## **Sonnets: Quantitative Inquiries**

by

Sergey Andreev Michal Místecký Gabriel Altmann

2018 RAM-Verlag Andreev, Sergey, Smolensk State University Místecký, Michal, University of Ostrava Altmann, Gabriel

## **Sonnets: Quantitative Inquiries**

© Copyright 2018 by RAM-Verlag Publisher: RAM-Verlag Stüttinghauser Ringstr. 44 D-58515 Lüdenscheid Germany RAM-Verlag@t-online.de http://ram-verlag.eu

The publisher cannot be held responsible for any linguistic errors in the book: Such responsibility is only up to the authors.

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	
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.