



NATIONAL RESEARCH UNIVERSITY  
HIGHER SCHOOL OF ECONOMICS

*Alexander B. Letuchiy*

# **LENGTH OF CONSTITUENT AS A RELEVANT FACTOR IN RUSSIAN SYNTAX**

**BASIC RESEARCH PROGRAM**

**WORKING PAPERS**

**SERIES: LINGUISTICS  
WP BRP 88/LNG/2019**

This Working Paper is an output of a research project implemented at the National Research University Higher School of Economics (HSE). Any opinions or claims contained in this Working Paper do not necessarily reflect the views of HSE.

## **LENGTH OF CONSTITUENT AS A RELEVANT FACTOR IN RUSSIAN SYNTAX<sup>2</sup>**

The article focuses on cases when the length of the constituent is relevant for Russian syntax and leads to ungrammaticality or partial acceptability of some structures. The list of phenomena under consideration includes ‘one-word restrictions’; differences between complement clauses of matrix verbs in indicative vs. imperative, include constructions with ‘predicatives’, special properties of nominalizations, structures with short embedded clauses, and so on. I show that the ‘length restrictions’ are in fact of different nature: some are related to discourse factors (deictic / non-deictic semantics, discourse coherence, imperative addressee orientation), while others are related to parsing needs (grammatical units are typically shorter than lexical ones, main clauses are more explicit than embedded ones). Another feature of length restrictions is that they are never holistic / integral: the particular rule usually holds for a particular type of structure, not for the whole domain of syntax (this is what makes our data different from purely syntactic restrictions, such as c-command requirements and island restrictions, which often cover a broad set of constructions and grammatical contexts).

Keywords: length of constituents, predicatives, complement clauses, mood, subject expression, one-word restrictions

JEL Classification: Z

---

<sup>1</sup> National Research University Higher School of Economics (Moscow, Russia). School of Linguistics. Associate Professor; e-mail: alexander.letuchiy@gmail.com

<sup>2</sup> The publication was prepared within the framework of the Academic Fund Program at the National Research University Higher School of Economics (HSE) in 2018-2019 (grant № 18-01-0032) and by the Russian Academic Excellence Project "5-100"



This paradox in principle, can be reduced to big syntax phenomena. One could assume that non-canonical (internally) phrases are also non-canonical externally: for instance, the position that NPs occupy in the argument structure is not the same as that of DPs. However, this point of view leads to counterintuitive results. The main of them is that a large number (potentially all) heads have two argument structure versions: one ‘canonical’, another one ‘non-canonical’, reserved for arguments that do not have the whole DP structure. Postulating these two structures leads to a significant increase of the lexicon.

I deal with another type of phenomena that lies between syntax and pragmatics: namely, cases when the length of the constituent influences grammatical processes or, vice versa, is affected by grammatical processes. Hawkins (1990) showed that the length of constituents influences the relative easiness of parsing of various word orders, which, in turn, can result in grammaticality contrasts if for some orders, the parsing is too complicated. Diessel & Schmidtke-Bode (2017) show the relevance of Hawkins’ approach for one domain of grammar, namely, complementation. They show, for instance, that the parsing account explains the tendency of finite complement clauses to occur after the main clause and nominal arguments and tendency of infinitive complements to change their position rather easily. I will apply the same approach to a number of Russian constructions and show that the length and parsing needs can together account for some tendencies of linear order in Russian and that some structures that are marginal from the point of view of length can become partially acceptable or completely ungrammatical. The sample of phenomena under analysis will be divided in two groups: (i) cases when grammatical phenomena influence the length (Section 2) and (ii) cases when the length of constituents affects the grammaticality of some constructions (Section 3).

In the cases when statistics is considered, it is based on the Russian National Corpus ([www.ruscorpora.ru](http://www.ruscorpora.ru)). Examples and figures are taken from the main corpus: although the sample from the main corpus may sometimes contain examples we did not look for (e.g., in some uses, *čto* is a pronoun, and not a complementizer), they are statistically significant, which is not always the case in the corpus with eliminated homonymy.

## **2. Imperative constructions as a factor of length**

First, consider cases when grammatical properties of constituents affect the length of clauses. It is represented by the mood opposition in the matrix clause of complex sentences, more precisely, the opposition of indicative vs. imperative. I show that the complement clause of a verb marked with imperative tends to be shorter / less explicit than the clausal complement of an indicative verb. It is well-known that the subject of imperative is often omitted. However, as I will see, the

relevance of length does not only cover the expression of subject. It manifests in three tendencies related to the embedded clause:

- 1) omission of the subject in the clause embedded under imperative is observed more often than under indicative in constructions with *čtoby*;
- 2) sometimes omission of the object of the main clause is better with a matrix verb in imperative than in indicative;
- 3) when the embedded subject with the indicative matrix clause is omitted, this usually takes place if the matrix subject is also omitted.

### 2.1. Imperative and subject omission

In constructions with the complementizer *čtoby*, the subject of the embedded clause can be omitted. The zero subject is co-referent to the DO or IO of the matrix clause. Omission is possible both if the main verb is in indicative and if it is in imperative, but with imperative, the proportion of examples with subject omission is greater. The relevance of the mood is shown in Table 1 where all relevant examples are considered, and Table 2, where the sample is restricted with constructions with a dative noun before the complementizer.<sup>3</sup>

Table 1. Proportion of embedded clauses with a nominative noun vs. indicative verb form in constructions with *čtoby*.

without dative	(S SPRO), nom + V,indic,praet	V,indic,praet
Indic	13109	2627 (20%)
Imper1	856	302 (35,3%)

Table 2. Proportion of embedded clauses with a nominative noun vs. indicative verb form in constructions with *čtoby* and a dative argument of the main verb.

with dative	(S SPRO), nom + V,indic,praet	V,indic,praet
Indic	795	243 (30,6%)
Imper1	147	140 (95,2%)

The difference in the proportion of clauses with subject omission is most illustrative in Table 2 when there is a dative argument in the matrix clause (the sample, thus, primarily includes speech act and object transmission predicates). However, without a dative argument, the difference is also rather significant.

Interestingly, in constructions with *čto*, the picture is different.

<sup>3</sup> Constructions with an explicit subject were searched by adding the requirement “a pronoun or a noun in dative immediately after the complementizer”. Constructions with a subject omission were, by contrast, searched using the requirement “a verb immediately after the complementizer”. Although there are some examples where the subject of the embedded clause is expressed but is situated after the embedded verb, examples where the subject is not expressed (omitted) seem to be more numerous in the search sample.

Table 3. Proportion of embedded clauses with a nominative noun vs. indicative verb form in constructions with *čto* and a dative argument of the main verb.

with dative	(S SPRO), nom + V,indic,praet	V,indic,praet
indic	6434	7994 (124,2%)
Imper1	946	569 (60,1%)

By contrast to *čtoby*, with this complementizer, the proportion of constructions without a subject (at least with no subject that immediately follows the complementizer) is more than two times lesser than with a subject. Does it result from some properties of *čto* or of the embedded clause? The difference becomes even greater if we take into account the fact that the search results of *čto* include many examples like (2) where *čto* is a subject pronoun, not a complementizer:

- (2) *Skaži-te mne čto proizoš-l-o.*  
 tell-IMV.2PL I.DAT what.NOM happen-PST-SG.N

‘Tell me what happened!’ [Надежда Трофимова. Третье желание // «Звезда», 2003]

*А теперь представьте себе, что появится в небе Москвы тоже одиночный, но не самолёт-разведчик, а носитель атомной бомбы.* [Владимир Губарев, Анатолий Савин. Академик Анатолий Савин: «Всю жизнь на разных фронтах» // «Наука и жизнь», 2009]

‘And now imagine that a single plane emerged in the Moscow sky, but it is not a scout plane, but an atomic bomb carrier.’

By contrast, cases of the real subject omission are rarely found with *čto* and imperative:

- (3) *??Pojmi čto ošib-Ø-sja.*  
 understand.IMV COMP make.a.mistake-PST.SG.M-REFL

Intended: ‘Realize that (you) made a mistake.’

- (4) *Ja ponja-l-Ø čto ošib-Ø-sja.*  
 I.NOM understand-PST-SG.M COMP make.a.mistake-PST.SG.M-REFL

‘I realized (it) that (I) made a mistake.’

The reason is that the two complementizers have different control properties. *Čto* allows a subject omission under subject control, while for *čtoby*, as mentioned above omission is possible in the object control situation (under subject control, *čtoby* changes to infinitive). If *čtoby* is used with imperative, this means that the speaker’s order is addressed indirectly to the dative argument, not only to the addressee. In this sense, the dative argument is influenced by the speaker. This fact is iconically reflected in the affinity of clauses: the embedded clause can easily be used with no explicit subject.

If *čto* used in the imperative, this means in the subject control case that the addressee is asked to understand, know, say or something like this about himself ('[you] understand that you made a mistake'). The degree of clause affinity is lesser here than under indicative. The reason is that only the embedded situation is realized (if the matrix verb is factive), and the main situation is not.

'I realized that I made a mistake': both 'I realized' and 'I made a mistake' are true.

'Realize that you made a mistake': 'Realize' is not true, and 'I made a mistake' is true.

The two events have different reality status, and their semantic difference is reflected in the fact that the subject is expressed independently in the embedded clause.

Moreover, the omission of the embedded subject in constructions with *čto* leads to a situation when both subjects (that of the imperative matrix verb and the embedded one) are omitted, which makes structures like (3) more problematic for parsing.

The tolerance of *čtoby* to structures with omitted subjects is also obvious from the fact that sometimes omission of the subject embedded under imperative is possible even if omission under indicative is dubious;

(5) - *Vasj-a*                      *zdes'*. – *Skaži*,              *čtoby*              *pozvoni-l-Ø*              *mne*.  
 Vasja-SG.NOM              here      tell.IMV              COMP              call-PST-SG.M              I.DAT  
 'Vasja is here. – Tell (him) to call me.'

(6) *Ty*                      *pozvoni*              *poka*      *ét-im*              *Dorožkin-ym*.  
 you.NOM              call.IMV              NOW      this-PL.DAT      Dorozhkin-PLDAT  
*Skaži*                      *čtoby*                      *side-l-i*              *doma*.  
 tell.IMV                      that                      sit-PST-PL              home

'Call these Dorozhkins, - ordered Gennady and turned to the road. – Tell (them) to stay at home and not to leave.' [Александра Маринина. Последний рассвет (2013)]

(7) *?Ja*                      *skažu*,                      *čtoby*                      *pozvoni-l-Ø*                      *tebe*.  
 I.NOM                      tell.IMV                      COMP                      call-PST-SG.M                      you.DAT  
 'I will tell (him) to call you.'

When the embedded subject with the indicative matrix clause is omitted, this usually takes place if the matrix subject is also omitted:

(8) *Skaza-l-Ø*,                      *čtoby*      *vyxodi-l-a*              *za*              *nego*                      *zamuž*.  
 tell-PST-SG.M                      COMP      go-PST-SG.F              for              he.ACC                      married  
 'He told me to marry me.' ... [Владимир Личутин. Любостай (1987)]

'[He loved only me and thought only about me all these years]. He told me to marry him (lit. 'that I would marry him.'

(9) *?On skaza-l-Ø, čtoby vychodi-l-a za nego zamuž.*  
 he.NOM tell-PST-SG.M COMP go-PST-SG.F for he.ACC married  
 ‘He told me to marry me.’

(10) *Skaza-l-Ø čtoby prisla-l-Ø mašin-u.*  
 tell-PST-SG.M COMP send-PST-SG.M car.ACC  
 ‘[I have just called my father]. (I) told (him) to send a car.’ [Андрей Геласимов. Год обмана (2003)]

(11) *?Ja skaza-l-Ø čtoby prisla-l-Ø mašin-u.*  
 I.NOM tell-PST-SG.M COMP send-PST-SG.M car.ACC  
 ‘I have just called my father. (I) told (him) to send a car.’

Exceptions are found but sometimes seem to be dubious:

(12) *ja i skaza-l-a čtoby še-l-Ø na vse čtyr-e storon-y.*  
 I.NOM and tell-PST-SG.F COMP GO-PST-SG.M on  
 all-PL.ACC FOUR-ACC side-PL.ACC

‘[In Strastnaya street he got entirely crazy], and I told him to hit the road.’ [Майя Кучерская. Тетя Мотя // «Знамя», 2012]

In (12), it seems preferable to express the subject (*čtoby on shel sebe na vse chetyre storony*).

## 2.2. Imperative and the average length of the embedded clause

Another tendency related to imperative the average length of the embedded clause is lesser under the imperative matrix verb than under the indicative matrix verb.

The tendency is primarily observable on the clause with *čto*.

Table 4. Number of embedded clauses of various length introduced by *čto*.

Imper	Total	1	2	3	4	5
3,94%	35556	1700	1982	2058	1891	1194
		4,78%	5,57%	5,79%	5,32%	3,36%

Indic	Total	1	2	3	4	5
100%	902458	13875	30203	40934	47177	35954
		1,54%	3,35%	4,54%	5,23%	3,98%

It turns out that the peak for indicative main clauses is on the length of four words. The value is less for 3-word, and much less for 2-word clauses. By contrast, with imperative main clauses, the peak is on 3-word embedded clauses, 2-word clauses are only a little bit rarer, while 4-word clauses are rarer than 4-word ones. The difference is also observable on 1-word clauses which

are extremely rare with indicative main clauses (1-word clauses are more than two times rarer than all other ones) and not so rare with imperative main clauses.

Surprisingly, the length of clauses with *čtoby* shows no correlation with the mood of the main clause. The proportion of clauses with each length is almost always equal for the main clause in imperative and in indicative (the most significant difference is in the proportion of 4-word clauses, but even here it is only 0,43%). Perhaps, the reason is that *čtoby* rarely allows omission of the embedded subject, which is mainly favoured by the imperative mood.

Table 5. Number of embedded clauses of various length introduced by *čtoby*.

	Total	1 word	2 words	3 words	4 words	5 words
Imper	4314	58	212	323	315	241
		1,34%	4,91%	7,49%	7,30%	5,59%
Indic	61 020	826	3007	4576	4714	3527
		1,35%	4,93%	7,5%	7,73%	5,78%

The difference between imperative vs. indicative matrix verbs with *čtoby*-clauses is not reducible to the difference between expressed and unexpressed subjects. We have seen that precisely in *čto*-clauses with the matrix verbs in imperative there is no tendency for the subject to be omitted – by contrast, with this complementizer, imperative makes subject expression more frequent than indicative.

The dependence between the mood of the main clause and the length of the embedded clause seems to be explicable if we assume that the two clauses of the complex sentence tend to share some syntactic and other characteristics. The word ‘share’ is used intentionally: this feature sharing is not reducible to either head-complement relations or agreement. The difference in the omissibility is statistical rather than strict grammatical, and no agreement pattern can be proposed (agreement tends to be observed in grammatical categories, and expression of an argument is not a category). Thus, we deal with a soft requirement of clause similarity. One of the tendencies is that imperative constructions should be short, rather than long, and argument omission in the embedded clause indirectly affects the length of the sentence.

Note that some other constructions, namely, adjunct clauses, also show argument accommodation regulated by explicitness / implicitness. In Russian, either both arguments of the verb in the clausal adjunct must be expressed or both omitted:

- (13) *Oni ego xvalj-at potomu\_čto ljubj-at (ego).*  
 they.NOM he.ACC praise-PRS.3PL because love-3PL he.ACC  
*Oni ego xvalj-at potomu čto oni ljubjat \*(ego).*  
 they.NOM he.ACC praise-PRS.3PL because love-3PL he.ACC

‘They praise him because they love him.’

The subject omission is compatible with explicit object, but the object omission requires the subject also to be omitted.

### 3. Length as a factor of grammaticality

#### 3.1. Nominalizations

Nominalizations occupy an intermediate position between complement clauses and nominal arguments. They have nominal properties but can retain most arguments of the base verb or predicative. The Russian data shows that we should distinguish ‘light’ nominalizations that prefer leaving the verb subject unexpressed in the co-reference contexts vs. ‘heavy’ ones that express them with the attributive pronoun *svoj* ‘own’ or other means. It seems that the distinction between ‘heavy’ and ‘light’ nominalizations is based both on pragmatic and lexical properties.

Consider the following examples with matrix verbs *stremit’sja* ‘seek’ and *nadejat’sja* ‘hope’:

(14) *Ivanov-Ø stremi-l-Ø-sja k (??svo-emu) vozvraščenij-u*  
Ivanov-NOM seek-PST-SG.M-REFL to (own-N.SG.DAT) return-SG.DAT  
*na rodin-u.*

to motherland-SG.ACC

‘Ivanov seeks to the return to Russia.’

(15) *Ivanov-Ø nade-ja-l-Ø-sja na (svo-e) vozvraščenij-e*  
Ivanov-NOM hope-PST-SG.M-REFL to (own- N.SG.ACC) return-SG.ACC  
*na rodin-u.*

to motherland-SG.ACC

‘Ivanov hoped for (his) return to Russia.’

The verb *stremit’sja* ‘seek’ does not take subject-referring possessive reflexives in nominalizations (this is why *svoj* sounds badly in (14)). This can be related to the fact that *stremit’sja* mainly occurs in control structures when used with a complement clause. By contrast, for *nadejat’sja* ‘hope’ the control structures do not unequivocally win over the non-control ones. This is why in constructions with the nominalization, the reflexive can be used to mark the co-reference of the main clause subject and the agent of nominalization.

In some cases, semantics of the matrix clause matters.

#### 3.2. ‘One word restrictions’

Outside imperative and nominalized constructions, length also matters. Hawkins (1990, 1991) shows that the length and syntactic weight of the constituent is often relevant for the choice of

the default word order in a particular language. At the same time, in the same language system, length can be relevant for the properties of some constructions and irrelevant for others.

I will list some restrictions that are related to constituent length but can be explained from other factors:

(1) in constructions with predicatives, main clause can rarely include only a predicative:

- (16) ??*Kogda my priezža-em v Adygej-u žarko.*  
 when we.NOM come-PRS.1PL to Adygheya-SG.ACC hot.PRAEDIC  
 ‘When we come to Adygheya, (it is) hot’

(2) constructions with *dlja* ‘for’ rarely combine with a predicative with no other dependent words:

- (17) \**Dlja menja xolodno.*  
 for I.GEN cold.PRAEDIC  
 Intended: ‘(It is) too cold for me’

(3) sometimes adjectives are used postpositionally, as markers of a property that is added to the information mentioned before. This position is mainly available for coordinate phrases but not single adjectives:

- (18) *Sneg-Ø, bel-yj i mjagkij, vse pada-l-Ø*  
 snow-SG.NOM white-M.SG.NOM and mild-M.SG.NOM PART fall-PST-SG.M  
*i pada-l-Ø.*  
 and fall-PST-SG.M  
 ‘The snow, white and mild, continued to fall’

- (19) ??*Sneg, mjagk-ij, vse padal i pada-l-Ø.*  
 snow-SG.NOM mild-M.SG.NOM PART fall-PST-SG.M and fall-PST-SG.M  
 Intended: ‘the snow, mild, continued to fall’

In fact, different restrictions are explicable from different factors. For instance, restrictions on the postposition of adjectives seems to result from parsing restrictions. It is well-known that preposed short constituents do not pose problems for parsing (because the constituent does not contain any dependent elements), while preposed long constituents do (after parsing the left boundary, the listener has to parse the whole coordinate structure before return to the head noun *sneg*). However, for Russian the coordinate preposed phrase is not blocked: the variant with preposition is also acceptable.

The same rule is valid for the postposed use of adjectives with explicit arguments. The contrast here is less sharp, but the postposition is better for heavy NPs:

- (19) a. *Mužčina, gord-yj svo-im syn-om, vyše-l-Ø.*

man-SG.NOM PROUD-M.SG.NOM own-M.SG.INS son-SG.INS go.out-PST-SG.M

‘The man, proud of his son, moved aside.’

- b. *Mužčina, gord-yj, vyše-l-ø.*  
 man-SG.NOM proud-M.SG.NOM go.out-PST-SG.M  
 ‘The man, proud, moved aside.’

This opposition remains of the English analogue – however, in English, only the postposition is possible for heavy NPs:

- (20) a. \**The proud of his son man moved aside.*  
 b. *The man, proud of his son, moved aside.*

By contrast, the restriction on **predicative** results from pragmatic factors and rules of argument expression in the main and the embedded clause. The pragmatic explanation says that the structures with predicatives and without a locative and temporal adverbials require a deictic interpretation. Isolated predicatives denote by default the state of arts at the moment of speech in the place of communication:

- (21) *Žarko.*  
 hot.PRAEDIC  
 ‘It is hot.’

When a locative and / or temporal modifiers are added, they shift the interpretation to the non-deictic ones:

- (22) *Na Lun-e xolodno.*  
 on moon-SG.LOC cold.PRAEDIC  
 ‘It is cold on the Moon.’

- (23) *Segodnja syro.*  
 today wet.PRAEDIC  
 ‘It is wet today.’

However, in (16), there is no temporal adverbials. Thus, the deictic interpretation of the predicative and the non-deictic interpretation of the embedded clause contradict each other. Another factor that may be responsible for the unacceptability of (16) is the tendency that the embedded clause should not be more explicit than the matrix one. See the following group of examples:

- (24) a. *Kogda prosnu-l-Ø-sja by-l-Ø ešče pjan-yj*  
 when wake.up-PST-SG.M-REFL be-PST-SG.M yet drunk-M.SG.NOM  
 ‘When (I) woke up, (I) was drunk yet.’  
 b. *??Kogda vozvraščaj-u-s’ domoj, pjan-yj.*  
 when return-PRS.1SG-REFL home drunk-M.SG.NOM

‘When (I) come home, (I am) drunk.’

In both examples, the subject is omitted both in the matrix and the embedded clause. However, the variant with an explicit copula is perfectly acceptable, while the one with a zero copula is dubious. The difference does not have to do with the expression of person: in (24a), the past tense forms *prosnulsja* ‘woke up’ and *byl* ‘was’ do not express person meanings either, while the number is calculable from the singular form of *pjanyj* ‘drunk (SG)’. The reason is that in (24b), the matrix clause does not contain an explicit verb, while the embedded clause does. In (24a), both clauses include verbal forms. Normally, matrix clauses are explicit, and omissions take place in embedded clauses.

It is very plausible that all sentences with zero predicate are considered as not explicit, though they (both in their binominative version, as in (24b), and in the impersonal version, as in (16)) do not have a more explicit variant. Thus, in both examples the embedded clause is a priori more explicit since it has an explicit verb. The only question remaining is why (25) is impossible:

(25) \**Kogda pyan-yj kak sumasšedšij.*  
 when drunk- M.SG.NOM like crazy-M.SG.NOM

‘When he is drunk he is like crazy.’

No clause in (25) has an explicit subject. Perhaps, another restriction is that at least one clause must have a verb expressed.

### 3.3. Independent clause length

We have shown that the components of complex sentences are subject to explicitness restrictions. However, some restrictions are imposed on independent clause. It turns out that in sequences of two sentences, the second one is worse if it contains only one phrase (in (27), it is an adverbial phrase). If the second sentence consists of two phrases (two AdvPs in (26)), the entire sequence becomes acceptable.

(26) *Menja zov-ut Vasj-a. Sečas v Moskv-e.*  
 I.ACC call-PRS.3PL Vasja-SG.NOM now in Moscow-SG.LOC

‘My name is Vasja. Now (I) am in Moscow.’

(27) ?*Menja zovut Vasja. V Moskv-e.*  
 I.ACC call-PRS.3PL Vasja-SG.NOM in Moscow-SG.LOC

Intended: ‘My name is Vasja. (I) am in Moscow.’

In principle it is possible to regard (27) as restricted by the deictic reading. However, the analysis valid for predicatives is doubtful for cases like this. Locative constructions without an explicit verb do not have an obvious deictic component, which predicatives have. Perhaps, here the length of the constituent directly affects the grammaticality. What is unusual about cases like

(26) and (27) is that the role of the length (or syntactic heaviness) is valid for a sequence of two sentences, rather than for one complex sentence.

### 3.4. Embedded clauses with subject omission

Although infinitive constructions are very often short, this is not the case with another type of embedded constructions – finite clauses with subject omission. Those tend to be long, and sometimes very short embedded clauses look strange:

- (28) *?Dumaj-u*                      *čto*    *ošibsja.*  
 think-PRS.1SG                      COMP    make.a.mistake.PST.SG.M-REFL  
 ‘I think that I made a mistake.’
- (29) *Dumaju*                      *čto*    *mog*                      *by*    *ét-o*                      *sdela-t’.*  
 think-PRS.1SG                      COMP    can.PST.SG.M    IRR    this-M.SG.ACC                      do-INF  
 make.a.mistake.PST.SG.M-REFL  
 ‘I think I could make it.’
- (30) *?Ja*                      *ne*    *ljublju*                      *kogda*                      *padaj-u*                      *v*    *luž-u.*  
 I.NOM                      NEG    love-PRS.1SG    when                      fall-PRS.1SG    in    paddle-SG.ACC  
 ‘I don’t like falling into paddles / I don’t like it when I fall in paddles.’
- (31) *Ja*                      *ne*    *ljublju*                      *kogda*                      *vdrug*                      *okazyva-u-s’*  
 I.NOM                      NEG    love-PRS.1SG    when                      suddenly                      get-PRS.1SG-REFL  
*v*    *durack-om*                      *položeni-i.*  
 in    stupid-N.SG.LOC                      situation-SG.LOC  
 ‘I don’t like suddenly getting into a stupid situation.’

The reason is the same as with prepositional constructions analyzed in 3.5 below. Subject omission makes the embedded clause lighter, and if there are few additional elements (e.g., no adverbials or objects), the length of the embedded clause and the complementizer becomes similar. This contradicts the general tendency saying that grammatical elements should be shorter than phrases they modify or introduce.

### 3.5. Complement clause length

In fact, complement clauses as a group are nonstandard in terms of length. However, the very CC status does not lead to the final position. As predicted by (Schmidtke-Bode & Diessel 2017), infinitives easily change their position, but finite complement clauses are also movable to the initial position.

At the same time, in some complementation strategies, the length is relevant. This is the case with the combinations “pronoun *to* + infinitive” which disfavor short complement clauses:

- (32)            *??Ja  znaj-u      o      t-om      čto      dela-t’.*  
                   I.NOM know-PRS.1SG about that-SG.LOC what.ACC do-INF  
                   ‘I know what to do.’ (lit. ‘I know about what to do’).

When the indirect question becomes longer, infinitives become more acceptable:

- (33)            *Ja      znaj-u      o      tom      čto      nam*  
                   I.NOM know-PRS.1SG about that-SG.LOC what.ACC we.DAT  
*sejčas      dela-t’.*  
 now          do-INF  
                   ‘I know what we should do now.’ (lit. ‘I know about what to do for us now’).

The reason is that the complementation marker in (32) and (33) is heavy. Thus, short complement clauses lead to the non-standard situation when the number of words in the marker and the embedded clause is almost equal.

#### 4. Non-integrality of restrictions

Although, as I have shown, length can influence the (un)acceptability of constructions of various types, restrictions motivated by length are never integral. We do not get restrictions like (34) based on length:

- (34) Main clause cannot include just one word.

They are mainly linked by a particular type of construction or grammatical context:

- (35) In constructions without an explicit verb, the main clause usually has more than one word.

- (36) In constructions where the matrix subject is omitted, the embedded clause with an omitted subject tend to be long.

This results from the fact that we mainly discussed those cases where length distinctions influence grammaticality judgments. Of course, integral tendencies will emerge if we consider statistical tendencies. For instance, it is a general tendency that the main clause is longer than the embedded one, but only in some cases does this tendency manifest itself in grammaticality.

It seems that length plays a role in grammaticality in what can be called ‘fragile contexts’ or ‘borderline contexts’. These are constructions peripheral by some features, and a minor change is enough to change its acceptability ratings. For instance, clauses without a verb are non-canonical, peripheral as compared to sentences with explicit verbs. Adding to this non-canonical feature a non-canonical pragmatic property (the fact that the main clause only contains one word) violates this equilibrium and may make the whole construction ungrammatical.

## 5. Conclusions

It turns out that the length of constituent is relevant for many syntactic processes in Russian. Sometimes the shortness of the constituent leads to ungrammaticality of the construction or the choice of one of the grammatical variants, while in other cases, grammatical oppositions (in our study we addressed the mood opposition) influence the length of constituents.

The extreme case is represented by complement clauses. They differ from standard nominal arguments with several properties: absence of case, syntactic heaviness, and length. In some cases, the CC length is relevant for the grammatical processes.

What is rather important is that complement clauses are linked to the main clause looser than nominal arguments. This seems another reason why they tend to be situated the rightmost of arguments.

Another tendency, observed with imperative matrix verbs shows that the length can be affected by the accommodation of two clauses. This tendency manifests itself also in other ways: accommodation of verb forms, change of complementation strategy and distribution of strategies under matrix clause negation.

The Russian system shows that restrictions motivated by the length of the constituent are usually relative what is important is the length of constituent with respect to another one, and not its absolute length. However, the precise content of the rule can be different:

- length can make some structures dubious, ungrammatical or awkward (e.g., for short finite complements with omitted subjects);
- length can favour a particular linear order of constituents (e.g., sentential vs. nominal arguments);
- length can serve as a link between the grammatical features of words (mainly heads, as for imperative vs. indicative opposition) and the discourse properties.

Finally, length never serves as a general grammatical factor that favors or disfavors general grammatical phenomena or groups of phenomena. For instance, the one word restrictions do not lead to a prohibition on any one-word main clauses: in constructions with *čto*, the main clause can contain one word. The same holds for the restriction on short embedded clauses with omitted subjects: in principle, one-word embedded clauses are possible. It seems that the length becomes relevant when a particular structure is marked with respect to others: for instance, one-word restrictions mainly act in constructions without an explicit predicate (with a zero copula), and the length of the embedded clause is important when the subject of the main clause is omitted. In other words, the length of the constituent becomes a relevant parameter in borderline grammatical contexts when a constituent is in a sense non-canonical with respect to the 'core' exemplars of the construction.

## References

- Bošković 2008 – Bošković Ž. What will you have, DP or NP? Proceedings of NELS 37. Elfner E., Walkow M. (eds.). Amherst, MA: University of Massachusetts, 2008. Pp. 101–114.
- Davies, William, and Stanley Dubinsky. 2009. On the existence (and distribution) of sentential subjects. In *Hypothesis A/hypothesis B: Linguistic explorations in honor of David M.*
- Letuchij, Aleksandr B. 2018. Glagoly s fiksirovannym porjadjkom dopolnenij v rusckom jazyke i svojstva sentencial'nyx aktantov [Verbs with fixed order of complements in Russian and the properties of complement clauses. *Russkij jazyk v naučnom osveščanii* 2 (36). 180-198.
- Lyutikova, Pereltsvaig 2015 – Lyutikova E.A., Pereltsvaig A.M. Struktura imennoj gruppy v bezrtiklevykh jazykah: universal'nost' i variativnost' [Structure of noun phrase in language without an article: universals and variance]. *Voprosy jazykoznanija*. 2015. Vol. 3. Pp. 216–244.
- Lyutikova 2017 – Lyutikova E.A. Struktura imennoj gruppy v bezrtiklevom jazyke [Structure of noun phrase in a language without an article]. Moscow: Jazyki slavjanskoj kul'tury, 2017.
- Perlmutter, ed. Donna B. Gerds, John C. Moore, and Maria Polinsky. Cambridge, MA: MIT Press, 2009. Pp. 111–128.
- Hawkins, John A. A Parsing Theory of Word Order Universals // *Linguistic Inquiry*. 21 (199). 1990. P. 223–261.
- Pekelis, Olga E. 2014. Infinitiv vs. pridatočnoe s sojuzoom *čtoby*: k voprosu o vybore sposoba oformlenija sentencial'nogo aktanta v rusckom jazyke [Infinitive vs. complement clause with *čtoby*: the choice of the type of complement clause in Russian]. *Voprosy jazykoznanija* 4. 13-45.
- Ross, John R. (1967). Constraints on variables in syntax. (Doctoral dissertation, Massachusetts Institute of Technology).
- Say S.S. On the nature of dative arguments in Russian constructions with “predicatives”. *Current Studies in Slavic Linguistics* [Studies in Language Companion Series, 146]. // Kor Chahine I. (ed.). Amsterdam: John Benjamins, 2013. Pp. 225–245.
- Schmidtke-Bode, Karsten & Holger Diessel. 2017. Cross-linguistic patterns in the structure, function, and position of (object) complement clauses. *Linguistics* 55(1): 1–38.
- Ross, H.J.
- Tsimerling, Anton V. Predikativy i kačestvennye narečija: klassy slov i napravlenija derivacii. *Rusistika na poroge XXI veka: problemy i perspektivy*. Moscow: Jazyki slavjanskoj kul'tury, 2003. Pp. 54–59.

Alexander B. Letuchiy

National Research University Higher School of Economics (Moscow, Russia). School of  
Linguistics. Associate Professor;

E-mail: [alexander.letuchiy@gmail.com](mailto:alexander.letuchiy@gmail.com)

**Any opinions or claims contained in this Working Paper do not necessarily  
reflect the views of HSE.**

© Letuchiy, 2019