Abstract

The article discusses the results obtained with the help of a computer program created for the automatic analysis of strophic syntax. The program was created using Boris Tomashevsky's method, which is 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 sign: i.e. a comma, a dash, a semicolon or a full pause / question / exclamation mark).

The analyzed texts (more than 18,000 stanzas) divided into two distinct groups, consistent with chronology. The continuity of Ariosto with respect to the lyric octave of Poliziano, as well as the difference between the octave of Tasso and the octave of Ariosto, was established. This positive result shows the necessity of a macrostudy that would include a much larger number of texts from different national traditions.

1 Introduction

Rhythmical-syntactic analysis is an essential step in any critical examination of stanzaic forms, because it allows us to establish the norms that define the relationship between meter, topic and style within the poetic text.

The concepts of line (verse) and line break—and, accordingly, the syntactic relation located on the border between two poetic lines—are issues of paramount importance for verse study. Therefore, verse scholars have created classifications of syntactic ties in accordance with their relative “strength” (Vinokur 1990: 170–171; Gasparov–Skulacheva 2004: 29–33; Yarkho 2006: 84–87; Shapir 2000: 164–167 (with examples); 2009: 11–13; Tomashevsky 1958: 116 and further; cf. Belousova 2011: 54–55, 2013).
The priorities of modern verse study in the field of poetic syntax were succinctly formulated in an article by I. A. Pilshchikov and A. S. Starostin (2009): 1) studying the distribution of syntactic ties within a line, 2) computerized calculation of the strength of interline ties, and 3) studying the syntactic organization of stanzas and quasi-stanzas. These tasks can be accomplished with the use of software, but many problems arise because of the complexity of constructing a functional syntactic model. However, the third task can be carried out without a syntactic analyzer.

This work is a step in that direction.

2 Method

In 1941, G. O. Vinokur in his classical study The Word and Verse in Eugene Onegin (Vinokur 1990: 146–195) analyzed the internal structure of the Onegin stanza in connection with the content of Pushkin's novel. 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) (Vinokur 1990: 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 distich, etc.), as well as describe many stylistic and poetic features of the novel. Later, these findings were confirmed by other scholars.

In particular, quite similar results were obtained by B. V. Tomashevsky using an entirely different method. In his Pushkin’s Strophics the scholar proposed to build a study of syntactic pauses on the borders of stanza lines based on punctuation (Tomashevsky 1958: 116 ff), without describing specifically 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. The sum of the numbers was then divided by the possible maximum amount (if all the verses 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 Yarkho, Vinokur, Gasparov–Skulacheva and Shapir) is its simplicity: all the decisions are unambiguous, and automatization is simplified, due to the lack of need for a functional syntax model. Tomashevsky’s method is very productive when working with a large number of texts, as well as when analyzing texts in different languages.

Let us once again emphasize that the results obtained using Tomashevsky’s method should be considered reliable. This has been shown not only by the data of the scholar himself (his study of the Onegin stanza and others), but also by other recent research. Thus, we used his method to study the rhythmical-syntactic organization of the Russian ottava rima (Belousova 2011): it allowed us to perceive different models for the assimilation of the ottava rima by Russian poets through relative averages. Furthermore, the data describing Pushkin’s House in Kolomna, however approximate,
Quantitative Approaches to Versification

allowed us to come to similar conclusions as Shapir (Shapir 2009: 60), who used his own accurate method.

Using Tomashevsky’s method, we created a tool (a computer program in the Python programming language), which enables computerized rhythmical-syntactic analysis of stanzaic texts. One can obtain preliminary descriptions of the rhythmical-syntactic structure of stanzas by different authors and in different epochs, compare these results, and find ways for further exploring one of the most important issues in versification.

3 Material

Our corpus consisted of the following Italian works: *Filostrato* (1336) and *Ninfale Fiesolano* (perhaps 1344–1345) by Giovanni Boccaccio, the anonymous poem *La Spagna* (XIV or XV), *Morgante* (1483) by Luigi Pulci, *Rime* by Poliziano (circa 1480), *Orlando Innamorato* by Matteo Maria Boiardo (1483–1495), *Orlando Furioso* (1516–1532) by Ludovico Ariosto and *Gerusalemme Liberata* (1581) by Torquato Tasso.

A total of 18,066 of stanzas were analyzed. Modern, philologically acceptable editions (listed in the bibliography) were used as the source texts. Turning to the oldest publications would have created problems due to the variability of punctuation norms. Our hypothesis is that modern publications quite adequately show the strength of the syntactic ties.

We examined the entire bodies of texts, except for the instances of “incorrect octaves”. Thus, additional quatrains were eliminated in 31 octaves in *Ninfale Fiesolano*, 4 seven-line octaves in *La Spagna* were omitted, and one seven-line octave was excluded from consideration in *Orlando Innamorato*.

Since the syntactic organization of the octave has long been studied by Italian researchers (for the bibliography see Soldani 1999; Praloran 2003, 2009; Giovine 2017; Juri 2017), it is possible to compare the data we obtained with the hypotheses of Italian scholars based on a qualitative analysis of the material.

Another advantageous feature of the Italian *ottava rima* is the great number of long texts written in this poetic form.

4 Data

TAB. 1 and FIG. 1 show the obtained data.

5 Interpretation of the data

The data we obtained allows us to divide the material into two groups: the *ottava rima* of Poliziano, Ariosto and Tasso, on the one hand, and all the others on the other.
<table>
<thead>
<tr>
<th>Text</th>
<th>No of stanzas</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Filostrato</td>
<td>713</td>
<td>17</td>
<td>32</td>
<td>24</td>
<td>31</td>
<td>23</td>
<td>31</td>
<td>21</td>
<td>97</td>
</tr>
<tr>
<td>2 Nin. Fies.</td>
<td>473</td>
<td>17</td>
<td>33</td>
<td>22</td>
<td>38</td>
<td>25</td>
<td>33</td>
<td>25</td>
<td>97</td>
</tr>
<tr>
<td>3 Spagna</td>
<td>1814</td>
<td>21</td>
<td>54</td>
<td>30</td>
<td>70</td>
<td>33</td>
<td>64</td>
<td>26</td>
<td>94</td>
</tr>
<tr>
<td>4 Morgante</td>
<td>3763</td>
<td>30</td>
<td>46</td>
<td>40</td>
<td>57</td>
<td>40</td>
<td>52</td>
<td>33</td>
<td>93</td>
</tr>
<tr>
<td>5 Orl. Innam.</td>
<td>4428</td>
<td>25</td>
<td>54</td>
<td>32</td>
<td>70</td>
<td>34</td>
<td>49</td>
<td>32</td>
<td>95</td>
</tr>
<tr>
<td>6 Rime</td>
<td>116</td>
<td>19</td>
<td>48</td>
<td>19</td>
<td>79</td>
<td>27</td>
<td>73</td>
<td>29</td>
<td>99</td>
</tr>
<tr>
<td>7 Orl. Fur.</td>
<td>4842</td>
<td>16</td>
<td>51</td>
<td>20</td>
<td>76</td>
<td>21</td>
<td>54</td>
<td>23</td>
<td>97</td>
</tr>
<tr>
<td>8 Ger. Lib.</td>
<td>1917</td>
<td>8</td>
<td>51</td>
<td>11</td>
<td>71</td>
<td>11</td>
<td>55</td>
<td>12</td>
<td>83</td>
</tr>
</tbody>
</table>

**TAB. 1: Strength of the syntactic ties after each octave line, %**

There is a general tendency to divide the octave into distiches, but the distribution of strong and weak pauses within the stanza varies from author to author. At the same time, one can observe a consistent tendency to increase syntactic orderliness.

Boccaccio’s *ottava rima* does not show an obvious tendency to pronounced syntactic segmentation. *La Spagna* and *Morgante* give higher rates, but even here the numbers show a weak syntactic order. Therefore, in both texts there are stanzas that do not conform to the general tendency of dividing the stanza into distiches, for example:

---

**Da ogni parte avea assai che fare;**
**ma que’ Pagani avien miglior partito**
**per che stavan di sopra a guerregiare;**
**e qual cogliea, che non fusse guernito,**
**di questa vita convenia passare**
**e render l’anima al padre gradito.**

*(La Spagna, XI, 12)*

---

**Tu cominciasti insino in Aspramonte**
**a dargli a intender che fussi gagliardo**
**e facessi gran cose a quella fonte.**
**Ma se non fussi stato il buon Gherardo,**
**io so che la vittoria era d’Almonte;**
**ma egli ebbe sempre l’occhio allo stendardo,**
**che si voleva quel dì coronarlo:**
**questo è colui c’ha meritato, Carlo.**

*(Morgante, I, 13)*

---
In pochi dì fur tutti battezzati.
L’abergator che ritenne costoro,
quanto poteva più gli ha ringraziati
Questa novella sentì il barbassoro
e gli altri che Rinaldo avea trovati:
alla città venien sanza dimoro;
e ’l barbassoro avea nome Balante,
e molto gaudio avea del re Vergante.

(Morgante, XIV, 21)

The Pulci octaves are weakly structured syntactically. The strongest syntactical pauses are located after the 4th and 6th lines (and they are almost equal), while the first quatrains does not divide into distiches—the pauses after the 2nd and 3rd lines are almost identical in strength. It should be noted that the difference between the maximum and minimum indices is not very large: the pauses after all the odd lines are relatively strong. This means that Pulci quite often ends the syntactic period after the 3rd, 5th and 7th lines, his syntactic preferences are not defined, and the syntactic movement of the stanza looks random.

The octave of Orlando Innamorato is organized in a much stricter way syntactically. Compared to the Pulci octave, the syntactic pauses are stronger, and the “rhythm of distiches” is more pronounced. Unlike Pulci, Boiardo often separates the first couplet.

The octave of Poliziano belongs to another genre: it is not a narrative but a lyrical octave, much more descriptive by its nature. The pause after the 4th line is very strong (79%), and the final couplet is almost autonomous syntactically (73%).

This rhythmical-syntactic distinctiveness compared to previous authors should not surprise us, since we are dealing with a completely different subject and rhetoric. Still, surprisingly, the data obtained for Ariosto are very close to Poliziano’s.

Italian scholars have noticed this fact previously. Gianfranco Contini wrote that Ariosto wanted “vincere questa scommessa: mantenere la conquista lirica del Poliziano e non rinunciare al carattere narrativo” (Contini 1974: 237). While Marco Praloran described this phenomenon as: “lo stile (...) è la prova della metamorfosi dell’apparato della tradizione lirica in un contesto diverso, appunto narrativo” (Praloran 2003: 15; see also Giovine 2017, and the bibliography, Giovine 2018: 125–126).

Let us describe the octave of Ariosto in detail. It demonstrates an extreme syntactic order, which can be immediately seen in the table. Some clear syntactic preferences are at once noticeable in the octave of Orlando Furioso. A strong syntactic pause after the 4th line becomes almost a constant; the index is extremely high: 76%. The first and the last distich are often autonomous. An even more interesting fact is the weakening of syntactic pauses after odd lines—none of the indices for them exceeds 25%, and after the 1st and 3rd lines it is lower than or equal to 20%! It can be said that Ariosto not only prefers to finish a syntagm after an even octave line, but that he also avoids strong pauses after the 1st and 3rd lines. The octave acquires a distinct rhythmical-syntactic cadence, thanks to which it is possible to predict the structure of the next stanza with great certainty.
Che vi fu tolta la sua donna poi:
ecco il giudizio uman come spesso erra!
Quella che dagli esperi ai liti eoi
avea difesa con si lunga guerra,
or tolta gli è fra tanti amici suoi,
 senza spada adoprar, ne la sua terra.
Il savio imperator, ch'estinguer volse
un grave incendio, fu che gli la tolse.

(Orlando Furioso, I, 7)

In the octave of Tasso, the syntactic tendencies of Ariosto’s octava rima are developed and strengthened. The strong syntactic pause after the 4th line is still very evident here (71%), but the first and last distich become autonomous more often. However, 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 enjambements (three times more than in Ariosto)—breaking between a noun and its adjective, a verb and an adverbial clause of place, etc., and located at the line border (Spoerri 1922: 9, Fubini 1948).

Quivi a lui d’improviso una donzella
tutta, fuor che la fronte, armata apparse:
era pagana, e là venuta anch’ella
per l’istessa cagion di ristorarse.
Egli mirolla, ed ammirò la bella
sembianza, e d’essa si compiacque, e n’arse.
Oh meraviglia! Amor, ch’a pena è nato,
già grande vola, e già trionfa armato.

(Gerusalemme Liberata, 1, 47)

L’un l’altro guarda, e del suo corpo essangue
su ’l pomo de la spada appoggia il peso.
Già de l’ultima stella il raggio langue
al primo albor ch’è in oriente acceso.
Vede Tancredi in maggior copia il sangue
del suo nemico, e sé non tanto offeso.
Ne gode e superbisce. Oh nostra folle
mente ch’ogn’aura di fortuna estolle!

(Gerusalemme Liberata, 12, 58)

TAB. 2 shows the standard deviation data for each text: the presence of two distinct groups is clear.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>25.82</td>
<td>25.47</td>
<td>25.76</td>
<td>19.99</td>
<td>23.79</td>
<td>30.85</td>
<td>30.02</td>
<td>30.72</td>
</tr>
</tbody>
</table>

TAB. 2: Standard deviation for each text

6 Conclusion

Our method, which is based on an elementary computer algorithm using Tomashevsky’s approach, allows for processing thousands of stanzas and hundreds of texts, demonstrating its applicability. After testing it on the Italian ottava rima—the syntax of which has been perhaps more thoroughly studied than that of any other stanza—we see that the results are consistent with the conclusions reached by Italian scholars with the help of selective quantitative and qualitative analysis.
The analyzed texts divided into two distinct groups, corresponding to the chronology. The continuity of Ariosto with respect to the lyric octave of Poliziano, as well as the difference between the octave of Tasso and the octave of Ariosto, was established. This positive result suggests the need for a large-scale study that should include many more texts from different national traditions.

Acknowledgment

This paper is part of a research project based at Lomonosov Moscow State University and supported by Russian Science Foundation Grant № 17-18-01701.

References


