Syntax and Pauses in a Verse Line: Statistical Analysis

Anastasia Kruglova  
Russian State University for Humanities  
nastel.kruglof@gmail.com

Olga Smirnova  
M.V. Lomonosov Moscow State University  
kisaolga@mail.ru

Tatyana Skulacheva  
V.V. Vinogradov Institute of Russian Language, Russian Academy of Sciences  
skulacheva@yandex.ru

Abstract

When Schwarzenegger tries to show that he is a robot or a cyborg he starts to speak with even pauses between words. In regular human speech pauses are of different lengths, and this phenomenon is based on the strength of syntactic breaks between words. This peculiarity of speech turns out to be of special importance for the structure of a poetic line, and by now this has become firmly established in a number of European languages (Russian, Spanish, French, English). It turns out that closer ties and shorter pauses normally occur close to the borders of a verse line (to form a contrast with the longest pauses and weakest syntactic ties between lines). Weak ties and long pauses within a line are concentrated in the middle of the line. This mirror-like opposition between close and loose connections between words at both the syntactic level and the corresponding phonetic one constantly occurs in verse and disappears in prose (various syllabic-accentual meters, various syllabic meters, and free verse were examined). In this article we concentrate on regularities observed in the Russian iambic tetrameter (A. S. Pushkin’s “Evgeniy Onegin”), as well as comparative data from other languages.

1 Introduction

The only feature which is preserved in verse up to the border with prose is the division into lines. At present it is clear that a verse line is not simply a graphic unit but has norms in its inner structure at all linguistic levels. In this article we shall deal with syntactic norms in the organization of a verse line and their stable correlation with corresponding phonetic features (different lengths of pauses in different parts of
a line). It has been shown for a number of European languages—Russian, English, French, Spanish—that there are very persistent norms present in verse, no matter the period (from the 18th through the 20th centuries), literary trend, or the individual styles (Gasparov 1981; Tarlinskaja 1984, 1987; Skulacheva 1989; Skulacheva 1996; Gasparov–Skulacheva 2004; Skulacheva 2014; Kruglova–Smirnova–Skulacheva 2017). These norms involve the words at the beginning and especially at the end of a line being connected by closer syntactic ties, while loose ties within a line concentrate in the middle of a line. The loosest syntactic connection is in the position between lines. The possible explanation for this phenomenon is that a close connection of words near the borders of a line helps to keep a line intact as an integral unit, while a loose syntactic tie between lines supports the division into lines—the most stable feature of a verse text.

2 Closer and looser syntactic ties and shorter and longer pauses at different positions of a verse line

We single out 10 types of syntactic ties derived from traditional grammar (with some minor adaptations for this type of analysis). The fact that we do not use any modern approaches (generative, the “Etap” system by Yu.D. Apresyan and others) is that so far nobody has managed to modify modern syntactic approaches in such a way that they could be instrumental for this type of work and show a good correlation between syntax and pauses. Normally, modern approaches offer very advanced and specialized tools for a completely different type of work. Traditional grammar is very simple and therefore works effectively for a much broader scope of phenomena. Still we plan to work actively with syntacticians proficient in different types of modern syntax to find out if a more formal approach may be suggested for describing this phenomenon. Here is our classification of syntactic ties, which turned out to reveal a significant dependence between syntax and pauses. The ties are enumerated going from the closest to the loosest.

**Close syntactic ties:**

- **a** – attributive tie
  
  solitary hall...;

- **d** – tie of a direct object
  
  holds a sword...;

- **i** – tie of an indirect object
  
  gives to Lara...;

---

m – tie of an adverbial modifier
rarely heard...;

Special types of tie:

p – predicative tie
Lara came...;

c – tie between similar, homogeneous parts of the sentence
knights and dames...;

Loose syntactic ties (loose ties are marked at both borders of the structure they introduce into the sentence, for these are the places where their pauses are normally realized):

cob – tie introducing unattached constructions
And yet they glide like happiness away;

bp – hypotactic tie between clauses of a complex sentence
The immortal lights that live along the sky...;

bc – paratactic tie between clauses of a complex sentence
The Serfs are glad through Lara's wide domain,
And slavery half forgets her feudal chain...;

f – border between sentences marked by a full period, exclamation or question mark
Why comes not Ezzelin? The hour is past...

We have also compared the data on distribution of close and loose syntactic ties within the line and between lines, as well as the distribution of longer and shorter pauses in the reading of the same text. For acoustic analysis, the text of the second and the third chapters of A. S. Pushkin’s “Evgeniy Onegin” were read by a woman with higher non-linguistic education, who is a skilled public reader. The duration of the recording is 60 minutes. Pauses were measured using the Praat program (version 6.0.36) (Boersma–Weenink 2017).

The graph for syntactic ties (FIG. 1) may differ a little in details from the one presented in Skulacheva 1989 and Kruglova–Smirnova–Skulacheva 2017, because in those studies we presented syntactic data for 1000 fully stressed lines, and here we have syntactic data for 284 fully stressed lines on the sound recording (the total number of lines in the recording is 1202). This is due to the fact that to date in phonetics smaller portions of material are normally used—because of the more complicated and time-consuming type of analysis, and because the physical capacity of a reader to participate in the experiment without getting tired is limited. Still the results show the same tendencies as in Skulacheva 1989 and Kruglova–Smirnova–Skulacheva 2017. Contact ties (ties between the last word of one line and the first word of the next) are shown for 898 lines (all other lines had distant ties between them—that is, no ties between the words at the very end of one line and the very beginning of another).

In FIG. 1–2 close syntactic ties are solid violet lines, loose syntactic ties are solid green, while the blue dashed line represents special types of tie. As we can see from FIG. 1, close syntactic ties occur at the beginning (between words of the first and the second
Quantitative Approaches to Versification

FIG. 1: Close and loose syntactic ties within a line in the iambic tetrameter of A.S. Pushkin

feet) and close to the end of a line (between the words of the third and the fourth feet). Loose syntactic ties show a mirror-like opposite distribution: they are not very numerous at the beginning, are extremely scarce near the end of a line, and they reach a maximum in the middle. We observed the same distribution in different periods of Russian (classical verse, dolnik, vers libre), English, and French verse (Skulacheva 1989, 1996). The reason for such a distribution may be the necessity to keep words of a line together as one unit. It is much easier to tear away one word at the beginning or the end of a line than to break a line at the middle by mistake. In classical verse we normally have inertia of line length: if all previous lines are iambic tetrameter we will be expecting that every subsequent line will be of approximately the same length, and if we have a loose syntactic tie in the middle of a line we still will not mistake it for the end of the line. Loose ties between lines support division into lines—the basic feature of a verse text. A long pause within a line close to the pause between lines will distract attention from the latter, which is of primary importance for verse structure.

In FIG. 2 we present the same data but also include the ties between lines (contact ties between a word in the 4th foot of one line and a word in the 1st foot of the next line). As we see, FIG. 2 shows that, as was once suggested by B. Yarkho, the position between lines may serve as a test position for the strength of a tie: loose ties reach the maximum level, while close ties go down to the minimum. This shows that division into lines is very important for verse and that syntax specially works to support it.

In FIG. 3 the pauses of different length at different positions within a line are presented. This actually is the same phenomenon as in FIG. 1, but on the phonetic level.
Quantitative Approaches to Versification

P0 is no pause, P1 stands for the shortest pauses, P2 pauses are a little longer and so on. P5 pauses are the longest. We will see the precise diapason of pauses a little later, when we come to the categorization of pauses by statistical methods. FIG. 3 shows that long pauses behave like the weak ties to which they correspond. The long pauses are not very numerous at the beginning, their maximum within a line occurs in the middle, and they are very infrequent closer to the end, between words in the last two feet of a line (3–4). The no pause (P0) distribution resembles that of close syntactic ties in FIG. 1. There is often no pause between words in the 1st and the 2nd feet of a fully stressed iambic tetrameter line, the absence of a pause is most typical for the position between words in the last two feet of a line (3–4), and the absence of pause falls to its minimum in the middle of a line. This shows that pauses are dependent on the strength of syntactic ties (see the statistics on their interdependence below) and that the reason for their distribution may be the same. The absence of pauses in positions close to the beginning and the end of a line helps to keep the line intact as one whole. Longer pauses are concentrated in the middle of a line, where they can not be mistaken for the end of a line because in classical poetry there is inertia of line length (a poem in iambic tetrameter maintains that meter throughout, as in “Evgeniy Onegin”). So the risk of breaking the line one word earlier is greater than that of taking the middle of a line for its end. Our previous data for Russian free verse (Skulacheva 1996) show that in verse, the length typical of a weak tie still occurs in the middle of a line, but such ties are much less numerous: the inertia of a line length disappears and using loose ties even in the middle of a line presents a greater risk to the integrity of a line.

FIG. 4 shows the same data as FIG. 3 but with the position between lines added. As with syntactic ties, the position between lines attracts the longest pauses, which reach...
FIG. 3: Pauses of different lengths within a verse line in the iambic tetrameter of A.S. Pushkin (5 types of pauses)

FIG. 4: Pauses of different lengths within a verse line and between lines in the iambic tetrameter of A.S. Pushkin (5 types of pauses)

FIG. 5: Pauses of different lengths within a verse line in the iambic tetrameter of A.S. Pushkin (3 types of pauses, P0—no pause)

FIG. 6: Pauses of different lengths within a verse line and between lines in “Evgeniy Onegin” (3 categories of pauses, P0—no pause)
their maximum in that position. The absence of pauses falls to its minimum between lines. As for enjambment, even when it is used often—in less strictly organized verse than Pushkin’s—it still exists as a stylistic device that manifests itself only against the background of the above described norm.

TAB. 1 shows cross tabulation of the type of syntactic tie and the categories of pauses by length.

<table>
<thead>
<tr>
<th>2-Way Summary Table: Observed Frequencies (Dataset “Onegin”). Include condition: fully stressed lines</th>
</tr>
</thead>
<tbody>
<tr>
<td>PD3 no</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>a</td>
</tr>
<tr>
<td>Row %</td>
</tr>
<tr>
<td>d</td>
</tr>
<tr>
<td>Row %</td>
</tr>
<tr>
<td>i</td>
</tr>
<tr>
<td>Row %</td>
</tr>
<tr>
<td>m</td>
</tr>
<tr>
<td>Row %</td>
</tr>
<tr>
<td>p</td>
</tr>
<tr>
<td>Row %</td>
</tr>
<tr>
<td>c</td>
</tr>
<tr>
<td>Row %</td>
</tr>
<tr>
<td>cob</td>
</tr>
<tr>
<td>Row %</td>
</tr>
<tr>
<td>bp</td>
</tr>
<tr>
<td>Row %</td>
</tr>
<tr>
<td>bc</td>
</tr>
<tr>
<td>Row %</td>
</tr>
<tr>
<td>f</td>
</tr>
<tr>
<td>Row %</td>
</tr>
<tr>
<td>Totals</td>
</tr>
</tbody>
</table>

TAB. 1: Contingency table (cross tabulation) of the type of syntactic tie and the categories of pauses by length (3 categories of pauses)

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Chi-square</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Chi-square</td>
<td>693.2958</td>
<td>df = 27</td>
<td>p = 0.0000</td>
</tr>
<tr>
<td>M-L Chi-square</td>
<td>760.7673</td>
<td>df = 27</td>
<td>p = 0.0000</td>
</tr>
<tr>
<td>Cramer’s V</td>
<td>0.5242044</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spearman Rank R</td>
<td>0.7704777</td>
<td>t = 35.009</td>
<td>p = 0.0000</td>
</tr>
</tbody>
</table>

TAB. 2: Statistics for this cross tabulationThe hypothesis that the strength of a syntactic tie and the length of a pause are independent is rejected with the probability of a type 1 error less than 0.00005.
Quantitative Approaches to Versification

Descriptive Statistics (3 pauses)

<table>
<thead>
<tr>
<th>Pauses</th>
<th>Valid N</th>
<th>Mean</th>
<th>Median</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short</td>
<td>332</td>
<td>161.03</td>
<td>147.94</td>
<td>33.32</td>
<td>302.24</td>
</tr>
<tr>
<td>Medium</td>
<td>828</td>
<td>445.13</td>
<td>446.74</td>
<td>304.37</td>
<td>578.95</td>
</tr>
<tr>
<td>Long</td>
<td>398</td>
<td>799.00</td>
<td>720.16</td>
<td>579.76</td>
<td>2555.90</td>
</tr>
</tbody>
</table>

TAB. 3: Descriptive statistics for the 3 classes of pauses: the result of formal classification by the K-means method

This is the rounded diapason for 3 types of pauses (P0 in the graphs is absence of a pause): short pauses—less than 300 ms with the mean value 150, medium pauses—300-600 (mean 450), long—more than 600 (mean 720). In this case we have chosen the median for mean value as providing more robust statistics that are less sensitive to outliers.

One can see that these groups correspond well with the peaks at the multimodal histogram of pause lengths (FIG. 7).

![FIG. 7: The pauses length frequency histogram; 3 types of pauses are marked](image)

The groups of pause lengths for the graphs with 5 categories of pauses are as follows (see TAB. 4).
Here are the rounded diapasons for 5 types of pauses: pauses shorter than 240 ms. with the mean value 130 (very short), 240–420 (350 ms, short), 420–580 (490 ms, medium), 580–820 (670 ms, long), longer than 820 with the mean 1000 ms—very long. P0—absence of pause at the particular position.

In general, categorization of pauses is very helpful in this type of analysis as it helps to neutralize differences in speech tempo between different readers. Normally there are 5 categories + P0 (no pause). So far results exhibiting these categories have been obtained for Russian verse and prose and for Spanish verse (Smirnova 2017; Krivnova–Smirnova 2018; Kruglova–Smirnova–Skulacheva 2017). In our case the reader seems to differentiate most distinctly three types of pauses, though differentiation of 5 groups is also possible.

### 3 Conclusion

Thus we can conclude that close syntactic ties occur at the beginning (between words of the first and the second feet) and close to the end of a line (between the words of the third and the fourth feet). Loose syntactic ties show a mirror-like opposite distribution: they are not very numerous at the beginning, are extremely few near the end of the line and they reach a maximum in the middle of the line. We observed the same distribution in different periods of Russian (classical verse, dolnik, vers libre), English, and French verse. The reason for this distribution may be the necessity to keep the words on a line together as an integral unit. Loose ties between lines support the division into lines—the basic feature of a verse text.

The length of a pause depends on the type of syntactic tie. Long pauses behave like the weak ties to which they correspond. Long pauses are not very numerous at the beginning, their maximum within a line is in the middle and they are very infrequent closer to the end, between words in the last two feet of a line (3–4). The (P0) distribution resembles that of close syntactic ties. There is often no pause between words in the 1st and the 2nd feet of a fully stressed iambic tetrameter line, an absence of pause is most typical for the position between words in the last two feet of a line (3–4), and the absence of pause falls to its minimum in the middle of the line. Absence of pauses in positions close to the beginning and the end of a line helps to keep the line intact as one whole. Longer pauses are concentrated in the middle of lines, where they can not
be mistaken for the end of a line, because in classical poetry there is consistency in line length. The longest pauses appear between lines, thus supporting the division into lines—the basic feature of verse. The position between lines in classical verse is normally the position for testing the strength of a tie: weak ties reach their maximums there, strong ties—their minimums. Also, the longer the pause, the more often it occurs between lines.

Acknowledgment

This work is supported by the grant of Russian Foundation for Basic Research N 19-012-00534 under the supervision of T.V. Skulacheva.

References


