Frequency of the V-N collocation as an explanatory factor in the selection for *čto*-clause complements of nouns in Russian: an experimental study

Mikhail Knyazev

Abstract The aim of this paper is to assess the previously suggested hypothesis that *čto*-clause complements of nouns are confined to restructuring contexts. The hypothesis was supported in a previous experimental study that found interaction between context type and clause type. This paper addresses the question whether this interaction might be caused by the frequency of the context relative to a given N.

Introduction It has been observed that infinitival complements of nouns in Russian show varying syntactic behavior depending on the syntactic environment in which the noun appears (Franks & Hornstein 1992, Lyutikova 2010). In particular, only those environments that license restructuring by virtue of being collocations and lacking overt functional structure in the DP, e.g., *vyražit' želanie* 'express desire', as opposed to *vyražit' svoe želanie* 'express one's own desire', allow wh-extraction and NOM-marking of secondary predicates (cf. also Davies & Dubinsky's (2003) discussion of *the make the claim* construction). Recently I suggested that *čto*-clause complements of nouns show similar sensitivity with respect to the selection of "bare" *čto*-clause complements vs. complements with an overt DP-shell (i.e. *to,čto*-clauses, see Hartman 2013). Thus the presence of the possessive in (1) blocks restructuring, leading to the preference for a *to,čto*-clause. Similarly, constructions like 'bring proof' in (2a), as opposed to 'ignore proof' in (2b) license restructuring, yielding preference for a *čto*-clause.

(1) On vyrazil **svoe** somnenie $^{??}$(v tom), [čto vloženija okupjatsja].
he expressed his$_{REFL}$ doubt in that$_{LOC}$ that investments pay off
'He had doubts that the money will pay off.'

(2) a. Èkspert **privel** **dokazatel'stva** (togo), [čto kartina poddelnaja].
expert brought proofs that$_{GEN}$ that painting fake
'The expert presented the proofs that the painting is fake.'

b. Oni **ignorirujut** **dokazatel'stva** $^{??}$(togo), [čto kartina poddelnaja].
they ignore proofs that$_{GEN}$ that painting fake
'They ignore the proof that presented that the painting is fake.'

Previous study This claim was supported in an exploratory scaled (1–5) acceptability judgment study (n = 282), which showed a general dispreference for *to,čto*-clauses in non-restructuring contexts. Specifically, 12 experimental sentences were tested in a 2 × 2 factorial design ((NON-)RESTRUCTURING × CLAUSE TYPE). A significant interaction between the presence of collocation and the clause type was found, i.e. a stronger decrease of acceptability of *čto*-clauses (wrt *to,čto*-clauses) in non-restructuring contexts than in restructuring ones. Interestingly, the analysis revealed that the interaction was robust only for the overall manipulation of the syntactic environment including the higher verb, as in (2), but not for the manipulation involving only functional structure, as in (1). This result is consistent with the hypothesis that the crucial factor distinguishing examples like (2a) from (2b) is the presence of the restructuring environment reflected in the possibility to form of a complex predicate coupled with the assumption that overt functional structure does **not** block restructuring in Russian and assuming that the interaction is caused by the ungrammaticality of *čto*-clauses in non-restructuring contexts (see Sprouse et al. 2016 for the reasoning).

Limitations of the previous study The overall manipulation of the syntactic environment in that study involved potentially multiple variables resulting in the interaction. It is plausible that the interaction is caused not by the restructuring nature of the context but by the relative infrequency of constructions like 'ignore proof' among constructions with 'proof' followed by a clausal complement. For example, it has been observed that speakers rate sentences with *il*-drop in colloquial French higher when they contain a high frequency verb compared to a mid/low frequency verb (Culbertson & Legendre 2014). Consequently, we may wonder whether "*to*-drop" in Russian clausal complements is more acceptable in frequent V-N combinations (for a given N). To test this hypothesis an exploratory acceptability study was conducted.

Experiment The aim of the study was to test whether speakers' preference for *čto*-clauses will be affected by the frequency of the collocation with which a particular noun appears with a complement clause. In a forced choice task participants (n = 157) were presented with 12 experimental sentences involving N-V collocations in two conditions corresponding to the choice of the verb. Verbs across conditions were matched in semantic content but differed by the frequency of their appearance with the complement-taking noun in the corpus. The materials were constructed using three nouns *nadežda* 'hope', *uverennost'* 'conviction' and *dokazatel'stvo* 'proof' used in the previous experiment. Frequencies were collected from the Russian National Corpus (texts starting from 1950) and included only examples with a V immediately followed by an N. Only Vs that appeared at least twice with either of the clause types were counted; counts

for the two clause types were added. Pairs of V-N collocations with similar meaning but different frequencies were chosen, as shown in Table 1. The classification procedure involved (informally) maximizing the difference in frequency ranks between the collocations while keeping the semantic difference minimal. These collocations (supplemented with 12 fillers with similar structures) were assigned to two groups of participants in a Latin Square design and were presented in pairs with a *to-* and a *čto*-clause. The participants were asked to choose the sentence that sounded more natural.

Table 1

	FREQUENT	Frequency	INFREQUENT	Frequency
1	*est' nadežda* '(there) is hope'	N=230 (out of 766), frequency 30%, rank 1 (27)	*ostaetsja nadežda* 'remains hope'	15 (766), 2%, 9 (27)
2	*pojavilas' nadežda* 'appeared hope'	44 (766), 5.74%, 3 (27)	*zarodilas' nadežda* 'was born hope'	3 (766), 0.4%, 20 (27)
3	*vyrazit nadeždu* 'express conviction'	201 (766), 27%, 2 (27)	*vyskazat' nadeždu* 'express conviction'	15 (766), 2%, 9 (27)
4	*davat' nadeždu* 'give hope'	28 (766), 4%, 6 (27)	*darit' nadeždu* 'give hope as a present'	3 (766), 0.4%, 20
5	*poter'at' nadeždu* 'loose hope'	44 (766), 6%, 3 (27)	*poxoronit' nadeždu* 'bury hope'	2 (766), 0.3%, 27 (27)
6	*est' uverennost'* '(there) is conviction'	113 (389), 29%, 2 (23)	*soxranjetsja uverennost'* 'remains conviction'	4, 1%, 15 (23)
7	*pojavilas' uverennost'* 'appeared conviction'	18 (389), 5%, 3 (23)	*voznikla uverennost'* 'emerged conviction'	4, 1%, 15 (23)
8	*vyrazit uverennost'* 'express conviction'	116 (389), 30%, 1 (23)	*vyskazat' uverennost'* 'express conviction'	18 (389), 5%, 3 (23)
9	*vselit' uverennost'* 'instill conviction'	15 (389), 4%, 6 (23)	*pridavat' uverennost'* 'add conviction'	5 (389), 1%, 12 (23)
10	*est' dokazatel'stva* '(there) is proof'	23 (88), 26%, 1 (14)	*imejutsja dokazatel'stva* '(there) exist proof'	3 (88), 3%, 8 (14)
11	*najti dokazatel'stva* 'find proof'	7 (88), 8%, 4 (14)	*obnaružit' dokazatel'stva* 'discover proof'	3 (88), 3%, 8 (14)
12	*predstavit' dokazatel'stva* 'present proof'	7 (88), 8%, 4 (14)	*predostavit' dokazat-va* 'produce proof'	3 (88), 3%, 8 (14)

Results & Discussion The results of the experiment were analyzed using logistic regression with a random intercept for subject and a random slope for item (the maximal model that converged). The analysis revealed a significant effect of frequency, i.e. less frequent V-N combinations are less likely to select *čto*-clauses, as shown in Figure 1 (the odds of a less frequent V-N to select a *čto*-clause are 0.70 the odds of a more frequent V-N, p = 0.001). However, a simple logistic regression performed for each item separately (to measure item variability) showed that the effect of frequency on clause type was significant only for Items 1 and 9 in Table 1 (with

Figure 1

the odds ratios for less frequent V-N being 0.5, p=0.03 and 0.38, p=0.004, respectively), see Table 2 for the frequencies of *čto*-clauses. These results show that the frequency of the collocation does not have a consistent effect on the preference for a *čto*-clause and that the effect might be confined to extremely frequent or infrequent collocations, such as Items 9 and 1, respectively, lending support to the idea that frequency per se cannot explain the interaction found in the earlier study and suggesting that a grammatical explanation might be correct, although more sophisticated frequency metrics (see, e.g., Divjak 2017) should be attempted in future work.
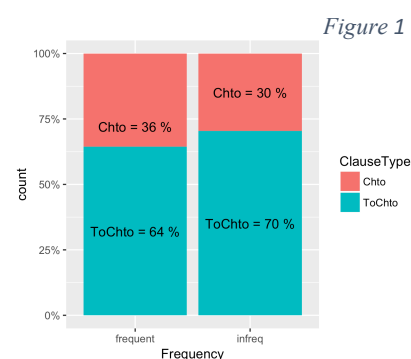
Table 2

Item #		1	2	3	4	5	6	7	8	9	10	11	12
% of *čto*-clauses	FRE-QUENT	0.58	0.54	0.42	0.21	0.25	0.44	0.41	0.31	0.51	0.22	0.16	0.22
	INFRE-QUENT	0.41	0.47	0.33	0.30	0.14	0.41	0.42	0.22	0.28	0.24	0.19	0.13