

**NOMINALIZED CLAUSES AND DISCOURSE-GIVENNESS:
EXPERIMENTAL EVIDENCE FROM RUSSIAN***

Mikhail Knyazev^{1,2} & Ekaterina Rudaleva²

¹*Institute for Linguistics Studies, Russian Academy of Sciences*

²*HSE University, St. Petersburg*

Abstract. In a seminal paper, Kiparsky & Kiparsky (1970) proposed a two-way correspondence between presuppositionality of clauses and nominal structure. The proposal remains highly relevant to current research (a.o. Kastner 2015, Bochnak & Hanink 2022), despite the existence of counterexamples in both directions. In this paper, we examine Russian nominalized clauses to show that presuppositionality is indeed neither necessary nor sufficient for nominalization. However, instead of completely discarding the correspondence between presuppositionality and nominalization, we argue for a weaker “preference” hypothesis, whereby presuppositional, or discourse-given, contexts are associated with a higher likelihood of nominalization compared to discourse-new contexts. We provide support for the preference hypothesis based on four experimental studies, a forced-choice and a givenness-rating study using matrix negation as a proxy for givenness and a forced-choice and a sentence completion study directly manipulating the discourse context. We suggest a tentative explanation for the preference hypothesis in terms of definiteness/familiarity marking.

1. Introduction

Nominalized clauses (NCs) have recently become an important topic in cross-linguistic research in the syntax-semantics interface (Kastner 2015, Bogal-Allbritten & Moulton 2018, Moulton 2020, Bondarenko 2022, Bochnak & Hanink 2022). The relevance of this topic stems from the fact that NCs can help clarify long-standing questions about how major semantic/pragmatic distinctions in the domain of clausal complementation are syntactically encoded, as well as the nature of this encoding.

In a seminal paper, Kiparsky & Kiparsky (1970) proposed a distinction between factive and nonfactive clauses (depending on whether the truth of the clause is presupposed), arguing that the former have an additional nominal structure (associated with a silent noun FACT).¹ Despite being highly influential, Kiparsky’s proposal has been criticized in many subsequent works (de Cuba 2007, de Cuba & Ürögdi 2010, Haegeman & Ürögdi 2010, Sheehan & Hinzen 2011).

*We wish to thank the audiences and reviewers at the 17th Conference on Typology and Grammar for Young Scholars (Saint Petersburg, Russia) and FASL31 (McMaster/online), where parts of this work were presented. We also thank the three anonymous reviewers for *Glossa*, to which the paper was originally submitted, as well as the two anonymous reviewers for *Studia Linguistica* for their thoughtful and insightful comments which greatly improved the paper. As usual, all errors are our own.

¹In Kiparsky & Kiparsky’s (1970) original proposal, nonfactive clauses were also nominal due to an NP-node above S. Subsequent work dropped this assumption as not essential to their account (cf. de Cuba 2007).

First of all, the relevance of factivity (as a lexical semantic property) for the syntactic behavior of clauses was questioned. For example, it was noted that nonfactive clauses that express information presupposed to be familiar to the addressee, such as complements of “response-stance” verbs like *agree* and *deny*, pattern with factive clauses with respect to wh-extraction, main clause phenomena, etc. (Cattell 1978; Hegarty 1992, cited in de Cuba 2007). This led to a reformulation of the factive/nonfactive distinction as one based on information structure, namely between “familiar” complements, including factive and response-stance complements, vs. “novel” complements of “volunteer-stance” predicates like *say* and *think*. It was also suggested that complements of nonfactive verbs like *believe* pattern with factive complements when they are contextually given (Kallulli 2006, 2010), suggesting that familiarity can depend on the discourse context. Still other approaches argued for a more abstract distinction in terms of “referentiality”, taken to be a precondition on familiarity/givenness (du Cuba 2007, de Cuba & Ürögdi 2010, Haegeman & Ürögdi 2010).

The above criticisms lead to a further issue: if the special syntactic properties of familiar/given complements no longer correspond to factivity as such, part of the motivation for their nominal character, which in Kiparskys’ proposal was tied to a silent noun FACT, is lost. Accordingly, a number of researchers tried to explain these properties in other ways, e.g. by an operator in Spec,CP (Melvold 1991, Hegarty 1992, Haegeman & Ürögdi 2010), truncated left-periphery (Haegeman 2006, de Cuba 2007) or a definiteness feature on C (Sheehan & Hinzen 2011).

The dissociation between factivity and a nominal status in these approaches was further motivated by independent distributional arguments (Haegeman & Ürögdi 2010, Sheehan & Hinzen 2011). First, factive complements (see below on sentential subjects) in general do not show the distribution of DPs but instead pattern like nonfactive complements. For example, both are normally disallowed as objects of P and, conversely, allowed in positions unavailable to DPs, e.g. after adjectival/passive predicates like *surprised*. Second, on Kiparskys’ view factive clauses are incorrectly predicted to be *strong* islands, on a par with complex NPs, whereas they are in fact *weak* islands (e.g. Hegarty 1992).²

Kastner (2015) presented an updated version of Kiparskys’ proposal, which dealt with some of the problems above. He proposed a distinction between “presuppositional” vs. “nonpresuppositional” clauses, with the former category subsuming factive and response-stance complements (as in Cattell 1978 and Hegarty 1992), and also sentential subjects, which typically refer to familiar/given information. Kastner’s main innovation was to recast the presuppositional vs. nonpresuppositional distinction in terms of Heim’s (1982) familiarity theory of definiteness. Specifically, he identified presuppositional clauses with definite descriptions, viewed in Heim’s theory as referring to familiar discourse referents (corresponding to existing filecards). By contrast, nonpresuppositional clauses (of verbs like *think*) were taken to introduce new discourse referents (corresponding to new filecards). This allowed Kastner to derive the nominal character of presuppositional clauses from the presence of a DP-layer headed by a definite (familiar) D, without the need for a silent FACT.³

²This problem was identified in earlier work (e.g. Zubizarreta 1982, Adams 1985), leading to a modification of Kiparsky’ proposal in terms of nominal features [+N] on Comp.

³Kastner (2015) still assumes that *overtly* nominalized clauses, such as *ze še*-clauses in Hebrew and *to oti* clauses in Greek (Roussou 1994, see also Moulton 2020) are complex NPs and hence

The syntactic implementation of Kastner’s proposal was based on an independent line of work which argued that sentential subjects (and potentially certain object CPs) are underlyingly DPs, and moreover are headed by a possibly silent D directly taking the CP as its complement (Davies & Dubinsky 1999, Han 2005, Takahashi 2010, Halpert & Schueler 2013). An advantage of this analysis is that it allows to avoid the problematic prediction that presuppositional clauses are strong islands (since they no longer need to be complex NPs). At the same time, it can still derive their weak island status from the presence of a definite D (Honcoop 1998, Kastner 2015).

In general terms, Kastner’s proposal can be formulated as in (1) and is referred to below as the *correspondence hypothesis*.

- (1) a. Presuppositional (familiar) clauses are DPs headed by a (possibly silent) definite/anaphoric D.
- b. Clausal DP arguments are interpreted as presuppositional (familiar).

Kastner’s proposal has been rather influential, but it has also been questioned both for English and as a cross-linguistic universal (Moulton 2015, Djärv 2019, 2021, Moulton, Bogal-Allbritten & Shimoyama 2020, Jarvis 2021). For example, it fails to provide a nonstipulatory solution to the problem of the distributional differences between presuppositional clauses and overt DPs mentioned above.⁴ Furthermore, it seems to be falsified by the fact that cognitive factive and response predicates in principle allow CP proforms like *so* (Bhatt 2010, Moulton 2015, Djärv 2019), contrary to Kastner 2015. In addition, weak island data, used by Kastner to support the presence of a definite D in presuppositional clauses, can also be potentially explained without invoking a DP-layer (Haegeman & Ürögdi 2010, Sheehan & Hinzen 2011, see also Jarvis 2021 and Djärv 2021).

The above problems may suggest that the correspondence hypothesis should be rejected in favor of alternative ways of encoding presuppositionality, as cited above or perhaps in favor of not encoding it at all. However, we think this conclusion is premature in view of the fact that the correspondence hypothesis receives cross-linguistic support from languages with overt “DP-shells”, such as Greek (Roussou 1994) and Hebrew, discussed by Kastner (2015), but also other languages with NCs such as Korean (Bogal-Allbritten & Moulton 2018, Moulton et al. 2020), Buryat (Bondarenko 2020) and Washo (Bochnak & Hanink 2022). In all these languages, the distribution of NCs broadly aligns with presuppositionality, suggesting that the correspondence hypothesis may be at least partly true.

The goal of this paper is to discuss experimental data from another language with NCs/DP-shells, namely Russian, in order to argue that presuppositionality plays a role in clausal nominalization (referred to below simply as “nominalization”) but in a much weaker form than in the correspondence hypothesis.

contain a silent N, based on their strong islandhood. But see Bondarenko 2022:328 for an alternative explanation of strong island data in terms of Anti-locality, which does not require overt NCs to be complex NPs (see Section 2.1).

⁴Kastner (2015) derives the difference between presuppositional clauses and overt DPs by taking the D that selects CP to be a specific determiner Δ such that verbs like *agree* can be taken to (lexically) select Δ , without selecting DPs in general. However, this lacks independent motivation. See Section 2.2 for further discussion.

We agree with Kastner 2015 that there is indeed an association between presuppositionality and nominalization, resulting from the fact that presuppositional clauses must be (or at least tend to be) syntactically marked. But unlike Kastner we do not assume that presuppositionality implies nominalization, and this is because crucially there are *other* ways of marking presuppositionality both across and within languages, such as deaccenting (Kallulli 2006, 2010, Rochemont 2016).⁵ We will not discuss such strategies in detail, what is important is that they render nominalization of presuppositional clauses generally optional (cf. Bogal-Allbritten & Moulton 2018 and Moulton et al. 2020).

Conversely, nominalization also does not imply presuppositionality. The reason for this is that at least in some languages, such as Russian, NCs can introduce new information (see Section 2.3), suggesting that familiarity need not be lexically encoded (“hardwired”) but can instead have the status of a defeasible inference. This view goes against the most straightforward interpretation of Kastner 2015, whereby NCs are introduced by a definite D with a built-in familiarity presupposition. However, as was suggested to us by a reviewer, this view is in principle compatible with a version of Kastner’s (2015) analysis in which the D in NCs has a weaker *definiteness* presupposition (e.g. uniqueness/identifiability in the sense of referring to a uniquely identifiable fact or possibility, cf. Hawkins 1978 and Hankamer & Mikkelsen 2022), which often, but crucially not always, implicates familiarity (cf. Abbott 2019).

Thus, although the correlation between presuppositionality and nominalization is only indirect, it still predicts that familiar (discourse-given) clauses should be associated with a higher preference for nominalization compared to discourse-new clauses. This is formulated in (2) as the *preference hypothesis*, which we present as an alternative to the correspondence hypothesis in (1) (see below on terminology).

(2) Nominalized clauses are more likely to occur in discourse-given contexts compared to discourse-new contexts.

The association between presuppositionality and nominalization, as captured by the preference hypothesis, should arise for the following reason. Although NCs will sometimes occur in discourse-new contexts for independent reasons, the frequency of NCs in discourse-given contexts should be higher because in such contexts NCs will in addition be used to mark presuppositionality at least some of the time, assuming all other factors are kept constant. Importantly, this does not imply that there will be a preference for NCs *relative to nonnominalized clauses*, because the preference hypothesis is consistent with the possibility that nominalization is not the most preferred way of marking presuppositionality (cf. footnote 5).

While the preference hypothesis might suggest that nominalization is *always* optional for presuppositional clauses, this need not be so. We allow for the possibility that in *certain* presuppositional environments nominalization can be *grammaticalized* (cf. Hawkins 2014), e.g. because other ways of marking presuppositionality are

⁵In fact, it is possible that deaccenting is (by default) obligatory or strongly preferred for presuppositional clauses in Russian (and possibly universally). On this view, nominalization should be viewed as predominantly an optional “reinforcing” marker of presuppositionality (which may, though, become grammaticalized, see below). However, since we do not systematically look at prosody, we leave the investigation of this possibility to future research. (See also Section 5.)

unavailable or dispreferred. A possible example are sentential subjects, which tend to be nominalized across languages, including Russian (see Section 2).

Before moving on, let us say a few words about terminology. We chose to formulate the preference hypothesis in terms of the notion “(discourse-)givenness” rather than “presuppositionality” or “familiarity”, which are more closely associated with Kastner 2015, in order to highlight that our experiments focus on *contextual* manipulation. More specifically, we use “(discourse-)given” roughly in the sense of Schwarzschild (1999) and Rochemont (2016) to refer to information that is explicitly mentioned in the previous discourse or entailed by such information, which is similar to Prince’s (1992) “discourse-old”. By contrast, we use (interchangeably) “familiar” (Hegarty 1992, Gundel, Hegarty & Borthen 2003) and “presuppositional” (Kastner 2015) in a wider sense, to refer to information that the hearer is presumed to know, whether due to previous mention (givenness proper) or as part of background knowledge/Common Ground (Stalnaker 1974). This corresponds to Prince’s (1992) “hearer-old”.

We tested the preference hypothesis by looking at predicates that alternate between bare and nominalized CPs and that are in principle compatible with both discourse-new and discourse-given contexts. The prediction is that such predicates should show a higher preference for nominalized CPs in discourse-given compared to discourse-new contexts. We conducted four experiments to test this prediction. Experiment 1b used matrix negation as a proxy for discourse-givenness (cf. Caplan & Djärv 2019), with Experiment 1a testing an auxiliary hypothesis regarding the association between negation and givenness, whereas Experiments 2a and 2b (using slightly different tasks) directly manipulated the discourse context. The results largely confirm the preference hypothesis, showing that discourse-givenness is indeed a factor in nominalization. At the same time, many questions remain, in particular, how to explain the preference hypothesis from a theoretical standpoint and how to account for cross-linguistic variation. Although we do not have definitive answers at this point, we offer some directions for future research (see Section 5).

The paper is structured as follows. Section 2 discusses the distributional properties of NCs in Russian in light of the correspondence hypothesis and also offers some motivation for the preference hypothesis. Section 3 presents Experiments 1a and 1b involving matrix negation. Section 4 presents Experiments 2a and 2b involving manipulation of discourse contexts. Section 5 provides a general discussion of the results from a more theoretical perspective.

2. The distribution of NCs in Russian

2.1. Some general remarks

In Russian, argument clauses can be nominalized by the case-marked form of the distal demonstrative *to* ‘that’, as shown in (3).⁶

⁶Subjunctive and interrogative CPs can also be nominalized. In this paper, we only look at declarative indicative clauses.

- (3) [To [čto on sbežal]] dokazyvaet / podverždaet ego vinu.⁷
 that.NOM COMP he ran.away proves confirms his guilt
 ‘(The fact) that he ran away proves/confirms his guilt.’

Following previous literature (Hartman 2012, Knyazev 2016, 2022, Bondarenko 2022), we analyze *to* as an overt realization of a definite D. We assume, after Knyazev 2016, 2022 and Bondarenko 2022 (see also Tatevosov & Lyutikova 2019) and contra Kastner 2015 (see footnote 3) that *to* directly takes CP as its complement (see also Hartman 2012).⁸ Thus, Russian NCs are structurally similar to presuppositional *that*-clauses in English on Kastner’s analysis. We also provisionally assume that *to* in NCs is a familiar/anaphoric definite, as in Kastner 2015 and also Kobozeva 2013, which fits with the fact that *to* otherwise functions as a demonstrative (we slightly revise this assumption below).⁹

Distributionally, Russian NCs initially seem to conform to the correspondence hypothesis (Khomitsevich 2007, Hartman 2012, Knyazev 2016, 2022). First, sentential subjects (which are presuppositional) are strongly preferred to be nominalized, as in (4a) (cf. 3).¹⁰ Conversely, (accusative) complements of prototypical *nonpresuppositional* verbs like *utverždat* ‘claim’ and *predpolagat* ‘suppose’ resist nominalization (in standard Russian), as in (4b) (unless they bear contrastive stress, as in (4c), in which case nominalization is obligatory).

- (4) a. ?? [Čto on sbežal] dokazyvaet / podverždaet ego vinu.
 COMP he ran.away proves confirms his guilt
 ‘(The fact) that he ran away proves/confirms his guilt.’
- b. Petja utverždaet / predpolagaet [(??)to] [čto èto maloverojatno]].
 Petya claims supposes that.ACC COMP this unlikely
 ‘Petya claims/supposes that this is unlikely.’
- c. Petja utverždaet (tol’ko) [TO / *Ø] [čto èto maloverojatno]].
 Petya claims only that.ACC COMP this unlikely
 ‘Petya only claims that this is unlikely.’

⁷We use the following abbreviations in the glosses: ACC = accusative case, COMP = complementizer, DAT = dative case, INF = infinitive, INS = instrumental case, NEG = negation, PREP = prepositional case, PRT = particle, SUBJ = subjunctive.

⁸For example, Bondarenko (2022:254–258) observes that Russian NCs cannot be anteceded by a gendered pronoun and cannot trigger plural agreement when coordinated, as Russian nouns usually do. Knyazev (2022) further notes that other determiner-like elements, e.g. *èto* ‘this’, cannot introduce NCs, suggesting that there must be a selectional relation between *to* and the CP (see also Tatevosov & Lyutikova 2019 and Hankamer & Mikkelsen 2022).

⁹Note that distal demonstratives in principle allow nonfamiliar (uniqueness) uses when combined with relative CPs (e.g. Jenks 2018, see also Tatevosov & Lyutikova 2019 for Russian), which may be related to the fact that NCs are also felicitous in discourse-new contexts (see Section 2.3). Bondarenko (2022:253) also observes that NCs allow indefinite interpretations in certain specific contexts.

¹⁰Sentential topics, which are also presuppositional, show a similar but weaker preference, as reported by Knyazev (2022).

However, there are two major problems with the correspondence hypothesis. First, complements of presuppositional predicates can but generally need *not* be nominalized (although specific verbs may require NCs), as in (5).

- (5) a. Ona {gorditsja (tem) / raduetsja (tomu)} što ona pobedila.
 she is.proud that.INS is.glad that.DAT COMP she won
 ‘She is proud/glad that she won.’
- b. Vasja {žaleet (o tom) / soglasilsja (s tem)} što on
 Vasya regrets about that.PREP agreed with that.INS COMP he
 sdelal ošibku.
 made mistake
 ‘Vasya regrets/agreed that he made a mistake.’
- c. Petja osoznaet / otricaet [(to) što on proigral].
 Petya realizes denies that.ACC COMP he lost
 ‘Petya realizes/denies that he has lost.’

Second, nonpresuppositional predicates generally *can* be nominalized as long as the NC is in the oblique/object-of-P position, as in (6a). Note also that overt Ps obligatorily require NCs independently of presuppositionality of the predicate, as in (6b).

- (6) a. On {nadeetsja (na to) / uveren (v tom)} što on prav.
 he hopes on that.ACC certain in that.PREP COMP he right
 ‘He hopes/is certain that he is right.’
- b. On nadeetsja na [to/ *Ø [što on prav]].
 he hopes on that.ACC COMP he right
 ‘He hopes that he is right.’

We discuss these problems in turn, showing that both undermine the strong version of the correspondence hypothesis (a two-way implication) but at the same time are compatible with the preference hypothesis.

2.2. Presuppositional complements without nominalization

To handle the first problem, some accounts (e.g. Knyazev 2022) argue that presuppositional complements are introduced by a null D and also a null P when they can occur in PP/non-DP positions with predicates like *soglasit'sja* ‘agree’ (cf. 5b).^{11,12}

However, an alternative account can also derive these facts if one drops the strict version of the correspondence hypothesis, allowing presuppositional clauses to be

¹¹Kastner (2015) does not postulate null P but tries to derive the fact that presuppositional clauses can appear in PP positions by assuming that PP-taking verbs may select Δ but not overt DPs (see footnote 4).

¹²To capture the fact that, in contrast to presuppositional complements, in sentential subjects (cf. 3), D must be *overt*, one may appeal to specific rules of Vocabulary Insertion (Bondarenko 2022:281) or independent principles of spanning lexicalization (Knyazev 2022).

optionally nominalized. On this view, bare presuppositional complements in PP/oblique positions can simply be analyzed as CPs, with the corresponding predicates (disjunctively) selecting for PP/oblique DP or CP (cf. Bondarenko 2022:338–340, see also footnote 13). In turn, to derive the obligatoriness of *to* in sentential subjects (cf. 3), one can adopt a *specific* (grammaticalized) constraint, as was suggested in the Introduction. An account without a null D/P is a priori preferable since, as far as we know, there is no convincing evidence for a null D (or for null P, for that matter) in presuppositional complements (at least for ordinary factive and response-stance predicates).¹³

Note that the above view that nominalization of presuppositional clauses is in general *optional* does not imply that the choice between bare and nominalized CPs is completely free. This is because presuppositionality may still render nominalization more likely. Specifically, if the preference hypothesis is true we expect complements of presuppositional predicates to show on average a higher preference for NCs compared to complements of nonpresuppositional predicates since the former should be more likely to occur in discourse-given contexts. This expectation is borne out.

Table 1 provides rough estimates of the relative frequencies of NCs for a number of oblique/PP-taking predicates, based on data from the Russian National Corpus (RNC, ruscorpora.ru).¹⁴

¹³Bondarenko (2022) argues that apparent bare CP complements of (accusative-taking) verbs like *ob'asnit'* 'explain', *utočnit'* 'clarify' are introduced by a null D when they have the presuppositional/theme reading ('what is explained/clarified') but are bare CPs when they have the nonpresuppositional/content-of-utterance reading ('what is said as an explanation/clarification of some fact'). Her main argument for the null D analysis comes from the correlation between the theme reading and strong island effects. However, she still analyzes bare CP complements of PP/oblique-taking factive (and response) predicates like *gordit'sja* 'be proud' / *žalet'* 'regret' (cf. 5a) as simply CPs, based on her assumptions about case realization (see footnote 30) (for supporting arguments, see Bondarenko 2022:338–340). Independent evidence for the CP analysis comes from a rating study conducted by the first author of this paper, which tested the acceptability of argument extraction from complements with and without *to* with presuppositional predicates like *gordit'sja* 'be proud' vs. nonpresuppositional predicates like *nadejat'sja* 'hope' (cf. 6). The results showed a strong overall effect of *to* but crucially no contrast between the two types of predicates, supporting the analysis without a null D for *gordit'sja* 'be proud' class predicates. Note that for Bondarenko the DP requirement for clausal complements is not directly tied to presuppositionality but has a purely semantic basis reflecting the status of a clause as *composing as an argument* (as opposed to modifier), which allows some presuppositional complements to be analyzed as CPs as long as they compose as semantic modifiers. (Space precludes a more detailed discussion of Bondarenko's (2022) account.)

¹⁴We used texts from the main subcorpus written after 1950 (date of access: May 2022). The queries involved the verb lemma followed by *čto* or by (P) + *to* + *čto*. We did not filter out sentences where *čto* functions as a wh- or a relative pronoun as that would be too time-consuming. The latter kind of sentences, however, are rather rare.

Table 1. Proportions of bare and nominalized CPs in RNC texts written after 1950

| verb | verb class | N <i>čto</i> | N <i>to čto</i> | % <i>to čto</i> |
|----------------------------------|---------------------|--------------|-----------------|-----------------|
| <i>gordit'sja</i> ‘be proud’ | presuppositional | 363 | 538 | 0.6 |
| <i>nastaivat</i> ‘insist’ | nonpresuppositional | 324 | 430 | 0.57 |
| <i>setovat</i> ‘lament’ | presuppositional | 194 | 125 | 0.39 |
| <i>udivlën</i> ‘surprised’ | presuppositional | 166 | 60 | 0.27 |
| <i>soglasit'sja</i> ‘agree’ | presuppositional | 922 | 335 | 0.27 |
| <i>dovolën</i> ‘content’ | presuppositional | 351 | 111 | 0.24 |
| <i>namekat</i> ‘hint’ | nonpresuppositional | 434 | 131 | 0.23 |
| <i>radovat'sja</i> ‘rejoice’ | presuppositional | 1103 | 305 | 0.22 |
| <i>rasčityvat</i> ‘count on’ | nonpresuppositional | 1027 | 277 | 0.21 |
| <i>somnevat'sa</i> ‘doubt’ | presuppositional | 2299 | 457 | 0.17 |
| <i>sožalet</i> ‘regret’ | presuppositional | 447 | 91 | 0.17 |
| <i>udivit'sja</i> ‘be surprised’ | presuppositional | 538 | 111 | 0.17 |
| <i>bespokoit'sja</i> ‘worry’ | nonpresuppositional | 149 | 22 | 0.13 |
| <i>xvastat'sja</i> ‘boast’ | nonpresuppositional | 153 | 23 | 0.13 |
| <i>verit</i> ‘believe’ | nonpresuppositional | 5069 | 711 | 0.12 |
| <i>žalet</i> ‘regret’ | presuppositional | 1231 | 123 | 0.09 |
| <i>ubeždën</i> ‘convinced’ | nonpresuppositional | 2933 | 271 | 0.08 |
| <i>rad</i> ‘glad’ | presuppositional | 2240 | 110 | 0.05 |
| <i>uveren</i> ‘certain’ | nonpresuppositional | 9895 | 504 | 0.05 |
| <i>žalovat'sja</i> ‘complain’ | nonpresuppositional | 915 | 47 | 0.05 |
| <i>podozrevat</i> ‘suspect’ | nonpresuppositional | 2355 | 91 | 0.04 |
| <i>nadejat'sja</i> ‘hope’ | nonpresuppositional | 7521 | 306 | 0.04 |

The predicates are ordered from the most to the least biased towards NCs. As can be seen, presuppositional predicates (in gray) tend to have a stronger bias for NCs, in line with the preference hypothesis. Note that there are also presuppositional predicates (e.g. *rad* ‘glad’ and *žalet* ‘regret’) that have a weaker bias for NCs than certain nonpresuppositional predicates (e.g. *rasčityvat* ‘count on’, *namekat* ‘hint’ and *nastaivat* ‘insist’), suggesting that factors beyond presuppositionality are at play (see the next section). The overall tendency, however, is clear. A similar tendency was observed in elicited production study by Knyazev & Ustinova (2023).

To summarize, the distributional and usage behavior of presuppositional complements in Russian argue against a strict implication from presuppositionality to nominalization but at the same time are consistent with the preference hypothesis.

2.3. Nominalized complements without presuppositionality

The second problem—i.e. that nonpresuppositional complements can be realized as NCs (with PP/oblique-taking predicates like *nadejat'sja* ‘hope’)—presents a challenge also for the implication from nominalization to presuppositionality. There are two potential ways to handle this problem. One option is to assume that NCs in object-of-P/oblique positions are generally ignored for the purposes of the correspondence hypothesis (Knyazev 2022). The idea would be that because NCs are obligatory in object-of-P/oblique positions regardless of their discourse status (cf. 6b), it is plausible

that their ordinary semantic/pragmatic contribution is stripped away from them in such positions.

However, this is difficult to reconcile with the fact noted in the previous section that presuppositional PP/oblique-taking predicates have a stronger tendency to occur with NCs compared to respective nonpresuppositional predicates because if NCs in the context of P/oblique marking are semantically/pragmatically inert their occurrence in such contexts should not depend on presuppositionality, which leaves the observed correlation unexplained.

Another way to maintain the implication from nominalization to presuppositionality is to say that when PP/oblique-taking predicates like *nadejat'sja* 'hope' combine with NCs (cf. 6a) they give rise to a presuppositional (familiar) interpretation, provided that presuppositionality is not built into the meaning of the predicate (Bogal-Allbritten & Moulton 2018, Bochnak & Hanink 2022). Such a view is suggested by Kobozeva (2013), who argues that in Russian NCs mark familiar, or "given" information (in the sense of being mentioned before or being part of Common Ground). Kobozeva supports this view by the contrast in (7a)–(7b) below, involving the cognitive factive *znat* 'know' / *uznat* 'learn' (note that despite being classified as presuppositional in Kastner 2015 complements of cognitive factives are compatible with new information, cf. Simons 2007, Djärv 2019).

(7) a. (Discourse-given context: We can no longer conceal the secret of Petya's birth from him.)

On uznal \ (to), što ego nastojaščij otec — N.
 he learned that.ACC COMP his real father N.
 'He [Petya] LEARNED that his real father is N.'

b. (Discourse-new context: A: Why did Petya fall out with his father?)

B: On uznal / (*to), što ego nastojaščij otec — N \.
 he learned that.ACC COMP his real father N
 'He [Petya] learned that his real father is N.'

(adapted from Kobozeva 2013: 147)

When the complement is given, as in (7a), nominalization is possible, though not obligatory. (The sentence also has a pitch accent on the matrix verb, indicated by a backslash, followed by a falling intonation and deaccenting of the complement.) Crucially, when the complement is discourse-new, as in (7b), nominalization is disallowed. (The sentence in (7b) has a rising intonation on the matrix verb and the pitch accent on the final constituent of the complement).¹⁵ The contrast is taken by Kobozeva to suggest that familiarity is a *necessary* but not a sufficient condition for nominalization (cf. Bogal-Allbritten & Moulton 2018 for a similar claim).

While we agree with Kobozeva's observations for *accusative*-taking verbs (compatible with NCs) like *uznat* 'learn' in (7) (and also *osoznat* 'realize', *vspomnit* 'remember (PFV)', *dokazat* 'prove', etc.), we doubt that these observations generally

¹⁵The contrast in (7) resembles the contrast in (i), involving the cognitive factive *notice* from Hegarty 1992 (cited in Haegeman & Ürögdi 2010:143).

- (i) I was talking to our agents in Russia yesterday,
 a. and they noticed that Max went to Moscow last week. ✓ NEW
 b. and they noticed it that Max went to Moscow last week. * NEW ✓ GIVEN

hold for *oblique/PP*-taking predicates like *nadejat'sja* 'hope' and *uveren* 'certain', at least on the standard understanding of familiarity that Kobozeva herself uses. Although the latter kind of predicates indeed sound rather marked with NCs when expressing the "main point of utterance" (Simons 2007), as in (8), and thus may be viewed as incompatible with discourse-novelty in the narrow sense, they are still compatible with new information in the sense that they need not belong to familiar information, i.e. they can be neither previously mentioned nor part of the Common Ground.

- (8) (Context: A: What will Vasya do on the weekend?)
 B: Katja nadeetsja {(??na to) / uverena (??v tom)}
 Katya hopes on that.ACC certain in t hat.PREP
 što on ostanetsja doma.
 COMP he will.stay home
 'Katya hopes / is certain that he will stay at home.'

This is suggested by the fact that there are many naturally-occurring examples where NCs with PP/oblique-taking predicates express new information, as illustrated by (9a)–(9b), from RNC (note that both sentences occur in the beginning of news reports). Moreover, such examples can be followed by material that explicitly denies that the complement had been previously asserted, believed or entertained (e.g. 'and he is/I am the first to make this claim' or 'even though no one has discussed this possibility before').

- (9) a. (Context: The International Academy of Vocal Art is going to be opened in Ufa. It was granted a license on March 13. This was communicated by the leader of Bashkiria Radij Xabirov.)

On uveren v tom, što v akademii budut vospityvat'sja
 he certain in that.PREP COMP in academy will.be raised
 solisty mirovogo urovnja. (RNC)
 soloists world of.level
 'He [Xabirov] is certain that the academy will bring up soloists of world level.'

- b. (Context: The star of the new *Spider-Man: Homecoming* Tom Holland came to Moscow, Interfax reports.)

"Ja nadejus' na to, što naša kartina prevratitsja
 I hope on that.ACC COMP our picture will.turn
 v analog franšizy o Garri Pottere", zajavil aktër
 in analog of.franchise about Harry Potter declare actor
 žurnalistam. (RNC)
 to.journalists.

"I hope that our movie will be similar to the Harry Potter franchise", the actor told the journalists.'

In view of examples like (9), even a one-way implication from nominalization to familiarity appears too strong. Thus, we need to revise our initial characterization of NCs as involving a definite D that lexically encodes familiarity (cf. Kastner 2015 and Section 2.1) in favor of a more nuanced view where familiarity is a defeasible inference from definiteness that is triggered in the presence of certain contextual/prosodic cues of

givenness (or perhaps by default) but does not arise in contexts that contain the opposite cues, such as (9) (cf. Abbott 2019).

We still need to account for Kobozeva's observations, namely why the familiarity implication is obligatory in accusative-marked NCs with verbs like *uznat* 'learn' (cf. 7b). While we do not have a definitive solution to this problem, this may be related to the fact NCs are *generally* marked in accusative object contexts (cf. 4b and also Khomitsevich 2007:10), suggesting that in such contexts NCs are licensed only when they serve a pragmatic function, i.e. givenness-marking.

Finally, we need to address the question why NCs should occur in discourse-new contexts like (9) to begin with, where they are pragmatically inert. We wish to suggest that the use of NCs can be driven by factors *other than presuppositionality*, including, but not limited to, lexical biases, individual speaker biases, stylistic choices, ease of processing, etc. Crucially, we assume that these factors operate *independently* of the discourse status and thus should be observed to the same extent in discourse-given contexts. This assumption ensures that the preference for NCs across discourse-given vs. discourse-new contexts should not be affected by factors beyond presuppositionality, as required by the preference hypothesis.

To sum up, Russian data show that presuppositionality is neither sufficient nor necessary for nominalization. At the same time, the data are compatible with the view that presuppositionality plays a role in nominalization as described by the preference hypothesis in (2). In the next two sections, we present experimental evidence to support this claim.

3. Experiments 1a and 1b: negation, givenness and nominalization

3.1. Hypotheses

The goal of Experiments 1a and 1b was to test the preference hypothesis in (2) by using sentences with matrix negation as a proxy for presuppositionality/discourse-givenness. A similar strategy was used by Caplan & Djärv (2019) (see Givón 1979:Ch.3 for an earlier claim), who hypothesized that complements of nonpresuppositional predicates are perceived as more likely to be given ("talked about before") in sentences with matrix negation compared to affirmative sentences.¹⁶ Their hypothesis was motivated by the observation that nonpresuppositional predicates, or more precisely, predicates compatible with new information are infelicitous in negative sentences in out-of-the-blue contexts since negation renders their complements given (on a par with complements of response and emotive factive predicates in affirmative contexts), as in (10). Using a rating task, they found that speakers indeed rated complements of nonfactive speech act/belief predicates as more likely to be talked about before in negative sentences like (10b) compared to affirmative sentences like (10a), whereas they did not find such contrast for emotive factive and response predicates, as in (10c).

(10) [Uttered out of the blue:]
Guess what — / You know what —

¹⁶Givón (1979:103) suggests that "negative sentences must be more presuppositional than their corresponding affirmatives". Zaliznjak (1988:119) observes, on the basis of Russian data, that "in the past tense most belief verbs acquire the factive implication in the context of negation".

- a. John thinks/told me that [Bill and Anna broke up].
 - b. #John doesn't think/didn't tell me that [Bill and Anna broke up].
 - c. #John appreciates/doesn't appreciate that [Bill and Anna broke up].
- (adapted from Caplan & Djärv 2019:103)

Since we designed our Experiments 1a and 1b before we became aware of Caplan & Djärv 2019, we did not take into account that negation may interact with presuppositionality. However, because at least 5 out of 6 predicates that we tested (see 13 below) are able to introduce new information, we assume that predictions concerning the effect of negation still hold for our experimental design. But we also conducted a post-hoc examination to see if there were any differences between the verbs.

Experiment 1a tested the auxiliary hypothesis in (11a), according to which in sentences with matrix negation clausal complements are perceived as more likely to be given compared to affirmative sentences. Experiment 1b tested the main hypothesis in (11b), according to which in sentences with matrix negation there is a higher preference for NCs compared to affirmative sentences, which is a prediction of the preference hypothesis in (2) under the assumption that the auxiliary hypothesis is true. Experiment 1a also tested an additional prediction of the preference hypothesis, namely that NCs are more likely to be given compared to bare CP, as in (11c). The prediction follows from the preference hypothesis since if givenness increases the likelihood of NCs (but not of bare CPs), they should be more likely to be given compared to bare CPs, all else being equal.

(11) **Experiments 1a and 1b: Hypotheses**

- a. Clausal complements are perceived as more likely to be given in negative sentences, compared to affirmative sentences.
- b. There is a higher preference for NCs in negative contexts compared to affirmative contexts.
- c. NCs are perceived as more likely to be given compared to bare CPs.

While our main hypotheses tested the effect of negation, we also wondered whether polar interrogatives might have a similar effect, given the infelicity of interrogative sentences like (12a)–(12b) in an out-of-the-blue context, although this effect may not be as systematic, as shown by (12c). Thus, we also included interrogatives in the experimental design.

(12) [Uttered out of the blue:]

- a. #Did you tell him that [Bill and Anna broke up]?
- b. #Do you believe that [Bill and Anna broke up]?
- c. Do you think that [Bill and Anna broke up]?

To test the hypotheses in (11), we used six predicates in (13), which were the same in both experiments: two nonfactive belief verbs (*nadejat'sja* 'hope', *rassčityvat* 'count on'), two nonfactive speech act verbs (*žalovat'sja* 'complain', *nastaivat* 'insist'), one cognitive factive (*priznat'sja* 'confess') and one emotive factive (*žalet* 'regret').

(13) *nadejat'sja* (na+ACC) 'hope', *rassčityvat'* (na+ACC) 'count on', *žalovat'sja* (na+ACC) 'complain', *nastaivat'* (na+LOC) 'insist', *priznat'sja* (v+LOC) 'confess', *žalet'* 'regret' (o+LOC)

The predicates in (13) were selected among PP-taking predicates, which, unlike e.g. accusative-taking predicates, show a regular alternation between bare and nominalized complements (inside PP) (see Section 2.1). Note that prototypical nonpresuppositional predicates such as *dumat'* 'think' and *govorit'* 'say' tend to take accusative complements (or no nominal complements at all) and therefore do not normally combine with NCs.¹⁷ As a result, predicates that we tested, as in (13), were mostly nonprototypical members of their respective classes. Otherwise, we tried to select predicates that are fairly common with both complement types (cf. Table 1) and also belong to different semantic classes (mostly for exploratory purposes). The number of the predicates tested (six), which is admittedly, rather small, was dictated by design considerations: it had to be a multiple of the number of conditions in Experiment 1a (3 × 2). Yet, we thought that testing a larger set of predicates (e.g. 12) would overburden participants as we also wanted to have at least two sentence sets per verb.

More importantly for our purposes, all of the predicates in (13), with the possible exception of *žalet'* 'regret' (but see footnotes 18 and 28), can introduce new information.¹⁸ Thus, we can expect that they can in principle be affected by negation in the way described by the hypothesis in (11b).

3.2. Experiment 1a

3.2.1. Design and materials

The experiment had a 3 × 2 design, crossing sentence type (affirmative vs. negative vs. interrogative) and complement type (bare vs. nominalized), as in (14) (see Appendix A for the full list of materials). For each verb in (13) two sentence sets consisting of 6 conditions were constructed. The 72 sentences (12 per condition) were distributed

¹⁷Several of such predicates, including *dumat'* 'think', *govorit'* 'say' and *predpolagat'* 'suppose', in principle allow NCs embedded in *about*-PPs. However, such constructions usually have a slightly different meaning (e.g. 'contemplate' or 'discuss') or are restricted to presuppositional contexts (in this they differ from ACC/PP-taking verbs in (19b) tested in Experiments 2a and 2b). This is in line with the preference hypothesis but makes these predicates unsuitable for this particular experimental design.

¹⁸A suitable new-information context is given in (i) (in the English translation) for verbs *žalovat'sja* 'complain', *nastaivat'* 'insist' and *priznat'sja* 'confess' and in (ii) for verbs *nadejat'sja* 'hope' and *rassčityvat'* 'count on'.

(i) A: What happened?

B: I just met someone who VERBS/VERBED that he was robbed.

(ii) A: What is Vasja going to do on the weekend?

B: Katja VERBS that Vasja is going to come to visit us.

While *žalet'* 'regret', as an emotive factive, is often taken to require familiar complements (Caplan & Djärv 2019), it is still compatible with new-information complements in some specific contexts such as embedded announcements (Abbott 2000:1430) (see also Section 4.1 and footnote 28).

across 6 lists in a Latin Square design such that each participant saw each sentence set only in one condition and 2 sentences per condition.

(14) a. **Affirmative / negative condition**

Pravitel'stvo nadeetsja / ne nadeetsja, (na to) što
 government hopes NEG hopes on that.ACC COMP
 peregovory o postavke syr'ja vozobnovjatsja.
 negotiations about supply of.raw.materials will.resume
 'The government hopes / does not hope that that negotiations on the supply of raw materials will resume.'

b. **Interrogative condition**

Context: An interviewer asks an official:
Nadeetes' li Vy (na to), što peregovory o
 hope PRT you on that.ACC COMP negotiations about
 postavke syr'ja vozobnovjatsja?
 supply of.raw.materials will.resume
 'Do you hope that that negotiations on the supply of raw materials will resume.'

There were 18 fillers (see Appendix A), all of which had subjunctive and infinitival clauses with or without *to*. Half had the same sentence types as experimental sentences and half other sentence types. The purpose of the fillers was to divert participants' attention from the experimental manipulation.

3.2.2. *Procedure and participants*

Participants were told that they were going to see sentences containing information which may or may not be familiar from a potentially preceding text or dialogue and were asked to rate on a 1–6 scale how likely this information (which was explicitly referred to in the question, typically using a lexical nominalization) was talked about before, as illustrated in (15) (in English translation). Below each sentence was a scale with the labeled endpoints.¹⁹

(15) In your opinion, was the resumption of negotiations on the supply of raw materials talked about in the (potentially) preceding text (dialog)?

| |
|--|
| The government hopes that negotiations on the supply of raw materials will resume. |
| 1 2 3 4 5 6 |

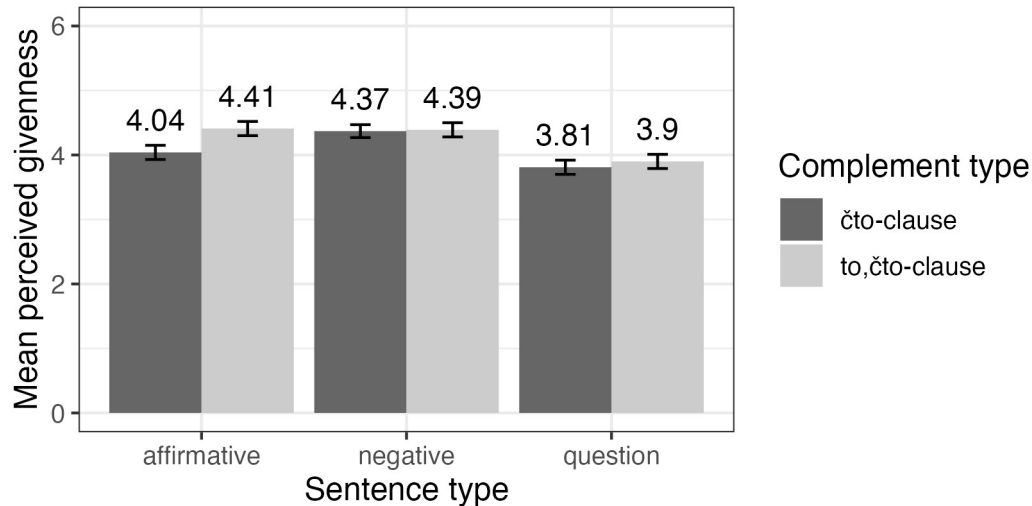
The experiment was hosted on Google Forms. The link to the experiment was distributed via social media. The experiment was completed by 125 participants, 67% of whom were aged 18–24.

¹⁹We used an even-number scale to prevent participants from always choosing the middle of the scale. The meaning of the midpoints was explained in the instructions as follows: 2 = *The information was probably introduced as new*; 3 = *The information was more likely to have been introduced as new rather than mentioned before*; 4 = *The information was more likely to have been mentioned before rather than introduced as new*; 5 = *The information was probably mentioned before*.

3.2.3. Results and discussion

Mean perceived givenness ratings per condition are given in Figure 1.

Figure 1. Mean rating of perceived givenness broken down by sentence type and complement type (Experiment 1a)



As can be seen, conditions did not strongly differ from each other. At the same time, in the bare CP condition negative sentences were associated with slightly higher givenness compared to affirmative sentences. The results were analyzed using linear mixed effects models as implemented in the lme4 (Bates et al. 2015) and the lmerTest (Kuznetsova et al. 2017) packages in R. We report the models with maximal random effects structure that allowed for convergence (Barr et al. 2013).

The model with sentence type (with affirmative as baseline), complement type (with bare CP as baseline) and their interaction as fixed effects and participant and item as random effects showed the main effect of sentence type ($\beta = 0.33$, $SE = 0.14$, $t = 2.37$, $p = 0.02$), such that that negative sentences were associated with higher givenness compared to affirmatives, in accordance with the hypothesis (11a). Interestingly, unlike negative sentences, interrogatives showed a marginally significant effect in the opposite direction ($\beta = -0.23$, $SE = 0.14$, $t = -1.70$, $p = 0.09$). (We do not have an explanation for this effect and do not discuss it further.) There was also a main effect of complement type ($\beta = 0.37$, $SE = 0.14$, $t = 2.68$, $p = 0.007$), such that NCs were associated with higher givenness compared to bare CPs, also in accordance with the hypothesis (11c). In addition, there was a marginally significant interaction between sentence type and nominalization ($\beta = -0.35$, $SE = 0.20$, $t = -1.78$, $p = 0.08$), suggesting a trend towards the reduction of the effect of nominalization in negative sentences.²⁰ (See Appendix C for the model outputs.)

²⁰This trend is consistent with the hypotheses (11a) and (11c), if we assume that each of the factors *alone* (i.e. the presence of negation and nominalization) is sufficient for comprehenders to infer givenness of the complement. Thus, in the presence of one factor the other becomes redundant and has no effect on perceived givenness.

As can be seen, the results are broadly in line with the hypotheses (11a) and (11c). However, the effects are numerically small and are not fully consistent across predicates and items (see Figure B1 in Appendix B for the by-item results). For example, in the bare CP condition (more strongly affected by negation, see Figure 1), with 2 out of 6 verbs (*nastaivat* ‘insist’, *priznat’sja* ‘confess’) the effect of negation, calculated as the difference in givenness between the negative and the affirmative condition, was negative or close to zero and with the remaining verbs it was rather weak (0.27–0.74). Furthermore, the latter group did not form a natural class. Thus, while the results are statistically significant, they provide only limited support for the hypotheses in (11).

Arguably, such weak results may be explained by the difficulty of the task, which required participants not only to reconstruct a possible discourse context for a given sentence but also to evaluate this context with respect to a rather abstract property (givenness). As suggested by a reviewer, this may have led to noise in the results.

A further question is why there is a discrepancy between our results and those of Caplan & Djärv (2019), who got a stronger and more consistent effect of negation (ca. 2–4 points for speech act/belief predicates) using a similar task. One difference between the two experiments is that we did not use baseline conditions for the discourse status.²¹ As a result, we had no way of excluding participants who may not have fully understood the task.²² Another difference is that Caplan & Djärv used more prototypical nonpresuppositional verbs (*say, tell, mention, think, believe, suspect*), which are arguably more strongly biased towards new-information complements and therefore should show a stronger effect of negation.

3.3. Experiment 1b

3.3.1. Design and materials

The experiment tested the hypothesis (11b), namely that negative sentences are associated with a higher preference for NCs compared to affirmative sentences. As in Experiment 1a, polar interrogatives were included for comparison. The dependent variable was complement type (nominalized vs. bare) and the only predictor was sentence type (affirmative, negative, question).

The same six verbs, as in Experiment 1a, were tested (cf. 13), but new stimuli sentences were constructed. For each verb, two sentence sets (in three experimental conditions each) were used. The 36 experimental sentences (see Appendix A) were

²¹Caplan & Djärv (2019) used the following baselines for discourse-new and discourse-given/old complements:

- (i) Was Joel leaving his wife talked about before?
 - a. *Discourse-new baseline (predict high ratings)*:
Guess what! Joel left his wife.
 - b. *Discourse-old baseline (predict low ratings)*:
Guess what! John thinks, like you do, that Joel left his wife.
(adapted Caplan & Djärv 2019:25).

²²A post-hoc examination showed that filler sentences (which contained subjunctive and infinitival complements) received on average lower givenness ratings compared to experimental sentences ($\beta = -0.67$, $SE = 0.10$, $t = -6.90$, $p < 0.001$). Assuming that subjunctive and infinitival clauses are less likely to be given than indicatives (since they refer to unrealized events), the experiment can still be taken to provide a valid measure of perceived givenness.

distributed among 3 lists in a Latin Square design such that each participant saw 12 sentences per condition.

There were 18 filler sentences, whose aim was to prompt participants to use both bare and nominalized CPs (and avoid strategic responses), as well as to exclude participants who were not paying attention to the task. One group of fillers contained sentences with predicates *svjazan* ‘be connected’ and *zaključat’sja* ‘be comprised’ which require a NC (inside PP), as in (16a). The second group contained sentences with verbs *rešit* ‘decide’ and *sčitat* ‘think’, which require bare CP or infinitival clauses, as in (16b). The unacceptable option for the second group was a NC inside an *about*-PP (*o tom*). The third group contained sentences that optionally allow NCs. In addition, half of the fillers were similar in structure to experimental sentences but had subjunctive or infinitive complements (cf. 16a), whereas the other half had declarative complements but a different overall structure (cf. 16b) (see Appendix A for the full list of fillers).

(16) a. Otstavka mèra svjazana *(s tem), čto-by izbežat’
 resignation of.mayor related with that.INS COMP-SUBJ avoid.INF
 v sostave upravljenja ljudej s pretenzijami
 in composition of.government of.people with complaints
 so storony pravooxranitel’nyx organov.
 from side law-enforcement of.authorities
 ‘The mayor’s resignation is related to avoiding people with complaints from law enforcement agencies on the board.’

b. Načal’nik možet sčitat’ (*o tom), čto generirovanie
 boss may think.INF about that.LOC COMP generation
 idej otvlečët potencial’nego rabotnika ot ispolnenija
 of.ideas will.distract potential worker from performance
 objazannostej.
 of.duties
 ‘The boss may think that generating ideas will distract the potential employee from the performance of his or her duties.’

Our choice of the unacceptable fillers deserves comment. Sentences like (16b), with *o tom*, sometimes occur in production, but they are considered nonstandard (Glovinskaja 2000). Thus, we assumed that participants would always opt for a fully acceptable alternative (in an FC task). Surprisingly, participants chose the nonstandard option rather frequently (ca. 30%), suggesting that it is genuinely acceptable at least for some speakers.²³ We decided to exclude such participants from the main analysis, but we also report the results including them.

3.3.2. Procedure and participants

A forced-choice (FC)/fill-in-the-blank task was used. Participants were presented with sentences containing blanks followed by two response options corresponding to the blank (whose order was randomized across trials). One option was “matrix verb + *čto*”,

²³This was confirmed by Experiment 2b, where participants sometimes produced such sentences (see Section 4.4).

the other was “matrix verb + P + *to* (in the required case) + *čto*” (see 17 below). (In the case of infinitival fillers, the verb was followed by *čtoby* + infinitive.) In the negative condition, illustrated in (17), the verb was preceded by the negative particle *ne*; in the interrogative condition the verb was followed by the question particle *li* and the 2nd person subject (cf. 14b).²⁴ Participants were instructed to choose the option that sounded more natural to them.

- (17) a. Štatnye artisty teatra _____ griměrki do six por
 staff actors of.theater _____ dressing.rooms until these times
 naxodjatsja v sostojanii remonta.
 are in state of.repair
 b. ne žalujutsja, čto
 NEG complain that
 c. ne žalujutsja na to, _____ čto
 NEG complain on that.ACC COMP
 ‘The staff artists of the theater (do not) complain that the dressing rooms are still in a state of repair.’

The experiment was hosted on Google Forms. The link to the experiment was distributed via social media. It was completed by 105 participants, 68% of whom were aged 18–24.

3.3.3. Results and discussion

The results were analyzed using mixed effects logistic regression, as implemented in the *lme4* and the *lmerTest* packages for R (see above). Response data were coded with 1 for a selection with *to* (NC) and with 0 for a selection without *to* (bare CP). In addition, proportions tests (chi-square tests of homogeneity) were performed at each level of complement type to determine whether the proportion of NC selections significantly differed from chance.

We first present the results for the fillers. The percentages of selections with *to* (with 95% confidence intervals/CI) are shown in Table 2.

Table 2. Percentages of selections with and without *to* by filler type (Experiment 1b)

| Filler type | % selections with <i>to</i> (NCs) | 95% CI |
|---|-----------------------------------|--------|
| obligatory (P +) <i>to</i> | 95 | 93–96 |
| optional (P +) <i>to</i> | 55 | 51–59 |
| disallowed/nonstandard <i>to</i> (<i>o tom</i>) | 28 | 24–32 |

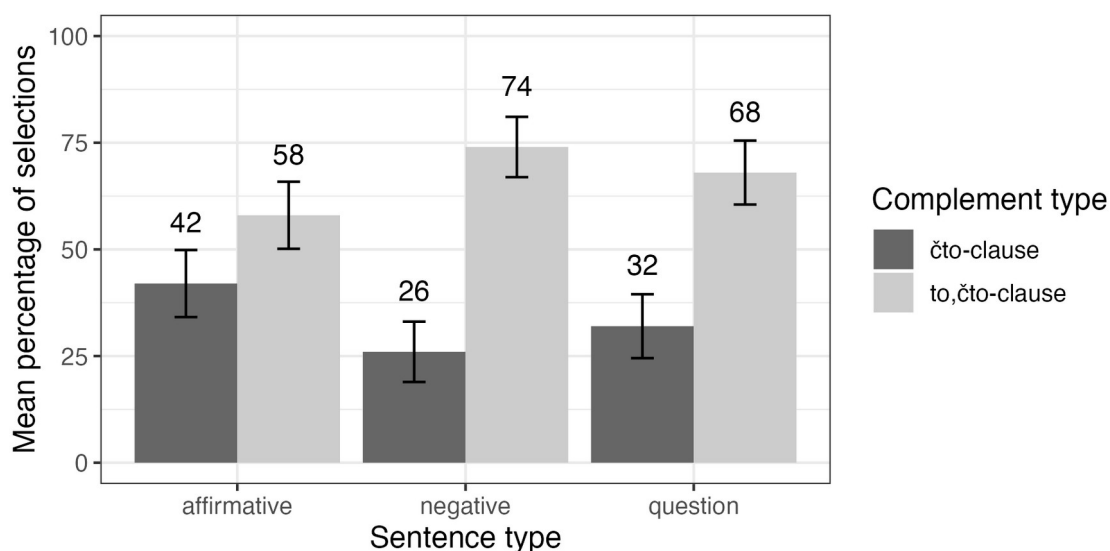
²⁴The inclusion of the negation in the answer options rather than in the prompt was not driven by any particular consideration. But note that separating the negation (proclitic) from the verb may disrupt the reading flow for participants. As for the question particle, its inclusion was due to the fact that it occurs between the verb and the complement.

Participants correctly chose the option with *to* for sentences with obligatory NCs 95% of the time. For sentences with disallowed NCs in standard Russian, there was a strong but not absolute preference for responses without *to*, with 28% of selections with *to*, presumably owing to their acceptability in nonstandard registers (see above). For sentences with optional NCs, selections with *to* were, as expected, close to 50%.

We excluded data from 18 participants who gave at least one unacceptable response (without *to*) for the fillers with obligatory NCs. We also excluded additional 47 participants who gave responses with unacceptable/nonstandard *to* (*o tom*). Thus, for the main analysis we included data from 40 participants (we also ran a separate analysis including the second group of participants).

The percentages of responses with and without *to* (with 95% CI) for the experimental conditions are summarized in Figure 2.

Figure 2. Mean percentages of *čto*- and *to,čto*-clause selections broken down by sentence type (Experiment 1b)

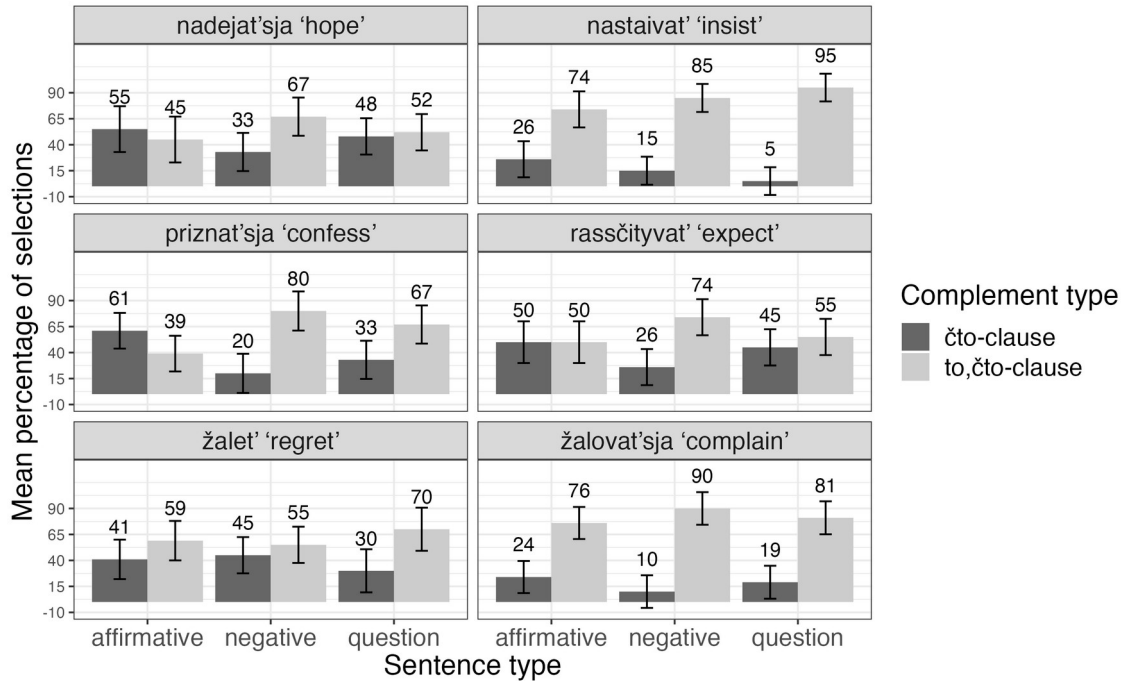


As can be seen, responses with *to* (NCs) were preferred in all three conditions. Crucially, this preference was higher in the negative (74% of selections with *to*) and also the interrogative condition (68%), compared to the affirmative condition (58%). This was confirmed by a statistical analysis. A model with sentence type as a fixed effect and participant and item as random effects showed that negative sentences were associated with a significant increase in the proportion of NCs compared to affirmatives ($\beta = 0.84$, $SE = 0.27$, $z(480) = 3.07$, $p = 0.002$). Interrogatives showed a similar but weaker effect ($\beta = 0.54$, $SE = 0.27$, $z(480) = 2.02$, $p = 0.04$). The proportions tests further showed that in negative and interrogative sentences NCs were significantly more frequent than chance ($\chi^2(1) = 35.2$, $p < 0.001$; $\chi^2(1) = 18.9$, $p < 0.001$). Affirmatives also showed a preference for NCs but it was just below the significance threshold ($\chi^2(1) = 3.9$, $p < 0.048$).²⁵ (See Appendix C for the model outputs.)

²⁵The analysis including accepters of disallowed/nonstandard *o tom* showed an even stronger preference for NCs (83%) in negative sentences ($\beta = 1.74$, $SE = 0.21$, $z(1044) = 8.23$, $p < 0.001$). Another difference is that all six verbs including *žalet'* 'regret' showed an effect of negation (see below). At the same time, interrogatives showed an equal preference for *to* (52%;

A by-verb analysis presented below in Figure 3 further confirmed this result.

Figure 3. Mean percentages of *čto*- and *to,čto*-clause selections broken down by sentence type and verb (Experiment 1b)



Five out of six verbs (except for *žalet* 'regret') showed a higher proportion of selections with NCs in negative sentences compared to affirmatives (see Figure B2 in Appendix B for the by-item results). Interestingly, *žalet* 'regret', which showed a roughly equal preference, was the only verb tested whose lexical semantic class (i.e. emotive factive) is strongly associated with familiar complements (cf. Caplan & Djärv 2019). Therefore, the lack of effect with this verb is in fact consistent with our theoretical expectations (although we did not make this prediction before the experiment).

As for the effect of interrogatives, it was not very consistent and, moreover, did not pattern with negation. It was clearly observed for three of the verbs, including *žalet* 'regret' (also *nastaivat* 'insist' and *priznat'sja* 'confess'), but was very weak for the remaining three (*nadejat'sja* 'hope', *rassčityvat* 'count on' and *žalovat'sja* 'complain'). Because this effect is difficult to interpret (especially in the context of Experiment 1a), we leave it for future study.

Overall, the results of Experiment 1b provide clear evidence for the view that negative sentences are indeed associated with a higher preference for NCs compared to affirmatives, in accordance with the hypothesis (11b). These results can, in turn, be taken to support the preference hypothesis on the assumption that matrix negation is associated with higher perceived givenness, as formulated by the hypothesis (11a). Although Experiment 1a, which specifically tested the latter hypothesis, did not provide unambiguous support for it, the overall pattern of results of Experiment 1a was still in line with it. Moreover, we provided some independent reasons for why Experiment 1a

$\beta = 0.03$, $SE = 0.16$, $z(1044) = 0.17$, $p = 0.87$), similarly to affirmatives (52%). (See Figures B3 and B4 in Appendix B.)

may have failed to show a clear effect of givenness (see Section 3.2.3). Given these considerations, as well as the existence of independent evidence for (11a) in Caplan & Djärv 2019, we take the results of Experiment 1b to provide support for the preference hypothesis (although we admit that the hypothesis (11a) requires further confirmation).

Before concluding this section, we need to discuss an interesting aspect of the results of Experiment 1b, namely that there was a relatively large proportion of responses with NCs in the baseline (affirmative) condition (58%). This may appear surprising if one assumes that the primary function of NCs is to mark presuppositionality. While it is true that there are other factors driving nominalization, as was suggested in Section 2.3, one may still wonder whether such factors are sufficient to render NCs the (slightly) preferred option in the absence of contextual cues of presuppositionality. We suggest that this is to be explained by the fact that the experimental sentences in Experiment 1b were presented *without context*. As a consequence, some participants may have reconstructed a *discourse-given* context at least for some of the sentences in the baseline condition, leading to an “inflated” rate of NCs. This is not implausible considering that some of the verbs tested may have a relatively weak bias towards new-information complements.

4. Experiments 2a and 2b: contextual givenness and nominalization

4.1. Hypotheses, design and materials

Experiments 2a and 2b provide a more direct test of the preference hypothesis in (2) by way of manipulating the discourse context. In the experiments, sentences with predicates that optionally take bare or nominalized clauses (inside PPs) were presented in two types of discourse contexts that render the complement new or given information. According to the preference hypothesis, NCs should be more likely to occur and more preferred in discourse-given compared to discourse-new contexts. We tested this prediction using an FC task (Experiment 2a) and a sentence completion task (Experiment 2b). Otherwise, the two experiments had the same design and used the same materials.

Following Rochemont (2016), we took givenness to arise from relations of either coreference or entailment between propositions. For the purposes of the experiments, however, we used a looser criterion. Namely we constructed target sentences in the discourse-given condition in such a way that they should be either roughly synonymous with some proposition in the presented discourse fragment (\approx coreference) or could be understood as a probable inference from such proposition (\approx entailment) (cf. 18 below).

We realize that such operationalization of givenness may have made it more difficult for participants to determine whether the complement was new or given (because it ultimately involves a gradient, as opposed to categorical, notion of givenness) and that a stricter operationalization of givenness might have been preferred. However, a full or substantial repetition of material between the two propositions, which might ensure a stricter notion of givenness, could have made the experimental manipulation too obvious for the participants and led to a certain artificiality of the task. Thus, we consider our methodological choice to be a reasonable trade-off. To compensate for this rather imprecise operationalization of givenness, we ran a separate

norming study in which we asked participants to rate the givenness of the complement in the two types of contexts (see Section 4.2).

In both Experiments 2a and 2b, the experimental materials consisted of the same short dialogue exchanges, as in (18) (all sentences were modified from actual news texts). (See Appendix A for the full list of materials.)

(18) A: Po zaveščaniju izvestnoj xudožnicy okazalos', što ona
by will famous of.artist. turned.out COMP she
poželala odat' vse svoi raboty Èrmitažu, daže
wished give.INF all her works to.Hermitage even
te, kotorye naxodjatsja v Tret'jakovskoj galeree. Èta novost'
those which are in Tretyakov gallery this news
vyzvala bol'šoj rezonans.
caused big resonance

'According to the will of the famous artist, it turned out that she wished to give all her works to the Hermitage, even those that are in the Tretyakov gallery. This news caused a great resonance.'

B: Byli li komentarii predstavitelej Èrmitaža?
were PRT comments of.representatives of.Hermitage
'Were there any comments from Hermitage representatives?'

A: Odin iz nix besedoval so mnoj. On byl očen' obradovan
one from them spoke with me he was very delighted
ètoj novost'ju.
this by.news

'One of them spoke to me. He was very happy about this news.'

A1. **Discourse-given condition**

Ved' on davno nadejalsja (na to), što kartiny
PRT he long hoped on that.ACC COMP paintings
xudožnicy iz Tret'jakovskoj galerii perejduť v fond
of.artist from Tretyakov gallery will.go to fund
Èrmitaža.
of.Hermitage

'After all, he had long hoped that the artist's paintings from the Tretyakov gallery would go to the Hermitage Fund.'

A2. **Discourse-new condition**

No on takže nadejalsja (na to), što èto ne privedët
but he also hoped on that.ACC that this NEG will.lead
k konfliktu meždu muzejami.
to conflict between museums

'But he also hoped that it will not lead to a conflict between the two museums.'

Each exchange started with speaker A introducing some news. Then followed a short question by A's interlocutor B asking about the reaction of some officials or other parties to this news. The exchange ended with A's reply which contained the target sentence with the complement clause (underlined).²⁶ In half of the replies (the discourse-

²⁶One sentence pair deviated from this scheme in that the first speaker reported the news and the second speaker gave the reply.

given condition, A1), the complement in the target sentence referred back to the information in A's introductory utterance by either paraphrasing it or by being inferred from it. In the other half (the discourse-new condition, A2), the complement was neither expressed by nor inferred from the information in the introductory utterance.

For example, in the discourse-given condition in (18) the proposition that a certain artist's painting will go from the Tretyakov gallery to the Hermitage (A1) can be inferred with some probability from the proposition that the artist stated in her will that all her paintings should be given to the Hermitage. The discourse-given status of (A1) is further supported by the possibility of placing the narrow focus on the verb and deaccenting the complement. By contrast, in the discourse-new condition, the proposition that the relocation of the paintings to the Hermitage will not lead to a conflict between the museums (A2) is neither expressed by nor inferred from the complement in the introductory utterance (accordingly, deaccenting the complement sounds odd).

Eight verbs that take optional NCs inside PP, as in (19), were tested. We used four verbs from Experiments 1a/1b (cf. 13), as in (19a), including two nonfactive belief/speech act verbs *nadejat'sja* 'hope' and *žalovat'sja* 'complain', one cognitive factive *priznat'sja* 'confess' and one emotive factive *žalet'* 'regret'.²⁷ Note that although *žalet'* 'regret', as an emotive factive, is usually taken to require discourse-given complements (Caplan & Djärv 2019), in fact it allows new-information complements when interpreted as a verb of announcing (see footnote 18).²⁸ Indeed, we were able to elicit this interpretation from our participants, as suggested by the fact that *žalet'* 'regret' patterned like others verbs in the norming study (and also in the experimental results).

(19) a. **PP-taking**

nadejat'sja 'hope', *žalovat'sja* 'complain', *žalet'* 'regret', *priznat'sja* 'confess'

²⁷The original design also manipulated factivity of the predicate. However, because this factor did not take into account the distinction between emotive and cognitive factives, which is relevant for the discourse status of the complement (Caplan & Djärv 2019), we discarded it from the analysis. This factor (as well as any interactions with it) was not significant in both Experiments 2a and 2b in all of the analyses that we report.

²⁸Cf. one of our discourse-new items with *žalet'* 'regret' in (i).

- (i) A: Včera byla transljacija čempionata mira po tenisu, v
 yesterday was broadcast of.championship of.world for tennis in
 ktorom odin iz izvestnyx tennisistov poterpel poraženie v finale,
 which one of famous tennis.players suffered loss in final
 a uže segodnja on dal interv'ju.
 and already today he gave interview

'Yesterday there was a broadcast of the world tennis championship, in which one of the famous tennis players lost in the final, and already today he gave an interview.'

B: Kak on prokomentiroval svoj proigryš?

How he commented his loss

'How did he comment on his loss?'

A: Sportsmen žalet (o tom), čto ne smog nakanune čempionata
 athlete regrets about that.PREP COMP NEG could before championship
 čašče poseščat' trenirovki.
 more.often visit.INF training.sessions

'The athlete regrets that he couldn't go to practice more often before the championship.

He apologizes to all the fans.'

b. ACC/PP-taking

ob''javit 'announce', *soobščit* 'report', *uslyšat* 'hear', *uznat* 'learn'

We also used four new verbs, as in (19b), which were included mostly for exploratory purposes. Two of these verbs were nonfactive speech act verbs (*ob''javit* 'announce' and *soobščit* 'report') and the other two cognitive factives (*uslyšat* 'hear', *uznat* 'learn'). These verbs, referred to as ACC/PP-taking verbs, also show an alternation between bare and nominalized CPs (inside *about*-PPs) with no clear difference in meaning, as in (20), much like PP-taking verbs in (19a). But they differ from the latter group in that they also allow (nonclausal) nominal complements in the accusative (cf. 20).^{29,30}

(20) Trener ob''javil {(o tom), čto zriteli ne smogut
 coach announced about that.PREP COMP spectators NEG will.be.able
 uvidet' vystuplenie figuristki na Olimpiade / èto / ob
 see.INF performance of.figure.skater on Olympics this about
 ètom}.
 this.PREP

'The coach announced that the spectators would not be able to see the figure skater perform at the Olympics / this.'

For each verb in (19) we constructed 2 matched pairs of exchanges (in the discourse-new and the discourse-given condition), as in (18). The resulting 32 exchanges were distributed among 2 lists in a Latin Square design, with each participant seeing 16 exchanges per condition.

There were 20 filler exchanges, divided into 3 groups according to the properties of the target sentence. In 8 fillers, target sentences had obligatory NCs inside a PP (cf. 16a), in another 8 they had disallowed/nonstandard NCs with *o tom* (except in one case) (cf. 16b), and in another 4 they had optional NCs. Half of the sentences in the first two groups had *čto*-clauses, and the other half had infinitival clauses, as did sentences in the third group. All sentences had discourse-new complements. The main goal of the fillers was to encourage participants to use both bare and nominalized CPs. nonalternating fillers were also used as an exclusion criterion. (See Appendix A for the full list of fillers.)

²⁹They also in principle allow NCs in the accusative (especially *uznat* 'learn' and *uslyšat* 'hear') (cf. 7b). However, such examples are exceedingly rare relative to examples with NCs inside *about*-PPs.

³⁰This difference becomes relevant on some null D accounts (e.g. Bondarenko 2022, see Section 2.2) which assume that null D is allowed only in the accusative, but not oblique, case. On such accounts, it becomes possible to analyze bare CPs with ACC/PP-taking predicates in presuppositional contexts as underlyingly DPs (in contrast to PP-taking predicates). This potentially predicts a lower rate of NCs in the discourse-given condition with ACC/PP-taking predicates compared to PP-taking predicates assuming that the former predicates can satisfy the DP requirement without the need for an overt D. The prediction was not confirmed as there were no differences between the two classes of verbs in either Experiments 2a or 2b, and we do not discuss it further. The null finding is consistent with the view that presuppositionality (at least understood as a contextual property) is marked by an *overt* D (see Section 2.2).

4.2. Perceived givenness norming study

To ensure that the experimental exchanges matched their intended discourse status, we asked a separate group of 15 participants to rate the perceived givenness of the complement in the target sentence. All target sentences contained bare CPs. There were no fillers. The 32 exchanges used in the experiments (cf. 18) were distributed among 2 lists (7 and 8 participants) in a Latin Square design so that participants saw 1 exchange per each pair. Participants were instructed to read each of the 16 exchanges carefully and to rate on a 1–7 scale whether the information in the complement was expressed in the preceding fragment. The replies with the target sentences were graphically separated from the rest of the exchanges, with the complements underlined, as illustrated in (21).

(21) Rate on a scale the truth of the following sentence: “*The underlined information in Fragment B was expressed in Fragment A.*”

| | |
|----------|--|
| <i>A</i> | ‘According to the will of the famous artist, it turned out that she wanted to give all her paintings to the Hermitage, even those that are in the Tretyakov gallery. This news caused a great resonance.’ ‘Were there any comments from Hermitage representatives?’ |
| <i>B</i> | ‘One of them spoke to me. He was very happy about this news. After all, he had long hoped <u>that the artist’s paintings from the Tretyakov gallery would go to the Hermitage Fund.</u> ’ |

was not expressed 1 2 3 4 5 6 7 was expressed
○ ○ ○ ○ ○ ○ ○

The results of the norming study showed that the perceived givenness of the complement was higher in the discourse-given ($M = 6.00$, $SD = 1.46$) compared to the discourse-new condition ($M = 1.78$, $SD = 1.44$), in accordance with our classification. This difference was confirmed by a linear mixed effects model with givenness, selectional class and their interaction as fixed effects (with discourse-new and PP-selecting as baselines) and participant, subject and by-item givenness slope as random effects. The model showed a significant effect of givenness such that in discourse-given contexts complements were perceived as more likely to be given ($\beta = 4.25$, $SE = 0.31$, $t = 13.52$, $p < 0.001$); selectional class and interaction were not significant ($p > 0.5$). (See Appendix C for the model outputs.) A further inspection of the items confirmed that all 16 sentence pairs showed a clear effect of givenness (calculated as the difference between the average perceived givenness of NCs in the discourse-given and discourse-new conditions). At the same time, there was also some variation among items, with the givenness effect ranging from 2.69 to 5.57 points. (See Figure B5 in Appendix B.)

4.3. Experiment 2a: Forced-choice task

4.3.1. Procedure and participants

Experiment 2a used an FC/fill-in-the-blank task. Participants were shown exchanges like (18) in which some material in the target sentence was replaced by a blank, as in

(22). They were followed by two response options in a randomized order. One option was “matrix verb + *čto*” as in (23a), the other was “matrix verb + P + *to* (in the required case) + *čto*”, as in (23b). (In the case of fillers with infinitival clauses, the verb was followed by either an infinitive or the preposition + *to* + *čto*by + infinitive.)

- (22) a. No on takže _____ èto ne privedět k konfliktu između
 but he also this NEG will.lead to conflict between
 muzejami.
 museums
- b. nadejalsja, _____ čto
 hoped COMP
- c. nadejalsja na _____ čto
 hoped on that.ACC COMP
- ‘But he also hoped that it would not lead to a conflict between the museums.’

Participants were asked to read the exchanges and fill in the blank by choosing the option that sounded more natural to them in the given situation. The experiment was conducted on Google Forms and was distributed via social media. It was completed by 53 participants, 70% of which were aged 18–24.

4.3.2. Results and discussion

We first present the results for the fillers. The percentages of selections with *to* (NCs) (with 95% CI) by filler type are shown in Table 3.

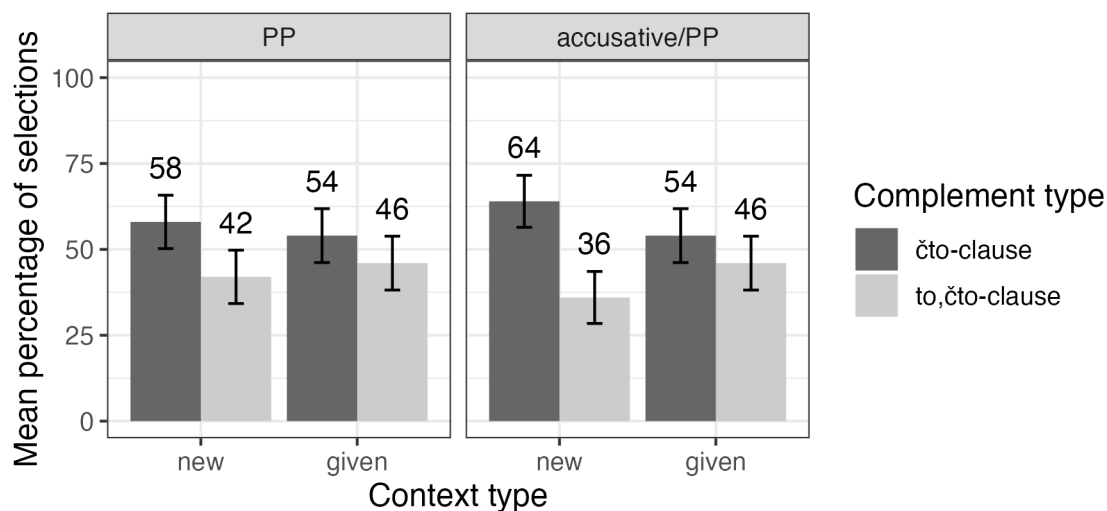
Table 3. Percentages of selections with and without *to* by filler type (Experiment 2a)

| Filler type | % selections with <i>to</i> (NCs) | 95% CI |
|---|-----------------------------------|--------|
| obligatory (P +) <i>to</i> | 98 | 96–99 |
| optional (P +) <i>to</i> | 34 | 28–41 |
| unacceptable/nonstandard <i>to</i> (<i>o tom</i>) | 4 | 2–6 |

Participants overwhelmingly chose the acceptable/standard option for nonalternating sentences. Interestingly, compared to Experiment 1b, there were much fewer selections with unacceptable/nonstandard NCs. 4 participants who chose the option without *to* for fillers with obligatory NCs, plus 8 additional participants who chose the option with *to* for fillers with unacceptable/nonstandard NCs, were excluded from the main analysis.

The mean percentages of selections with bare and nominalized CPs in the two conditions are shown in Figure 4.

Figure 4. Mean percentages of *čto*- and *to,čto*-clause responses broken down by discourse context and selectional class (Experiment 2a)



As can be seen, for verbs of both selectional classes there was a small preference for bare CPs in both conditions. At the same time, the percentages of selections with NCs were slightly higher in the discourse-given (46% for both selectional classes) compared to the discourse-new condition (42% and 36% for the PP- and ACC/PP-taking predicates), in accordance with the preference hypothesis.

The results were analyzed using mixed effects logistic regression, as in Experiment 1b. Models were compared using the `anova` function from the `lmerTest` package. We first fit a model which included givenness as a fixed effect and participant, item and by-item givenness slope as random effects. This model did not show a significant effect of givenness ($\beta = 0.34$, $SE = 0.22$, $z(656) = 1.51$, $p = 0.13$).³¹ However, a model with a simplified random effects structure without the slope but with the same goodness of fit ($\chi^2(2) = 3.52$, $p = 0.17$) did show an effect of givenness ($\beta = 0.35$, $SE = 0.17$, $z(656) = 2.05$, $p = 0.04$), suggesting that the effect may have been difficult to detect because of the variation in the items.³² The difference between the discourse-new and the discourse-given conditions also emerged in the proportions tests, which showed that in the discourse-new condition bare CPs were generally more frequent than NCs ($\chi^2(1) = 15.4$, $p < 0.001$), whereas in the discourse-given condition there was an equal preference for both ($\chi^2(1) = 1.91$, $p = 0.17$).

We also fit a model with givenness, selectional class (with PP-selecting as baseline) and their interaction as fixed effects and participant, item and by-item givenness slope as random effects. This model did not show any significant effects or interactions ($p > 0.4$) and similarly for a simpler model without the slope but with the same goodness of fit ($p > 0.3$), suggesting that verbs of the two selectional classes did not differ from each other.³³ (See Appendix C for the model outputs.)

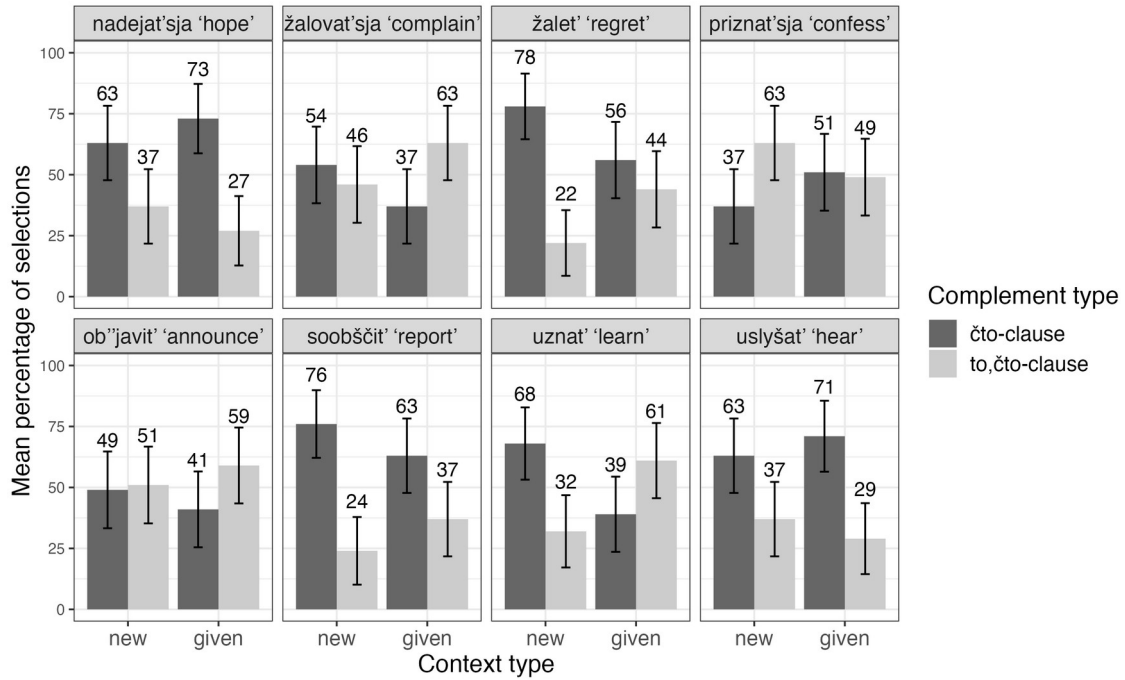
³¹A model including accepters of unacceptable/nonstandard *o tom* did show a significant effect of givenness ($\beta = 0.46$, $SE = 0.21$, $z(656) = 2.21$, $p = 0.03$).

³²The simplified model without the by-item slope was also the preferred model of the model selection algorithm implemented by the `model.sel` function of the `MuMIn` package for R.

³³A similar model including accepters of unacceptable/nonstandard *o tom* also did not show significant effects ($p > 0.3$).

A by-verb inspection (see Figure 5) showed that a positive effect of givenness (averaging across both items) was observed for 5 out of 8 verbs (*žalovat'sja* ‘complain’, *žalet* ‘regret’, *ob'javit* ‘announce’, *soobščit* ‘report’ and *uznat* ‘learn’), whereas 3 verbs (*nadejat'sja* ‘hope’, *priznat'sja* ‘confess’ and *uslyšat* ‘hear’) showed a negative effect. A further inspection of individual items showed a positive effect for 11 out of 16 items (see Figure B6 in Appendix B).

Figure 5. Mean percentages of *čto*- and *to,čto*-clause selections broken down by discourse context and verb (Experiment 2a)



Although the results provide mixed evidence for the preference hypothesis, we nevertheless take them to be suggestive, in view of the numerical trend in the predicted direction, as well as statistical significance in some analyses. We attribute the mixed results to features specific to the materials and to the experimental task, as we explain below.

As we saw in Section 4.2, there was some variation among items in the effect of givenness in the norming study. Assuming that items that showed a weaker effect in the discourse-given condition were also less likely to be perceived as given in the experiment, it is possible that some of such items may not have reached the required threshold for givenness for the purposes of the preference hypothesis. Evidence for this explanation comes from a statistically significant correlation ($r(14) = 0.63$, $p = 0.009$) between the effect of givenness in the norming study and in the experiment (calculated as the difference between the proportion of responses with NCs in the discourse-given and the discourse-new conditions).

However, we might still wonder why the experiment failed to yield a consistent effect of givenness compared to Experiment 1b, which tested the effect of matrix negation (using the same FC task) but showed a more robust effect. As suggested by a reviewer, the answer could be related to the nature of the experimental manipulation. Specifically, in Experiment 1b givenness was indicated by matrix negation, which is a

salient feature of the sentences, immediately perceived by participants. By contrast, in Experiment 2a (and 2b) givenness had to be *inferred* from the content of multi-sentence fragments, which presumably required much more engagement from participants. At the same time, a relative simplicity of the FC task itself (i.e. selecting one of the two acceptable options) may have led some participants not to properly attend to the accompanying context (despite attending to the target sentences, as we can verify by the results of the fillers), leading, in turn, to somewhat noisy results.³⁴

To summarize, despite certain caveats we take Experiment 2a to provide at least indirect evidence for the preference hypothesis. This conclusion is strengthened by Experiment 2b, which used arguably a more sensitive task and showed a stronger and more consistent effect.

4.4. Experiment 2b: Sentence completion task

4.4.1. Procedure and participants

The experiment had the same design and materials, as Experiment 2a, but used the sentence completion task.³⁵ Participants saw exchanges like (18) in which the target sentence appeared in an incomplete form, without functional material introducing the embedded clause (such as a complementizer, D (*to*) or P). The matrix predicate was separated by a slash from the clause and was enclosed in parentheses along with some material from the embedded clause, as in (23).

- (23) No on takže (nadejalsja / èto ne privedët k konfliktu) meždu
but he also hoped this NEG will.lead to conflict between
muzejami.
museums
'But he also hoped // it would not lead to a conflict between the museums.'

Participants were asked to read the exchanges and complete the target sentences by inserting any necessary function words (and punctuation) in place of the slash without changing the words inside the parentheses. Participants were instructed to enter into the answer field the completed parenthesized portion of the sentence. Participants were also given the option of entering a dash in case they found the sentence complete; this option was required for fillers with infinitives with unacceptable/nonstandard *to* (*o tom*) + *čtoby*. Participants were also provided with two completed sample exchanges.

The experiment was conducted on Google Forms and was distributed via social media. It was completed by 66 participants, 67% of which were aged 18–24.

³⁴Because in this (and the next) experiment we had no comprehension questions and also no baselines to test how participants assess whether information is given/new in clear cases, we cannot verify whether all participants complied with the task as intended. We thank a reviewer for pointing this out.

³⁵We could not ensure that there would be no overlapping participants between Experiments 2a and 2b. However, because the two experiments were separated by a one-month interval we assume that it should not have interfered with the results. We thank a reviewer for raising this point.

4.4.2. Coding

Three participants provided only sequences of function words (e.g. *čto* or *o tom, čto*). Such responses were accepted if they led to acceptable completions. Responses where speakers omitted or simplified embedded clause material (which were rather common) were also accepted. Typos and punctuation errors were ignored. Responses where participants provided anything other than the required functional material were coded as “other”. Table 4 lists percentages of “other” responses by category.³⁶

Table 4. Percentages of “other” responses by example category (Experiment 2b)

| Example category | % “other” responses | 95% CI |
|--|---------------------|--------|
| fillers: obligatory <i>to</i> | 20 | 16–23 |
| fillers: disallowed/nonstandard <i>to</i> (<i>o tom</i>) | 5 | 3–7 |
| fillers: optional <i>to</i> | 2 | 0–4 |
| experimental sentences | 3 | 2–4 |

As can be seen, in categories other than fillers with obligatory *to*, including experimental sentences, “other” responses were rare. Most “other” responses contained complex NPs headed by a lexical nominalization or a *fact*-like noun. Further common types of “other” responses included: changing the infinitival structure to an indicative one (with verbs that allow both), replacing indicative clauses with *because*-clause or embedded questions, using a different C (e.g. *budto* instead of *čto*) or P and leaving the embedded clause without *čto*.³⁷

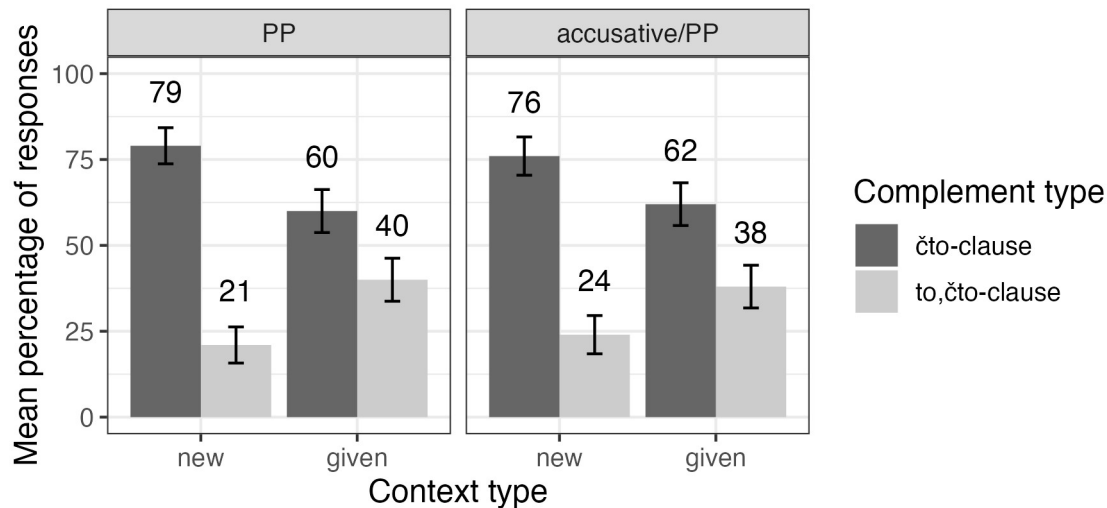
4.4.3. Results and discussion

Responses coded as “other” were excluded. Data from 3 participants who used nonstandard *o tom* in responses to fillers were also excluded (the results including those participants were very similar). The percentages of responses with bare and nominalized CPs (inside PPs) for the two conditions are given in Figure 6.

³⁶One sentence in the optional *to* category had a wrong structure due to an error and was excluded from the analysis.

³⁷Interestingly, a few “other” responses contained structures that were classified as unacceptable/nonstandard in Experiments 1b and 2a, such as omitting obligatory P + *to* (cf. 16a) and using nonstandard *o tom* (cf. 16b).

Figure 6. Mean percentages of *čto*- and *to,čto*-clause responses broken down by discourse context and selectional class (Experiment 2b)



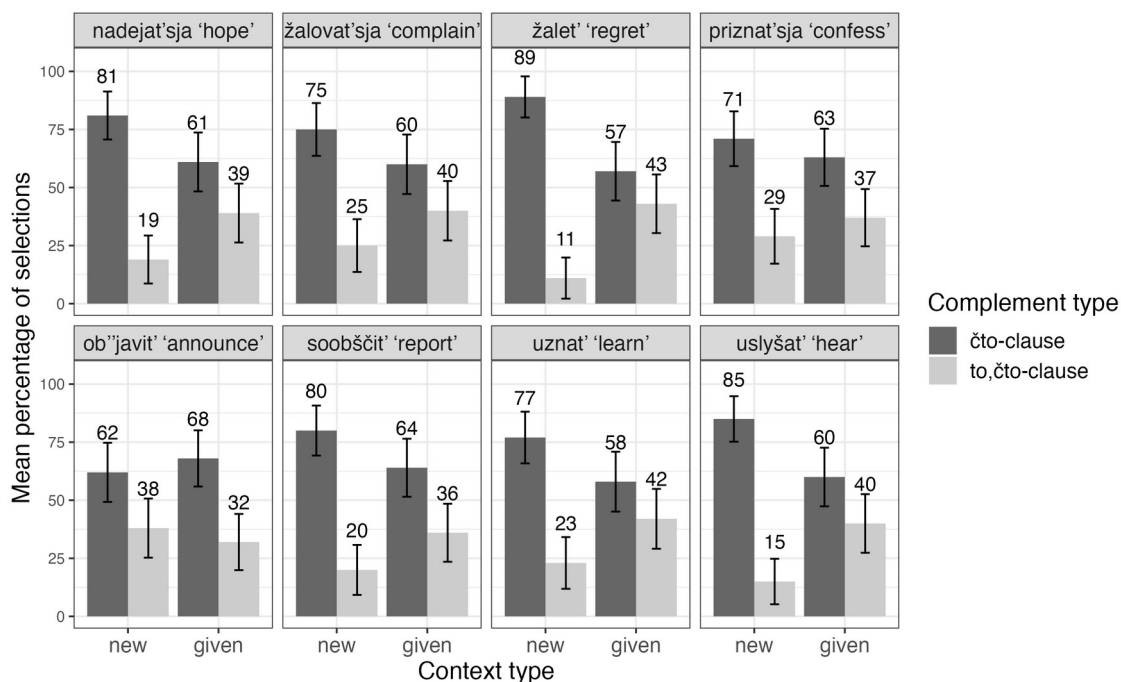
Although NCs were overall less frequent than bare CPs, they were more frequent in the discourse-given (40% and 38% for PP- and ACC/PP-taking verbs) compared to the discourse-new condition (21% and 24%), in accordance with the preference hypothesis.

The effect of givenness was confirmed by a statistical analysis. Mixed effects logistic regression with givenness as a fixed effect and participant, item and by-participant slope as random effects showed a significant effect of givenness ($\beta = 0.93$, $SE = 0.23$, $z(980) = 4.00$, $p < 0.001$). To assess the effect of selectional class, we also fit a model with selectional class and their interaction as fixed effects and participant, item and by-participant givenness slope as random effects. This model also showed an effect of givenness such that responses with NCs are more likely to occur in discourse-given compared to discourse-new contexts ($\beta = 1.11$, $SE = 0.22$, $z(980) = 4.97$, $p < 0.001$). Neither selectional class nor its interaction with givenness were significant ($p > 0.4$). (See Appendix C for the model outputs.) The proportions tests (collapsing the two selectional classes) showed that there was an overall preference for bare CPs in both conditions (77%, $\chi^2(1) = 146.1$, $p < 0.001$; 61%, $\chi^2(1) = 23.5$, $p < 0.001$).

A further by-verb analysis (see Figure 7) showed that only 1 out of 8 verbs (*ob''javit* 'announce') did not show the effect of givenness. Moreover, only 2 out of 16 items (1 with *ob''javit* 'announce' and 1 with *priznat'sja* 'confess') did not show the expected effect, with the remaining 14 items showing the effect of at least 9%. (See Figure B7 in Appendix B.) Note that these 2 items were also the ones that showed the smallest effect in the norming study, confirming the above suggestion that givenness may require a certain threshold to trigger nominalization (see Section 4.3.2).³⁸

³⁸There was also a marginally significant correlation between the effect of givenness in the experiment and the norming study ($r(14) = 0.44$, $p = 0.085$).

Figure 7. Mean percentages of *čto*- and *to,čto*-clause responses broken down by discourse context and verb (Experiment 2b)



A question one might ask is why the effect of givenness in this experiment was much clearer than in Experiment 2a, with the FC task. The answer, we think, is that the sentence completion task in Experiment 2b had a stronger production component than the FC task. Specifically, in the sentence completion task participants had to actually produce (by way of typing) the functional material themselves, as opposed to mentally insert the already produced options in the prompt, as in the FC task (although we admit that that the latter may also involve a bit of production). As a result, the task in Experiment 2b should be more similar to natural speech production and therefore should more strongly encourage participants to pay attention to the discourse context.

The higher sensitivity of the sentence completion task to the experimental manipulation (compared to the FC task) is also consistent with the fact that there was a lower rate of NCs in discourse-new contexts in Experiment 2b (on average, 23%) compared to Experiment 2a (39%). This is because if participants are more likely to take into account the discourse context, this should affect their identification of *both* discourse-new and discourse-given contexts. Accordingly, if the choice of NCs in discourse-new contexts is at least sometimes due to failure of participants to attend to the context, this should lead to a lower rate of NCs in such contexts in Experiment 2b.³⁹

Yet, as noted by a reviewer, the proportion of NCs in discourse-new contexts in Experiment 2b is still somewhat large in absolute terms. Given concerns about the operationalization of givenness and between-item variation in the norming study, it is unclear how accurate this estimate is. However, as it stands it argues against a strict

³⁹Note that, by the same token, we may expect that the proportion of NCs in discourse-new contexts should be smaller in Experiment 2a (39%) compared to Experiment 1b (58%) (which is in fact the case) because participants of the former experiment are provided with the discourse context in contrast to the latter experiment.

implication from nominalization to presuppositionality, confirming the view that factors beyond presuppositionality play a role in nominalization (see Section 2.3).

Another aspect of Experiment 2b is that despite the fact that givenness was shown to increase the preference for nominalization, it did not lead to a higher preference of NCs *compared to bare CPs* in discourse-given contexts. Thus, its effect can be characterized as *reducing* the preference for bare CPs, without reversing or neutralizing it.⁴⁰ Although such qualitatively weak effect may appear surprising, it is consistent with our view that nominalization is an *optional* marker of presuppositionality and further that other markers such as deaccenting may be more preferred (see the Introduction).

To summarize, Experiment 2b provides unambiguous support for the preference hypothesis, at the same time showing that presuppositionality is neither a necessary nor sufficient condition for nominalization. In the next section, we will attempt to place these findings in a larger theoretical context.

5. General discussion and conclusion

The starting point for our study was the correspondence hypothesis, according to which there is a two-way implication between presuppositionality and nominalization (Kastner 2015). Using distributional data from Russian, we showed that the hypothesis is too strong in both directions: first, presuppositional complements (of response and emotive factive predicates) can but need not be expressed by NCs and, second, NCs in general need not be presuppositional and can in principle express new information.

However, instead of rejecting the hypothesis, we suggested that it might hold in a weaker form (the preference hypothesis), whereby nominalization is a possible but not the only strategy of marking presuppositionality and, conversely, that presuppositionality is a possible but not the only or necessary source of nominalization. Crucially, in this weak form the hypothesis still predicts a higher preference for NCs in discourse-given compared to discourse-new contexts (all else being equal).

We conducted four experiments to test this prediction. Overall, despite certain caveats, the prediction was confirmed. The most straightforward support for it comes from Experiment 2b, which directly manipulated contexts in the sentence completion task and showed a reliable increase in the proportion of NCs in the discourse-given compared to the discourse-new condition. Experiment 2a, using the same design but the FC task, showed a similar trend (which, though, was significant only on some analyses). However, given that Experiment 2a presumably had a less sensitive task than Experiment 2b and given the post-hoc correlation between the experimental results and the norming study, we take it to provide indirect support for the preference hypothesis. Further evidence comes from Experiment 1b, which used matrix negation as a proxy for givenness, and showed a clear effect of negation. While independent testing of the effect of negation on perceived givenness in Experiment 1a yielded only a weak (though significant) result, given the noisy nature of the givenness rating task and also the absence of alternative explanations for the results of Experiment 1b, we interpret the latter results as also indirectly supporting the preference hypothesis. Finally, Experiment 1a showed that NCs are perceived as slightly more likely to be given

⁴⁰Similarly, in Experiment 2a discourse-givenness tended to neutralize the preference for bare CP, otherwise observed in discourse-new contexts, but did not reverse it.

compared to bare CPs (in affirmative sentences). We believe that taken together these results provide convincing evidence for the preference hypothesis.

We admit that there are important limitations of our experiments (especially 2a/2b) of which we highlight two (we thank a reviewer for clarifying these to us). First, although we had acceptable and unacceptable controls in Experiments 2a/2b to filter out inattentive participants and to prime participants with both bare and nominalized CPs, we did not have baselines to test how participants assess whether information is given or new in simpler sentences (although integrating such baselines would presumably make the task less straightforward). Second, we used a somewhat imprecise operationalization of givenness in Experiments 2a/2b, relying on participants' ability to infer givenness from quasi-naturalistic discourse fragments. We agree with the reviewer that using a more unambiguous operationalization might have led to a stronger effect.

Let's now turn to theoretical implications of our results. Perhaps the most relevant comparison for our study is Bogal-Allbritten & Moulton's (2018) and Moulton et al.'s (2020) analysis of a special kind of NCs in Korean and Japanese (so-called *ta-nun-kes/to-yuu-no* clauses). As the authors argue, such NCs are licensed by givenness, namely, they presuppose that the previous discourse contains an assertion (or, more precisely, an abstract contentful individual such as rumors, ideas, etc., including objects of assertion as a special case, cf. Moulton et al. 2020) that has the same content as the NC (or whose content entails the content of the NC).

Their results are similar to ours in three respects. First, they show that presuppositionality need not be lexically encoded by the predicate but can be contextually determined, which they demonstrate by manipulating the discourse context (with the predicate 'believe') (cf. Bochnak & Hanink 2022).⁴¹ Second, they tie presuppositionality to the presence of a definite D head (which they further analyze as encoding familiarity/anaphoricity). Third, they show that presuppositionality does not imply nominalization because *ta-nun-kes/to-yuu-no* clauses are generally not obligatory in discourse-given contexts. The main difference from our results is that Korean/Japanese NCs are *disallowed* in discourse-new contexts, suggesting that in those languages nominalization *does* imply presuppositionality, which is not the case for Russian NCs (with oblique/PP-taking predicates), as shown by the results of Experiments 2a/2b (see also 9 in Section 2.3).

A natural way to account for this difference is in terms of the different status of the familiarity implication in the two cases. Whereas in Korean/Japanese familiarity can be analyzed as a hardwired presupposition of the definite D (Moulton et al. 2020), in Russian, as we suggested in Section 2.3, it can be viewed as a defeasible inference from definiteness triggered in contexts entailing or implying familiarity but cancelled in other contexts. This is consistent with the fact that in our experimental data NCs were produced/selected in discourse-new contexts with nonnegligible frequency.

This inference can be compared to the familiarity implication associated with the English definite article, at least according to Abbott (1999, 2019), who argues, contra Heim 1982 (see also Roberts 2003), that the latter is not a hardwired presupposition

⁴¹Their argument assumes that the response-stance 'believe' (Cattell 1978), which they elicit in discourse-given contexts, is the same verb as the volunteer-stance 'believe' used in discourse-new contexts. The assumption is nontrivial, given that *ta-nun-kes/to-yuu-no* clauses are lexically selected and are obligatory with unambiguous response predicates like 'reject'/'accept'.

because it can be defeated, as in so-called “referent-establishing” (Hawkins 1978) uses.⁴² Another potentially relevant comparison is the factive inference of “optionally factive” verbs like *announce*, *report*, *acknowledge* (Kiparsky & Kiparsky 1970, Degen & Tonhauser 2022) triggered in some contexts (i.e. those that entail the truth of their complement) but not others, leading to these verbs being called “part-time/contextual” triggers (Abrusán 2011, Schlenker 2021).⁴³

Having dealt with the optionality of the implication from nominalization to presuppositionality, which is specific to Russian NCs, let us move on to the optionality of the implication from presuppositionality to nominalization, which is shared by Russian and Korean/Japanese NCs. How can we understand the latter from a theoretical standpoint? Before addressing this question, we need to ask what is the mechanism that explains the tendency to nominalize presuppositional clauses in the first place. As suggested by a reviewer, a natural way to think about this mechanism is in terms of Heim’s (1991) *Maximize Presupposition!* (MP) principle (Sauerland 2008), according to which when two variants have the same assertive meaning but differ in presuppositions, the more specialized variant, with stronger presuppositions, becomes obligatory, blocking the other one. A classic application of MP is the blocking of the indefinite article by the presuppositionally stronger definite article in uniqueness/familiarity contexts (assuming the nonuniqueness/nonfamiliarity presupposition is not inherent in the lexical entry for the former).

However, as pointed out by the same reviewer, if we apply this logic to the choice between bare vs. nominalized CPs in presuppositional (discourse-given) contexts, we will run into problems as we would expect at a minimum a preference for NCs over bare CPs in such contexts, given that the former but not the latter have the potential to contribute the relevant presuppositional inference (i.e. familiarity). (Here we are assuming that the fact that familiarity is not lexically encoded but only contextually triggered by Russian NCs is immaterial for the purposes of MP.) Yet, as we saw in the results of Experiments 2a/2b, there is no such preference since bare CPs are preferred in *both* discourse-new and discourse-given contexts (albeit more strongly in the latter).

One possible solution to this problem may be as follows. Throughout this paper we have been assuming that nominalization of presuppositional clauses is optional due to the fact that presuppositionality can be signaled by other means, e.g. deaccenting. Now, suppose that in presuppositional contexts bare CPs *must* be deaccented (cf. footnote 5), and suppose further that deaccenting comes with the feature

⁴²We find Roberts’s (2003) view that familiarity is a presupposition *conventionally* associated with definites problematic, at least when extended to the analysis of Russian NCs. In order to explain away apparent counterexamples involving unfamiliar nominals, such as (i), Roberts must resort to *presupposition accommodation*, which involves adding the presupposed content to the context (Stalnaker 1974, von Stechow 2008).

(i) (The department chair, speaking at a faculty meeting:)

At a college meeting yesterday, the Dean informed us of the possibility that the budget will be expanded still further for the next academic year. (Roberts 2003:302)

However, it is not clear to us how this can handle cases like (9) where the familiarity implication can be explicitly denied, as that should lead to a contradiction (cf. Abbott 2019:122).

⁴³Schlenker (2021) analyzes optional factive inferences of verbs like *announce* as presuppositions on his general view of presuppositions as possibly arising from contextual entailments (via some triggering algorithm). As far as we can tell, such a perspective is fully compatible with how we view the familiarity inference associated with Russian NCs.

[given/presupposed] (Kallulli 2006, 2010). On this view, what looks like the optionality of NCs in discourse-given contexts could in fact be obligatory blocking of bare nonpresuppositional CPs by either NCs or CPs with the feature [given/presupposed], in accordance with MP. As a consequence, the preference for bare (presuppositional) CPs over NCs in discourse-given contexts would no longer be problematic as long as the former is the more preferred variant for independent reasons. This account is of course rather sketchy and must be tested in future work, especially focusing on prosody.

To conclude, we have shown, based on experimental evidence from Russian, that although there is no strict correspondence between presuppositionality and nominalized clauses (i.e. presuppositionality is neither a necessary nor sufficient condition for nominalization), presuppositionality still plays a role in nominalization in the form of the preference hypothesis. While a formal theoretical account of the preference hypothesis still needs to be fleshed out, we hope that our empirical findings will be useful for the study of the syntax-semantic interface and will advance our understanding of the marking of presuppositionality.

Competing interests

The authors have no competing interests to declare.

References

- ABBOTT, B. 1999. Support for a Unique Theory of Definite Descriptions. *Proceedings of the Ninth Conference on Semantics and Linguistic Theory*. T. Matthews & D. Strolovitch (eds.), 1–15. Ithaca: Cornell University.
- ABBOTT, B. 2000. Presuppositions as nonassertions. *Journal of Pragmatics* 32.10, 1419–1437.
- ABBOTT, B. 2019. Definiteness and familiarity. *The Oxford Handbook of Reference*, ed. J. Gundel & B. Abbott (eds.), 117–129. Oxford University Press.
- ABRUSÁN, M. 2011. Predicting the presuppositions of soft triggers. *Linguistics and Philosophy* 34.6, 491–535.
- ADAMS, M. 1985. Government of empty subjects in factive clausal complements. *Linguistic Inquiry* 16.2, 305–313.
- BARR, D. J., LEVY, R., SCHEEPERS, C. & TILY, H.J. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68.3, 255–278.
- BATES, D., MAECHLER, M., BOLKER, B. & WALKER, S.. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67.1, 1–48.
- BHATT, R. 2010. Comments on “referential CPs and DPs: An operator movement account”. *Theoretical Linguistics* 36.2/3, 173–177.
- BOCHNAK, M. R. & EMILY A. HANINK, E.A. 2022. Clausal embedding in Washo: Complementation vs. modification. *Natural Language & Linguistic Theory* 40.4:979–1022.
- BOGAL-ALLBRITTEN, E. & MOULTON, K. 2018. Nominalized clauses and reference to propositional content. *Proceedings of Sinn und Bedeutung* 21, ed. R. Truswell, C. Cummins, C. Heycock, B. Rabern & H. Rohde, 215–232.

- BONDARENKO, T. 2020. Factivity from pre-existence: Evidence from Barguzin Buryat. *Glossa: A Journal of General Linguistics* 5.1, 1–35.
- BONDARENKO, T. 2022. Anatomy of an attitude. MIT. Ph.D. diss.
- CAPLAN, S. & K. DJÄRV. 2019. What usage can tell us about grammar: Embedded Verb Second in Scandinavian. *Glossa: a journal of general linguistics* 4.1, 1–37.
- CATTELL, R. 1978. The source of interrogative adverbs. *Language* 54.1, 61–77.
- DAVIES, W. & DUBINSKY, S. 1999. Functional structure and a parameter account of subject properties. *Proceedings of the 15th eastern states conference on linguistics*, ed. R. Daly and A. Riehl, 48–59. Ithaca: Cornell University, CLC Publications.
- DE CUBA, C. & ÜRÖGDI, B. 2010. Clearing up the ‘facts’ on complementation. *Proceedings of the 33rd Annual Penn Linguistics Colloquium (University of Pennsylvania Working Papers in Linguistics (PWPL))* 16, ed. J. S. Stevens, 41–50. Philadelphia, PA: ScholarlyCommons.
- DE CUBA, C. 2007. On (Non)Factivity, clausal complementation and the CP-field. Stony Brook University. Ph.D. diss.
- DEGEN, J. & TONHAUSER, J. 2022. Are there factive predicates? An empirical investigation. *Language* 98.3, 552–591.
- DJÄRV, K. 2019. Factive and Assertive Attitude Reports. The University of Pennsylvania. PhD diss.
- DJÄRV, K. 2021. Factivity, assertion, and clausal definiteness. *Proceedings of the 38th West Coast Conference on Formal Linguistics (WCCFL)*, ed. R. Soo, U. Y Chow, and S. Nederveen, 169–179. Somerville, MA: Cascadilla Proceedings Project.
- FEATHERSTON, S. 2005. The decathlon model of empirical syntax. *Linguistic evidence*, ed. M. Reis & S. Kepser, 187–208. Mouton de Gruyter.
- GIVÓN, T. 1979. *On understanding grammar*. New York: Academic Press.
- GLOVINSKAJA M. JA. Aktivnye processy v grammatike (na materiale innovacij i massovyx jazykovyx ošibok) [Active processes in the grammar (on the basis of innovations and widespread language errors)]. *Russkij jazyk konca XX stoletija (1985–1995)*, ed. E. A. Zemskaja, 237–302. Moscow: Jazyki russkoj kul'tury.
- GUNDEL, J.K., HEGARTY, M. & BORTHEN, K., 2003. Cognitive status, information structure, and pronominal reference to clausally introduced entities. *Journal of Logic, Language and Information* 12, 281–299.
- HAEGEMAN, L. & ÜRÖGDI, B. 2010. Referential CPs and DPs: An operator movement account. *Theoretical Linguistics* 36.2/3, 111–152.
- HAEGEMAN, L. 2006. Conditionals, factives and the left periphery. *Lingua* 116.10: 1651–1669.
- HALPERT, C. & SCHUELER, D. 2013. That sentential subjects are (linked to) DPs is explained herein. *Snippets* 28.
- HAN, J. H. 2005. A DP/NP-shell for subject CPs. *Proceedings of the 31st annual meeting of the Berkeley Linguistics Society: general session and parasession on prosodic variation and change*, ed. R. T. Cover and Y. Kim, 133–143. Berkeley: University of California, Berkeley Linguistics Society.
- HANKAMER, J. & MIKKELSEN, L. 2021. CP Complements to D. *Linguistic Inquiry* 52.3, 473–518.
- HARTMAN, J. 2012. Varieties of clausal complementation. MIT. PhD diss.
- HAWKINS, J. 1978. *Definiteness and indefiniteness: A study in reference and grammaticality prediction*. London: Croom Helm.

- HAWKINS, J. A. 2014. Cross-linguistic variation and efficiency. Oxford: Oxford University Press.
- HEGARTY, M. 1992. Familiar complements and their complementizers: On some determinants of A' – locality. Unpublished manuscript: University of Pennsylvania.
- HEIM, I. 1982. The Semantics of Definite and Indefinite Noun Phrases. University of Massachusetts Amherst. PhD diss.
- HEIM, I. 1991. Artikel und Definitheit. *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung*, ed. A. von Stechow and D. Wunderlich, 487–535. Berlin: de Gruyter.
- HEIM, I. 1992. Presupposition projection and the semantics of attitude verbs. *Journal of Semantics* 9, 183–221.
- HONCOOP, M. 1998. Dynamic excursions on weak islands. Leiden University. PhD diss.
- JARVIS, R. 2021. Presuppositionality and syntactic nominalization in finite clausal complements. *Proceedings of WCCFL 39*. [lingbuzz/006055](https://doi.org/10.1017/S0000000000000000)
- JENKS, P. 2018. Articulated definiteness without articles. *Linguistic Inquiry* 49.3, 501–536.
- KALLULI, D. 2006. Triggering factivity: Prosodic evidence for syntactic structure. *Proceedings of the 25th West Coast Conference on Formal Linguistics*, ed. D. Baumer, D. Montero and M. Scanlon, 211–219. Sommerville, MA: Cascadilla Proceedings Project.
- KALLULLI, D. 2010. Belief will create fact: on the relation between givenness and presupposition, and other remarks. *Theoretical Linguistics* 36.2/3, 199–208.
- KARTTUNEN, L. 1971. Some observations on factivity. *Papers in Linguistics* 4.1, 55–69.
- KASTNER, I. 2015. Factivity mirrors interpretation: The selectional requirements of presuppositional verbs. *Lingua* 164, 156–188.
- KHOMITSEVICH, O. 2007. Dependencies across Phases: From sequence of tense to restrictions on movement. Utrecht University. PhD diss.
- KIPARSKY, C. & KIPARSKY, P. 1970. Fact. *Progress in Linguistics*, ed. M. Bierwisch & K. E. Heidolph, 143–173. The Hague: Mouton.
- KNYAZEV, M. 2016. Licensing clausal complements. The case of Russian čto-clauses. Utrecht University. PhD diss.
- KNYAZEV, M. 2022. Spanning complement-taking verbs and spanning complementizers: On the realization of presuppositional clauses. *Journal of Linguistics* 58.4, 887–899. doi:10.1017/S0022226722000160
- KNYAZEV, M. & USTINOVA, I. 2023. The role of case and presuppositionality in the distribution of *to čto*-clauses in colloquial Russian: an experimental study. *Russian Linguistics* 47.1, 87–121. <https://doi.org/10.1007/s11185-023-09271-2>
- KOBOZEVA, I. M. 2013. Uslovija upotreblenija «to» pered pridatočnym iz 'jasnitel'nyh s sojuzom «čto» [Conditions for the use of “to” before explicative clauses with the subordinate conjunction “čto”]. *Du mot au texte. Études slavo-romanes*, ed. O. Inkova, 129–148. Bern: Peter Lang.
- KUZNETSOVA, A., BROCKHOFF, P. B. & CHRISTENSEN, R. H. B. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82.13, 1–26.
- LYUTIKOVA, E. & S. TATEVOSOV. 2019. On the Structure of Russian Relative Clauses. *Zeitschrift für Slavische Philologie* 75.1, 65–94.

- MELVOLD, J. 1991. Factivity and definiteness. *MIT Working Papers in Linguistics* 15, 97–117.
- MOULTON, K. 2015. CPs: Copies and compositionality. *Linguistic Inquiry* 46.2. 305–342.
- MOULTON, K. 2020. Remarks on propositional nominalization. *Nominalizations: 50 Years on from Chomsky's Remarks*, ed. H. Borer and A. Alexiadou, 255–276. Oxford: Oxford University Press.
- MOULTON, K., BOGAL-ALLBRITTEN, E., & SHIMOYAMA, J. 2020. Things we embed. Talk given at the *Brussels Conference on Generative Linguistics (BCGL)* 13. <http://individual.utoronto.ca/keirmoulton/MoultonBCGL.pdf>.
- ROBERTS, C. 2003. Uniqueness in definite noun phrases. *Linguistics and Philosophy* 26.3: 287–350.
- ROCHEMONT, M. 2016. Givenness. *The Oxford handbook of information structure*, ed. C. Féry & S. Ishihara, 41–63. Oxford: Oxford University Press.
- ROUSSOU, A. 1994. The syntax of complementisers. University College London. Ph.D. diss.
- SAUERLAND, U. 2008. Implicated presuppositions. *Sentence and Context: Language, context, and cognition*, ed. A. Steube, 1–28. Berlin: de Gruyter.
- SCHLENKER, P. 2021. Triggering Presuppositions. *Glossa: a journal of general linguistics* 6.1. doi: <https://doi.org/10.5334/gjgl.1352>
- SCHWARZSCHILD, R. 1999. GIVENNESS, AVOIDF and other constraints on the placement of accent. *Natural Language Semantics* 7.2, 141–177.
- SHEEHAN, M. & HINZEN, W. 2011. Moving towards the edge. *Linguistic Analysis* 37.3/4, 405–458.
- SIMONS, M. 2007. Observations on embedding verbs, evidentiality, and presupposition. *Lingua* 117.6, 1034–1056
- STALNAKER, R. 1974. Pragmatic presuppositions. *Semantics and philosophy*, ed. M. Munitz & P. Unger, 197–213. New York: New York University Press.
- TAKAHASHI, S. 2010. The hidden side of clausal complements. *Natural Language and Linguistic Theory* 28.2, 343–380.
- VON FINTEL, K. 2008. What is presupposition accommodation again? *Philosophical Perspectives* 22.1, 137–170.
- ZALIZNJAK, A. A. 1988. O ponjatii implikativnogo tipa (dlja glagolov s propozicionalnym aktantom) [On the notion of implicative type (for verbs with propositional arguments)]. *Logičeskij analiz jazyka. Znanie i mnenie* [Logical analysis of language: Knowledge and opinion], ed. N.D. Arutjunova, 107–121. Moscow: Nauka.
- ZUBIZARRETA, M. L. 1982. Theoretical implications of subject extraction in Portuguese. *Linguistic Review* 2.1, 79–96.

Supporting information

Appendix A: <https://osf.io/9gqrh>

Appendix B: <https://osf.io/pxvhb>

Appendix C: <https://osf.io/4g9hk>