RELATIVE CLAUSES IN SPOKEN RUSSIAN AND ELSEWHERE: A CORPUS APPROACH

V. I. Podlesskaia (podlesskaya@ocrus.ru)

Russian State University for the Humanities, Moscow,
Russian Federation

The paper addresses the problem of discrepancy between syntactic and prosodic grouping in Russian relative clauses. Basing on oral corpora systematically annotated for prosodic details, the paper demonstrates structural and prosodic "autonomy" of relative clauses from their heads, which previously remained unnoticed in the literature on relativization based mainly on written data.

Key words: relative clauses, syntactic grouping, prosodic grouping, prosodic details.

1. Introduction

In the paper, I discuss the systematic mismatch between syntactic and prosodic boundaries in Russian relative clauses[1]. An example below demonstrates the pattern in question, square brackets show syntactic constituents, bold curly brackets show intonation phrases:

(1)

    {*Mne nravitsja* [*plat'je***}** **{**[*kotoroe ona iz sitca sšila*]]**}**
    I   like   dress   REL   she of printed.cotton made
    'I like the dress that she made of printed cotton '

Syntactically, the head noun 'dress' and the relative clause are within one constituent (are governed by the same maximal projection), but the most common prosodic boundary of this sentence comes between the head noun and the relative clause. And what is still more important; the intonation phrase **{***Mne nravitsja* [*plat'je***}** 'I like the dress' doesn't form a constituent, that is, it is not a full grammatical unit. Bill Croft in his paper on the relation between intonation units and grammatical units (Croft 1995: 847–848) paid attention to the similar phenomenon in spoken English and made the following statement:

"Final relative clauses, embedded as well as adjoined, often make up their own intonation unit. ... This means that in the standard phrase-structure analysis, the clause preceding RC is not a complete constituent: it is missing a modifier of a subconstituent... In contrast, clause-internal RCs are NEVER separated from their head NPs, although the NP+RC grammatical unit is itself often split off from the rest of the clause... None of other syntax-prosody mismatches exhibits this categorical yet schizophrenic behavior."

I must say that in Russian this mismatch is not as categorical as in English, but it is schizophrenic enough to have this Croft's statement as a headline for the rest of my paper.

Basing on corpus data, I will put and try to answer the following research questions:

- What are possible prosodic phrasings in Russian relative clauses?
- What types of phrasing are actually attested in discourse? The answer will be — that most commonly attested prosodic phrasing place a boundary between the head noun and the relative clause (RC).
- How this most common prosodic phrasing functions in natural discourse?
- What are the prosodic symptoms of the boundary between the head noun and the relative clause? I will look at pausation patterns to view one of such cues.
- What are possible syntactic operations that can aggravate the separation of the relative clause from the head noun?
- How can Russian data be mapped to what we know about the interrelationship between syntax and prosody?

The paper is based on the following data:

1. Two oral corpora systematically annotated for prosodic details, incl. pausation and pitch movements: "Night Dream Stories", 1h 50 min of sounding (Kibrik, Podlesskaya 2009) and "Stories about presents and skiing", 35 min of sounding (Xurshudjan 2006)

2. National Corpus of Russian (www.ruscorpora.ru)

2. Three possible prosodic phrasings in Russian relative clauses

Russian allows three main types of prosodic phrasing in relative clauses described below as patterns A, B and C.

Phrasing A. The prosodic boundary comes between the head noun and the RC, this is the phrasing demonstrated above, in (1).

Phrasing B. No mismatch between syntax and prosody — the prosodic boundary comes in front of the head noun. There is only one (!) example with a full noun head in our oral corpora:

(2)   Night Dream Stories[2]

..(04)   i   /uvidel iz   /temnoty, mal'en'kuju ten'   kotoraja \molitsja
           and   saw   from   darkness small   shadow REL   prays
{And [he] saw from the darkness a small shadow that was praying**}**

Phrasing C. Also no mismatch between syntax and prosody — the whole sentence is pronounced as a single intonational phrase without an internal prosodic boundary. This is possible, for instance, for short RCs with contrastive focus. No such cases are attested in our oral corpora, but (3) is a constructed example with the focused intensifier *sama* 'herself' (shown in bold):

(3)

*Mne   nravitsja plat'je kotoroe ona **sama** sšila*
I   like   dress   REL   she herself made
{'I like the dress that she made herself'**}** <and not the one she bought>

In sum: actual prosodic phrasings attested in the corpus follow only the pattern A — putting the break between the head noun and the RC — neglecting the two other possibilities.

3.   Discourse functions of pattern A

In natural discourse, the prosodic phrasing pattern A is used to accomplish one of the three following functional tasks:

Task 1. To detach the head noun and wrap it in a topical constituent of the sentence, while the RC forms the comment constituent:

(4)   Night Dream Stories

i   my   uvideli ..(0.2) /ploščadku,
and   we   saw   ground
..(0.3) na   kotoroj   bylo   množestvo raznyx \červjakov.
on   REL   were   lots   various   worms
'**{**And we saw the ground**}**$_\text{Topic}$ **{**on which there were lots of various worms**}**$_\text{Comment}$

Information structure of (4) is very close to that of:

(5)

{On the ground**}**$_\text{Topic}$ **{**there were lots of various worms**}**$_\text{Comment}$

---

[2]   Breaking into intonational phrases is shown in the original text — by breaking into lines, and also in translation — by bold curly brackets; slashes iconically show the direction of the pitch movement, the nuclear pitch of the intonational phrase is shown by underlying the respective syllable

In relativization of this type, the matrix clause introduces an entity (referenced to by the head noun), while the RC contains a statement "about" this entity. RCs of this type comprise 30% of the total number of RCs in our corpora.

Task 2. To produce the RC as a separate utterance added to the matrix clause as an afterthought:

(6) Night Dream Stories
...(1.4)Togda ..(0.1) moj /komandir /menja /nagradil zolotoj   \medalju.
then   my commander me   awarded golden.INSTR medal.INSTR
··(0.3)   Kotoraja stoila   dvesti   dollarov
              REL   costed   two.hundred   dollars
'Then my commander awarded me with a gold medal. That cost two hundred dollars.'

In relativization of this type, the matrix clause can be articulated with the falling pitch movement in the nuclear pitch, i.e. as an independent clause projecting no continuation in the unfolding discourse. Hence, the relative clause is added after the speaker recognizes the just produced portion of discourse as insufficient and requiring elaboration.

RCs of this type comprise 45% (!) of the total number of RCs in our corpora. The figure 45% being unexpectedly high by itself , becomes even more significant, if we look into other types of postnominal subordinate clauses and see how often they are produced as fragmented afterthoughts. For the Night Dream Stories corpus, the counts published in [Korotaev 2009] show the ratio of clauses used an afterthought to the total number of postpositional subordinate clauses of the given type:
Of the total amount of
- complement clauses — 26.3% are produced as afterthoughts
- adverbial clauses — 50% are produced as afterthoughts
- relative clauses — 44.9% are produced as afterthoughts

So, RCs appear as afterthoughts much more often than complement clauses, and almost as often as adverbial clauses.

Task 3. To produce RCs as parentheticals — semantically, these RCs provide background information; prosodically, they are articulated with lowered pitch, narrower pitch range, reduction in loudness, increased tempo etc.:

(7) Stories about presents and skiing (the parenthetical RC is shown in bold curly brackets)
...(0.5) On ..(0.1) priexal v /\ avtosalon,
          he   came   to car.shop
uh(0.2) vybral bol'šuju krasivuju   /–mašinu
         chose big.ACC beautiful.ACC car.ACC
{kotoruju rešil'   podarit'   svoej   — žene},
REL   decided   to.present his.DAT wife.DAT

...(0.6) /vot,
well
\no um(0.3)   kogda uh(0.2)   on   uznal   /cenu   etoj   mašiny,
but   when   he   found.out   price.ACC this.GEN car.GEN
on   /užasnulsja,
he   was.shocked
i   rešil   čto   \net.
and   decided   that   no
'He went to a car shop, chose a big beautiful car {which [actually] he decided
to give his wife as a present}, but when he found out the price, he was shocked
and decided "NO".'

RCs of this type comprise 25 % of the total number of RCs in our corpora.

In sum: in our spoken corpus, RCs are used (1) as comment constituents, (2)
as afterthoughts, or (3) as parentheticals. These uses are distributed as 30 % : 45 % :
25 %. Thus, in natural discourse, RCs tend to be desubordinated, and show symptoms
of communicative autonomy.

4. Prosodic symptoms of the boundary between the head noun and the relative clause: pausation patterns

For the Night Dream Stories corpus, Table 1 shows how the distribution of pauses
depends on the type of postnominal subordinate clause (counts are based on [Koro-
taev 2009])

Table 1. Number and length of pauses on the left edge of (=in front of)
postpositional subordinate clauses

	complement clauses	adverbial clauses	relative clauses
Number of postpositional subordinate clauses	182	53	52
Mean duration of pauses on the left edge	0.14 sec	0.20 sec	0.24 sec
The ratio of zero pauses to the total number of left edges	74.7 %	66.0 %	55.8 %
The ratio of pauses longer than 0.5 sec to the total number of left edges	11.5 %	15.1 %	19.2 %

As Table 1 shows, pauses that come in front of RCs are much longer than pauses
in front of postpositional complement clauses, and even longer than pauses in front
of postpositional adverbial clauses. Then, the percentage of cases with no pauses at all
is maximal for complement clauses and minimal for relative clauses. Here again,
relative clauses appear to be even more strongly detached than adverbials. Finally,

we consider medial and long pauses (for oral stories, these are, typically, half-second or longer). The table shows that medial and long pauses occur more often in front of relative clauses than in front of two other types of relative clauses.

In sum: pausation patterns convincingly demonstrate the strong prosodic break on the left edge of RSc.

5. Syntactic operations that can aggravate the separation of the relative clause from the head noun

The default most common word order in relative clauses requires that the head noun is immediately followed by the relative pronoun. There are however at least two processes that result in intervening material between the head noun and the relative pronoun; they are schematically shown in (8):

(8)

	Matrix clause	**||**	**Relative clause**
Default RC:	head noun	||	relative pronoun
Discontinuous RC:	head noun ←	||	relative pronoun
RC with pied-piping:	head noun	||	→ relative pronoun

The first process moves the head noun inside the matrix clause to the left from the boundary; the resulting phenomenon is known as "extraposed", or "discontinuous" relatives

The second process moves the relative pronoun to the right from the boundary inside the RC, as a result of a phenomenon known as pipe-piping.

An example of a discontinuous relative given in (9), which is actually a modified pattern example (1):

(9)

{Mne /plat'je nravitsja} {{kotoroe ona iz \sitca sšila}
I dress like REL she of printed.cotton made
'I like the dress that she made of printed cotton '
Lit. '{I dress like}$_{Topic}$ {that she of printed cotton made}$_{Comment}$'

The "detached" head noun (in (9), it is *plat'je* 'dress') usually retains the main phrasal accent (typically, with the rising pitch movement) responsible for signaling the topical status of the first intonation phrase and its non-final status in the sentence, i.e. signaling the continuation of the unfolding discourse. The RC forms a comment constituent and retains its prosodic autonomy as a separate intonation phrase.

Extraposition of RCs is optional in Russian. In the National corpus of Russian approximately 1 % of RCs are discontinious.

Pipe-piping of RCs, on the other hand, is forced in Russian, especially with prepositions, in the National corpus of Russian approximately 30 % of RCs are pipe-piped.

An example of a pipe-piped relative is given in (10):

(10)

> {*Mne   nravitsja   /plat'je*}_{Topic1}
> I   like   dress
>
> {*árukava /kotorogo*ñ_{Topic2} *ásšity iz \sitca*ñComment2}_{Comment1}
> sleeves   REL   made of printed.cotton
>
> Lit. '{I like dress} {< sleeves [of] which> < are made of printed cotton>} '

Of pipe-piped RCs in the National corpus of Russian: 74 % are within preposi-tional phrases, 21.9 % are within NPs, like the one in example (10), other 4.1 % in-clude groups headed by infinitives, comparative forms of adjectives and even finite verb forms, like in (11):

(11)

> {*Eto byla /armija*} {*<komandoval /kotoroj> <general \Samsonov>*}
> this was army   lead   which   general Samsonov
> 'This was the army which general Samsonov lead'
> Lit. '{this was army} {<lead **which>** <general Samsonov>}'

Previous studies (Zaliznjak, Padučeva 1979, Ljutikova 2009 inter alia) have convincingly demonstrated the effect of pied-piping in Russian RCs: if a relative pronoun appears inside a smaller constituent within the RC, the whole this constit-uent is fronted rather than the single relative pronoun. What remained unnoticed so far is that the fronted constituent and the rest part of the relative clause can be articulated as separate intonation phrases (shown with angle brackets in (10) and (11)). Within the fronted constituent the pied-piped relative pronoun, other-wise strictly atonic, may acquire the main phrasal accent. The accent (typically, with the rising pitch movement) signals the topical status of this intonation phrase and its non-final status within the relative clause, which thus acquires its inter-nal information structure. The internal intonation and information structuring, in its turn, increases the autonomy of the RC and levels it with an independent predication.

In sum: discontinuous RCs and RCs with the pipe-piping effect can aggravate the separation of the relative clause from the head noun both syntactically and prosodically.

6. Conclusions: How can Russian data be mapped to what we know about the interrelationship between syntax and prosody

Sun-Ah Jun in her paper on "Prosodic Phrasing and Attachment Preferences" (Jun 2003) reports an experiment in which native speakers of English, Greek, Span-ish, French, Japanese and Korean were asked to produce a sentence meaning 'John chased/saw the dog that bit the cat' to check the most natural prosodic phrasing. The results are summarized in (12):

(12) based on Jun 2003

English:	*{John chased the dog} {that bit the cat}*
Greek:	*{O Giannis kinigise to skilo} {pu dagose ti gata}*
Spanish:	*{Juan vio al perro} {que persiguió al gato}*
French:	*{John a poursuivi le chien} {qui a mordu le chat}*

Japanese: *John-ga   neko-ni kamitzuita inu-o   oikaketa*
John-nom   cat-at   bit   dog-acc   chased
{Johnga} {nekoni kamitzita inuo} {oikaketa}

Korean: *John-i   koyangi-lul mun   kangaji-lul ccochatta*
John-nom cat-acc   bit-that puppy-acc   chased
{Johni} {koyangilul mun} {kangajilul ccochatta}

As shown in (12), speakers of English, Greek, Spanish and French put a prosodic break between the head noun and the relative clause, exactly as in Russian. Japanese, on the other hand, shows a different pattern: the prosodic break comes after the head noun (*inu-o*), so that the relative clause and the head noun form one intonation phrase.

English Greek, Spanish, French and Russian are right-branching with postpositional relatives, while Japanese is a left-branching language with prepositional relatives. So, one could hypothesize that the difference in prosody is conditioned by word-order. But this hypothesis makes a wrong prediction for Korean: in Korean, which is also left-branching with prepositional relatives, the prosodic break comes before the head noun (*kangaji-lul*), so the head noun and the relative clause appear again in separate intonation phrases.

Jun and Chisato (2008) suggest that this may be due to a morpho-syntactic fact, that, unlike Korean, Japanese has no complementizer marking the boundary of a relative clause. If there is a prosodic break after the RC, the verb could be interpreted as a sentence final verb. In (12), this would result in false understanding, like, 'John bit the cat <and then> chased the dog'.

The rule, however, can be overridden and the prosodic break after the RC is favored when the head noun is complex, having an internal nominal modifier, like in the sentence *John chased his friend's dog // that bit the cat*.

Even this short illustration shows that prosodic phrasing is a multifactorial process influenced by various structural and speech production parameters. It is not clear yet, to what extent these parameters are universal. Regarding relative clauses, so far, very little is known about their cross-linguistic prosodic variation. There is some literature on the prosody of relative clauses in individual languages, but it is mostly restricted to the two topics: first, the prosodic difference between restrictive and non-restrictive relatives, and second, the prosodic phrasing in relative clauses with a complex head noun (otherwise known as the problem of distinguishing between "early" and "late" closure).

The absolute majority of this literature is based on experimental, laboratory material, rather than on natural data. However, only natural corpus data can help not only in discovering actual prosodic patterns, but also in understanding why some possible prosodic phrasing patterns remain underrepresented, while others are favored

in particular discourse settings. In this paper, basing on data from the prosodically annotated corpus, I have shown that Russian systematically demonstrates syntactic and prosodic autonomy of the relative clause, favoring the prosodic break between the head noun and the relative clause, thus systematically desubordinatings RCs.. Our corpus data allows to hypothesize that there might be a strong discourse reason for this — namely, speakers tend to produce discourse in such a way, that each clause appears as a separate intonation phrase. But certainly, the typological validity of this hypothesis is to be further checked against natural data from other languages.

So far, this work is intended as an empirical study, but, hopefully, the one that can form the foundation for further theoretical and typological development, for better understanding the nature of syntax-prosody mapping in sentential embedding constructions.

Notes

References

1. *Croft W.* 1995. Intonation units and grammatical structure. Linguistics, 33 : 839–852.
2. *Jun, Sun-Ah.* 2003. Prosodic Phrasing and Attachment Preferences. Journal of Psycholinguistic Research, 32 (2) : 219–49.
3. *Jun, Sun-Ah, Koike Chisato.* 2008. Default Prosody and Relative Clause Attachment in Japanese. Japanese-Korean Linguistics, 13 : 41–53.
4. *Korotaev N. A.* 2009. Syntax and Prosodia in the Systems of Discourse Connection Means [Sintaksis i Prosodiia v Sisteme Sredstv Diskursivnoi Sviaznosti Teksta].
5. *Liutikova E. A.* 2009. Relative Sentences with the Word 'KOTORYI'[Otnositel'nye Predlozhenia s Soiuznym Slovom 'KOTORYI']. Korpusnye Issledovaniia po Russkoi Grammatike : 436–511.
6. *Zalizniak A. A., Paducheva E. V.* 1979. Syntactic Characteristics of the Pronoun 'KOTORYI' [Sintaksicheskie Svoistva Mestoimeniia 'KOTORYI']. Kategoriia Opredelennosti-Neopredelennosti v Slavianskikh i Balkanskikh Iazykakh : 289–329.