

Rhythm, Syntax, Punctuation: A Distant Analysis of the European Sonnet

Anastasia Belousova, Juan Sebastián Páramo Rueda, Paula Ruiz Charris*

Abstract: Elaborating on an analysis of a corpus of more than 1200 sonnets by Italian, French, Spanish, English and Russian authors, this article describes the general rhythmic-syntactic arrangement of thirteenth- and fourteenth-century Italian sonnets, European Petrarchist sonnets, and several experiments with this form in the nineteenth and twentieth centuries. It presents results obtained with the help of a computer program developed for the automated analysis of strophic syntax. The program was created using Boris Tomashevsky's method based on analyzing the punctuation at the end of poetic lines (the strength of the syntactic pause is evaluated depending on the absence or presence of a punctuation sign: i.e., a comma, a dash, a semicolon, or a full stop / question mark / exclamation mark). We supplemented this with two more indices also based on punctuation. The first characterizes the length of sentences (the percentage of sentences in one line, two lines, three lines, etc.), and the second characterizes the number of sentences that end with a full stop, which comes in the middle of a line followed by the beginning of the next sentence in the same line (or, which is the same, the number of such lines). This study demonstrates that both the number of lines with a strong pause in the middle and the number of short sentences have increased over time.

Keywords: sonnet, rhythm, syntax, strophic syntax, stanzaic syntax, digital humanities, quantitative methods

* Authors' addresses: Anastasia Belousova, Universidad Nacional de Colombia, Carrera 30 No. 45-03, Facultad de Ciencias Humanas, Edif. 224, Of. 3048, Bogotá D.C., Colombia, e-mail: abelousova@unal.edu.co; Juan Sebastián Páramo Rueda, Universidad Nacional de Colombia, Carrera 30 No. 45-03, Facultad de Ciencias Humanas, Edif. 214, Of. 221, Bogotá D.C., Colombia, email: jsparamor@unal.edu.co; Paula Ruiz Charris, Universidad Nacional de Colombia, Carrera 30 No. 45-03, Facultad de Ciencias Humanas, Edif. 214, Of. 221, Bogotá D.C., Colombia, email: plruizc@unal.edu.co.

1. Introduction

Strophic arrangement is closely related to the logical, linguistic, and literary aspects of poetic speech. In traditional stanzas, and especially in fixed verse forms, the connection between metrical segments and the progression of a theme (or a poetic thought) is particularly evident. As a simple sequence of verse lines with a given rhyme pattern, a strophe is governed by certain historically established rhythmic, rhythmic-syntactic, and thematic organizational principles. The study of these principles is possible through analysis of the rhythmic-syntactic structure of the strophic forms, i.e., of the relations that are established between the boundaries of verse segments (lines, quatrains, etc.) and the syntactic units of a poetic text (inter-line syntactic ties and their distribution within a stanza). Such analysis makes it possible at the next stage to reveal regularities between the meter, theme, and style of a poetic work in its organic unity.

While there is a considerable body of observations on the poetic syntax of individual poets, literary eras, and poetic traditions, large-scale generalized descriptions of evolutionary trends in the historical development of European stanzaic syntax are still scarce.

Among the successful attempts of creating such a description based on the material of a single national tradition is Maksim Shapir's article "Three Reforms of Russian Poetic Syntax (Lomonosov – Pushkin – Joseph Brodsky)" (2003). Shapir describes three evolutionary types of syntactic organization in Russian poetry between the eighteenth and twentieth centuries based on a detailed classification of syntactic links according to their strength. During that time period, Shapir argues, "three types of poetic speech arrangement crystallized one after another, due to the systematic transformation of poetic language: 'syntactic', 'anti-syntactic' and, finally, 'parasyntactic'; these three types may also "be interpreted as 'classical', 'romantic' and 'modernist'" (Shapir 2003: 66).

The first type highlighted by Shapir – "syntactic" or "classical" – is characteristic of Mikhail Lomonosov. The end of a line in his odes usually tends to be coincident with the weakest syntactic link, ideally with the end of a sentence. Of course, this coincidence is not always achievable, as a sentence may be longer than a line. In this case, the synchronization of grammatical and verse segmentation shifts from the line to the stanza as a whole. The result is a sentence in which the syntax rather closely corresponds to the stanza rhyme structure. In the odic decima with the AbAbCCdEEed rhyme pattern, there are clearly defined strophic positions with strong links and others with weaker links. A distinct syntactic rhythm is evident in the stanza. It is created by the

regular alternation of ties: strong – weak – strong – weak – medium – strong – weak – medium – strong – weak – medium – strong – weak. In the following example, the syntactic rhythm is emphasized by the anaphor:

Тогда от радостной Полтавы
 Победы Росской звук гремел, |
Тогда не мог Петровой славы
 Вместить вселенныя предел, |
Тогда Вандалы низложены
 Главы имели преклоненны
 Еще при пеленах твоих; |
Тогда предъявлено судьбою,
 Что с трепетом перед Тобою
 Падут полки потомков их.

(Mikhail Lomonosov, “Ode on the birthday of Her Majesty, the Sovereign Empress Elisaveta Petrovna, Autocrat of all Russia, in the year 1746”)

This principle of syntactic organization has been maintained almost unchanged by Russian poets after Lomonosov.

According to Shapir, the second type of poetic speech organization, which he refers to as “anti-syntactic” or “romantic”, originated with Alexander Pushkin. However, the majority of Pushkin’s writings remained to a large extent within the syntactic system bequeathed by Lomonosov, and its imperative was not conclusively overcome until the 1830s. Pushkin’s last poem, “Mednyj vsadnik” (“The Bronze Horseman”, 1833, published in 1837), in which the grammatical coherence of lines reached unprecedented levels, played a significant part in this process.

Previously, weaker syntactic connections between the lines predominated, as in the case with Lomonosov and other poets of the eighteenth century and the first third of the nineteenth century, including Pushkin himself. In “The Bronze Horseman”, however, stronger ties between the lines prevail throughout the entire poem. The coherence of lines in “The Bronze Horseman” is enhanced by an abundance of enjambments (run-on lines):

| Несчастный
 Знакомой улицей бежит
 В места знакомые. | Глядит,
 Узнать не может. | Вид ужасный!
 Всё перед ним завалено;

Что сброшено, | что снесено;
 Скривились домики, | другие
 Совсем обрушились, | иные
 Волнами сдвинуты; | кругом,
 Как будто в поле боевом,
 Тела валяются. | Евгений
 Стремглав, не помня ничего,
 Изнемогая от мучений,
 Бежит туда, | где ждет его
 Судьба с неведомым известьем,
 Как с запечатанным письмом.
 И вот бежит уж он предместьем,
 И вот залив, | и близок дом...
 Что ж это?... |
 Он остановился.

Compare a similar arrangement in “Lines Written a Few Miles above Tintern Abbey” (1798) by William Wordsworth:

While here I stand, not only with the sense
 Of present pleasure, but with pleasing thoughts
 That in this moment there is life and food
 For future years. | And so I dare to hope
 Though changed, no doubt, from what I was, when first
 I came among these hills; | when like a roe
 I bounded o'er the mountains, by the sides
 Of the deep rivers, and the lonely streams,
 Wherever nature led; | more like a man
 Flying from something that he dreads, than one
 Who sought the thing he loved. | For nature then
 (The coarser pleasures of my boyish days,
 And their glad animal movements all gone by,
 To me was all in all.— | I cannot paint
 What then I was. | The sounding cataract
 Haunted me like a passion: | the tall rock,
 The mountain, and the deep and gloomy wood,
 Their colours and their forms, were then to me
 An appetite: | a feeling and a love,
 That had no need of a remoter charm,
 By thought supplied, or any interest

Unborrowed from the eye.— | That time is past,
 And all its aching joys are now no more,
 And all its dizzy raptures. | Not for this
 Faint I, nor mourn nor murmur; | other gifts
 Have followed, for such loss, I would believe,
 Abundant recompense. |

And in “L’infinito” (“The Infinite”, 1819) by Giacomo Leopardi:

Ma sedendo e mirando, interminati
 Spazi di là da quella, e sovrumani
 Silenzi, e profondissima quiete
 Io nel pensier mi fingo; | ove per poco
 Il cor non si spaura. | E come il vento
 Odo stormir tra queste piante, io quello
 Infinito silenzio a questa voce
 Vo comparando: | e mi sovvien l’eterno,
 E le morte stagioni, e la presente
 E viva, e il suon di lei. | Così tra questa
 Immensità s’annega il pensier mio:
 E il naufragar m’è dolce in questo mare.

Why did Shapir name this organization of poetic speech “anti-syntactic”? It is a system that emphasizes the misalignment of linguistic and metrical units by playing with their “displacement” in relation to one another, thus shattering the old, “classical” system of verse syntax.

According to Shapir, the most recent radical transformation of Russian verse syntax is associated with the name of Joseph Brodsky, who succeeded in overcoming the harmonization of syntax and rhythm by creating a para-syntactic, or modernist, system. Compare his “Peñe bez muzyki” (“Singing without music”, 1970):

Когда ты вспомнишь обо мне
 в краю чужом — хоть эта фраза
 всего лишь вымысел, а не
 пророчество, о чем для глаза,
 вооруженного слезой,
 не может быть и речи: даты
 из омута такой лесой
 не вытащишь — итак, когда ты

за тридцать земель и за
 морями, в форме эпилога
 (хоть повторяю, что слеза,
 за исключением былого,
 все уменьшает) обо мне
 вспомнешь все-таки в то Лето
 Господне и вздохнешь — о не
 вздыхай! — обзревая это
 количество морей, полей,
 разбросанных меж нами, ты не
 заметишь, что толпу нулей
 возглавила сама.

В гордыне [...]

Another attempt at a systematic study of the evolution of poetic syntax within a particular national tradition was made in recent decades by Italian versologists. Scholars associated with the journal *Stilistica e metrica italiana* (Marco Praloran, Arnaldo Soldani, Laura Facini, Leonardo Bellomo, Amelia Juri, and others) have been developing a detailed description of the evolution of the metrical and rhythmic organization of traditional stanzas of Italian poetry. Soldani's fundamental work on sonnet syntax (2009) and recent monographs and collections of essays on the sonnet, the *ottava rima*, and the *terza rima* have resulted from this endeavour (Soldani, Facini 2017; Facini 2018; Facini et al. 2020).

Poetic syntax does not develop only within the narrow confines of a national tradition. In his analysis, Shapir notes (referring Wachtel 1998: 59ff.) that the transition to the “anti-syntactic” system in Russian verse was influenced by German poetry. Thus, he looked at the evolution of poetic syntax from a comparative perspective:

The transition to the anti-syntactic system of verse, which occurred in some of Pushkin's later works, is evident in the use of blank iambic pentameter. This metrical and semantic form originated in Vasily Zhukovsky with his translations from Johann Peter Hebel. Clusters of enjambments, impermissible in rhymed verse, were tolerated there. (Shapir 2003: 70)

Italian verse scholars have also recently expressed some thoughts on the influence of Petrarch's syntax on European poetry. For example, in her work on Juan Boscán's sonnets, Laura Facini demonstrates how Boscán reproduces Petrarchan syntactic models:

In fact, the *Fragmenta* are taken as a model thematically and rhetorically, as well as stylistically, not only as a reservoir of new themes and motifs, images and rhetorical figures, syllabic measures, and metrical patterns, but also as an organic structure in which all levels of formal elaboration correspond: from the macro-structural plane of the Canzoniere form to the relationship between metric and syntax of each individual line. (Facini 2015: 72)

Sergio Bozzola and Allison Steenson (2018) discussed the possible influence of Petrarch's rhythmic-syntactic experiments on Spanish, French and English Renaissance authors.

In this paper, using the methodological perspectives of comparative poetics and distant reading,¹ we analyze the rhythmic-syntactic organization of a corpus of 1239 European sonnets written from the thirteenth to the twentieth centuries. Our aim was to establish whether the evolutionary types described by Shapir are present in other national poetic systems and whether they are automatically identifiable. In addition, we wanted to see what other syntactic trends computerized analysis of strophic syntax can reveal.² We examined sonnets in Italian, Spanish, English, French and Russian by Guido Cavalcanti, Dante, Petrarch, Pietro Bembo, Ronsard, Du Bellay, Lope de Vega, Francisco de Quevedo, Shakespeare, Foscolo, Baudelaire, Federico García Lorca, and Joseph Brodsky.³

¹ Also compare the priorities of modern verse study in the field of poetic syntax formulated by Igor Pilshchikov and Anatoli Starostin (2009): a) studying the distribution of syntactic ties within a line, b) computerized calculation of the strength of interline ties, and c) studying the syntactic organization of stanzas and strophoids.

² Partial preliminary results of this research have previously been reported in Belousova, Paramo 2019 and Belousova 2019.

³ For Italian authors, we used the digitized editions available on the website *La biblioteca della letteratura italiana* (BLI). For Spanish authors, we used texts from *Biblioteca virtual Miguel de Cervantes* (BVMC); for Russian authors, materials came from the poetic sub-corpus of the Russian National Corpus (RNC). In cases where we used other sources, this is specified in a footnote.

2. Method

In 1941, in his classic study “The Word and Verse in *Eugene Onegin*”, Grigorij Vinokur analyzed the internal structure of the Onegin stanza in connection with the content of Pushkin’s novel (Vinokur 1990: 146–195). To this end, he studied the nature of the syntactic pauses on the borders between parts of the Onegin stanza (after the 4th, 8th and 12th lines) (Ibid.: 170–171). The data he obtained allowed Vinokur to identify the main features of Pushkin’s stanzaic structure (syntactic autonomy of the first quatrain, special compositional role of the closing couplet, etc.), as well as describe stylistic and poetic features of the novel. Later, these findings were confirmed by other scholars.

Similar results were obtained by Boris Tomashevsky using an entirely different method. In “Pushkin’s Stanzaics”, he proposed to build a study of syntactic pauses on the borders of stanza lines based on punctuation (Tomashevsky 1958: 116 ff.) without specifically describing the nature of the interline connection. The strength of each pause was assigned a number: 0 – no pause; 1 – pause corresponding to a comma; 2 – pause corresponding to a colon or semicolon; 3 – pause corresponding to a period (full stop). The sum of the numbers was then divided by the possible maximum amount (if all the lines ended with periods), and the result was expressed as a percentage. The percentage indicator thus showed the relative average strength of the syntactic pause after each line of the stanza or quasi-stanza. The main advantage of this method in comparison with the others proposed by Boris Yarkho (2006: 84–87), Vinokur (1990: 170–171), Mikhail Gasparov and Tatiana Skulacheva (2004: 29–33), and Shapir (2000: 164–167, with examples; 2009: 11–13) is its simplicity: all the decisions are unambiguous, and automation is simplified due to the lack of need for a functional syntax model. At the same time, punctuation, which serves as the basis of this approach, only partially reflects the actual syntactic phenomena of a text and can vary over time. Our hypothesis, however, is that punctuation in modern publications adequately reflects the relative strength of the syntactic ties, especially when it comes to identifying general trends (cf. Oras 1960; Bruster 2015). Any distant approach to literature and culture inevitably leads to simplifications. On the other hand, in order to interpret any concrete phenomenon, it is necessary to imagine the big picture.

Using Tomashevsky’s method, we created a computer program that enables computerized rhythmical-syntactic analysis of stanzaic texts. We supplemented it with two more indices based on punctuation. The first characterizes the length of sentences (the percentage of sentences in one line, two lines, three

lines, etc.),⁴ and the second characterizes the number of sentences that end with a full stop, which comes in the middle of a line followed by the beginning of the next sentence in the same line (or, which is the same, the number of such lines).

3. Standard rhythmic-syntactic arrangement

New European lyric poetry, starting with the troubadours, is characterized by the general coherence of its syntactic, rhythmic, and semantic structure (see, for example, Switten 1985: 53–63). This is probably due to its origin since the new lyrical forms of European poetry are related to music and round dance.

For example, a canzone stanza is both a two-part and a three-part structure of the XX/Y type, where each element consists of at least two lines. The first two elements are identical (the same rhymes in the same order, for example, ABA ABA) or similar in structure (for example, the same rhymes but in the inverted order: ABC CBA). The third part differs in the number of lines, the rhyming scheme, or both (Wilkins 1915: 148–149). Such three-part stanzaic structures probably originate in Romance folk songs, which accompanied round dances (Gasparov 1996: 149). In the round dance, the chorus moved first to one side for half a circuit, then back to the initial position to the same rhythm, and finally made a full circuit to a different rhythm. This structure is reflected in such Romance strophic forms as canzone, sonnet, ballade, virelai, triolet, and others. Twelfth-century troubadour poetry saw the transition of such structures from folk dances and songs to literature, where the repetitions and refrains disappeared. Mikhail Gasparov describes the successive estrangement from the original folk song:

First, the opening burden was fully repeated after each stanza (as in the virelai). Later, there was only a partial repetition of the initial lines of the song at the end of each stanza (as in the triolet, the rondel, the rondeau). Later still, even this partial echoing of the beginning disappears, and only the final line recurs in every stanza (as in the ballade). Finally, the repetition of the final line in every stanza also becomes effaced: the stanzas follow each other without any repetitions, and they do not have to turn around the same motif semantically

⁴ If a sentence finishes in the middle of the poetic line, we round up the number of lines. For example, if a sentence occupies 1.5 poetic lines, we consider it to be of two lines. If it occupies 3.5 lines, we consider it to be of four lines. For the two last indices, we considered pauses corresponding to a colon or semicolon as equivalent to the end of a sentence.

(compare the canzone). More archaic forms, with burdens and refrains, were still mostly used in the North, in France; newer forms, without them, developed mostly in the South, in Italy. (Gasparov 1996: 149–150)

Gasparov typologically correlates the strophic structure of the New European lyrics with Greek choral poetry, in which the parts are also syntactically autonomous (Ibid.: 62–63). This theory of the origin of European strophic forms explains their syntactic circularity, which is necessary if the text is not read but performed.

In the case of the sonnet, the relation of the form with the music is a subject of intense debate (see Magro, Soldani 2017: 21–22). Although the name directly connects the poetic form with the melody, it is certain that in Tuscany the word was already used in the new metrical sense, not the melodic one. However, for the Sicilian School, where the sonnet appears, the situation is unclear. Modern critics following Gianfranco Contini's opinion (1951: 176) have predominantly considered that it was in this historical moment that the “divorce” between poetry and music occurred. Nevertheless, Pietro G. Beltrami (1999) once again has called this thesis into question, arguing that the canzone and perhaps even the sonnet were conceived as texts that have in their context of origin courtly representation and a musical destination. In any case, from the morphological point of view, the sonnet from its origin has been characterized by synchronization between the rhyming scheme and the syntactic arrangement. This basic principle and its gradual variation and shattering can be observed in our data.

4. Data

Tables 1, 2, and 3 show the data obtained in this study.

Table 1. Strength of the syntactic ties after each sonnet line, %

	№ of sonnets	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Cavalcanti	35	14.29	30.48	20.0	70.48	19.05	47.62	18.1	81.9	20.0	23.81	78.1	20.95	25.71	100.0
Dante	76	15.79	39.04	20.18	80.7	19.74	45.18	17.11	87.28	22.81	26.32	79.82	30.7	30.7	100.0
Petrarch	320	14.9	32.81	18.44	79.69	20.21	36.35	23.23	92.19	22.81	22.81	79.9	27.19	28.65	100.0
Bembo	144	16.67	28.47	16.44	69.21	11.81	31.02	21.06	84.03	22.45	23.61	81.48	20.83	31.02	100.0
Ronsard	54	22.22	50.0	27.78	90.12	27.16	39.51	29.63	96.3	41.98	45.06	67.9	63.58	39.51	100.0
Du Bellay	196	26.02	32.65	26.02	84.69	25.85	39.12	28.74	91.84	29.25	34.01	71.6	34.86	33.16	100.0
Lope	48	20.14	37.5	25.0	86.81	26.39	38.19	26.39	90.28	27.08	31.94	81.25	29.86	34.72	100.0
Quevedo	112	20.54	40.77	24.11	96.43	23.21	41.37	23.51	97.62	24.7	28.87	91.67	37.8	32.14	100.0
Foscolo	12	13.89	19.44	5.56	72.22	11.11	27.78	13.89	83.33	16.67	19.44	52.78	38.89	16.67	100.0
Baudelaire	49	31.97	42.86	20.41	82.31	31.29	45.58	34.01	91.16	25.17	42.18	57.14	36.05	31.29	100.0
Lorca	19	28.07	63.16	15.79	91.23	8.77	52.63	28.07	96.49	24.56	21.05	96.49	29.82	35.09	100.0
Brodsky	20	38.33	40.0	31.67	60.0	46.67	50.0	41.67	61.67	50.0	43.33	33.33	58.33	43.33	100.0
Shakespeare	154	19.26	52.16	24.89	87.01	21.21	44.59	21.86	93.51	22.51	46.32	24.89	92.64	31.82	100.0

Table 2. Percentage of sentences of determinate length (1 line, 2 lines, 3 lines, etc.)

	1	2	3	4	5-8	9-13	14
Cavalcanti	15.33	13.14	29.93	22.63	0.73	0.0	0.0
Dante	8.98	24.55	31.44	25.15	2.1	0.0	0.0
Petrarch	16.78	17.69	30.7	24.34	1.68	0.14	0.0
Bembo	11.32	14.47	32.1	23.93	0.74	0.37	0.74
Ronsard	13.99	25.1	19.75	32.51	2.06	0.0	0.0
Du Bellay	11.92	9.24	24.59	33.23	0.89	0.0	3.13
Lope	20.65	26.32	24.7	23.89	0.4	0.0	0.0
Quevedo	28.46	29.82	23.19	16.42	0.6	0.0	0.0
Foscolo	24.56	24.56	17.54	21.05	3.51	0.0	0.0
Baudelaire	33.46	27.57	11.76	18.01	1.84	0.37	0.0
Lorca	37.29	23.73	22.88	13.56	0.0	0.0	0.0
Brodsky	42.42	30.3	13.64	9.09	3.03	0.0	0.76
Shakespeare	19.59	49.2	5.13	23.46	0.34	0.11	0.0

Table 3. Percentage of lines with a full stop in the middle of the line

Cavalcanti	0.61
Dante	1.5
Petrarch	2.66
Bembo	1.69
Ronsard	1.98
Du Bellay	1.28
Lope	2.98
Quevedo	3.7
Foscolo	7.74
Baudelaire	6.71
Lorca	3.01
Brodsky	13.57
Shakespeare	2.5

5. Interpretation of the data

5.1. Cavalcanti, Dante, Petrarch

Chronologically, the earliest sonnets in our corpus are by Guido Cavalcanti (from *Rime*, before 1301), Dante (from *Vita nova* and *Rime*, approx. 1283–1308) and Petrarch (from *Canzoniere*, 1336–1374).

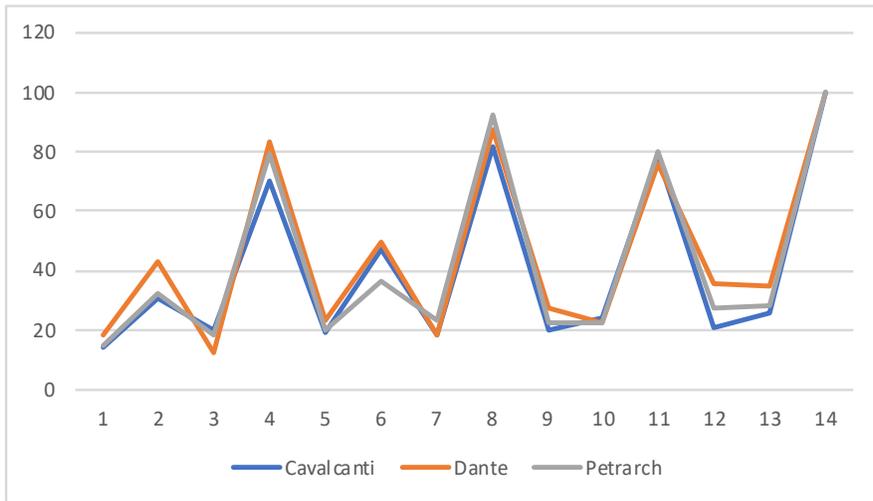


Figure 1. Strength of the syntactic ties after each sonnet line (Cavalcanti, Dante, Petrarch), %

The lines on the chart are very similar for all three poets. The sonnet is clearly divided into two parts [(4+4) + (3+3)]. The strongest pause separates the quatrains from the tercets, while other pauses are organized hierarchically. The index for the pause after line 8 exceeds 85% for Dante and Petrarch and 80% for Cavalcanti. After lines 4 and 11, it is similar to or exceeds 75%. The indices for the weak positions are generally about 20%. The quatrains follow a basic syntactic rhythm of couplets; meanwhile, the tercets clearly do not. The basic distribution of syntactic ties is as follows: strong – medium – strong – weak – strong – medium – strong – weak – strong – strong – weak – strong – strong – weak.

The index of Cavalcanti for the 8th line is lower than that of Dante and Petrarch because sometimes Cavalcanti does not respect the division in *fronte* and *sirma*, as occurs in the sonnet “S’io fosse quelli che d’amor fu degno...”:

S'io fosse quelli che d'amor fu degno,
 del qual non trovo sol che rimembranza,
 e la donna tenesse altra sembianza,
 assai mi piacerea siffatto legno.

E tu, che se' de l'amoroso regno
 là onde di merzé nasce speranza,
 riguarda se 'l mi' spirito ha pesanza:
 ch'un prest' arcier di lui ha fatto segno //

e tragge l'arco, che li tese Amore,
 s' lietamente, che la sua persona
 par che di gioco porti signoria.

Or odi meraviglia ch'èl disia:
 lo spirito fedito li perdona,
 vedendo che li strugge il suo valore.

The rhythm of couplets in the octet is present but not distinctly pronounced. If we compare the index for the second line of the first and second quatrain with the strong even-rhythm of *ottava rima* (for example, in Poliziano's *Rime*, Ariosto's *Orlando Furioso* and Tasso's *Gerusalemme liberata*), the difference becomes clear (see Belousova, Páramo 2019: 26): the indices for odd and even lines of the classical octave display a much greater difference (the difference between even and odd lines is 30–50%, not 20%, as it is here). In this classically arranged octave, the basic syntactic structure is (2+2)+(2+2), and the alternation of syntactic ties is as follows: strong – weak – strong – weak – strong – weak – strong – weak. The sonnet, in comparison, presents medium ties in positions 2 and 6.

This fact possibly can be explained through the medium length of the sentences. In Table 2, we see that sentences that occupy 3 lines are the most common for the three Italian authors, accounting for about 30% of all the sentences. Sentences of 4 lines are also quite common and amount to about 23%. The percentage of interrupted lines (Table 3) is different for each of the three authors, and it is the highest in Petrarch: 0.61% in Cavalcanti, 1.5% in Dante and 2.66% in Petrarch.

5.2 Petrarchism

The principles of poetic organization of speech were developed within the framework of Italian Renaissance classicism and assimilated throughout Europe along with the Petrarchism movement significantly promoted by Pietro Bembo. The sonnets of the Petrarchists – of Pietro Bembo himself (sonnets from *Rime*, published in 1530), Ronsard (*Le Second Livre des Sonnets pour Hélène*,⁵ 1578), Du Bellay (*Les Regrets*,⁶ 1553–1557, published in 1558), and later, Lope de Vega (sonnets from *Rimas*, published in 1602) and Francisco de Quevedo (*Parnaso español*, published in 1648) – are organized in a similar way to those of the Italian Duecento and Trecento. The trends are sometimes even stronger: for example, the index for the pause after line 8 generally exceeds 90%.

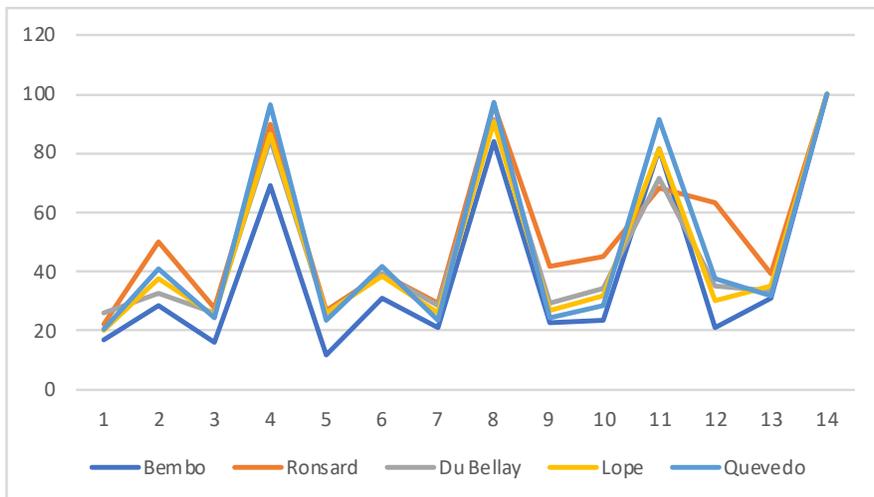


Figure 2. Strength of the syntactic ties after each sonnet line (Bembo, Ronsard, Du Bellay, Lope de Vega, Francisco de Quevedo), %

The most striking feature that deserves separate consideration is the behavior of Ronsard's tercets and, specifically, of the second tercet's first line, which is often syntactically autonomous. It seems that in his sonnets, Ronsard avoids sentences of 1 and 3 lines. Although the lines in the chart are similar, the distribution of phrases of different lengths is quite dissimilar for each of the five authors. The most common length in Bembo is 3 lines (like in Cavalcanti,

⁵ We used the digitized edition Ronsard 1921.

⁶ We used the digitized edition Du Bellay 1910.

Dante and Petrarch), 4 lines for Ronsard and Du Bellay, and meanwhile, Lope and especially Quevedo prefer shorter phrases. Lope and Quevedo also demonstrate quite a large number of interrupted lines (Lope 2.98%, Quevedo 3.7%).⁷ Compare a quatrain by Quevedo:

Lloro mientras el sol alumbra, y cuando
descansan en silencio los mortales
torno a llorar; | renuévanse mis males,
y así paso mi tiempo sollozando.

It seems that the large numbers of one-line phrases and lines with a full stop in the middle of the line are symptoms of rhythmical-syntactic progress.

In Shakespeare's sonnets (published in 1609),⁸ we can see how the syntactic structure changes as the rhyme pattern changes, continuing to follow the principles of meter-syntax coherence. Shakespeare's distribution of pauses clearly follows the rhyme structure of the English sonnet (abab cdcd efef gg): 4+4+4+2.

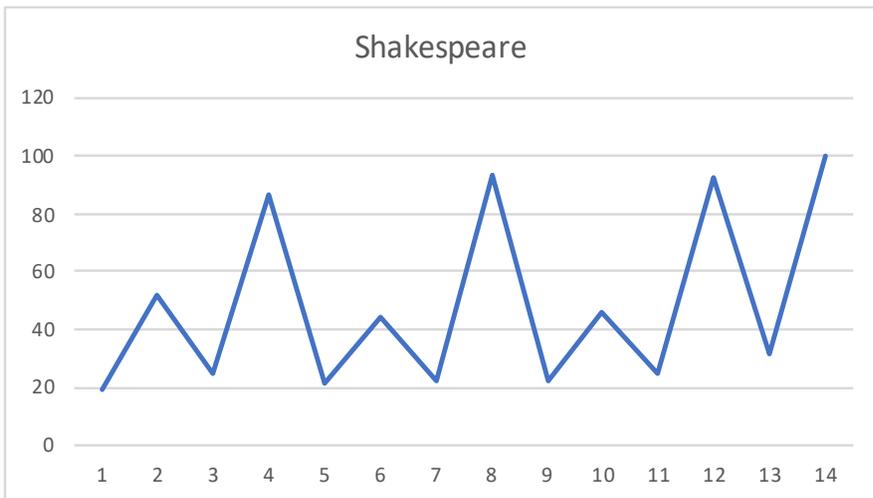


Figure 3. Strength of the syntactic ties after each sonnet line (Shakespeare), %

⁷ Compare the data on enjambments in Spanish sonnets (Ruiz Fabo et al. 2021: i74–i75).

⁸ We used the digital text from The Folger Shakespeare Library (Shakespeare n. d.).

The form of the English sonnet allows the tendency toward an even syntactic rhythm and the hierarchization of the pauses to be fully realized, and the distribution of pauses is as follows: weak – medium – weak – strong – weak – medium – weak – strong – weak – medium – weak – strong – weak – strong. Shakespeare prefers 2-line phrases (49.2%) and avoids the 3-line sentences (5.13%) that are characteristic for the Italian authors.

In our corpus, there are sonnets that do not follow the general trends. In particular, this is because the presence of a fixed rhythmic-logical structure allows one to play with the reader's expectations (cf. Bozzola 2018). The use of "rare" syntactic structures (in particular, sonnets "a fronte aperta"; that is, those in which the phrase continues between the eighth and the ninth line) in Petrarch and his European followers is the subject of a recent study by Sergio Bozzola and Allison Steenson (2018). In this work, the authors link the emergence of some "non-classical" European sonnets to the direct influence of Petrarch's poetic techniques. However, it may be that it is not necessarily a question of direct influence and the more careful, unsimplifying reception that the researchers suggest (Bozzola, Steenson 2018: 151) but of the universal mechanisms of poetic speech built on the creation and frustration of rhythmic (as well as syntactic, semantic and other) expectations. As Baudelaire writes, "le rythme et la rime répondent dans l'homme aux immortels besoins de monotonie, de symétrie et de surprise" [rhythm and rhyme meet man's immortal needs for monotony, symmetry and surprise].

5.3 Non-classical sonnets

We also looked at smaller collections of sonnets written in the nineteenth and twentieth centuries. Figure 4 shows the variety of arrangement types.

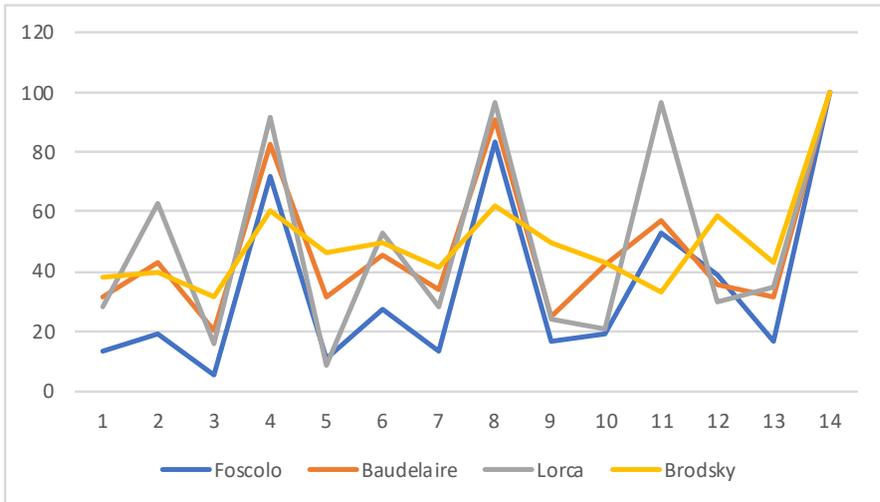


Figure 4. Strength of the syntactic ties after each sonnet line (Foscolo, Baudelaire, Lorca, Brodsky), %

If we compare sonnets by Petrarch and Foscolo (published in 1802–1803; see Fig. 5), we see that generally, Foscolo follows the normal rhythmic-syntactic logic of the sonnet.

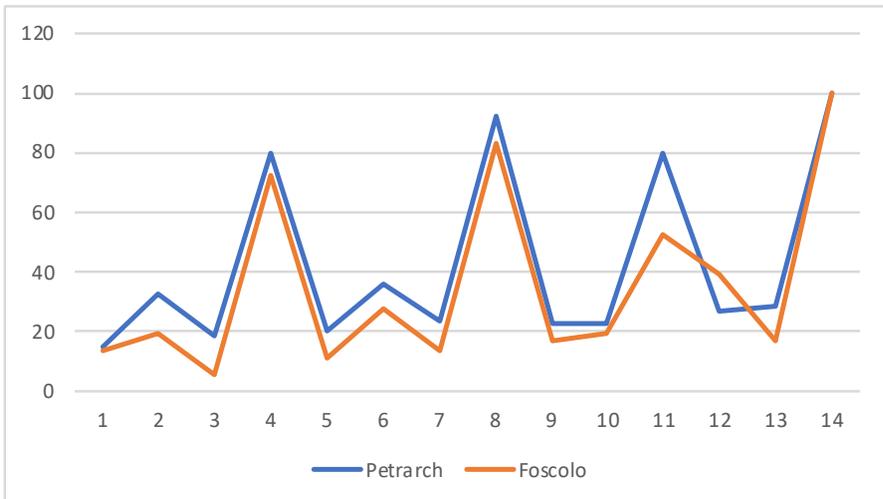


Figure 5. Strength of the syntactic ties after each sonnet line (Petrarch and Foscolo), %

At the same time, there are some important transformations. The pauses after the 11th and the 12th lines are weaker and stronger than the “classical” ones. Also, the total indices for Foscolo’s sonnet are significantly lower. This is a sign of a greater number of run-on lines. Previously, we identified the same pattern when comparing Ariosto’s and Tasso’s octave:

In the octave of Tasso, the syntactic tendencies of Ariosto’s *ottava rima* are developed and strengthened. [...] the more striking development is what happens with the pauses after odd lines. The highest index only reaches 12%! This phenomenon should be interpreted as the actual prohibition of strong pauses after odd lines. At the level of the actual poetic syntax, this is manifested in a large number of strong enjambments (three times as frequent as in Ariosto) – breaking between a noun and its adjective, a verb and an adverbial clause of place, etc., and located at the line border. (Belousova, Páramo 2019: 28)

In Foscolo’s sonnets, the index for the third line is only 5.56% and 11.11% for the fifth line. Compare:

E tu ne’ carmi avrai perenne vita //
sponda che Arno saluta in suo cammino //
partendo la città che dal latino //
nome accogliea finor l’ombra fuggita.

Né più mai toccherò le sacre sponde
ove il mio corpo fanciulletto giacque,
Zacinto mia, che te specchi nell’onde //
del greco mar da cui vergine nacque //

Già dal tuo ponte all’onda impaurita //
il papale furore e il ghibellino //
mescean gran sangue, ove oggi al pellegrino //
del fero vate la magion si addita.

Venere, e fea quelle isole feconde
col suo primo sorriso, onde non tacque
le tue limpide nubi e le tue fronde
l’inclito verso di colui che l’acque //

Per me cara, felice, inclita riva
ove sovente i pie’ leggiadri mosse //
colei che vera al portamento Diva //

cantò fatali, ed il diverso esiglio //
per cui bello di fama e di sventura
baciò la sua petrosa Itaca Ulisse.

in me volgeva sue luci beate,
mentr’io sentia dai crin d’oro commosse //
spirar ambrosia l’aure innamorate.

Tu non altro che il canto avrai del figlio,
o materna mia terra; | a noi prescrisse //
il fato illacrimata sepoltura.

The sentences are shorter in Foscolo in comparison with other Italian authors (about 50% consist of phrases of 1 or 2 lines). The percentage of interrupted lines is high (7.74%).

Comparing Baudelaire (sonnets from *Les Fleurs du mal*, published in 1857)⁹ to Ronsard, we can see a similar tendency: the indices of the odd lines are generally lower in Baudelaire. The number of interrupted lines is high (6.71%), and the sentences are normally 1 or 2 lines (they comprise about 60% of all sentences, while in Ronsard, 4-line sentences comprise 32.51% of all phrases).

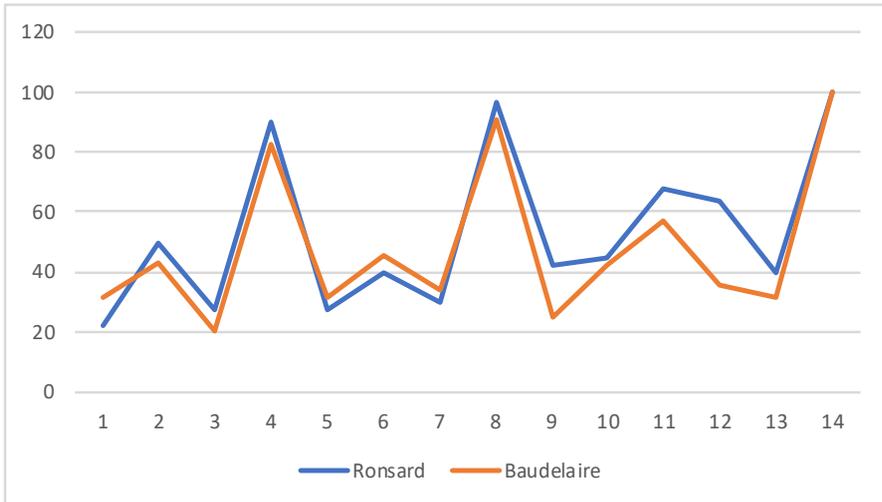


Figure 6. Strength of the syntactic ties after each sonnet line (Ronsard and Baudelaire), %

Compare:

Deux guerriers ont couru l'un sur l'autre; | leurs armes //
 Ont élaboussé l'air de lueurs et de sang.
 Ces jeux, ces cliquetis du fer sont les vacarmes //
 D'une jeunesse en proie à l'amour vagissant.

Les glaives sont brisés! | comme notre jeunesse,
 Ma chère! | Mais les dents, les ongles acérés,
 Vengent bientôt l'épée et la dague traîtresse.
 — Ô fureur des cœurs mûrs par l'amour ulcérés!

⁹ We used the digitized edition Baudelaire 1861.

Dans le ravin hanté des chats-pards et des onces //
 Nos héros, s'entreignant méchamment, ont roulé,
 Et leur peau fleurira l'aridité des ronces.

— Ce gouffre, c'est l'enfer, de nos amis peuplé!
 Roulons-y sans remords, amazone inhumaine,
 Afin d'éterniser l'ardeur de notre haine!

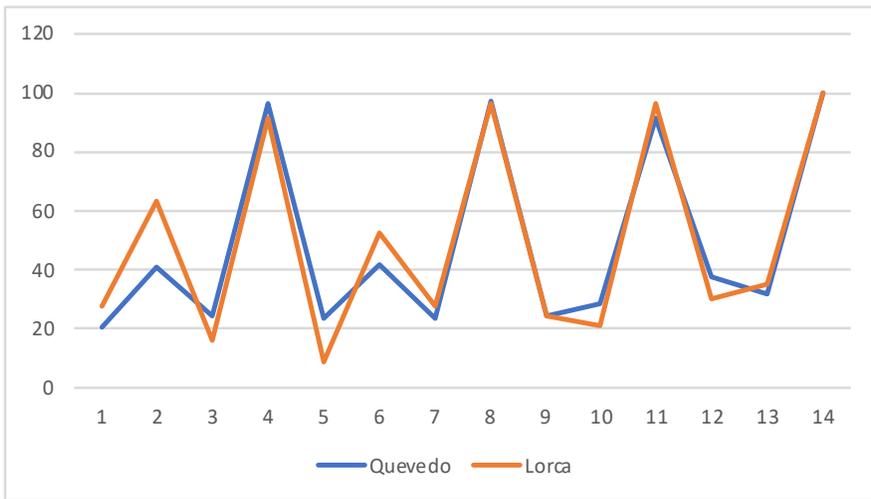


Figure 7. Strength of the syntactic ties after each sonnet line (Quevedo and Lorca), %

Comparing García Lorca's sonnets (before 1937)¹⁰ to Quevedo, we see similar tendencies mostly in the tercets, while the quatrains seem to be quite different, especially after the second and the fifth line. The most striking feature of Lorca's sonnet is his pronounced preference for one-line phrases (37.29%). Compare:

Esta luz, este fuego que devora.
 Este paisaje gris que me rodea.
 Este dolor por una sola idea.
 Esta angustia de cielo, mundo y hora.

¹⁰ *Sonetos del amor oscuro*, "En la muerte de José de Cíña y Escalante", "A Manuel de Falla", "A Carmela Condon", "Adán", "Soneto", "Epitafio a Isaac Albéniz", "En la tumba sin nombre de Herrera y Reissig en el cementerio de Montevideo", "A Mercedes en su vuelo". We used the edition García Lorca 2009.

Este llanto de sangre que decora
lira sin pulso ya, lúbrica tea.
Este peso del mar que me golpea.
Este alacrán que por mi pecho mora.

Son guirnalda de amor, cama de herido,
donde sin sueño, sueño tu presencia
entre las ruinas de mi pecho hundido.

Y aunque busco la cumbre de prudencia
me da tu corazón valle tendido
con cicuta y pasión de amarga ciencia.

The rhythmic-syntactic arrangement of Brodsky's sonnets (*Twenty Sonnets to Mary Stewart*, 1974) is the most "non-classical" in our corpus. Apparently, the poet is unaware of the traditional structure he uses. The pause after the 8th line is only 61.67%, by far the lowest index of our authors. The couplets practically disappear. The number of interrupted lines is the highest, and the percentage of short sentences (consisting of 1 and 2 lines) is more than 70%. Brodsky systematically closes the sentence after the 9th line and situates an enjambment between the quatrains and the tercets.

Я вас любил. Любовь еще (возможно,
что просто боль) сверлит мои мозги,
Все разлетелось к черту, на куски.
Я застрелиться пробовал, но сложно //
с оружием. | И далее, виски:
в который вдарить? | Портила не дрожь, но
задумчивость. | Черт! все не по-людски!
Я Вас любил так сильно, безнадежно, //
как дай Вам бог другими — — — но не даст!
Он, будучи на многое горазд,
не сотворит — по Пармениду — дважды //
сей жар в груди, ширококостный хруст,
чтоб пломбы в пасти плавилась от жажды
коснуться — "бюст" зачеркиваю — уст!

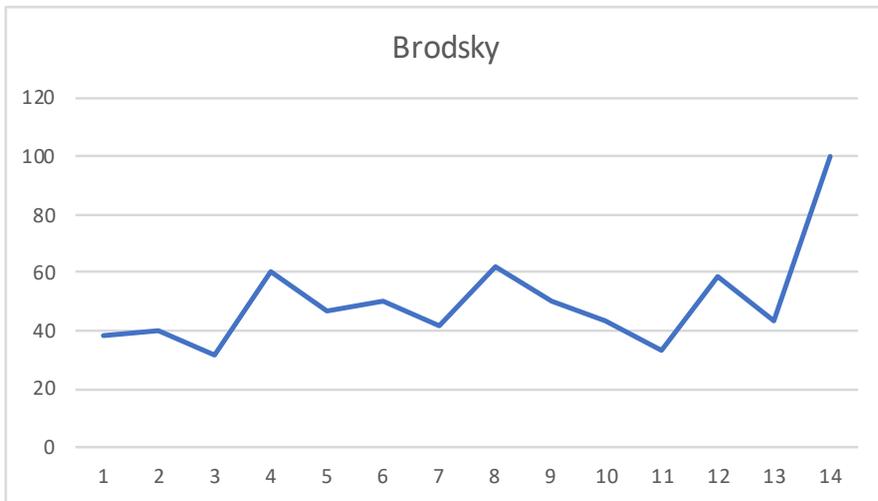


Figure 8. Strength of the syntactic ties after each sonnet line (Brotsky), %

6. Conclusion

In this study, we hoped to learn whether the evolutionary syntactic types described by Maksim Shapir for Russian poetry are present in other national poetic systems and whether they are automatically identifiable. In addition, we wanted to determine what other syntactic trends computerized analysis of strophic syntax can reveal.

We applied Boris Tomashevsky's method, which is based on analyzing the punctuation at the end of poetic lines, to a corpus of European sonnets (more than 1200 texts). Then we supplemented our data with two more indices based on punctuation: the length of sentences (the percentage of sentences of 1 line, 2 lines, 3 lines, etc.), and the percentage of lines with a full stop in the middle of the line.

As a result, we 1) described the general rhythmic-syntactic arrangement of Italian sonnets of the thirteenth and fourteenth centuries, the European Petrarchist sonnet, and some experiments with the form in the nineteenth and twentieth centuries; 2) demonstrated that the number of lines with a strong pause in the middle of the line and the number of short sentences have been increasing over time. We have demonstrated that the chosen method 1) illuminates general trends in the strophic syntax of different authors, languages, and

literary epochs; 2) reveals some striking features of each individual author's style; and 3) allows us to distinguish between texts of "classical", "romantic" and "modernist" orientations.

Since the indicators of sentence length and the number of strong pauses in the middle of a line do not depend on the stanza, they can and should be applied further to the non-stanzaic texts. Further study is needed to determine the genesis not only of the "classical" system but also of the Romantic and Modernist arrangements. In what literatures and genres did they form to then spread throughout Europe?¹¹

References

- Baudelaire, Charles 1861. *Les Fleurs du mal*. 2^e édition. Paris: Poulet-Malassis et de Broise. [https://fr.wikisource.org/wiki/Les_Fleurs_du_mal_\(1861\)](https://fr.wikisource.org/wiki/Les_Fleurs_du_mal_(1861))
- Belousova, Anastasia; Páramo Rueda, Juan Sebastián 2019. Macroanalysis of the Strophic Syntax and the History of the Italian Ottava Rima. In: Plecháč, Petr; Scherr, Barry P.; Skulacheva, Tatiana; Bermúdez-Sabel, Helena; Kolár, Robert (eds.), *Quantitative Approaches to Versification*. Prague: The Institute of Czech Literature of the Czech Academy of Sciences, 23–31. <https://versologie.cz/conference2019/proceedings/belousova-paramo-rueda.pdf>
- Belousova, Anastasija Sergeevna 2019. Strofičeskij makrosintaksis: oktava, tertšina, sonet (Rossija i Italija). In: *Rhema* 4, 9–20. <https://doi.org/10.31862/2500-2953-2019-4-9-20>
- Beltrami, Pietro Giorgio 1999. Osservazioni sulla metrica dei Siciliani e dei Siculo-toscani. In: Coluccia, Rosario; Gualdo, Riccardo (eds.), *Dai Siciliani ai Siculo-toscani: Lingua, metro e stile: Atti del Convegno (Lecce, 21–23 aprile 1998)*. Galatina: Congedo, 187–216.
- BLI = *La biblioteca della letteratura italiana*. <http://www.letteraturaitaliana.net/>
- Bozzola, Sergio 2018. Definizione ed esperienza dell'attesa metrica nella forma sonetto. In: Peron, Gianfelice; Sangiovanni, Fabio (ed.), *L'attesa: Forme, retorica, interpretazioni: Atti del XLV Convegno Interuniversitario (Bressanone, 7–9 luglio 2017)*. Padova: Esedra, 115–130.

¹¹ We are grateful to Igor Pilshchikov and Vera Polilova for their comments and suggestions.

- Bozzola, Sergio; Steenson, Allison 2018. Occasioni del Petrarca minore nel sonetto rinascimentale europeo: Boscán e Garcilaso, Ronsard e il primo petrarchismo inglese. In: Gregori, Elisa (ed.), *Rinascimento fra il Veneto e l'Europa Questioni, metodi, percorsi*. Padova: Cleup, 149–204.
- Bruster, Douglas 2015. Shakespeare's Pauses, Authorship, and Early Chronology. In: *Studia Metrica et Poetica* 2(2), 25–47. <https://doi.org/10.12697/smp.2015.2.2.03>
- BVMC = *Biblioteca virtual Miguel de Cervantes*. <https://www.cervantesvirtual.com/>
- Contini, Gianfranco 1951. Preliminari sulla lingua del Petrarca. In: *Paragone* 2(16), 3–26.
- Du Bellay, Joachim 1910. Les Regrets. In: *Œuvres complètes*. T. III: Les Antiquités; Les Regrets; Les jeux rustiques. Avec un commentaire historique et critique par Léon Séché. Paris: Revue de la Renaissance, 26–116. https://fr.wikisource.org/wiki/Livre:Du_Bellay_-_Œuvres_complètes,_édition_Séché,_tome_3
- Facini, Laura (ed.) 2018. *Nuove prospettive sull'ottava rima: [Atti del Convegno tenuto a Ginevra nel 2017]*. Edizione a cura di Laura Facini. Lecce; Rovato: Pensa multimedia.
- Facini, Laura 2015. La costruzione sintattica del sonetto di Juan Boscán. In: *Il confronto letterario* 63, 49–74.
- Facini, Laura, et al. (eds.) 2020. *Nuove prospettive sulla terza rima: Da Dante al Duemila*. Edizione a cura di Laura Facini, Jacopo Galavotti, Arnaldo Soldani, Giovanna Zoccarato. Padova: libreriauniversitaria.it.
- García Lorca, Federico 2009. *Tutte le poesie e tutto il teatro*. A cura di Claudio Rendina e Elena Clementelli; edizioni integrali con testo spagnolo delle poesie a fronte. Roma: Grandi Tascabili Economici Newton.
- Gasparov, Mikhail Leonovich 1996. *A History of European Versification*. Translated by Gerald Stanton Smith and Marina Tarlinskaja. Edited by Gerald Stanton Smith and Leofranc Holford-Strevens. Oxford: Clarendon Press.
- Gasparov, Mikhail Leonovich; Skulacheva, Tatiana Vladimirovna 2004. *Stat'i o lingvistike stikha*. Moskva: Jazyki slavianskoj kul'tury.
- Magro, Fabio; Soldani, Arnaldo 2017. *Il sonetto italiano. Dalle origini a oggi*. Roma: Carocci.
- Oras, Ants 1960. *Pause Patterns in Elizabethan and Jacobean Drama: An Experiment in Prosody*. Gainesville: University of Florida Press.

- Pilshchikov, Igor' Alekseevich; Starostin, Anatolij Sergeevich 2009. Problemy avtomatizatsii bazovykh protsedur ritmiko-sintaksicheskogo analiza sillabotonicheskikh tekstov. In: Plungian, Vladimir Aleksandrovich (ed.), *Natsional'nyj korpus russkogo jazyka: 2006–2008: Novye rezul'taty i perspektivy*. Sankt-Peterburg: Nestor-Istorija, 298–315.
- RNC = The *Poetic subcorpus of the Russian National Corpus*.
<https://ruscorpura.ru/new/search-poetic.html>.
- Ronsard, Pierre de 1921. *Sonnets pour Hélène*. Introduction et notes de Roger Sorg. Paris: Bossard. [https://fr.wikisource.org/wiki/Livre:Ronsard_-_Sonnets_pour_Hélène_-_1921.djvu](https://fr.wikisource.org/wiki/Livre:Ronsard_-_Sonnets_pour_H%C3%A9l%C3%A8ne_-_1921.djvu)
- Ruiz Fabo, Pablo; Bermúdez Sabel, Helena; Martínez Cantón, Clara; González-Blanco, Elena 2021. The Diachronic Spanish Sonnet Corpus: TEI and linked open data encoding, data distribution, and metrical findings. In: *Digital Scholarship in the Humanities* 36 (Supplement 1), i68–i80. <https://doi.org/10.1093/llc/fqaa035>
- Shakespeare, William (n. d.). Shakespeare's Plays, Sonnets and Poems. In: *The Folger Shakespeare*. Ed. by Barbara Mowat, Paul Werstine, Michael Poston, and Rebecca Niles. Washington, DC: Folger Shakespeare Library.
<https://shakespeare.folger.edu/shakespeares-works/shakespeares-sonnets/>
- Shapir, Maksim Il'ich 2000. *Universum versus: Jazyk – stikh – smysl v russkoj poezii XVIII–XX vekov*. Vol. 1. Moskva: Jazyki russkoj kul'tury.
- Shapir, Maksim Il'ich 2003. Tri reformy russkogo stikhotvornogo sintaksisa (Lomonosov – Pushkin – Iosif Brodsky). In: *Voprosy jazykoznanija* 3, 31–78.
- Shapir, Maksim Il'ich 2009. *Stat'i o Pushkine*. Moskva: Jazyki slavjanskoj kul'tury.
- Soldani, Arnaldo 2009. *La sintassi del sonetto: Petrarca e il Trecento minore*. Firenze: Sismel – Edizioni del Galluzzo.
- Soldani, Arnaldo; Facini, Laura (eds.) 2017. *Otto studi sul sonetto: Dai Siciliani al Manierismo*. Padova: libreriauniversitaria.it.
- Switten, Margaret 1985. *The Cansos of Raimon de Miraval: a study of poems and melodies*. Cambridge, MA: Medieval Academy of America.
- Tomashevsky, Boris Viktorovich 1958. Strofika Pushkina. In: Alekseev, Mikhail Pavlovich (ed.), *Pushkin: Issledovanija i materialy*. Vol. 2. Moskva; Leningrad: Izdatel'stvo AN SSSR, 49–184.
- Vinokur, Grigorij Osipovich 1990 [1941]. *Filologicheskie issledovanija: Lingvistika i poetika*. Moskva: Nauka.

- Wachtel, Michael 1998. *The Development of Russian Verse: Meter and Its Meanings*. Cambridge: Cambridge University Press.
- Wilkins, Ernest Hatch. 1915. The derivation of the “canzone”. In: *Modern Philology* 12(9), 135–166. <https://doi.org/10.1086/386980>
- Yarkho, Boris Isaakovich 2006. *Metodologija tochnogo literaturovedenija: Izbrannyye trudy po teorii literatury* (Philologica russica et speculativa 5). Moskva: Jazyki slavjanskikh kul'tur.