Comma Effect in Reading Russian Sentences with Syntactic Ambiguity

Mikhail S. Vlasov, Elena B. Trofimova and Ulyana M. Trofimova*
Shukshin Altai State Humanities Pedagogical University
53 Korolenko Str., Biysk, 659333, Russia

Do punctuation marks facilitate sentence readability? Tasks consisting of sentences with syntactic ambiguity are suitable for experimental studies. In such studies the subjects deal with a specific pragmatic problem of relative clause (RC) attachment with complex noun phrase (NP) and choose early (N1 modification) or late (N2 modification) closure of ambiguous constructions in Russian. Our experiment showed that the presence or absence of a comma on a RC boundary had different effects on individual interpretation strategies of a certain sentence as well as speech tempo in reading. The experiment showed that the role of the punctuation factor in reading sentences in Russian with late closure prime was negligible. Null prime generally facilitated early closure preference, but there were no significant differences in tempo pronouncing of sentence segments. In addition, there was no correspondence between a pause and a comma on a RC boundary. Comma absence in the sentence with early closure prime caused tempo slowing in pronouncing N1 and reducing preferences of early closure from 100% to 80%. The experiment revealed gender differences in tempo pronouncing of N1 depending on the punctuation factor: females tend to read N1 slower than males. This effect becomes stronger when a comma precedes a RC.

Keywords: Russian language, syntactic ambiguity, relative clause attachment, early closure, late closure, priming, reading aloud, speech tempo, comma effect.

The study is supported by the grant of the President of the Russian Federation for young Russian scholars state support - Project No. MK-5819.2016.6 (Project manager – M.S. Vlasov).


Research area: philology.

Introduction

It is known that linguistic determinants of syntactic ambiguity in a Russian sentence are grammatical word forms, variants of lexical compatibility, omission of some sentence parts, word order, punctuation and other factors. The punctuation factor was studied in previous works on the material of Russian and English ambiguous sentences (Vlasov, 2008).

Syntactical disambiguation in languages of different structure is one of the most topical psycholinguistic problems. For example, such type of syntactic ambiguity as relative clause (RC) attachment (Someone shot the servant [N1]

© Siberian Federal University. All rights reserved
* Corresponding author E-mail address: vlasov_mikhailo@mail.ru; eltrofimova@mail.ru; umt2005@rambler.ru
of the actress [N2] who was on the balcony) has been studied in a number of languages (Fodor, 1998, pp. 285-319), (Fedorova et al., 2007), (Hemforth et al., 2015). It has been proved that the speakers of Russian, Polish, Japanese, Afrikaans, Greek, Dutch, German and French prefer early closure (N1 modification: the servant stood on the balcony). On the contrary, the speakers of English, Arabic, Norwegian, Romanian, Swedish usually prefer late closure (N2 modification: the actress stood on the balcony).

Initially, the choice of RC attachment (N1 vs. N2 modification) was explained by universal parsing principles. It is assumed that the speakers of different languages interpret syntactically ambiguous sentences with one strategy (Fodor et al., 1974). Subsequently, Cuetos & Mitchell (1988) disproved this and showed that the speakers of different languages used different strategies in RC attachments. The universal parsing principle was rejected and a new idea was put forward: the mental process of a RC attachment depends on the language. In addition to these two hypotheses, one more hypothesis was proposed: the choice of a RC attachment strategy depends on subjects’ individual characteristics (Pearlmutter, MacDonald, 1995). Thus, there is still no single point of view on this problem.

Exploring the priming effect in a RC attachment, Iudina (2010) admits a number of factors, influencing RC attachment preferences in Russian (with statistically high early closure preference). They are as follows:

1. The RC length factor implies that a short RC usually modifies N2, but when a RC is long its attachment preferences depend on other factors.

2. Linguistic Tuning Hypothesis, firstly proposed by Mitchell (Mitchell, Cuetos, 1991). In Russian the basic idea is that, when faced with RC attachment ambiguity, readers employ statistical preferences based on the most frequent solutions in general language (Iudina 2010). According to (Fedorova et al, 2009), a Russian RC prefers early closure as the statistically significant tendency.

3. Animate noun factor implies that if a complex NP contains both animate and inanimate nouns, readers tend to choose animate nouns in any position.

4. Context effect: if the previous information is biased towards N1 modification, the readers prefer early closure. According to Iudina (2010) if the context is biased towards N1 modification, 91% of Russian readers prefer early closure, but if the context is biased towards N2 modification, only 60% of subjects choose early closure.

5. Syntactic priming effect predicts 57% of N1 modification preferences after early closure prime and only 46% of N1 modification preferences after late closure prime (Fedorova, 2009).

6. Predicate type effect implies that sentences with verbs of sensory perception and verbs of motion will have more percentage of early closure in reading while sentences with verbs of thinking and intelligence will have more percentage of late closure (Iudina, 2010).

7. Perceived or unperceived ambiguity recognizing can also determine RC attachment (Fedorova et al., 2007).

The analysis of recent works on other languages revealed that prosodic and intonation effects should be considered as the possible predictors of RC attachment. This paper studies the role of prosody and punctuation in RC attachment in Russian, since there are no special works on this topic.

The Role of Prosody and Punctuation in RC Attachment

Traditionally, punctuation and intonation strongly interact in reading sentences with syntactic ambiguity (Fodor, 2002).

Prosodic cues to early closure (N1 modification) tend to be robust across languages.
There are a number of recent works on the role of prosodic boundaries in syntactic disambiguation in RC attachment (White et al., 2012; Jun, Bishop, 2015; Bishop et al., 2015; Cruz-Pavia, Elordieta, 2015), but we cannot find any studies of this topic in Russian. The only work of Podlesskaia (2011) proved structural and prosodic autonomy of a Russian RC from its heads, so we can assume that role of prosody and punctuation could be robust in RC attachment.

On the material of English, Lee & Watson (2011) tested the role of prominence of N1 vs. N2 attachment by holding boundary placement constant. The subjects read the sentences where a boundary occurred late (after N2) and the accent status of N1 and N2 was manipulated. However, prominence patterns (N1 vs. N2 attachment) cannot be held constant when boundary locations vary: in English, the phonological prominence of a head noun is closely related to the boundary location (Bishop et al., 2015).

In written speech, in reading aloud tasks the experimenter can control the boundary locations with punctuation marks. According to the Russian punctuation standard, a comma before RC is obligatory and it reflects the syntactic structure as well as the prosodic boundary. If we admit that prosody and punctuation interact in syntactic disambiguation process in reading (according to Implicit Prosody Hypothesis - IPH), we can answer the following questions: what is the role of punctuation in RC attachment in Russian? How do Russian speakers disambiguate these sentences under different punctuation conditions? Do the comma and prosodic boundary facilitate early closure preference in Russian?

Our experimental study verifies the punctuation factor (comma vs. no comma condition) in RC attachment in Russian in reading aloud tasks.

The Present Study

Material

Taking into account several studies on RC attachment (early / late closure) in Russian, we try to reveal the role of punctuation and prosody in RC attachment preferences. The questionnaire and the reading aloud task were used in the experiment. Six sentences of equal structure were constructed for the experiment, including three ambiguous (test items) and three unambiguous sentences (fillers). The sentences were presented in two experimental conditions: a) comma condition and b) no comma condition (before RC). The null hypothesis was that there were no differences in RC attachment preferences in choosing N1/N2 answers and tempo strategies in reading, depending on punctuation factor. Experimental block is presented below, where (a) – is comma condition list and (b) – is no comma condition list (N1 and N2 in test items are underlined):

(a)

1. Na polu bylo mnogo vody, kotoraya postoianno pribyvala i pribyvala. На полу было много воды, которая постоянно прибывала и прибывала.

2. Prestupnik zastrelil sluzhanku aktrisy, kotoraya otkazalas` pustit` ego v kvartiru. Преступник застрелил служанку актрисы, которая отказалась пустить его в квартиру.

3. Malchik shvyrial v vorob`ev kamniami, kotorye razletalis` vo vse storony. Мальчик швырял в воробьев камнями, которые разлетались во все стороны.

4. Eto byla derevnia, v kotoroi ia tak liubil byvat` v detstve. Это была деревня, в которой я так любил бывать в детстве.

5. Zasedanie bylo posviashchено rassmotreniu zajavlenii rabochikh, kotorye postupili za poslednii mesiat. Заседание было посвящено рассмотрению заявлений рабочих, которые поступили за последний месяц.

(b)
1. Na polu bylo mnogo vody kotoraia postoianno pribyvala i pribyvala. На полу было много воды которая постоянно прибывала и прибывала.

2. Prestupnik zastreli sluzhanku aktrisy kotoraia otkazalas` pustit` ego v kvartiru. Преступник застрели служанку актрисы которая отказалась пустить его в квартиру.

3. Mal`chik shvyrial v vorob`ev kamniami kotorye razletalis vo vse storony. Мальчик швырял в воробьёв камнями которые разлетались во все стороны.

4. Eto byla derevnia v kotoroi ia tak liubil byvat` v detstve. Это была деревня в которой я так любил бывать в детстве.

5. Zasedanie bylo posviashcheno rassmotrenii zaiaavlениi rabochikh kotorye postupili za poslednii mesiats. Заседание было посвящено рассмотрению заявлений рабочих которые поступили за последний месяц.

6. Na perron pribyl poezd kotoryi zhdali uzhe dvoe sutok. На перрон прибыл поезд который ждали уже двое суток.

The test items were: 2, 3, 5. The fillers were: 1, 4, 6.

The Test items had different semantic priming effect (statistically high preference to one of the two types of closure). This effect is well known in a number of RC attachment studies (Scheepers, 2003), (Iudina, Fedorova, 2009), (Traxler, 2014).

In sentence 2, the RC tends to attach N1 (sluzhanka) in 62% and N2 (aktrisa) in 48% of cases (Iudina, 2010), i.e. there is no robust semantic priming effect in this sentence (null prime).

Sentence 3 admits late closure semantic priming effect (late closure prime): the Russian verb razletat`sia is combined more often with inanimate nouns (kamni). According to the Russian National Corpus there are 1879 occurrences of such combination of words with a distance from 1 to 3 items, but the combination with animate nouns (vorob`i) has 504 occurrences only. Such preferences distribution confirms our introspective hypothesis (kamni are combined with the verb razletat`sia more often than vorob`i). This N2 modification preference with late closure semantic priming effect is proved for the Russian language in 60% of all cases (Ibid., 2010).

Sentence 5 has early closure semantic priming effect (early closure prime) with N1 modification preference (zaiaavlениiia). Under these conditions Russian speakers prefer this modification in 91% of cases (Ibid., 2010).

Participants and Procedure
Twenty adult Russian speakers, without philological education, (n=20, males=10, females=10) were asked to read all the sentences aloud (without any training reading) and then answer the questions on RC attachment preferences.

Six sentences were presented to the subjects one by one, including three ambiguous items and three unambiguous fillers. The fillers were needed to inhibit syntactic priming effect, i.e. subject’s self-adjusting to N1 or N2 modification as described in (Iudina, Fedorova, 2009).

All the subjects were divided into two groups: 10 subjects received the sentences with comma condition and 10 subjects - with no comma condition. The latter condition was implicit for the subjects: the second group was not informed that the sentences contained punctuation errors (they were asked to read all the sentences without any training). In the reading aloud task the subjects
of all groups didn’t know about the syntactic ambiguity of the sentences.

The tasks were presented one by one:

1) read the sentences aloud in your ordinary speech tempo;

2) answer the questions on three ambiguous sentences:

Prestupnika otkazalas’ vpusstit’ v dom: aktrisa (N2), b) sluzhanka (N1). Преступника отказала впустить в дом: а) актриса (N2), b) служанка (N1).

V raznye storony razletalis’: a) vorob’i (N1), b) kamni (N2). В разные стороны разлетались: а) воробьи (N1), b) камни (N2).

Za poslednii mesiats postupili: a) rabochie (N2), b) zaiavleniia (N1). За последний месяц поступили: а) рабочие (N2), b) заявления (N1).

For the subjects’ speech recording we used Philips digital recorder (bit rate 384 kbps, sample rate 24 kHz, WAV format). For speech visualization and analysis we used WaveSurfer software (Beskow, Sjolander, 2000-2011). We recorded the following acoustic data:

1) individual speech tempo (in syllables per second);

2) absolute and relative N1 vs. N2 tempo, the difference between these variables;

3) duration of pauses before RC and between the sentences.

As a result, we received 120 interpretations of all the sentences, including 60 interpretations of the test items.

Results and Discussion

Early vs. late closure preferences and punctuation effect

The results of the interpretation task, when the subjects answered the questions about RC attachment preferences, are presented in Table 1.

The hypothesis about semantic priming effect was confirmed: we found moderate early closure preference for sentence 2, moderate late closure preference for sentence 3 and robust early closure preference for sentence 5.

In the second task (reading aloud), as other dependent variables the following acoustic characteristics of subjects’ responses were registered: 1) N1 tempo (in syllables per second); 2) N2 tempo (in syllables per second); 3) the difference between N1 and N2 tempo values (ΔN tempo, in syllables per second); 4) N1 relative tempo to sentence tempo (%); 5) N2 relative tempo to sentence tempo (%); 6) the difference between N1 and N2 relative tempo (ΔN relative tempo, %); 7) duration of the pause before the sentence (in seconds); 8) duration of the pause before RC (in seconds).

The punctuation factor was tested as an independent variable, using SPSS nonparametric procedures. The rates of early vs. late closure responses in comma vs. no comma conditions were verified by Mann–Whitney U–test. There was no significant main effect of the punctuation factor in all test items. There were no differences in speech tempo for

<table>
<thead>
<tr>
<th>Sentence No (test items)</th>
<th>Comma Condition</th>
<th>No Comma Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Early closure (N1 modification), %</td>
<td>Late closure (N2 modification), %</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>40</td>
</tr>
<tr>
<td>3</td>
<td>40</td>
<td>60</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 1. The results of interpretation task (questionnaire)
early vs. late preferences in sentences 2 and 3. Perhaps, the results stem from a low number of sample sentences.

Sentence 5 showed strong preference to early closure (N1 modification), as expected. This confirms the hypothesis of semantic priming and early closure preference in Russian. The strong punctuation factor was revealed in this sentence parsing: N1 was read faster in comma condition with average tempo at 8.65 syllables per second, but slower in no comma condition with average tempo at 7.33 syllables per second. This difference was quasi-significant (Z = -1.89, p = 0.059).

According to the questionnaire, this sentence has robust early closure preference in two conditions. But there is also robust punctuation effect in reading aloud. We assume that N1 tempo tends to be slower in no comma condition and it is caused by parsing difficulty of this ambiguous sentence, i.e. RC attachment difficulty.

### Gender factor and punctuation effect on RC attachment strategies

In the interpretation task, gender preferences for early and late closure were distributed in equal proportions: males as well as females had 50% of each type of closure.

In the reading aloud task, we found gender differences in reading all the test sentences. They are presented in Table 3.

Using Mann–Whitney U–test for independent samples (males vs. females) we revealed significant gender differences in N1 average tempo of all test items. Males tend to read N1 with the average tempo at 6.78 syllables per second, but females – at 5.81 syllables per second (Z=-2.011, p=0.044). N2 average tempo differs to a greater extent: 6.36 syllables per second for males and 5.74 syllables per second for females (Z=-2.824, p=0.005). The average sentence tempo was statistically different between genders: 7.01 syllables per second for males and 6.5 syllables per second for females (Z=-2.225, p=0.026).
Mikhail S. Vlasov, Elena B. Trofimova, ... Comma Effect in Reading Russian Sentences with Syntactic Ambiguity

Table IV. Gender differences in reading test sentences (no comma condition)

<table>
<thead>
<tr>
<th>Speech Parameter</th>
<th>No Comma Condition</th>
<th>Nonparametric Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Males</td>
<td>Females</td>
</tr>
<tr>
<td>N1 average tempo, syllables per second</td>
<td>6.37</td>
<td>5.88</td>
</tr>
<tr>
<td>N2 average tempo, syllables per second</td>
<td>6.40</td>
<td>5.78</td>
</tr>
<tr>
<td>(\Delta N) average tempo, syllables per second</td>
<td>-0.03</td>
<td>0.10</td>
</tr>
<tr>
<td>Sentence average tempo, syllables per second</td>
<td>6.92</td>
<td>6.65</td>
</tr>
<tr>
<td>N1 relative average tempo to sentence tempo, %</td>
<td>0.91</td>
<td>0.89</td>
</tr>
<tr>
<td>N2 relative average tempo to sentence tempo, %</td>
<td>0.94</td>
<td>0.87</td>
</tr>
<tr>
<td>(\Delta N) relative average tempo, %</td>
<td>-0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>Duration of the pause before the sentence, seconds</td>
<td>0.73</td>
<td>0.44</td>
</tr>
<tr>
<td>Duration of the pause before the RC, seconds</td>
<td>0.10</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Table V. Gender differences in reading test sentences (comma condition)

<table>
<thead>
<tr>
<th>Speech Parameter</th>
<th>Comma Condition</th>
<th>Nonparametric Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Males</td>
<td>Females</td>
</tr>
<tr>
<td>N1 average tempo, syllables per second</td>
<td>7.18</td>
<td>5.73</td>
</tr>
<tr>
<td>N2 average tempo, syllables per second</td>
<td>6.32</td>
<td>5.71</td>
</tr>
<tr>
<td>(\Delta N) average tempo, syllables per second</td>
<td>0.86</td>
<td>0.03</td>
</tr>
<tr>
<td>Sentence average tempo, syllables per second</td>
<td>7.10</td>
<td>6.35</td>
</tr>
<tr>
<td>N1 relative average tempo to sentence tempo, %</td>
<td>1.00</td>
<td>0.89</td>
</tr>
<tr>
<td>N2 relative average tempo to sentence tempo, %</td>
<td>0.90</td>
<td>0.91</td>
</tr>
<tr>
<td>(\Delta N) relative average tempo, %</td>
<td>0.10</td>
<td>-0.02</td>
</tr>
<tr>
<td>Duration of the pause before the sentence, seconds</td>
<td>0.54</td>
<td>0.51</td>
</tr>
<tr>
<td>Duration of the pause before the RC, seconds</td>
<td>0.07</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Apparently, these differences reflect different semantic and syntactic processing speed of these test sentences: females tend to parse and choose RC attachment slower than males.

Tables 4 and 5 show punctuation effect on reading strategies of males and females.

Thus, there were no gender differences in N1 average tempo in no comma condition (6.37 in males vs. 5.88 in females). In comma condition, quasi-significant difference appeared: men read N1 at 7.18 syllables per second, women – at 5.73 syllables per second (\(Z = -1.888, p = 0.059\)). That is the comma before RC facilitated faster N1 reading (25%) in males compared with females. The most significant gender differences revealed in average tempo of all test sentences in comma condition: males tend to read them faster (at 7.1 syllables per second) than females (at 6.35 syllables per second). That is, females preferred slower reading of the test sentences. This difference had a high statistical significance (\(Z = -2.136, p = 0.033\)). In no comma condition, no such effect was observed.

**Conclusion**

The comma on Russian RC boundary with two possible NPs can influence sentence tempo in certain conditions.
According to our experiment, there was negligible punctuation effect in sentence 3 (with late closure prime). In sentence 2 (with null prime), no comma condition facilitated early closure preferences, but there were no tempo differences in reading N1, N2 and the whole sentence. There was no congruence between commas and pauses on RC boundary in reading aloud. In other words, no prosodic disambiguation cues depending on the punctuation factor were revealed.

Taking into account the early closure preference in RC attachment in Russian (proved by O. Fedorova), in sentence 5 we revealed that no comma condition has an inhibitory effect on N1 average tempo and reducing early closure preferences from 100% to 80%. Such effect was not revealed in other sentences.

There were gender differences in N1 average tempo depending on punctuation factor. Females tend to read N1 slower than males. This difference increases in comma condition.

Acknowledgments

The study was supported by the grant of the President of the Russian Federation for young Russian scholars state support (Project No. MK-5819.2016.6).

References


«Эффект запятой»
при чтении предложений
с синтаксической неоднозначностью
на русском языке

М.С. Власов, Е.Б. Трофимова,
У.М. Трофимова

Алтайский государственный гуманитарно-педагогический университет имени В.М. Шукшина
Россия, 659333, Бийск, ул. Короленко, 53

Влияют ли пунктуационные знаки на удобочитаемость предложения? Для экспериментального исследования данного вопроса целесообразно использовать предложения с синтаксической неоднозначностью. Например, испытуемым предлагается решить конкретную коммуникативно-прагматическую проблему – выявить вершину определительного придаточного при его отнесении к одному из имен сложной именной группы (ИГ) в главной клаузе. В данном случае возможны две стратегии: предпочтение раннего закрытия (придаточное присоединяется к первому имени ИГ) и позднего закрытия (придаточное присоединяется ко второму имени ИГ). Наше экспериментальное исследование позволило установить «эффект запятой» (на границе главной и придаточной клауз) на уровне индивидуальных стратегий интерпретации и темпа прочтения предложений в разных контролируемых условиях.
семантического прайминга. Так, показано, что «эффект запятой» при чтении предложения с праймом позднего закрытия оказался незначительным. В предложении с отсутствующим прайминг-эффектом пунктуационный фактор в целом повлиял на усиление приоритета раннего закрытия (характерного в целом для русскоязычных носителей), но значимых различий в темпе прочтения разных сегментов предложения выявлено не было. Помимо этого, по результатам эксперимента не было установлено соответствий между наличием паузы и запятой перед придаточной клаузой. «Эффект отсутствия запятой» в предложении с праймом раннего закрытия проявился в значимом замедлении темпа прочтения первого имени сложной ИГ и снижении предпочтения раннего закрытия на 20 %. «Эффект запятой» проявился и на уровне гендерных различий в темпе прочтения первого имени сложной ИГ. Установлено, что мужчины читают первое имя значительно быстрее, чем женщины, при этом при наличии запятой на границе главной и придаточной клаузы данный эффект усиливается.

Ключевые слова: русский язык, синтаксическая неоднозначность, присоединение определительного придаточного, раннее закрытие, позднее закрытие, прайминг, чтение вслух, темп прочтения, эффект запятой.

Исследование выполнено при финансовой поддержке Министерства образования и науки РФ в рамках гранта Президента Российской Федерации для государственной поддержки молодых российских ученых – кандидатов наук, проект № 5819.2016.6 (руководитель – М.С. Власов).

Научная специальность: 10.00.00 – филологические науки.