a = 137.7668 b = 2.3965 c = 0.4946 R^2 = 0.9974	
Length	**H S 5**		**H S 6**		**H S 7**		**H S 8**	
	Frequ.	**Comp.**	**Frequ.**	**Cp.**	**Frequ**	**Comp.**	**Fr.**	**Comp.**
1	11	13.94	7	10.18	4	11.14	6	11.58
2	94	93.80	98	97.89	84	83.62	87	86.70
3	53	53.65	55	55.34	69	69.56	60	60.60
4	17	10.45	14	8.21	19	10.31	17	9.87
5	2	4.14	1	3.11	3	3.85	5	3.76
6	1	2.19						
	a = 179.9165 b = 2.3844. c = –0.4012 R^2 = 0.9915		a = 545.7072 b = 2.4182 c = 0.1955 R^2 = 0.9929		a = 322.3249 b = 2.4688 c = 0.2781 R^2 = 0.9780		a = 257.3364 b = 2.4377 c = –0.3120 R^2 = 0.9825	

Table 11.16
Fitting the Lorentzian function to syllable lengths in sonnets:
Hviezdoslav's *Krvavé sonety*

L	**H KS 1**		**H KS 2**		**H KS 3**		**H KS 4**		**H KS 5**	
	Fr.	**Comp.**	**Fr.**	**Comp.**	**Fr.**	**Comp.**	**Fr.**	**Comp.**	**Fr.**	**Comp.**
1	12	14.89	11	11.48	8	17.40	5	10.50	4	7.62
2	74	73.64	66	65.96	63	60.63	71	70.64	72	71.88
3	40	41.17	55	55.06	53	55.90	49	49.69	57	57.20
4	17	10.07	11	10.41	29	16.22	15	8.84	12	7.05
5	3	4.19	4	4.02	3	6.76	7	3.41	1	2.58
6									1	1.33
	a = 99.9526 b = 2.3334 c = 0.5578 R^2 = 0.9818		a = 138.2271 b = 2.4600 c = –0.4394 R^2 = 0.9998		a = 94.2396 b = 2.4670 c = –0.7485 R^2 = 0.9009		a = 165.1383 b = 2.4314 c = 0.3730 R^2 = 0.9761		a = 1833.7711 b = 2.4705 c = –0.0950 R^2 = 0.9919	

Table 11.17
Fitting the Lorentzian function to syllable lengths: selected Russian sonnets

L	T 1		T 2		T 3		T 4		T 5	
	Fr.	**Cp.**	**Fr.**	**Cp.**	**Fr.**	**Cp.**	**Fr.**	**Cp.**	**Fr.**	**Cp.**
1	8	9.85	12	14.41	9	12.99	10	10.48	7	9.94
2	98	97.94	78	77.73	87	86.72	99	98.98	94	93.90
3	65	65.13	65	65.43	61	61.58	54	54.08	58	58.27
4	11	8.55	18	13.05	18	10.97	10	8.29	13	8.32
5			1	5.08	1	4.23	1	3.16	2	3.13
6			2	2.66						
	a = 1710.3397 b = 2.4467 c = 0.1101 R^2 = 0.9984		a = 150.2663 b = 2.4590 c = –0.4751 R^2 = 0.9959		a = 201.8410 b = 2.4329 c = 0.3757 R^2 = 0.9862		a = 452.8212 b = 2.4105 c = –0.2171 R^2 = 0.9989		a = 613.3345 b = 2.4325 c = 0.1839 R^2 = 0.9950	

Tables 11.18 and 11.19
Fitting the Lorentzian function to syllable lengths: selected Hungarian and German sonnets

L	BM (Hu)		FF (Hu)		GTA (Ge)		GTD (Ge)		RA (Ge)	
	Fr.	**Cp.**	**Fr.**	**Lor**	**Fr.**	**Cp.**	**Fr.**	**Cp.**	**Fr.**	**Cp.**
1	6	7.32	6	6.83	5	8.22	1	13.74	9	11.01
2	63	62.82	38	37.92	49	48.72	51	47.75	54	53.74
3	73	72.84	59	59.03	60	60.19	64	65.80	58	58.24
4	6	7.70	9	8.46	12	9.09	27	18.00	14	11.56
5					3	3.33	5	7.04	3	4.41
6					1	1.70	1	3.64		
	a = 312483.4 b = 2.5400 c = 0.0070 R^2 = 0.9995		a = 134.7784 b = 2.5852 c = –0.3662 R^2 = 0.9995		a = 157.8555 b = 2.5402 c = 0.3609 R^2 = 0.9942		a = 83.6851 b = 2.6246 c = 0.7199 R^2 = 0.9285		a = 110.7713 b = 2.5203 c = 0.5051 R^2 = 0.9956	

L	RATA (Ger)		RBH (Ge)		RB (Ge)		RDS (Ge)		RJG	
	Fr.	**Cp.**	**Fr.**	**Cp.**	**Fr.**	**Cp.**	**Fr.**	**Cp.**	**Fr.**	**Cp.**
1	2	12.38	5	9.85	6	9.94	6	12.05	10	11.42
2	51	49.01	42	41.07	47	46.72	47	45.69	62	61.85
3	63	64.16	77	77.31	60	60.19	56	56.91	52	52.24
4	22	15.07	19	14.87	14	11–59	20	14.28	13	10.36
5	7	5.79	1	5.34	2	4.34	3	5.58	2	4.03
6	1	2.99	1	2.68			2	2.90		
	a = 92.2697 b = 2.5867 c = –0.6245 R^2 = 0.9532		a = 106.0221 b = 2.6736 c = 0.5356 R^2 = 0.9862		a = 104.7175 b = 2.5644 c = 0.5065 R^2 = 0.9943		a = 80.1761 b = 2.5741 c = 0.6630 R^2 = 0.9717		a = 120.5292 b = 2.4500 c = –0.4723 R^2 0.9955	

Table 11.20
Fitting the Lorentzian function to syllable lengths: József's sonnets

L	J S 1		J S 2		J S 3		J S 4		J S 5	
	Fr.	**Cp.**	**Fr.**	**Cp.**	**Fr.**	**Cp.**	**Fr.**	**Cp.**	**Fr.**	**Cp.**
1	6	9.93	9	7.18	12	6.90	10	9.67	16	13.46
2	61	60.68	61	61.03	57	57.21	54	54.03	63	63.34
3	67	67.26	73	72.88	74	73.85	73	72–98	59	58.60
4	14	10.43	5	7.62	3	7.51	11	11.24	10	12.80
	a = 187.4127		a = 233562.48		a = 80828.29		a = 167.9780		a = 111.8245	
	b = 2.5195		b = 2.5222		b = 2.5319		b = 2.5600		b = 2.4786	
	c = –0.3595		c = 0.0084		c = –0.0142		c = 0.3856		c = 0.5471	
	R^2 = 0.9904		R^2 = 0.9972		R^2 = 0.9869		R^2 = 0.9999		R^2 = 0.9937	

Table 11.21
Fitting the Lorentzian function to syllable lengths: Czech sonnets

L	M LS 1		M LS 2		M LS 3		M LS 4		M LS 5	
	Fr.	**Cp.**	**Fr.**	**Cp.**	**Fr.**	**Cp.**	**Fr.**	**Cp.**	**Fr.**	**Cp.**
1	8	10.68	3	7.13	13	18.84	3	11.68	12	12.66
2	67	66.79	66	65.50	67	65.52	71	70.27	70	69.92
3	56	56.29	61	61.05	42	45.23	46	47.62	51	51.19
4	13	9.81	10	6.95	27	13.59	24	9.44	13	10.56
5			2	2.52	2	5.89	1	3.72	1	4.17
	a = 162.6828		a = 298768.18		a = 79.5148		a = 132.3544		a = 125.2305	
	b = 2.4657		b = 2.4906		b = 2.3468		b = 2.4134		b = 2.4251	
	c = –0.387		c = 0.0073		c = 0.7504		c = 0.4398		c = –0.4780	
	R^2 = 0.9934		R^2 = 0.9936		R^2 = 0.9068		R^2 = 0.9158		R^2 = 0.9953	

The Lorentzian can be used in any case though in several cases, the parameter *a* acquires very great values. But since it is merely an equilibrating parameter, the function can be used here until it is necessary to apply other ones. The parameter *b* is quite constant, the parameter *c* is very small, and can even go below zero.

11.2 Comparisons

The comparison of syllabic structures of individual sonnets or of collections can be performed in two ways: (1) One compares the frequencies of identical types either directly, or by means of a non-parametric test; and (2) one simply compares the rank-frequency distributions. Consider, e.g., the syllabic structure of the first sonnet by Goethe in German, and the first one by József in Hungarian. Having put the above data in a common table, we obtain the data listed below.

Table 11.22
Syllable types in Goethe's and József's first sonnets

Syllable Types	Frequency in Goethe	Frequency in József
CVC	62	66
CV	44	47
VC	16	14
CVCC	7	10
CCVC	6	2
VCC	5	–
CCV	4	1
CVCCC	4	–
V	2	6
CCVCC	1	2
CVCCC	1	–
CVCCCC	1	–
CCCVC	1	–

Since sonnets have approximately the same length (in terms of word numbers), the comparison can be performed in many ways. Here, we compare the distance of the two vectors considering the cosine between them – i.e., we compute

$$\cos(\alpha) = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2}\sqrt{\sum y_i^2}}, \tag{11.7}$$

which yields, in the above case,

$$\cos(G,J) = \frac{62 * 66 + 44 * 47 + 16 * 14 \pm \cdots + 1 * 2}{\sqrt{62^2 + 44^2 + 16^2 + \cdots + 1^2} * \sqrt{66^2 + 47^2 + 14^2 + \cdots + 2^2}} = 0.9920\,;$$

therefrom, the distance is

$$\arccos(0.9920) = 0.1258\,.$$

The final number can be considered the distance. The *arccos* moves in the interval (0, 1.57).

Comparing by means of a chi-square test or a ranking test can yield a probability of identity, but in our case, one must manipulate the data because many frequencies are too small. The above way of evaluation can be used for any classification of sonnet elements.

Further problem concerning syllables in the sonnet are:

(1) Runs. The equal syllable types may follow one another, but they can be also separated by other types. In strongly inflecting languages, the number of runs is much greater than, e.g., in agglutinating or isolating languages. One can

conjecture that at least for a given poet, the results will be similar. However, it must be tested whether the authors are different, whether the language changes, what the difference between a strongly inflecting language and strongly agglutinative one is, etc.

(2) Distances between equal syllable types. This aspect must be correlated with the one of runs. If the runs are long, the mean distances are smaller. Again, it depends on the author and language.

(3) The syllabic similarity of subsequent lines of the sonnet. That means the subsequent pairs will be compared. The result of the comparison may be tested, or simply expressed by an indicator. There is a possibility of discovering a sequential development.

(4) Testing the Skinner hypothesis, conjecturing that lines lying nearer to one another are more similar than more distant ones. Hence, the results from point (3) can be used as distance 1, all the other ones must be added. Even if the author "corrects", i.e. changes something in the sonnets, similarities can appear on various levels. The syllable structure is merely one of many ways.

Syllable structures may be investigated also bidimensionally, as shown by Zörnig and Altmann (1993). One subdivides the syllable in two parts: the one in front of the vowel, and the one behind the vowel. One obtains the bivariate Conwell-Maxwell-Poisson distribution, but here, we shall omit this modelling.

12. Weighting

Whatever aspect of the reality is taken into account, one can always weight the properties, as it is done in natural sciences. Usually one speaks about quantification. Here, the individual degrees of a property are defined quantitatively, though sometimes – especially in the social sciences – one can use the famous Osgood scaling, depending on estimations of test persons. It is to be remarked that there is no gnoseological difference to mechanical measurement. Usually, one quantifies a property by an exact definition of the indicator and the scale. Then, each element obtains a value, and the whole text can be expressed as a sequence of numbers. Its properties can be modelled.

We shall mention some types of weighting at all levels presented above.

Phonemes

As is well-known, sounds have their acoustic frequencies, but only phoneticians can examine them. Another way is the measurement of muscular effort, but one needs some machines in order to be able to measure it. The simplest way is the scaling of the position of articulation, the opening of the mouth and the position of the tongue. One can find many works about these properties, one can combine them and rewrite the sonnet in terms of articulation difficulty. One obtains the sonnet lines in form of numbers, which can be evaluated. One can make dichotomic decisions, one can scale discretely, or one can measure the situation using machines.

Syllables

One usually writes the syllables as sequences of V and C. The weight of a syllable can be expressed in terms of consonant numbers in it, but this can be done also by measuring the length (i.e., length – 1). Another way is the use of sound weights presented above, and taking the mean phonetic weight of the syllable as its own weight. Again, the text can be transcribed as a sequence of syllable weights, the sonnet has its syllabic weight distribution, etc.

Morphemes

Morphemes are the smallest grammatical entities. They have both a phonetic weight and a grammatical/semantic weight. One can scale them according to the form, the expression of grammatical categories, their position, the number of their forms (morphs), the state of their independence – some of them exist as independent words, other ones only in combination with other words (e.g., personal affixes in German), still other ones may be written together or separated from the word (e.g., verbal prefixes in German or in Hungarian), etc. One should take into account all forms of morphemes and devise respective scales.

Words

Each of the ca. 200 definitions of the word can be used for weighting. Here, phonetics, grammar, and semantics may be separated or taken together. The number of possible scales is practically infinite. None is better than than another; all of them are merely our concept formations. It is not known as yet which one should be used for sonnets or for texts in general. Words can be simple or complex, they can have a special function in the sentence, they belong in the text to a special part-of-speech class, they have a position, they have a special function (e.g., adnominal or valency), etc. Every book concerning linguistics contains some weighting possibilities for words. This is, of course, only the first step. The next step would be finding dependencies of the given property on other ones, a direction initiated by G. K. Zipf and systematized by R. Köhler. But this must be done not only for one language or one text type. The field is open. Words can be ranged into a great number of classes, and we search for regularities holding for each class.

Hrebs

As soon as one begins to classify the words, one opens a way to infinity. As an example, we may mention the parts-of-speech. If one ascribes to each word its POS-character, in some languages, one may see that some of them belong to several classes at once – e.g., German verbs are all also nouns; German adjectives can be used as adverbs, etc. Either one sets up a new classification, or "weights" the appurtenance. Let us suppose that one has obtained a classifiation; then the individual words can be classified to semantic classes (e.g., verbs of motion, thinking, natural phenomena, eating, sensual activity, etc.) about which many books have been written. Now, each of these semantic classes can be further subdivided according to various views – e.g., verbs in all classes according to their activity, necessity, etc.; then, activity can be scaled, etc. There is no end on this way. If one ordered the morphemes and words in hrebs, one can see that one word is the main word of a herb, and the other ones are related to it. The type of relation may be weighted. As an example, we show below a possibility for the hreb consisting of all words of a text related to the hreb head. We do not weight them, showing merely the ways.

A hreb may contain a head, its synonym, its less intensional hyperonym, its more intensional hyponym, its cohyponym as opposed to a metaphor, a personal pronoun, a relative pronoun, a demonstrative pronoun, a possessive pronoun, a conversion, a derivative affix, presence in a compound, suppletivism. All of them can obtain different weights within the hreb, hence a text consists of hrebs which, in turn, consist of weighted elements. The same method can be used in describing the Belza-chains, and rewriting the text as a sequence of weights, one can set up quantitative motifs.

Evidently, we are situated at the beginning of quantitative evaluations. In classical linguistics, one considered the work as finished if some kind of classification was presented. In quantitative linguistics, any classification is a beginning of further hard work.

Motifs, Belza-chains, Frumkina's passages

These units are more abstract because they are formed from sequences of some entities. But whatever their form, their properties can be weighted. Motifs can be constructed from any entities, and need not have a semantic part; Belza-chains are defined on a semantic basis, but every researcher can delimit them differently. The decision about the adequateness of some kind of quantification and measurement can be attained as soon as one finds the associations of the given Belza-chains or motifs with other properties. Frumkina's passages may concern any of the above properties. But here, one subdivides the text in equally long parts and studies the occurrence of the respective entitity in individual passages. This is merely a way that makes everything comparable. It is a quite different aspect if one compares the situation in a sonnet and in a scientific article. But once turned into Frumkina passages, everything is possible and reasonable.

Indicators

Some textual indicators move in a very small interval because the sonnets are short. They may be associated with the given writer, with the given language, with the given time or with all of the possible influences. In order to examine the associations, many restricted investigations will be necessary.

Further problems will arise automatically – e.g., concerning phrases, clauses, sentences, speech acts, but one does not need to begin with their description in sonnets. One can consider other text types first.

13. Nominal Valency

Valency is a well known concept in linguistics. Mostly verb valencies are taken into account, but it depends on our view of text or grammar. In poetic texts, one can expect lower valencies than in other texts. The line does not play any role, and in sonnets, we can consider the complete sonnet as a whole. We shall study here merely the nominal valency in 46 Russian sonnets. Since the sonnets are short, one cannot expect a regular distribution in individual ones; however, if one considers the results from all sonnets of a language – i.e., one examines all nouns having 0, 1, 2, … valencies –, one must obtain a regular sequence. The valencies of Russian nouns in 46 sonnets are presented in Table 13.1.

Table 13.1
Number of nouns with the given valency X in Russian sonnets

X	**T1**	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**	**T8**	**T9**	**T10**	**T11**	**T12**	**T13**	**T14**
0	2	13	17	12	2	12	18	25	11	14	13	4	0	0
1	9	7	8	4	10	4	6	6	12	10	8	9	11	17
2	3	1	1	3	1	–	–	–	1	0	2	4	1	3
3	–	–	–	–	–	–	–	–	–	0	–	–	–	0
4	–	–	–	–	–	–	–	–	–	1	–	–	–	1
X	**T15**	**T16**	**T17**	**T18**	**T19**	**T20**	**T21**	**T22**	**T23**	**T24**	**T25**	**T26**	**T27**	**T28**
0	T	8	3	8	9	5	13	12	11	7	12	9	7	11
1	14	12	19	13	14	17	8	16	16	10	11	4	8	13
2	6	2	–	5	–	1	1	1	1	3	–	–	1	4
3	3	–	–	1	–	–	–	–	–	–	–	–	1	0
4		–	–	–	–	–	–	–	–	–	–	–	–	1
X	**T29**	**T30**	**T31**	**T32**	**T33**	**T34**	**T35**	**T36**	**T37**	**T38**	**T39**	**T40**	**T41**	**T42**
0	7	22	14	4	6	16	8	11	8	10	11	7	8	7
1	11	7	5	18	17	7	3	12	6	11	13	8	13	5
2	4	–	5	4	1	1	–	2	4	4	–	3	5	2
3	–	–	–	–	1	–	–	–	–	–	–	1	–	–
4	–	–	–	–	–	–	–	–	–	–	–	–	–	–
X	**T43**	**T44**	**T45**	**T46**										
0	17	11	14	11										
1	3	11	7	6										
2	–	1	2	2										
3	–	1	–	1										
4	–	–	–	–										

Adding all numbers in the respective lines, we obtain the results presented in Table 13.2.

Table 13.2
Fitting the Lorentzian function to nominal valencies in 46 Russian sonnets

X	Fr.	Lorentzian function
0	452	451.89
1	460	460.32
2	87	76.19
3	9	28.51
4	3	14.70
a = 1221.9384, b = 0.5037, c = -0.3858, R^2 = 0.9971		

The sequence is concave, hence we were forced to apply a function capturing this property. Maybe, in other languages, authors or in the historical development, different results will be obtained. Nevertheless, this way of investigation could bring a number of surprises.

Nominal valency is only one of the possible ways to evaluate various kinds of valency. Köhler (2015: 92 ff.) studied the verbal valency, the same as Čech and Mačutek (2010); further, Čech, Pajas and Mačutek (2010) analyzed full valency, and the study may be even extended (cf. also Liu 2011; Song et al. 2015; Sanada 2015).

14. Conclusions

The aim of the book was to show ways of various quantitative investigations of sonnets; it explains the metrics and demonstrates sample results, leaving in-depth analyses for separate articles and specialized collections of papers. The main focus was paid to supporting the fact that many language properties may be modelled via a set of mathematical functions. It has been shown that the study of sonnets has both advantages and disadvantages. They are as follows:

(a) The texts are short and can be quickly processed.

(b) At the same time, this is a disadvantage because statistical tests hold true for an appropriate size only. Unfortunately, up to now, nobody said what the appropriate size is.

(c) There is a size restriction in sonnets allowing easy comparability – namely the 14 lines of the sonnet. Other texts have very different lengths. Even the number of words does not move in a great interval. It means that texts can be compared even without using statistical tests.

(d) The sonnet can be considered a special text type, and the results can be used for textology and literary studies. Although there are no sonnets in many languages, in Euroamerican literatures, it is a traditional genre appearing in various forms and epochs.

(e) Even languages can be compared from many views, but each property must be considered separately. If comparisons are to be carried out, scatter plots and network graphs can be made use of. Again, the properties are our concepts and change with the evolution of science.

(f) The formulas expressing some properties are simple and can be inserted in the Köhlerian control circuit – i.e., for each property of sonnets, one should find the associated properties.

(g) Considering further properties, one can begin to set up a theory of sonnets, in which all formulas are derived from the same theory. This is the aim of all sciences; however, it will last a long time until one obtains good results.

We restricted ourselves to a few properties well-known in textology, in order to find a system of investigations. Needless to say, many other properties can be analysed, and in the course of time, one will extend the studies to other languages, too.

References

Altmann, G., Köhler, R. (2015). *Forms and Degrees of Repetitions in Texts. Detection and Analysis*. Berlin/Munich/Boston: de Gruyter Mouton.

Altmann, G., Lehfeldt, W. (1980). *Einführung in die Quantitative Phonologie.* Bochum: Brockmeyer.

Andreev S. (2018). Distribution of Syllables in Russian Sonnets. *Glottometrics 41, 13-23*.

Andreev, S., Fan, F., Altmann, G. (2017). Adnominal aggregation. *Glottometrics* 40, 63–76.

Belza, M. I. (1971). K voprosu o nekotorych osobennostjach semantičeskoj struktury svjaznych tekstov. In: *Semantičeskie problemy avtomatizacii informacionnogo potoka*, 58–73. Kiev.

Boschtan, A., Best, K.-H. (2010). Diversification of simple attributes in German. *Glottotheory* 3(2), 5–9.

Čech, R., Mačutek, J. (2010). On the quantitative analysis of verb valency in Czech. In: Grzybek, P., Kelih, E., Mačutek, J. (eds.). *Text and Language. Structures, Functions, Interrelations. Quantitative Perspectives.* Wien: Praesens Verlag, 21–29.

Čech, R., Pajas, P., Mačutek, J. (2010). Full Valency. Verb Valency without Distinguishing Complements and Adjuncts. *Journal of Quantitative Linguistics* 17, 291–302.

Delmonte, R. (2016). Exploring Shakespeare's Sonnets with SPARSAR. *Linguistics and Literature Studies 4(1),: 61-95*
DOI: 10.13189/lls.2016.040110

Gao S., Zhang. H., Liu, H. (2014). Synergetic Properties of Chinese Verb Valency. *Journal of Quantitative Linguistics 29(1): 1-21.*

Herfindahl, O. C. (1950). *Concentration in the steel industry*. Diss. New York: Columbia University.

Hřebíček, L. (1992). *Text in communication: supra-sentence structures*. Bochum: Brockmeyer.

Jacobs, A.M., Schuster, S., Xue, s., & Lüdtke, J. (2017). What's in the brain that ink may character …: A Quantitative Narrative Analysis of Shakespeare's 154 Sonnets for Use in (Neuro-)Cognitive Poetics. *Scientific Study of Literature, 6(2*).

Kernot, D., Bossomaier, T. & Bradbury, R. (2017) Novel Text Analysis for Investigating Personality: Identifying the Dark Lady in Shakespeare's Sonnets. *Journal of Quantitative Linguistics, 24:4, 255-272*, DOI: 10.1080/09296174.2017.1304049

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

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

Liu H. (2011). Quantitative properties of English verb valency. *Journal of Quantitative Linguistics 18(3), 207-233.*

Liu, H., Liang, J. (eds.) (2017). *Motifs in language and text*. Berlin/Boston: de Gruyter.

Lupea, M., Rukk, M., Popescu, I.-I., Altmann, G. (2017). *Some Properties of Rhyme*. Lüdenscheid: RAM-Verlag.

Místecký, M. (2018). Belza Chains in Machar's *Letní sonety*. *Glottometrics* 41, 45-56.

Místecký, M. (2018). Counting Stylometric Properties of Sonnets: A Case Study of Machar's *Letní sonety. Glottometrics* 41,1-12.

Pan, X., Chen, X., Liu, H. (2018). Harmony in diversity: The language codes in English-Chinese poetry translation. *Digital Scholarship in the Humanities. 33(1), 128–142.*

Pan, X., Liu, H. (2014). Adnominals in Modern Chinese and their Distribution Properties. *Glottometrics 29, 1-30.*

Popescu, I.-I. (2007). The ranking by the weight of highly frequent words. In: Grzybek, P., Köhler, R. (eds.). *Exact Methods in the Study of Language and Text*, 555–565. Berlin/New York: Mouton de Gruyter.

Popescu, I.-I., Altmann, G.. (2015). A simplified lambda indicator in text analysis. *Glottomerics* 30, 19–44.

Popescu, I.-I., Čech, R., Altmann, G. (2011). *The Lambda Structure of Texts*. Lüdenscheid: RAM–Verlag.

Popescu, I.-I., Kelih, E., Mačutek, J., Čech, R., Best, K.–H., Altmann, G. (2010). *Vectors and Codes of Text*. Lüdenscheid: RAM-Verlag.

Sanada, H. (2015). A co-occurrence and an order of valency in Japanese sentences. In: A. Tuzzi, J. Mačutek & M. Benešová (Eds.), *Recent Contributions to Quantitative Linguistics: 139−152.* Berlin: Walter de Gruyter.

Skorochoďko, E. F. (1981). *Semantische Relationen in der Lexik und in Texten.* Bochum: Brockmeyer.

Tuzzi, A., Popescu, I.-I., Altmann, G. (2010). *Quantitative Analysis of Italian Texts.* Lüdenscheid: RAM-Verlag.

Uhlířová, L. (1997). O vztahu mezi délkou slova a jeho polohou ve větě. *Slovo a slovesnost* 58, 174–184.

Wimmer, G., Altmann, G. (2005). Unified derivation of some linguistic laws. In: Köhler, R., Altmann, G., Piotrowski, R. G. (eds.). *Quantitative Linguistics. An International Handbook*, 791–807. Berlin: de Gruyter.

Yesypenko, N. (2009). An integral qualitative-quantitative approach to the study of concept realization in the text. In: Kelih, E., Levickij, V., Altmann, G. (eds.). *Methods of Text Analysis*, 302–327. Černivci: Černivec'kij nacional'nyj universitet.

Zhang, H., Liu, H. (2017). Motifs of Generalized Valencies. In: Liu, H., Liang, J. (eds), *Motifs in Language and Text; 231-260.* Berlin/Boston: de Gruyter.

Zhang, K.T., Zhang, Zh. (2010) Shakespearean Sonnets versus Shakespearean Canon, *Journal of Quantitative Linguistics, 17:2, 81-93.*

DOI: 10.1080/09296171003642983

Ziegler, A., Altmann, G. (2002). *Denotative Textanalyse*. Wien: Edition Praesens.

Zörnig, P., Altmann, G. (1993). A model for the distribution of syllable types. *Glottometrika* 14, 190–196.

Zörnig, P., Stachowski, K., Popescu, I.-I., Mosavi Miangah, T., Chen, R., Altmann, G. (2016). *Positional Occurrences in Texts: Weighted Consensus Strings.* Lüdenscheid: RAM-Verlag.

Sources

Averina, T. Available at: < http://www.stihi.ru/2014/07/12/9160>.

Babits, M. *A lírikus epilógja.* Available at: <http://www.birtalan.hu/balazs/olvasokonyv/babits_alirikusepilogja.htm>.

Baudelaire, Ch. Available at: < https://fr.wikisource.org/wiki/Les_Fleurs_du_mal>.

Beljaeva, N. Available at: < http://nelli123.avtor.me/menu/notes/7>.

Bouška, S. Available at: <http://sonety.blog.cz/0906/sigismund-bouska-1867-1942-pet-sonetu-basnika-katolicke-moderny-ze-sb-pietas-1987>.

Čelakovský, F. L. Available at: <http://sonety.blog.cz/0707/frantisek-ladislav-celakovsky-znelky-1821-u-vltavy-8-sonetu-ze-smisenych-basni-1822>.

Coleridge, S. T. Available at: <https://en.wikipedia.org/wiki/Sonnets_on_Eminent_Characters>.

Corbière, T. Available at: <https://fr.wikisource.org/wiki/Les_Amours_jaunes>.

Cros, C. Available at: <https://fr.wikisource.org/wiki/Le_Coffret_de_santal>.

Chmelenský, J. K. Available at: <http://sonety.blog.cz/1009/josef-krasoslav-chmelensky-lkani-basne-1822-sedmibasnovy-sonetovy-cyklus>.

Faludi, F. *A pipárul*. Available at: <http://mek.oszk.hu/00600/00646/00646.htm#>.

Goethe, J. W. v. *Sonette.* Available at: <http://gutenberg.spiegel.de/buch/sonette-3649/1>.

Gorodeckij, S. Available at: < http://m.rupoem.ru/poets/gorodeckij/ya-dolzhen-vse>.

Hlaváček, K. Available at: <http://sonety.blog.cz/0706/karel-hlavacek-sokolske-sonety>.

Hvězdy, J. z. Available at: <http://sonety.blog.cz/1409/jan-z-hvezdy-pet-sonetu-ze-sb-pisne-i-jine-drobne-basne-1843-a-balady-romance-povesti-a-legendy-1843>.

Hviezdoslav, P. O. *Sonety.* Available at: <http://zlatyfond.sme.sk/dielo/113/Orszagh-Hviezdoslav_Sonety/9#ixzz52HxkjNjZ>.

Hviezdoslav, P. O. *Krvave sonety*. Available at:

<https://zlatyfond.sme.sk/dielo/112/Orszagh-Hviezdoslav_Krvave-sonety/>.

József, A.. *A kozmosz éneke* 1923. Available at: <http://old.sztaki.hu/~blb/irodalom/jattila/kozmosz.html>.

Karásek J. ze Lvovic. *Pokušení.* Available at: <://archive.org/stream/basnickespisyj00karuoft/basnickespisyj00karuoft_djvu.txt>.

Karásek J. ze Lvovic. *Propast.* Available at: <https://www.databazeknih.cz/knihy/sodoma-13167>.

Karásek J. ze Lvovic. *Všechny sonety ze sbírky Zazděná okna.* Available at: <http://sonety.blog.cz/0810/jirikarasek-ze-lvovic-vsechny-sonety-ze-sb-zazdena-okna-1894>.

Keats, J. Available at: < http://keats-poems.com/poems/sonnets/>.

Kollár, J. Available at: <https://zlatyfond.sme.sk/dielo/142/Kollar_Slavy dcera/2>.

Kvapil, J. Available at: <http://sonety.blog.cz/0906/jaroslav-kvapil-1868-1950-intermezzo-sonety-7-sonetu-ze-sbirky-ruzovy-ker-1890>.

Mácha, K. H. Available at: <http://sonety.blog.cz/0707/karel-hynek-macha-vsechny-sonety-znelky-vcetne-uvodniho-sonetu-z-maje-v-cestine-a-v-anglickem-prekladu-jamese-naughtona>.

Machar J. S. Available at: <://www.rodon.cz/admin/files/ModuleKniha/1100-Ctyri-knihy-sonetu.pdf>.

Mallarmé, S. Available at: <https://fr.wikisource.org/wiki/Poésies_(Mallarmé,_1914,_8e_éd.)>.

Mařík, R. Available at: <http://sonety.blog.cz/0906/ruda-marik-1878-pisek-1933-dablovy-sonetty-pozoruhodny-epicky-cyklus-v-nemz-dabel-hovori-v-sonetech-ze-sb-z-potulek-herce-1904>.

Meredith, G. Available at: < https://en.wikisource.org/wiki/Modern_Love>.

Musset, A. de. Available at: <http://www.wikipoemes.com/poemes/alfred-de-musset/index.php>.

Nezval, V. Available at: <http://sonety.blog.cz/1205/vitezslav-nezval-vsech-16-sonetu-ze-sb-skleneny-havelok-1932>.

Prokoshin, V. Available at: <http://magazines.russ.ru/ra/2009/8/pro3.html>.

Prudhomme, S. Available at: <https://fr.wikisource.org/wiki/Catégorie:Sonnets_de_Sully_Prudhomme>.

Rilke, R. M. Available at: <http://www.sonett-archiv.com/r/Rilke.htm>.

Rimbaud, J.-A. Available at: <http://poesie.webnet.fr/lesgrandsclassiques/poemes/arthur_rimbaud/arthur_rimbaud.html>.

Rossetti, D. G. Available at: <https://www.poemhunter.com/dante–gabriel–rossetti/poems/>.

Seifert, J. Available at: <http://sonety.blog.cz/0706/jaroslav-seifert-venec-sonetu-praha>.

Sel'vinskij, I. Available at: <https://45parallel.net/ilya_selvinskiy/sonet_obyvatel_verit_mode.html>.

Sládek, J. V. Available at:
<https://cs.wikisource.org/wiki/Selsk%C3%A9_p%C3%ADsn%C4%9B_a_%C4%8Desk%C3%A9_zn%C4%9Blky>.
Sovalin, V. S., Velikanova, L. O. (eds.) [1986]. Russkij sonet XVIII – natchalo XX veka. Moscow: Moskovskij rabotchij.
Šolc, V. Available at: <http://sonety.blog.cz/0902/vaclav-solc-sonety-z-oddilu-svatovaclavske-pisne-ze-sb-prvosenky-1872>.
Swinburne, A. C. Available at: < http://www.sonnets.org/swinburn.htm>.
Tennyson, A. Available at: <https://www.poets.org/poetsorg/poem/kraken>.
Trakl, G. Available at: <http://www.sonett-archiv.com/tu/Trakl.htm>.
Verlaine, P. Available at:
<http://poesie.webnet.fr/lesgrandsclassiques/poemes/paul_verlaine/paul_verlaine.html>.
Wordsworth, W. Available at: <www.sonnets.org/wordsworth.htm>.

List of Abbreviations

BM	M. Babits: *A lírikus epilógja*
Comp., Cp.	Computed (Values)
FF	F. Faludi: *A pipárul*
Fr., Freq., Frequ.	Frequency
Ft.	Fitting
G S	J. W. von Goethe: *Sonette*
GTA	G. Trakl: *Afra* (*Gedichte*, 1913)
GTD	G. Trakl: *Dämmerung* (*Gedichte*, 1913)
H KS	P. O. Hviezdoslav: *Krvavé sonety*
H S	P. O. Hviezdoslav: *Sonety*
J S	A. József: *A kozmosz éneke*
M LS	J. S. Machar: Letní sonety
RA	R. M. Rilke: *Adam*
RATA	R. M. Rilke: *Archaischer Torso Apollos*
RB	R. M. Rilke: *Buddha*
RBH	R. M. Rilke: *Blaue Hortensie*
RDS	R. M. Rilke: *Dame vor dem Spiegel*
RJG	R. M. Rilke: *Das jüngste Gericht*
T 1	V. K. Trediakovskij: *Sonet*
T 2	V. K. Trediakovskij: *Sonet iz seja grecheskija rechi*
T 3	M. M. Heraskov: *Sonet i jepitafija*
T 4	M. M. Heraskov: *Kol' budu v zhizni ja nakazan nishchetoj* (…)
T 5	A. A. Rzhevskij: *Sonet, zakljuchajushchij v sebe tri mysli* (…)
T 6	A. A. Rzhevskij: *Sonet, tri raznye sistemy zakljuchajushchij* (…)
T 7	I. I. Dmitriev: *Sonet*
T 8	V. A. Zhukovskij: *Sonet*
T 9	A. Del'vig: *N. M. Jazykovu*
T 10	A. Del'vig: *Vdohnovenie*
T 11	A. Del'vig: *Ja plyl odin s prekrasnoju v gondole* (…)
T 12	E. A. Baratynskij: *My p'jom v ljubvi otravu sladkuju* (…)
T 13	E. A. Baratynskij: *Hotja ty malyj molodoj* (…)
T 14	N. M. Jazykov: *K. K. Janish*
T 15	N. M. Jazykov: *Na prazdnik vash prines ja dva priveta* (…)
T 16	A. S. Pushkin: *Sonet*
T 17	A. S. Pushkin: *Pojetu*
T 18	A. S. Pushkin: *Madona*
T 19	V. G. Benediktov: *Priroda*
T 20	V. G. Benediktov: *Kometa*
T 21	V. G. Benediktov: *Vulkan*

T 22	V. G. Benediktov: *Groza*
T 23	V. G. Benediktov: *Cvetok*
T 24	V. G. Benediktov: *Krasavica, kak rajskoe viden'e* (…)
T 25	V. G. Benediktov: *Kogda vdali ot suety vsemirnoj* (…)
T 26	F. K. Sologub: *Sonet*
T 27	V. J. Brjusov: *Sonet*
T 28	V. J. Brjusov: *Egipetskij rab*
T 29	A. A. Blok: *Ne ty l' v moih mechtah, pevuchaja, proshla* (…)
T 30	V. I. Ivanov: *Pritcha o devah*
T 31	V. I. Ivanov: *Hramina chuda*
T 32	M. A. Voloshin: *Venok sonetov. Sonet 1*
T 33	M. A. Voloshin: *Venok sonetov. Sonet 2*
T 34	I. V. Severjanin: *Sonet*
T 35	A. Belyj: *Prosti*
T 36	N. S. Gumilev: *Popugaj*
T 37	N. S. Gumilev: *Roza*
T 38	S. Esenin: *Moej carevne*
T 39	K. D. Bal'mont: *Mikel' Andzhelo*
T 40	K. D. Bal'mont: *Leonardo da Vinchi*
T 41	K. D. Bal'mont: *Marlo*
T 42	S. Gorodeckij: *Mudrost'*
T 43	I. Sel'vinskij: *Sonet*
T 44	V. Prokoshin: *Vozvraschenije*
T 45	T. Averina: *Stremis' k mechte*
T 46	N. Beljaeva: *Kto mozhet zapretit' tebja lubit'*

Author Index

Altmann, G. 8,32,35,38-44,67,79,80, 95,97,112
Andreev, S. 41,44
Averina, T. 49
Babits, M. 5,70,77,84
Bal`mont, K. 49,58
Baratynskij, E. 48
Baudelaire, Ch. 5,13,14,20,72,77,78,81, 82,85, 90,91,94
Beljaeva, N. 46
Belyj, A. 49
Benediktov, V. 49
Blok, A. 49
Bouška, S. 45
Brjusov, V. 49
Browning, E.B. 5
Busemann, A. 49,66,73,74,76,77,79, 81,85,101
Byron, G.G. 5-7,85,90,94
Čech, R. 80,117
Čelakovský, F.L. 45-47
Cheraskov, M. 4,48,92
Chmelenský, J.K. 45
Coleridge,S.T.71,75,77,81,82,85,89,94, 97
Corbière, T. 72,78
Cros, Ch. 5
Del'vig, A. 48
Delmonte, R 1
Dmitriev, I. 48
Esenin, S. 49
Faludi F. 5,78
Fan, F. 41
Frumkina, R.M. 115
Gao, S. 118
Goethe, J.W.v. 5,11-17,26,28,33-38,66-68,73,75,76,80,88,93,94,96-99,101
Gorodeckij, S. 49
Gumilev, N. 49
Herfindahl, O.C. 95-97
Hlaváček, K. 45
Hřebíček, L. 32
Hviezdoslav, P.O.4,5,9-13,16,17,23-26, 29,30,62,63,68,69,73,74,78,80,84,87 88,93,94,96,97,100,101,104,107,110
Ivanov, V. 49
Jacobs, A.M. 1
Jan z Hvězdy 45
Jazykov, N. 48
József, A. 5,29,36,39,70,71,73,74,80, 84,88,94,96,97,100,110
Karásek ze Lvovic, J., 42,43,45,47,48, 51,53,56,57, 75
Keats, J. 5
Kernot, D. 1
Köhler, R. 6,40,67,95,97,114,117
Kollár, J. 42,43,45
Kvapil, J. 42,43,45,53-55,57,58,78
Lehfeldt, W. 97
Liang, J. 6
Liu, H. 6,41,58,118
Lupea, M. 8,41
Mácha, K.H. 45,47,51,53,55,56
Machar, J.S. 5,6,18,20,26-28,31,42,43, 45,47,69,70,73,74,78-80,84,88,89, 94,96,97
Mačutek, J. 97
Mallarmé, S. 72,73,78,82,86,91,92,94
Mařík, R. 45
McIntosh, R.P. 97
Meredith, G. 13,14,20,21,82
Musset A. de 6
Nezval, V. 42,43,45,48,51,53
Pajas, P. 117
Pan, X. 1,41,
Petrarca, F. 1
Popescu, I.-I. 8,41,71,80
Prokoshin, V. 49
Prudhomme, S. 6.82
Pushkin, A.S. 48,49
Rzhewskij, A. 4,48
Rimbaud, J.-A. 82,86,91,94
Ronsard, P. de 1
Rosetti, G.D. 71,72,77,82,85,90,94,97

Rukk, M. 8
Sanada, H. 118
Seifert, J. 45
Sel`vinskij, I. 49
Severjanin, I. 49
Shakespeare, W. 1,24
Shannon, C. 95
Shelly, P.B. 5
Šolc, V. 45
Sládek, J.V. 45,47
Sologub, F. 49
Swinburn, A.C. 82
Tennyson, A. 5
Trediakovskij, V. 3,4,6,7,48,86,92,94
Tuzzi, A. 79
Verlaine, P. 6,82,86,91,94
Voloshin, M. 49
Waring, E. 38
Wimmer, G. 8,9
Wordsworth, W. 71,72,82,84,88,94,97
Zhang, H. 1,58,118
Zhang, K.T. 1
Zhang, Z. 1
Zhukovskij, V. 48
Ziegler, A. 32,35,38,39
Zipf, G.K. 114
Zörnig, P. 1,51,62,66,112

Subject Index

activity 1,46,54,65-78,114
adnominals 1,41-59,114
arc 83
association 23-28,115
attributiveness 45-50
Belza-chains 1,6,22-32,40,114,115
Busemann's coefficient 4,5,66-75,77, 80,95
change 29,41,42,45,52,64,107,112,118
coherence 32
compactness 39
concentration 2,35,79,80,95-97
consensus strings 1,62-65
control circuit 118
Conway-Maxwell-Poisson d. 99
diffuseness 32,36,38,39
distance 1,32,40,80,102,112
entropy 11,83,95
equality 1,40
exponential f. 1,2,4,13,15,16,18,20,21, 24-26,30,35,36,42,44,50,60,61,83,87, 92,95,98-100,102,-106
h-point 2,79-83,93-95
hreb 1,32-39,114
inertia 11
lambda 2,79-83,92-95
languages
  Czech 2,4,5,13,26,41-43,45-47,49-51 53,67,97,106,110
  English 2,4,5,8,13,20,22,66,71,73,74, 82,97
  French 2,4,5,12,20,72,73,75,81,82
  German 2,4,5,8,11,13,15,32,33,67, 93,99,100,105,109,111,113, 114
  Hungarian 2,4,5,22,28,36,67,70,97, 99,100,109,111,113 105
  Russian 2-4,6,41,43,46,48-50,59,61, 104,109,116,117
  Slovak 2,4,5,9,12,13,63,66,67
law 1,2,4,8,12,15,55,77,92
length 1,6,7,8,15-18,20,21,23,25-27 , 29-32,39,41,50,51,59,61,64,81,83, 98-100,106-111,113,118
Lorentzian f. 6-8,28,60,100,102,106-108,110,117
mean 6,23-26,28,39,66,112,11
modeling 6,8,41,54,55,72,112
morpheme 22,23,32,3536,40,113,114
motif 1,6-8,14-22,30,31,40,58-61,80, 114,115
Ord's criterion 83
ornamentality 41,45
parts-of-speech (POS) 2-21,39,62-64, 79,80,114
phoneme 98,99,113
phrase 41,44,115
Piotrowski-law 55,77
power law 4
range 6
rank 3-6,5,10-18, 20,30,42-44,49,50, 60,61,73-75,79,87,98-100,102-105, 111
repeat rate 11,83,95-97
rhyme 1,6,8-21,62,64,79,80
runs 50-53,112
semantic organization 1
Skinner's hypothesis 102,112
spectrum 2,87,92,93
speech act 115
syllable 40,41,64,98-100,102-113
valency 2,58,114,116,117
vocabulary richness 2
weighting 2,113-115
Zipf-Alekseev f.. 8,10,13,17,26,35,99, 100,102,106
Zipf-Mandelbrot f. 8,11,83,102

The RAM-Verlag Publishing House edits since 2001 also the journal ***Glottometrics*** **–** up to now 41 issues – containing articles treating similar themes. The abstracts can be found at http://www.ram-verlag.eu/journals-e-journals/glottometrics/.

Herausgeber – Editors of Glottometrics

The contents of the last issue (41, 2018) is as follows:

Michal Místecký

Counting Stylometric Properties of Sonnets: A Case Study of Machar's *Letní sonety* 1 - 12

Sergey Andreev

Distribution of Syllables in Russian Sonnets 13 - 23

Jieqiang Zhu, Haitao Liu

The Distribution of Synonymous Variants in Wenzhounese 24 - 39

Haruko Sanada, Gabriel Altmann

Word Length and Polysemy in Japanese 40 - 45

Michal Místecký

Belza Chains in Machar's *Letní sonety* 46 - 56
Andrij Rovenchak, Olha Rovenchak

Quantifying Comprehensibility of Christmas and Easter Addresses from the Ukrainian Greek Catholic Church Hierarchs 57 - 66

Gabriel Altmann

Some Properties of Adjectives in Texts 67 - 79

History

Antoni Hernández-Fernández, Ramon Ferrer-i-Cancho 80 - 86
José María de Oleza Arredondo, S.J. (1887-1975)

Book Reviews

Haitao Liu, Junying Liang (eds.) (2017), *Motifs in Language and Text.* Berlin/ Boston: De Gruyter Mouton, pp. 271. *(Quantitative Linguistics Vol. 71*). Reviewed by **Hanna Gnatchuk** 87 - 90

Mikhail Kopotev, Olga Lyashevskaya, & Arto Mustajoki (Eds.) (2017). *Quantitative Approaches to the Russian Language*. New York: Routledge. ISBN:978-1-138-09715-5, 220 pp. Reviewed by **Heng Chen** 91 - 95