Glottometrics 48, 2020, 1-8

Syllabic Identity of Verse Lines in Russian Long Poems: Skinner's Hypothesis

Sergey Andreev[1]

Abstract. The article is devoted to the analysis of distances between identical lines in 10 Russian long poems. The identity is measured by the number of motifs which are modeled in each line from syllable sequences. The data obtained were used to test Skinner's hypothesis and to analyze the distribution of distances between identical lines. Exponential function and exponential function with added 1 were used to fit the distribution of distances between lines. The results proved Skinner's hypothesis and showed good fitting of the formulas.

Keywords: *verse, line, syllables, motif, exponential function, Skinner's hypothesis.*

The main feature of a poetic text (verse) which distinguishes it from prose is its specific structure consisting of related fractions – verse lines. It was observed that a mere division of a prose text into lines gives the reader an impression that this graphically reorganized text is a poem (Gasparov, 2001: 14).

As a basic unit of verse, lines are characterized by semantic, syntactic and rhythmic completeness, i.e. they are often to a large extent semantically, syntactically and rhythmically independent, though in most cases closely correlated with one another. Semantically, lines break the poem's contents into segments and increase the tightness of links between words (Tynianov, 1924), syntactically, lines may be characterized by their own pattern of syntactic links (Gasparov, Skulacheva 2004). Rhythmically, a line is separated from the next line by a pause (coinciding with the end of a syntagm), the transfer of this syntactic pause from the end of the line to some other place creates the phenomenon called enjambement, which brings about a break in the automaticity of perception.

A regular repetition of lines influences the reader in such a way that he keeps in memory not only the preceding words and grammatical links, but also words and patterns (rhythmic, syntactic, phonetic) of preceding lines. This raises the question of how long the memory keeps in mind such information and what kind of relations between the lines originate due to it. According to Skinner's hypothesis since human memory reserves are limited, a speaker tends to place in his speech identical elements (words, syntactic constructions, etc.) at small distances from one another (Skinner 1957).

For the Russian language this hypothesis was tested and proved on the data of adnominal patterns in Russian prose (Andreev, Popescu, Altmann, 2017). In this article Skinner's hypothesis will be tested on the data-basis of the syllabic patterns of verse lines.

The data-base for the analysis includes 10 fragments from 10 long poems of Russian poets of the 60s (20th century) and the 2000s. Each fragment has 100 lines, taken from the beginning of a poem. The list of these poems is given in the Appendix.

Syllabic division is based on the principles of sonority, for the Russian language worked out by R. I. Avanesov and A. A. Reformatsky (Avanesov 1954, Reformatskij 1950). According to this theory syllabic division should be made in such a way as to achieve in noninitial syllables a rise in sonority (less sonorous sounds should precede sound which are more sonorous). In our research of the analyzed poetic fragments the following types of sylla-

[1] Smolensk state university, e-mail smol.an@mail.ru.

bles were found (V is used for a vowel, C for a consonant). V, CV, CCV, CVC, CCVC, CVCC, CCCV, VC, CCVCC, CCCVCC, CCCVC, VCC, CCCCV, CCCCVC, CVCCC.

To using sequences of these concrete syllabic types in order to model syllabic patterns which would help to classify lines is rather problematic because in this case the number of patterns of line structure will be too high. To overcome this problem it is possible to resort to units of a higher level of abstraction – motifs, which were introduced into linguistics by R. Köhler (2008, see also Köhler, Altmann 2014) and demonstrated very good results (Kelih et al. 2016). A quantitative motif is understood as a sequence with non-descending values of the selected linguistic property (Köhler, Naumann 2010). Besides such quantitative motifs another type of qualitative motifs can be modeled, using not quantitative sequences, but categorical variables (Köhler, Naumann, 2016). In our research both types of motifs will be used.

At the first stage of analysis, syllables are transformed into qualitative motifs, which is done according to the following rule: a motif is understood as a sequence of different types of syllables, so the repetition of the syllable which was already found in the preceding motif marks the beginning of a new motif. There is one more condition – the following motif cannot contain two syllabic types from the previous one.

Thus the first line from T1 which in terms of syllabic types can be represented as CCV,CVC,CVC,CV,CVC,CCVC,CV,CV,CV,CV is divided into 7 motifs (motifs are given in brackets):

[CCV,CVC],[CVC,CV],[CVC,CCVC],[CV],[CV],[CV],[CV].

It should be noted that motifs are modeled in every line separately. This means that a motif cannot contain syllables from the previous or the next lines.

According to the number of syllables in a motif it is possible to classify them by length obtaining quantitative motifs: 1-member motifs (possesses 1 syllable), 2-member motifs (2 syllables), etc. The maximum length of a motif in these poems is 6.

Each line is characterized by the number of motifs it contains. As an example let us take the first 12 lines from T1 with the number of motifs they possess.

Table 1

The number of length-motifs in the first 12 lines of T1

Line number	The number of motifs in the line
1	7
2	6
3	7
4	7
5	5
6	7
7	5
8	8
9	7
10	7
11	7
12	5

Distances are established by the number of steps, needed to move from one line to another with the same number of motifs. According to the data in Table 1 there are three lines, containing 5 motifs (# 5, 7, 12), seven lines with 7 motifs (# 1, 3, 4, 6, 9, 10, 11). One line has 6 motifs (# 2) and also one line (# 8) 8 motifs.

The distances between lines with 7 motifs (7-element lines) are as follows. Two steps are needed to move from line 1 to line 3 (distance 2), 1 – from line 3 to line 4, (distance 1). Other distances between 7-member lines are 2 (between lines 4 and 6), 3 (between lines 6 and 9), 1 between lines 9 and 10 and between lines 10 and 11. Distances between lines with 5 motifs (5-element lines) in this extract are 2 (between lines 5 and 7) and 5 (between lines 7 and 12). Lines 6, 8 in this extract have no identical line counterparts. Thus we obtain distances 2, 1, 2, 3, 1, 1, 2, 5. Their total is given in Table 2 in ascending order of distance lengths.

Table 2
Distances between identical lines

Distance	Frequency
1	3
2	3
3	1
5	1

Following this procedure all distances between identical lines were calculated for all 10 poems. The observed frequencies are given in Table 4.

To test Skinner's hypothesis the frequencies of the first 10 distances (D) between identical lines in the texts of all the authors were added (Table 3).

Table 3
Total number of frequencies between identical lines

Distance	Frequency
1	437
2	209
3	104
4	61
5	43
6	26
7	13
8	8
9	14
10	8

As can be seen from this table short distances prevail. The ratio of distance 1 to all longer distances (D2 – D10) shows the big predominance of D1 (0.899) over the rest, the ratio of distance 2 against longer distances (3–10) is 0.755. All these facts prove a highly significant inertia in forming syllable structures in adjacent lines. It is interesting that if we take individual authors, the strongest cohesion is observed in T 5 in which the ratio of D1 to all the other distances is 1.743 and the ratio of D2 to all longer distances is 0.75; T8 (1.317; 0.64), T3 (1.00; 1.234) and T2 (0.899; 0.754). The weakest connectivity of adjacent lines is observed in T6 where the ratio of D1 to other distances is only 0.34, though it is slightly compensated for by the number of 2-step distances (the ratio of D2 to longer distances is 0.861).

In order to analyze the distribution of the distances the formula y = 1 + a*exp(-b*x) was used (Andreev, Popescu, Altmann, 2017). The result showed a very good fitting, the determination coefficient is $R^2 = 0.9958$ (the parameters a = 854.9861; b = 0.6840).

At the next stage of analysis we test if there is any order in the distribution of identical lines, using the above-mentioned formula of the exponential function with added 1. The results of the fitting for individual texts are given in Table 4.

Table 4
Fitting the exponential function with added 1 to distances between lines with equal number of motifs

T1			T2		
Distance	Observed	Predicted	Distance	Observed	Predicted
1	37	37.208	1	49	48.646
2	21	21.229	2	20	21.799
3	15	12.302	3	13	9.768
4	4	7.314	4	2	4.377
5	6	4.528	5	1	1.962
6	2	2.971	6	3	0.879
7	2	2.101	7	2	0.394
8	1	1.615	9	2	0.079
9	1	1.344	11	3	0.016
10	1	1.192	13	1	0.003
11	1	1.107	24	1	0
12	1	1.060			
13	1	1.033			
14	1	1.019			
16	1	1.006			
35	1	1.000			
a = 64.809, b = 0.582, $R^2 = 0.986$			a = 108.556, b = 0.803 $R^2 = 0.9861$		

T3			T4		
Distance	Observed	Predicted	Distance	Observed	Predicted
1	43	42.667	1	47	47.676
2	22	22.233	2	26	23.444
3	9	11.820	3	10	11.792
4	10	6.514	4	5	6.190
5	5	3.810	5	3	3.495
6	2	2.432	6	3	2.200
7	1	1.730	12	1	1.015
9	2	1.189	21	1	1.000
11	1	1.049			
93	1	1.000			
a = 81.767, b = 0.674 $R^2 = 0.9858$			a = 97.070, b = 0.732 $R^2 = 0.9933$		

T5			T6			T7		
Distance	Observed	Pre-dicted	Distance	Ob-served	Pre-dicted	Distance	Ob-served	Pre-dicted
1	61	60.891	1	23	28.730	1	42	40.672
2	15	15.676	2	31	19.831	2	16	20.565
3	5	4.596	3	13	13.787	3	13	10.649
4	3	1.881	4	9	9.683	4	8	5.759
5	5	1.216	5	4	6.896	5	5	3.347
6	4	1.053	6	3	5.004	6	2	2.157
7	1	1.013	7	1	3.719	7	1	1.571
9	2	1.001	8	1	2.846	8	1	1.282
35	1	1.000	9	2	2.254	9	1	1.139
			10	3	1.851	11	1	1.034
			12	1	1.393	17	2	1.000
			15	1	1.123	21	1	1.000
			16	1	1.083	23	1	1.000
			21	1	1.012			
			29	1	1.001			
a = 244.417, b = 1.406 R^2 = 0.9837			a = 40.837, b = 0.387 R^2 = 0.8450			a = 80.442, b = 0.707 R^2 = 0.9768		

T8			T9			T10		
Dis-tance	Ob-served	Predict-ed	Dis-tance	Ob-served	Predict-ed	Dis-tance	Ob-served	Predict-ed
1	54	53.151	1	43	42.822	1	38	38.296
2	16	19.472	2	20	20.487	2	22	20.115
3	7	7.543	3	11	10.080	3	8	10.797
4	13	3.317	4	2	5.231	4	5	6.021
5	3	1.821	5	5	2.972	5	6	3.573
10	2	1.005	6	5	1.919	6	2	2.319
32	1	1.000	7	2	1.428	7	3	1.676
42	1	1.000	8	2	1.199	8	3	1.346
			9	2	1.093	9	2	1.178
			11	3	1.020	10	2	1.091
			18	1	1.000	11	1	1.047
						13	1	1.012
						22	1	1.000
						26	1	1.000
a = 147.2385, b = 1.0379 R^2 = 0.9510			a = 89.753, b = 0.764 R^2 = 0.9808			a = 72.769, b = 0.668 R^2 = 0.9830		

The fitting is very good. The only exception when the determination coefficient is R^2 <0.9 is the poem called *RU* by Voznesensky (T6) which has a very specific form, imitating internet chats and sometimes even including pictures to depict some words. (These rebus words were not analyzed). "Ru" is the internet country code top-level domain for Russia. This unusual form of the poem may have to some extent influenced the results and raises the question of the degree to which Skinner's hypothesis is relevant for internet discourse. But even for this poem, if the distance frequencies are ranked (downward ranking) and we use ranks, the same formula gives a very good fitting.

Table 5
Ranked distances of Text 6

Distances	Observed	Predicted
1	31	32.22
2	23	20.49
3	13	13.17
4	9	8.60
5	4	5.74
6	3	3.96
7	3	2.85
8	2	2.15
9	1	1.72
10	1	1.45
12	1	1.28
15	1	1.18
16	1	1.11
21	1	1.07
29	1	1.04

In the above cases the analysis was aimed at capturing the relationship of the text structure by taking into account the total number of distances between all identical lines (between 1-member lines plus between 2-member lines, etc.). But it is also possible to analyze distances between identical lines with the given number of motifs. Thus we counted the distances between lines with 2 motifs and distances between lines with 3 motifs separately. Other types of lines (with the number of motifs 1, 4, etc.) are rather rare.

To analyze the distribution of distances between lines with the above mentioned number of motifs we used the exponential function y=a*exp(-b*x). In Table 6 the determination coefficients are given which demonstrate the goodness of fit results.

Table 6
Fitting the exponential function to distances between identical lines with 2 and 3 motifs

	Distances between lines with 2 motifs R^2	Distances between lines with 3 motifs R^2
T1	0.955	0.989
T2	0.816	0.993
T3	0.947	0.976
T4	0.928	0.991
T5	0.853	0.995
T6	0.679	0.786
T7	0.983	0.856

T8	0.874	0.940
T9	0.853	0.994
T10	0.729	0.993

On the whole the results are rather good. In several cases $R^2 < 0.9$ which is probably caused by the shortness of the text and the small number of such distances in it. But again, as above, if one ranks the data and analyses the distribution of the ranks, the fitting becomes still better. Thus for two-member lines the results are as follows:

T2: $R^2 = 0.914$
T5: $R^2 = 0.931$
T6: $R^2 = 0.953$
T8: $R^2 = 0.918$
T9: $R^2 = 0.919$
T10: $R^2 = 0.948$

For lines with 3 motifs in T6 and T7 after ranking the distances, the determination coefficients also show good fitting, respectively, R^2=0.908 and R^2=0.933.

On the whole it is possible to conclude that Skinner's hypothesis certainly holds for the data of syllabic patterns and besides there is a certain order in the distribution of identical lines, indicating that syllabic structure of poems, represented by motifs, is controlled by a law.

References

Andreev, S., Popescu, I.-I., Altmann, G. (2017). Skinner's hypothesis applied to Russian adnominals. In: *Glottometrics 36. 32–69.*

Avanesov, R.I. (1954). O slogorazdele i stroenii sloga v russkom yazyke (On syllable division and syllable composition in the Russian language). In: *Voprosy yazykoznaniya. № 6. 88–101.*

Gasparov, M.L. (2001). *Russkiy stikh nachala XX veka* (Russian verse of the beginning of the 20th century). Moscow: Fortuna Ltd.

Gasparov, M.L., Skulacheva, T.V. (2004). *Stat'i o lingvistike stikha* (Papers on the linguistics of verse). Moscow: Yazyki slovyanskoy kul'tury.

Köhler, R. (2008). Sequences of Linguistic Quantities. Report on a New Unit of Investigation. *Glottotheory* 1 (1). 115–119.

Köhler, R., Altmann, G. (2014) *Problems in Quantitative Linguistics*, 4. Lüdenscheid: RAM-Verlag.

Köhler, R., Naumann, S. (2010). A syntagmatic approach to automatic text classification. Statistical properties of F- and L-motifs as text characteristics. In: Peter Grzybek, Emmerich Kelih and Yán Mačutek (eds). *Text and Language, structures, functions, interrelations, quantitative perspectives.* Wien: Praesens. P. 81–89.

Köhler R., Naumann, S. (2016). Syntactic Text Characterisation Using Linguistic S-Motifs. *Glottometrics 34, 1–8.*

Kelih, E., Knight R., Mačutek, J., Wilson, A. (2016) (Eds). *Issues in quantitative linguistics 4. (= Studies in quantitative linguistics 23).* Lüdenscheid: RAM-Verlag.

Reformatskij, A.A. (1950). *Metodicheskie ukazaniya i rukovodstvo po sovremennomu russkomu yazyku dlya studentov-zaochnikov*. Moskva: Moskovskij gosudarstvennyj pedagogicheskij institut (in Russian).

Skinner, B.F. (1957). *Verbal behaviour.* Acton: Copley Publishing Group.

Tynianov Y.N. (1924). *Problemy stikhotvornogo yazyka (The problems of verse language)*. Leningrad: Academiya.

Appendix

Text number	Author	Titles of long poems
T1	Brodsky	Zofia
T2	Brodsky	Felix
T3	Rozhdestvensky	Poehma o raznyh tochkah zreniya
T4	Yevtushenko	Bratskaya GEHS
T5	Yevtushenko	Pushkinskij pereval
T6	Voznesensky	RU
T7	Yemelin	Pechen'
T8	Yemelin	Poehma truby
T9	Stepanova	Proza Ivana Sidorova
T10	Kalinina	Peterburggo