

Pasternak lyrics: part of speech structure

Sergey Andreev^{1*} 

¹ Smolensk State University

*Corresponding author's email: smol.an@mail.ru

DOI: https://doi.org/10.53482/2021_51_391

ABSTRACT

The article is devoted to the study of the stability and variability of part of speech structures in the collections of lyrical poems by B. Pasternak, the Nobel Prize winner for literature. The analysis is based on the methodology proposed by Gabriel Altmann in his studies. The database includes 7 collections of Pasternak's lyrics, published by him over the period of more than 40 years. The study was carried out on the material of both individual poems and the framework of entire collections. The results obtained showed that in Pasternak's lyrics, nominality of texts is very high. Within the framework of each separate collection a high stability of the general structure of parts of speech was observed. Dynamic description was found to prevail over static description. It was found that both types of description are guided by the tendency to compensation when the growth of one of them causes a decrease in the other. It was discovered that the distribution of parts of speech within each collection of lyrics is very well fitted by the Zipf-Alekseev function. Using the Euclidean distances between the collections of lyrical poems, published during different periods of the author's creative work, assumptions were made about possible stages of the author's style evolution.

Keywords: arts of speech, Pasternak, lyrics, dynamic and static description, compensation, variability, Zipf-Alekseev function, Euclidean distances.

1 Introduction

One of the most frequently discussed questions regarding the style of Boris Pasternak, the Nobel prize winner in literature, is the degree of its variability. According to two Russian prominent poetesses M. Tsvetayeva and A. Akhmanova, Pasternak created his own style from the very beginning of his work and never changed it (Tsvetayeva 1986). According to their opinion it seemed that all his poems were written on one and the same day (Bayevsky 1993, p. 66). Pasternak viewed the changes of his style in a different way, saying that his lyrics did change very much and may be divided into two big parts – before and after 1940 (Pasternak et al. 1990).

Philologists, dealing with this issue, single out different periods of Pasternak's poetry (Bayevsky 2001). The periodization as a rule is based on biographic facts, on the type of genre preferred by the author

at various periods of life (lyrics, long poems, translations), on types of composition in his works, images and characteristic metaphors, clarity of content. The latter relates to the generally accepted fact that his earlier poems are vague (“dark”), representing a stream of consciousness, images with unusual and unclear interconnections whereas his later works are “clearer” thanks to a more accurate description of poetic world and more explicit relationship between images. In such cases the research does not provide sufficient information about the intensity of style changes and the extent to which they affected the style at different times – the issues which require a quantitative approach.

In those not very numerous studies which use exact methods the research is mostly focused on poetic features: strophic, rhythmic and rhyme structures, the intensity of the use of tropes and images (Bayevsky 1993, 2001; Gasparov 2012). Recognizing the importance of such characteristics for poetry, nevertheless it is absolutely necessary to pay close attention to linguistic markers of the variation of the poet’s style.

This study is devoted to quantitative analysis of the use of one of the basic morphological parameters – parts of speech (PS) with extra attention to those PS which express dynamic description of topics and PS, used to convey static visualization of the author’s poetic world. The importance of using these characteristics has been shown in a number of studies devoted to the investigation of different aspects of style (Andreev et al. 2018; Naumann et al. 2012).

2 Material and Features

The material includes 7 collections, published at different times, with a total volume of lines equal to 5962. Table 1 contains the data about the titles of these 7 collections of lyrical poems as well as their short designations and the number of analyzed poems and lines in each collection.

Table 1: Research material.

Short Designation	Collection	Year of publication	Analyzed	
			Poems	Lines
TC	Bliznets v Tuchakh (Twin in the Clouds)	1914	21	462
OB	Poverkh Baryerov (Over the Barriers)	1917	28	1052
MSL	Sestra Moya – Zhizn' (My Sister – Life)	1922	50	1340
IT	Nachal'naya Pora (Initial Time)	1928	14	247
SB	Vtoroye Rozhdeniy (The Second Birth)	1932	27	1155
ET	Na Rannikh Poyezdakh (On Early Trains)	1943	27	967
PYZ	Stikhotvoreniya Yuriya Zhivago (The Poems of Yuri Zhivago)	1957	25	986

The following parts of speech were counted: nouns (N), verbs (V), adjectives (AJ), substantive pronouns (PNS), adjectival pronouns (PNA), two types of participles (PT-1, PT-2), adverbs (AD), others (OTH). Some of these classes need clarification.

- *Nouns* (N) This class includes both common and proper nouns.
- *Verbs* (V) This class includes personal forms, infinitive, *deeprichastiye* (the form of the verb denoting action additional to the main action).
- *Adjectives* (AJ) include qualitative, relative types, ordinal numerals and adjectivized participles (*шагающий экскаватор* “walking excavator”).
- *Substantive Pronouns* (PNS) include the reflexive pronoun *себя* “oneself”, interrogative, relative and negative pronouns.
- *Adjectival Pronouns* (PNA) include demonstrative, possessive, qualitative, negative, interrogative, relative and indefinite types of pronouns.
- *Participles-1.* (PT-1). This includes participles in attributive function (*Они в неубранном бору* “They in an *uncleared* forest”).
- *Participles-2.* (PT-1). Participles which are used in attributive participial constructions (*Как спущенной шторы бесплодые, / Вводящей фиалку в обман* “Like drawn barren curtains / Introducing the violet into deception”).
- *Adverbs* (AV). Here belong adverbs of manner and place.
- *Others* (OTH). This class includes all other parts of speech (conjunctions, prepositions, particles, interjections, numerals).

The total number of words analyzed in 7 books is more than 27000. PoS-tagging was done manually.

3 Results and Interpretation

The counts of PS allowed us to find out their percentages in each collection. These data are presented in Table 2. The values are given in percents.

Table 2: Percentage of PS types in 7 collections.

Collection	N	V	AJ	AV	PNS	PNA	PT-1	PT-2	OTH
TC	38.84	13.67	8.65	4.19	6.01	3.16	3.37	1.04	21.1
OB	37.17	15.78	6.90	2.96	5.08	2.58	1.08	0.81	27.6
MSL	36.47	17.76	6.34	2.95	4.14	2.54	1.00	0.68	28.1
IT	37.56	15.25	6.22	4.68	7.41	3.24	1.70	0.85	23.1
SB	36.14	14.19	6.83	4.69	6.21	3.70	0.72	0.97	26.5
ET	38.31	12.67	8.32	4.29	5.48	3.54	1.03	0.39	26.0
PYZ	36.80	14.29	6.62	4.42	5.85	4.21	0.68	0.53	26.6
Average	37.33	14.80	7.13	4.03	5.74	3.28	1.37	0.75	25.57

3.1 Variation of PS

Analyzing the data in Table 2, the first thing that attracts attention is the obvious similarity of the percentage of PS in different collections. This is especially noticeable in nouns whose percentage representation is very similar in all 7 collections. But generally speaking, the similarity also manifests itself for the entire percentage structure. To establish the degree of such similarities the variation coefficient was used:

$$(1) \quad CV = \frac{\sigma}{k} * 100$$

where σ is the standard deviation and k is the mean.

Table 3 shows the results of this analysis.

Table 3: The variation coefficient of PS in all collections.

PS	CV	PS	CV
N	2.63	PNS	17.64
V	11.16	PNA	18.24
AJ	13.55	PT-1	68.83
AV	18.81	PT-2	31.05
		OTH	10.00

The smallest variation is observed in N class. This variability is extremely small by any standards and since nouns in poetry are representatives of themes (topics), this demonstrates a high stability in introducing the number of topics relative to the size of poems.

The description of topics is expressed in most cases by verbs and adjectives (dynamic and static description respectively). Comparison of these types of descriptions shows the following. Both dynamic (CV = 11.16) and static (CV = 13.55) description vary very little. It should be noted that a rather weak variation also takes place in the class which includes functional words (class OTH).

On the other hand, strong variation is observed in the use of both types of participles. This is actually the only indicator of the differences between the poems within one and the same collection.

Thus we see that in each collection Pasternak uses the same part of speech model, regardless of the time when the poems were written and the collection was published.

3.2 Fitting the Distribution of PS

In the previous section different collections were compared. In this section each collection will be analyzed separately. In other words if earlier Table 2 was viewed vertically, now the horizontal direction will be used. In this case the PS percentages form sequences and the distribution of their elements will be analyzed.

To do this the percentages of PS in each collection were ranked in descending order and afterwards the Zipf-Alekseev function was used (Hřebíček 2002):

$$(2) \quad f_x = f_1 * x^{a+b*\ln x},$$

where f_1 is the maximum frequency of the biggest score, a and b – parameters, x – the given PS type.

Zipf-Alekseev distribution is one of the most popular as well as successfully functioning models which reflect downscale frequencies of various language units (Pan and Liu 2014; Best and Altmann 2018; Hřebíček 2002). This function proved to be successful also for fitting the distribution of ranked frequencies of various linguistic units in the studies of styles of different poets (cf. Andreev 2020; Místecký 2018).

The results of the fitting are shown in Table 4 in which observed (Obs.) and predicted (Pred.) values are given.

Table 4: Fitting of the Zipf-Alekseev function to the distribution of PS in 7 collections of lyrics.

Rank	TC		OB		MSL		IT	
	Obs.	Pred.	Obs.	Pred.	Obs.	Pred.	Obs.	Pred.
1	38.84	38.84	37.17	37.17	36.47	36.47	37.56	37.56
2	21.08	21.46	27.64	27.59	28.12	28.70	23.08	23.43
3	13.67	13.11	15.78	15.28	17.76	15.78	15.25	14.13
4	8.65	8.65	6.90	8.34	6.34	8.42	7.41	8.90
5	6.01	6.04	5.08	4.69	4.14	4.60	6.22	5.86
6	4.19	4.39	2.96	2.73	2.95	2.60	4.68	4.01
7	3.37	3.30	2.58	1.65	2.54	1.52	3.24	2.83
8	3.16	2.54	1.08	1.03	1.00	0.92	1.70	2.06
9	1.04	2.00	0.81	0.66	0.68	0.57	0.85	1.53
		a = -0.628; b = -0.328; R ² = 0.9985			a = 9.218; b = -0.935; R ² = 0.9975			a = 0.366; b = -1.027; R ² = 9929
								a = 0.323; b = 0.517; R ² = 0.9959

Rank	SB		ET		PYZ		
	Obs.	Pred.	Obs.	Pred.	Obs.	Pred.	
1	36.14	36.14	38.31	38.31	36.80	36.80	
2	26.54	25.66	25.96	24.99	26.60	25.82	
3	14.19	14.96	12.67	14.35	14.29	14.85	
4	6.83	8.76	8.32	8.45	6.62	8.59	
5	6.21	5.31	5.48	5.18	5.85	5.15	
6	4.69	3.33	4.29	3.31	4.42	3.20	
7	3.70	2.16	3.54	2.18	4.21	2.05	
8	0.97	1.44	1.03	1.48	0.68	1.37	
9	0.72	0.98	0.39	1.03	0.53	0.92	
		a = 0.034; b = -0.762 R ² = 0.9913			a = 0.143; b = 0.683; R ² = 0.9945		
					a = 0.026; b = 0.776; R ² = 0.9905		

As seen from the table the fitting is simply excellent – the determination coefficient is is very high: $R^2 > 0.99$ in all cases. In the first place, this may serve as evidence that the distribution of parts of speech in collections obeys a certain rule. Choosing poems for his collections, changing and rewording them many times, Pasternak who was guided by his artistic taste and the specificity of the creative manner that was inherent in him at that time, subconsciously followed one and the same pattern, hidden from direct observation.

In the formula one can consider the parameter a as the constant of the language, and the parameter b as the individual impact of the writer (Ráková et al. 2019). In Figure 1 the values of b -parameter for different collections are represented.

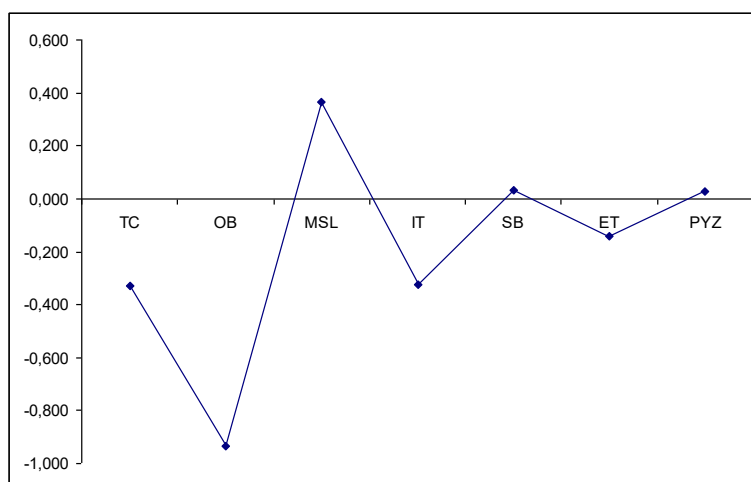


Figure 1: Parameter *b* values.

First of all, it is interesting to note that each subsequent collection by this parameter is different from the previous one. The graph can be split into two parts. The first part includes the first 4 collections, which are characterized by large differences in the value of the parameter *b*. The two collections with the largest deviations in the *b* parameter are OB and MSL which form a strong opposition to each other.

The collections included in the second group differ much less from one another. These works in case of a three-part periodization of the poet's creative work, are usually regarded as belonging to the second (SB) and the third (ET, PYZ) periods.

3.3 Static versus dynamic style

To assess whether the author uses for description more adjectives (decorative or static description) or verbs (dynamic description) Busemann's coefficient (B) is used. Its formula is:

$$(3) \quad B_A = \frac{A}{A+V},$$

where A stands for the number of adjectives, V – for the number of verbs.

To check whether the difference is significant several tests were proposed (Zörnig et. al 2015, pp. 4-19). A simpler formula (chi-square) was suggested by G. Altmann and R. Köhler (2015):

$$(4) \quad \chi^2 = \frac{(A-V)^2}{A+V}$$

The coefficient is statistically significant with 1 degree of freedom and $p < 0.05$, if $\chi^2 > 3.4$.

The following scheme of the interpretation of the results was proposed (Popescu et al. 2014; cf. Andreev et al. 2018, p. 67). In our case it looks like this:

- SD – significantly dynamic ($B > 0.55$, $\chi^2 > 3.84$);
- AC – dynamic ($B > 0.55$, $\chi^2 < 3.84$);
- BAL – balanced, ($0.45 < < 0.55$);
- ST – static ($B < 0.45$, $\chi^2 < 3.84$);
- SST – significantly static ($B < 0.45$, $\chi^2 > 3.84$).

Using this approach, the relationship between dynamic and static descriptions was established in all collections. These data are given in Table 5.

Table 5: Busemann's coefficient of static relative to dynamic description.

Collection	B	TYPE
TC	0,39	SD
OB	0,30	SD
MSL	0,26	SD
IT	0,29	SD
SB	0,33	SD
ET	0,40	SD
PYZ	0,32	SD

The results show a fairly strong predominance of dynamic description in almost all except for TC and ET collections in which there is a certain tendency towards a balanced description type. The maximum dynamics is noted in MSL, which is to some extent unexpected, because this is a highly lyrical work about the poet's love for Elena Vinograd and therefore one could expect a greater decorativeness of style. One of the possible explanations for such tendency to intensify dynamic description may be accounted for by the tense social situation in society caused by two revolutions in 1917 in Russia when most of the collection's poems were written. It should be mentioned that in the poems of this collection a new feature in the style of the poet originated – participation of inanimate phenomena of the world in the life of society together with people (rallies, debates, etc.).

3.4 Distances between collections of poems

To obtain a better picture of the differences of the collections Euclidian distances were calculated between them. In this case the percentage values for each collection are considered as its vector. Thus TC and OB are represented by the following vectors:

$$A_{TC} = 38.84, 13.67, 8.65, 4.19, 6.01, 3.16, 3.37, 1.04, 21.10;$$

$$B_{OB} = 37.17, 15.78, 6.90, 2.96, 5.08, 2.58, 1.08, 0.81, 27.6.$$

The Euclidean distance is calculated as follows:

$$(5) \quad d_{(p,q)} = \sqrt{\sum_{k=1}^n (p_k - q_k)^2}$$

In our case this is

$$d_{(p,q)} = \sqrt{(38.84 - 37.17)^2 + (13.67 - 15.78)^2 + \dots + (21.10 - 27.6)^2} = 7.38$$

where p and q are points in n -dimensional space.

Table 6 shows the distances between all the collections.

Table 6: The Euclidean distances between 7 collections of poems

Book	TC	OB	MSL	IT	SB	ET	PYZ
TC	0	7.83	9.40	4.37	6.95	5.61	6.91
OB		0	2.42	5.56	3.24	4.35	3.03
MSL			0	6.90	4.94	6.53	4.75
IT				0	4.25	4.95	4.31
SB					0	3.28	1.07
ET						0	2.99
PYZ							0

Of all distances, the most interesting for the purposes of our study are the distances between those collections which are adjacent in time of publishing:

- TC – OB: $d = 7.83$;
- OB – MSL: $d = 2.42$;
- MSL – IT: $d = 6.90$;
- IT – SB: $d = 4.25$;
- SB – ET: $d = 3.28$;
- ET – PYZ: $d = 2.99$.

Several groups stand out quite clearly here. First, it is TC that has a very big distance from the next collection OB. Another group includes OB and MSL. Collection IT takes an isolated position, being quite far from both adjacent MSL and SB. Three collections SB, ET and PYZ form one more cluster.

The results obtained, with the exception of one case, seem quite understandable and can be accounted for. A fairly strong change from the first to the second collection is due to the development of style at an early age. *Twin in the Clouds* was the first published book by Pasternak. Unlike his other collections of poems there are few superemotional means in the depiction of relationships and feelings. The next collection (*Over the Barriers*) dates back to 1916-17, when the author had already begun to develop his own style. This process went on with his new collection of poems *My Sister – Life. Initial Time* is, to some extent, a reissue of *Twin in the Clouds*, though some alternations were made (changes in poems and the inclusion of several new poems, written in the 1920s).

The grouping of *On Early Trains* and *The Poems of Yuri Zhivago* into the same class is also understandable, since according to Pasternak himself they constitute a completely new stage in his work whose lyrics he considered to be much better than all his former poems. The only unexpected result is the grouping of *The Second Birth* together with the two above mentioned poems. This collection included the poems written in 1930-1931 and marked the return of Pasternak to lyrics from which he almost departed during the period of crisis and artistic searches in the field of epic forms.

4 Conclusions

The study of parts of speech allowed a number of conclusions regarding the PS model used by the author in his lyrical works within the framework of collections of lyrics.

First of all, it corroborated the opinion that these collections may be regarded as self-contained units. The variability within the collections is very small which is especially noticeable for nouns forming a topic base of poems. The Zipf-Alekseev function provides a very good fitting of the distribution of parts of speech in every collection of lyrics.

Low variability of PS percentages in all collections testifies to the fact that the scheme for constructing collections is maintained by the author throughout his entire life.

Based on the Euclidean distances of the morphological structure of the collections of lyrics, the following main groups are distinguished:

1. The first stage (TC);
 - 1.1. Transitory (IT);
2. The second stage (OB, MSL);
3. The third stage (SB, ET, PYZ).

Initial Time, being to some extent a reissue of *Twin in the Clouds* from the first period, partly retains its features (PS model, PS distribution), at the same time acquiring new features of the other periods.

My Sister – Life by its morphological properties significantly stands out from all the other collections. Though it is classified into the same class with *Over the Barriers*, it differs markedly from the latter, as, indeed, from all the other collections in greater completeness and coherence, being actually a lyrical novel based on the real story of the poet's love. Some literary critics as well as the admirers of Pasternak's poetry consider it as the best achievement of Pasternak's lyrics – the opinion, not supported though by many other philologists.

There is one unexpected fact in the above division into periods – the inclusion of *The Second Birth* into the same class with *On Early Trains* and *The Poems of Yuri Zhivago*.

It should be kept in mind that the conclusions drawn and the regularities noted were observed at the level of morphology. They, of course, should be verified in at least two following directions – by analyzing Pasternak's work on a more extensive material with the involvement of his epic poems and the use of syntactic and phonetic features. One more direction of research is to analyze collections of other authors in order to establish whether the noted trends are characteristic only for Pasternak or reflect a general trend in poetry.

References

- Altmann, G. (2015). *Problems in Quantitative Linguistics 5. Studies in Quantitative Linguistics 21*. Lüdenscheid: RAM-Verlag.
- Altmann, G., Köhler, R. (2015). *Forms and Degrees of Repetitions in Texts. Detection and Analysis*. Berlin/Munich/Boston: de Gruyter Mouton.
- Andreev, S. (2020). Adnominal valency in modern Russian. In: Kelih, E., Köhler, R. (Eds.). *Words and Numbers. In Memory of Peter Grzybek (1957-2019)*, pp. 104-119. Lüdenscheid: Ram-Verlag.
- Andreev, S., Místecký, M., Altmann, G. (2018). *Sonnets: Quantitative Inquiries. Studies in Quantative Linguistics*, 29. Lüdenscheid: RAM-Verlag.

- Baevskij, V. S.** (2001). *Lingvisticheskie, matematicheskie, semioticheskie i kompyuternye modeli v istorii i teorii literatury*. Moskva: Yazyki slavyanskoj kultury.
- Bayevsky, V. S.** (1993). *B. Pasternak - lyric*. Smolensk: Trast-imakom.
- Best, K. H., Altmann, G.** (2018). Word Length with G. Herdan. *Glottometrics*, 42, pp. 86-90.
- Gasparov, M. L.** (2012). “Temniye” stihi i yasniye stihi: Tropy v “Sestre moyey zhizny” B. Pasternaka. In: *Izbrannye trudy. Lingvistika stixa. Analizy i interpretacii*, pp. 23-35. Moskva: Yazyki slavyanskoj kultury.
- Gnatciuc, A., Gnatchuk, H.** (2020). Identification of English Styles on the Basis of Parts of Speech: a Case of Principal Component Analysis and Factor Analysis. *Glottometrics*, 48, pp. 52-66.
- Hřebíček, L.** (2002). Zipf's law and text. *Glottometrics*, 3, pp. 27-38.
- Melka, T. S., Místecký, M.** (2020). On Stylometric Features of H. Beam Piper's *Omnilingual*. *Journal of Quantitative Linguistics*, 27(3), pp. 204-243.
- Místecký, M.** (2018). Counting Stylometric Properties of Sonnets: A Case Study of Machar's *Letni sonety*. *Glottometrics*, 41, pp. 1-12.
- Naumann, S., Popescu, I.-I., Altmann, G.** (2012). Aspects of nominal style. *Glottometrics*, 23, pp. 23-55.
- Pan, X., Liu, H.** (2014). Adnominal Constructions in Modern Chinese and their Distribution Properties. *Glottometrics*, 29, pp. 1-30.
- Pasternak, E.B., Pasternak, E.V.** (1990). *Perepiska Borisa Pasternaka*. Moskva: Khudozhestvennaya Literatura.
- Popescu, I.-I., Čech, R., Altmann, G.** (2014). Descriptivity in special texts. *Glottometrics*, 29, pp. 70-80.
- Ráková, A., Zörnig, P., Altmann, G.** (2019). Syllable Structure in Romani: A Statistical Investigation. *Glottometrics*, 46, pp. 41-60.
- Tzvetayeva, M.I.** (1986). *Sochineniya v dvukh tomakh*. Moskva: Khudozhestvennaya Literatura.
- Wilson, A.** (2020). Lengths and L-motifs of Rhythmical Units in Formal British Speech. *Glottometrics*, 48, pp. 37-51.
- Zörnig, P., Stachowski, K., Popescu, I.-I., Miyangah, T. M., Mohanty, P., Kelih, E., Chen, R., Altmann, G.** (2015). *Descriptiveness, Activity and Nominality in Formalized Text Sequences. Studies in quantitative linguistics 20*. Lüdenschied. RAM-Verlag.