

What micro-diachronic analysis can tell us about unmarked indefinites: Evidence from Russian

© 2023

Olga E. Pekelis

Russian State University for the Humanities, Moscow, Russia; HSE University, Moscow, Russia;
opekelis@gmail.com

Abstract: The paper deals with the evolution of Russian unmarked indefinite pronouns, also known as bare interrogatives, since the 18th century. The Russian National Corpus data suggest that the distribution of unmarked indefinites both in modern Russian and in the Russian language of the 18th–19th centuries is not in line with previous proposals. Surprisingly, unmarked indefinites are more frequent in some specific contexts compared to non-specific contexts such as imperatives or subjunctive clauses. I argue that the distribution of Russian unmarked indefinites can be accounted for by some sort of economy principle: due to the lack of a marker of indefiniteness, unmarked indefinites tend to be used in contexts that are strongly biased toward a certain interpretation of an indefinite in terms of specificity (specific known vs. specific unknown vs. non-specific). This analysis, as I suggest, correctly predicts two typological tendencies reported on unmarked indefinites in the literature.

Keywords: corpus linguistics, diachrony, pronouns, indefiniteness, Russian

Acknowledgements: The work has been supported by the Ministry of Science and Higher Education of the Russian Federation within Agreement No. 075-15-2020-793. I am profoundly thankful to Ekaterina Rakhilina, Yuri Lander, and the anonymous reviewers of *Voprosy Jazykoznanija* for their insightful comments and corrections. All errors remain mine.

For citation: Pekelis O. E. What micro-diachronic analysis can tell us about unmarked indefinites: Evidence from Russian. *Voprosy Jazykoznanija*, 2023, 1: 19–53.

DOI: 10.31857/0373-658X.2023.1.19-53

Неопределенные местоимения без маркера неопределенности через призму микродиахронического анализа: данные русского языка

Ольга Евгеньевна Пекелис

Российский государственный гуманитарный университет, Москва, Россия;
Национальный исследовательский университет «Высшая школа экономики»;
Москва, Россия; opekelis@gmail.com

Аннотация: Статья посвящена эволюции неопределенных местоимений без маркера неопределенности в русском языке с XVIII в. до наших дней. Данные Национального корпуса русского языка свидетельствуют о том, что дистрибуция таких местоимений и в современном языке, и в языке XVIII–XIX вв. не отвечает предлагавшимся ранее описаниям. Вопреки ожиданиям, местоимения без маркера более частотны в некоторых референтных контекстах, чем в таких нереферентных контекстах, как клаузы с императивом и с союзом *чтобы*. Дистрибуцию местоимений без маркера предлагается объяснять действием своего рода принципа экономии: в отсутствие маркера немаркированные местоимения тяготеют к контекстам, которые навязывают местоимению тот или иной тип референтности (референтные слабоопределенные vs. референтные неопределенные vs. нереферентные). Демонстрируется, что такой подход позволяет учесть две типологические тенденции, известные из литературы о неопределенных местоимениях.

Ключевые слова: диахрония, корпусная лингвистика, местоимения, неопределенность, русский язык

Благодарности: Работа выполнена при поддержке гранта Министерства науки и высшего образования № 075-15-2020-793 «Компьютерно-лингвистическая платформа нового поколения для цифровой документации русского языка: инфраструктура, ресурсы, научные исследования». Автор глубоко признателен Е. В. Рахилиной и Ю. А. Ландеру, а также рецензентам журнала «Вопросы языкознания» за советы и замечания.

Для цитирования: Pekelis O. E. What micro-diachronic analysis can tell us about unmarked indefinites: Evidence from Russian. *Voprosy Jazykoznanija*, 2023, 1: 19–53.

DOI: 10.31857/0373-658X.2023.1.19-53

1. Introduction

With *unmarked indefinites* (more commonly known as *bare interrogatives*, see [Haspelmath 1997: 170 ff.]), I refer to indefinite pronouns that lack an overt marker of indefiniteness and are formally identical to interrogative pronouns. Cf. an instance of the Russian indefinite pronoun *kto* ‘someone’, formally identical to the interrogative *kto* ‘who’, within a *yes/no*-question in (1).

- (1) *Už ne umer li kto?*
 PTCL NEG die.PST.SG.M Q someone.NOM
 ‘Has someone died?’ [RNC]

According to Tret'yakova [2004; 2009a; b], Russian unmarked indefinites (henceforth UIs) are allowed in questions (as in (1)), conditionals, disjunctive ‘or’-clauses, purpose clauses, clauses with imperatives and several other non-specific contexts, spanning the ‘irrealis non-specific’, ‘question’ and ‘conditional’ functions on the semantic map of indefinite pronouns proposed by Haspelmath [1997]; cf. Fig. 1.

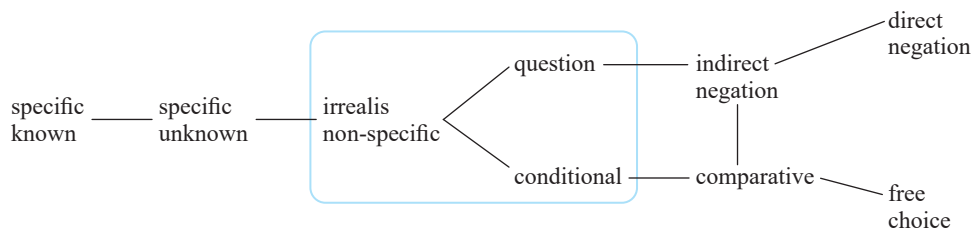


Figure 1. Boundaries of the Russian UIs according to Tret'yakova [2004; 2009a; b], on the semantic map by Haspelmath [1997]

The Russian National Corpus [RNC] data suggest, however, that the presentation in Fig. 1 is imprecise. It turns out that in present-day Russian, UIs are, with few exceptions, absent from the ‘irrealis non-specific’ contexts, including imperatives and purpose clauses, but are admissible in a few types of ‘specific unknown’ contexts. Furthermore, in the Russian language of the 18th–19th centuries UIs appear to have been used more extensively than they are today. In particular, they were allowed in ‘irrealis non-specific’ contexts from which they are almost excluded in modern Russian.

Based on both synchronic and diachronic corpus data, the main questions that I address in this paper are as follows: what factors lay behind the evolution of Russian UIs since the 18th century and why did UIs evolve in this particular way, i.e. toward the elimination of ‘irrealis non-specific’ contexts? I argue that the scenario revealed by the corpus data can be accounted for by some sort of economy principle, according to which UIs largely rely on context in the type of indefiniteness

they express in terms of specificity ('specific known' vs. 'specific unknown' vs. 'non-specific'). In this regard, UIs differ from marked indefinite pronouns, for which it is the marker itself that serves to express the type of indefiniteness in the first place. Furthermore, I suggest that those contexts in which UIs are rare or absent in modern Russian (e.g., imperatives and subjunctive clauses) are less clearly biased toward a certain interpretation of an indefinite pronoun than they include than those contexts in which UIs are relatively frequent (e.g., conditionals and questions).

The structure of my argument is twofold. First, I show that the analysis I propose allows one to explain several facts concerning the distribution of UIs both in present-day Russian and in the Russian language of the 18th–19th centuries. Second, I suggest that this analysis offers an account of several typological tendencies in the use of UIs.

UIs are not widely used in modern Russian — they are a rather marginal alternative to several series of pronouns with a marker of indefiniteness. Therefore, I evaluated the frequency of UIs not by itself but as compared to the frequency of marked indefinites. There are eight major markers: *koe-*, *-to*, *-nibud'*, *-libo*, *by to ni bylo*, *ni-*, *ljuboj* and *ugodno*. Except for *ljuboj*, they attach to interrogative pronouns of different ontological categories, cf. *kto* 'who' vs. *kto-to* 'somebody', *čto* 'what' vs. *čto-nibud'* 'anything' etc. The distribution of the markers on the semantic map of indefinite pronouns is as in Fig. 2, according to [Haspelmath 1997: 65].

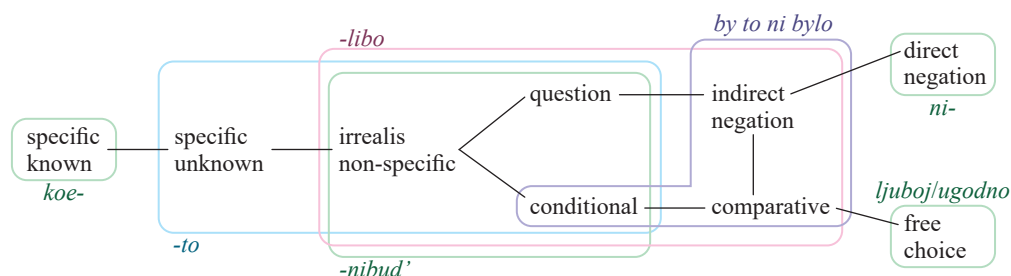


Figure 2. Boundaries of Russian marked indefinites according to Haspelmath [1997: 65]

Comparing Fig. 2 with Fig. 1 and my own corpus data on UIs suggests that UIs overlap in their distribution with *to-*, *nibud'*-, *libo*- and (in conditionals) *by to ni bylo*-series. However, *by to ni bylo* is in fact very marginal in conditionals in modern Russian: only nine instances of a *by to ni bylo* indefinite immediately following *esli* 'if' (the major conditional subordinator) have been attested in the Main corpus of the RNC, seven of which date to before 1950.¹ *By to ni bylo* series is also excluded from the 'conditional' function in the semantic map of Russian indefinite pronouns proposed by Tatevosov [2002: 141], who elaborates on the map by Haspelmath.² Therefore, in the corpus part of my study, I did not take into account the distribution of *by to ni bylo*, i.e. I assessed the frequency of UIs in different contexts as compared to the total frequency of *to-*, *nibud'*- and *libo*-series in the same contexts.

Throughout the corpus data analysis, I considered as UIs not only bare interrogatives as such, but also combinations thereof with the words like *drugoj / inoj* 'another' or *eščë* 'else'. Although, as noted by Tret'yakova [2009a: 30 ff.], such combinations may be qualified as series of indefinite pronouns on their own, bare interrogatives within them can always be substituted with a marked pronoun, cf. (2). Since in my corpus study, I compared pronouns with and without a marker in frequency, the contexts in which both types of pronouns are admissible have all been included in the sample. For the same reason, I excluded combinations of bare interrogatives with

¹ The following query was used to retrieve the data: **если на расстоянии 2 от бы на расстоянии 1 от кто на расстоянии 1 от ни на расстоянии 1 от было**.

² Tatevosov [2002: 141] introduced a few further modifications of the map in Fig. 2. I do not discuss them here since they do not regard the functions that are involved in the distribution of UIs.

quantificational words like *redko* ‘rarely’ or *malo* ‘few’ from the analysis. They not only convey their own (quantificational) semantics [Tatevosov 2002: 138] but are also not interchangeable with indefinite pronouns bearing a marker [Tret’yakova 2009a: 29], cf. (3).

- (2) *Osinskij ili kto* (^{OK}*kto-to*) *drugoj podderžal Axmatovu,*
 Osinsky.NOM or someone.NOM who.NOM-INDF else support.PST.M.SG Akhmatova
ne pomnju.
 NEG remember.PRS.1SG
 ‘Was it Osinsky who supported Akhmatova, or someone else — I don’t remember.’ [RNC, 1985–2002]³

- (3) *Malo kto* (**kto-nibud’*, **kto-to*, **kto-libo*)
 few someone.NOM who.NOM-INDF who.NOM-INDF who.NOM-INDF
ljubit kino tak kak ja.
 love.PRS.3SG cinema so as I.NOM
 ‘Few people love cinema as much as I do.’ [RNC, 1988]

The RNC consists of several subcorpora. Along with its Main corpus, which includes predominantly literary prose from the 18th to 21st centuries, the Newspaper corpus was also used. It contains media texts from the 1990–2000s, which may be assumed to reflect the modern usage more accurately than the fiction texts, the latter being more biased toward imitation of the older usage (see more on the use of the Newspaper corpus in Section 3.1). The methodology and the study in general are part of a research project on the Russian language of the 19th century, implemented at the National Research University Higher School of Economics (see [Rakhilina et al. 2016]).

A final caveat on terminology is in order. Although, as mentioned above, “bare interrogatives” is a wide-spread term for the phenomenon under discussion, I prefer the term “unmarked indefinites” because I focus on comparing indefinites with and without a marker. Since in Russian both marked and unmarked indefinites go back to interrogative pronouns (with the only exception being the free-choice marker *ljuboj*), whether the marker is present or absent seems to be the most crucial point, which is thus reflected in the term.

The rest of the paper is structured as follows. First, in Section 2 I consider previous proposals on the distribution of Russian UIs. Then, in Sections 3 and 4 I present the RNC data concerning the distribution of UIs across contexts and with respect to some other factors, namely the ontological category of the pronoun, register and emphasis, that have been assumed to be relevant for Russian UIs. In Section 5, I show that the boundaries of UIs on the semantic map differ from those suggested in the previous works both for present-day Russian and the Russian language of the 18th–19th centuries, and put forward several hypotheses to provide an account for the facts observed. In Section 6, I discuss typological consequences of my account. Section 7 concludes.

2. Distribution of Russian UIs: previous proposals

According to Tret’yakova [2004; 2009a: 120 ff.], the following contexts are compatible with UIs in Russian. In all these contexts, the UI is interchangeable with a “marked” indefinite of at least one series (*-to*, *-nibud’*, or *-libo*).

— Both independent and embedded *yes/no*-questions (4), including those in which the speaker expects a positive answer (5) (on the relevance of the speaker’s expectations for the choice of an indefinite pronoun, see Section 5.1).

³ Here and below, the number next to the source of the example indicates the year or period the text was created.

- (4) *Kraem glaza ja staralsja uvidet',*
 corner.INS eye.GEN I.NOM try.PST.M.SG see.INF
net li kogo poblizosti ot mašiny.
 be.NEG.PRS Q anyone.GEN near from car.GEN
 'Out of the corner of my eye, I tried to see if anyone was near the car.' [RNC, 2001]

- (5) *Molites' za upokoj duši.*
 pray.IMP.2PL for rest.ACC soul.GEN
 — *Umer kto? — sprosila niščenka, vzdrognuv.*
 die.PST.M.SG someone.NOM ask.PST.F.SG beggar.woman.NOM start.CNV
 'Pray for the rest of the soul. "Did someone die?" asked the beggar woman with a start.'
 [RNC, 1997]

— Disjunctive contexts (6).

- (6) *Najdutsja pokupateli, — prodam, a net,*
 happen.to.be.FUT.3PL buyer.NOM.PL sell.FUT.1SG and no
tak ostavlju ili podarju komu.
 then keep.FUT.1SG or give.FUT.1SG someone.DAT
 'If there will be buyers, I will sell it, but if not, I will keep it for myself or give it to someone.' [RNC, 2003]

— Optative contexts (7).

- (7) *Xot' by knigi kakie doma byli, ènciklopedii ...*
 PTCL SBJV book.NOM.PL some.NOM.PL at.home be.SBJV.PL encyclopaedia.NOM.PL
 'If only there were some books at home, encyclopaedias...' [RNC, 2009]

— Contexts with the subordinators *budto*, *kak budto*, *točno* and *slovno*, which express the simulative 'as if' meaning (8).

- (8) *Vnezapno As'ka povernula golovu, budto kakoj zvuk uslyšala.*
 suddenly As'ka.NOM turn.PST.F.SG head.ACC as.if some.ACC sound.ACC hear.PST.F.SG
 'Suddenly, As'ka turned her head as if she heard some sound.' [RNC, 2015]

— The protasis of conditionals (9).

- (9) *Esli kakie problemy s podružkami,*
 if some.NOM.PL problem.NOM.PL with girlfriend.INS.PL
ona nam tože govorit.
 she.NOM we.DAT also tell.PRS.3SG
 'If there are any problems with girlfriends, she also tells us.' [RNC, 2004]

— Purpose clauses, typically under negation within the embedded or superordinate clause.

- (10) *Mne nužno bylo rjadom s nim naxodit'sja,*
 I.DAT necessary COP.PST near with he.INS be.INF
čtoby čego ne slučilos'.
 so.that anything.GEN NEG happen.SBJV.N.SG
 'I needed to be with him so that nothing happened.' [RNC, 2002]

— Imperative contexts, typically under negation.

- (11) *Ty tam ne izmaž' čego.*
 you.NOM there NEG smear.IMV.2SG anything.GEN
 'Don't smear anything there.' [RNC, 1998]

— Contexts with epistemic modals *možet* ‘maybe’, *vrjad li* ‘hardly’ and some others.

- (12) *Možet, čto putnoe i vyjdet iz tebjja.*
 maybe something.NOM worthwhile PTCL come.out.FUT.3SG from you.GEN
 ‘Maybe something worthwhile will come out of you.’ [RNC, 2015]

Oskol’skaya and Zevakhina [2013] suggest that Russian UIs are allowed in a much wider range of contexts. They mention both direct and indirect negation, as well as comparative and free-choice functions from the semantic map in Fig. 2 as compatible with UIs. However, the indirect negation use of UIs is illustrated with examples as in (13), where a subjunctive complement clause is embedded under a negated matrix predicate. For Russian at least, such cases may be attributed to the ‘irrealis non-specific’ function with no less or even more reason than to ‘indirect negation’, since the pronouns of the *by to ni bylo* series, which are common in the latter function (cf. Fig. 2), can be used in such contexts only marginally.

- (13) *Ne pomnju, čto by baby kogo*
 NEG remember.PRS.1SG COMP woman.NOM.PL someone.ACC
 (*“kogo by to ni bylo”*) *xoronili...*
 anyone bury.PST.PL
 ‘I don’t remember that women buried someone.’ [RNC, 2006]

The direct negation function is exemplified with the depreciative, hence stylistically marked, use of the UIs, as in (14).⁴

- (14) *Ja že ne zver’ kakoj...*
 I.NOM PTCL NEG beast.NOM some.NOM
 ‘I’m not some kind of a beast.’ [RNC, 1999]

Furthermore, both in comparative and free-choice functions UIs can only be used in combination with another marker: *drugoj* or *inoj* in comparatives (15) and *ugodno* in free-choice contexts (16). This is in contrast with, for example, conditionals, in which UIs can be used both with and without *drugoj* or *inoj*.

- (15) *V laskax Zininyx videlos’ emu bol’she žalosti,*
 in caress.LOC.PL of.Zina.LOC.PL see.PASS.PST he.DAT more pity.GEN
*čem čego *(drugogo).*
 than anything.GEN other.GEN
 ‘In Zina’s caresses he saw more pity than anything else.’ [RNC, 1979]
- (16) *Vojti sjuda mog kto *(ugodno).*
 enter.INF here can.PST.M.SG who.NOM any
 ‘Anyone could enter here.’ [RNC, 2001]

⁴ According to an anonymous reviewer, imperatives under negation could also be associated with the ‘direct negation’ function. However, a UI as well as a *nibud*’-pronoun used with a negated imperative do not seem to be in the scope of negation, contrary to pronouns of the *ni*-series, which are typically used in the ‘direct negation’ function (cf. Fig. 2). The following examples are illustrative: (ii) is a negation of (i), but (iii) is not. For a reason that I do not discuss here, (iii) does not allow an interpretation in which the UI or the *nibud*’-pronoun are out of the scope of negation; hence, (iii) is infelicitous.

- (i) *Slušaj kakie-nibud’ pesni.*
 listen.IMV.SG which.ACC.PL-INDF song.ACC.PL
 ‘Listen to some kind of songs.’ (Google)
- (ii) *Ne slušaj nikakix pesen.*
 NEG listen.IMV.SG no.GEN.PL song.GEN.PL
 ‘Don’t listen to any songs.’
- (iii) *“Ne slušaj kakix-nibud’/ kakix pesen.*
 NEG listen.IMV.SG which.GEN.PL-INDF some.GEN.PL song.GEN.PL
 Intended meaning: ‘Don’t listen to any songs.’

Thus, the contexts listed by Oskol'skaya and Zevakhina [2013] but excluded by Tret'yakova [2004; 2009a] are marginal for Russian UIs in one way or another. For this reason, they will not be considered in my corpus study (see Sections 3 and 4).

It is also noteworthy that not all 'irrealis non-specific' contexts, which are basic for Russian UIs according to Tret'yakova [Ibid.], are compatible with UIs. Those that are not include future contexts (17) and modal contexts expressing necessity, permission, or obligation (18) [Oskol'skaya, Zevakhina 2013].

- (17) *On najdët kakoj-nibud' (kakoj) sposob ispravit' položenie.*
 he.NOM find.FUT.3SG which.ACC-INDF some.ACC way.ACC rectify.INF situation.ACC
 'He will find some way to rectify the situation.' [RNC, 2003]

- (18) *Zdes' dolžen byt' kakoj-nibud' (kakoj) zapasnoj vyxod.*
 here must be.INF which.NOM-INDF some.NOM emergency.NOM exit.NOM
 'There must be some kind of emergency exit.' [RNC, 1997]

A comment is due on the source of the data provided by Tret'yakova [2004; 2009a] and Zevakhina and Oskol'skaya [2013]. Both studies are based on corpus data (Zevakhina and Oskol'skaya use the RNC, Tret'yakova uses three smaller corpuses of Russian). However, none of them draws a line between the 19th—early 20th century data and modern Russian, considering the 19th-century data to be representative of the distribution of UIs in modern Russian. I suppose this to be the crucial reason for the discrepancies between their results and my own corpus data presented in the next sections.

Along with the type of the context discussed so far, the following three factors have been mentioned in the literature as possibly relevant to the distribution of UIs in Russian and/or cross-linguistically: ontological category of the pronoun, register (colloquial vs. formal), and emphasis.

The role of the ontological category of the pronoun may apparently be generalized as follows: UIs that belong to an ontological category associated with adjuncts (place, time, reason, purpose, etc.) are rarer than UIs that, based on their ontological category, are associated with arguments. According to Belyaev and Haug [2014: 4], many languages do not allow indefinite readings of counterparts to *why* and *how*. Tret'yakova [2009a: 150–158] argues for Russian that the interrogatives *kogda* 'when' and *gde* 'where' are restricted in use as UIs compared to *kto* 'who' and *čto* 'what'.

Furthermore, it has been assumed that Russian UIs are only used in colloquial texts or fiction imitating the spoken discourse ([Yanko-Trinititskaya 1977: 73; Kuz'mina 1989: 189], inter alia). However, this view is disputed by Tret'yakova [2004]. Based on a corpus study, she argues that UIs are more frequent in written narrative discourse than in direct-speech contexts in the texts of both 19th and 20th centuries.

Finally, emphasis may serve as a further factor in Russian that makes the use of a UI more appropriate. As Tret'yakova [2009a: 156] observes, *kogda* 'when' is used as a UI very rarely and must be accompanied by an emphatic particle, such as *i* in (19):

- (19) *Možet, kogda i otkrojut formulu ljubvi.*
 maybe someday PTCL discover.FUT.3PL formula.ACC love.GEN
 'Maybe someday the formula of love will be discovered.' [Tret'yakova 2009a: 156]

3. Distribution of UIs across contexts according to the RNC data

In this section, two UIs, *kto* 'somebody / anybody' and *kakoj* 'some / any', are analysed based on the RNC data. I compare their frequency in different contexts in three subcorpora of the RNC: the Newspaper corpus, the 20th-century corpus, and the 18th–19th-century corpus. *Kto* is used

most typically as a subject or object and is thus expected to be more frequent than UIs associated with the position of adjuncts (see Section 2). *Kakoj* may be an (adjunct) constituent within the subject or object phrase. It also differs from *kto* in that the former is an adjective phrase, while the latter is a noun phrase. *Kakoj* was thus chosen as a UI that is maximally contrasting to *kto* but at the same time is not an undisputed adjunct.⁵

I consider *kto* in Section 3.1 and *kakoj* in Section 3.2. Section 3.3 concludes.

3.1. *Kto*

As the first step, I analysed the frequency of *kto* compared to indefinite pronouns of the *ni-bud'*-, *to*- and *libo*-series in the Newspaper corpus of the RNC. I considered all types of contexts listed by Tret'yakova [2004; 2009a] as compatible with UIs in present-day Russian (see Section 2):

- the protasis of conditionals introduced by the subordinator *esli* 'if';
- *yes/no*-questions marked by the question particle *li* 'whether';
- imperatives;
- contexts with the subordinator *čtoby*, which introduces both purpose clauses and subjunctive complement clauses (hereinafter, both are referred to as subjunctive clauses);
- disjunctive contexts with *ili* 'or';
- contexts with the markers *kak*, *kak budto*, *slovnno*, and *točno*, which most commonly introduce simulative 'as if'-clauses but may also introduce complement clauses, in which case they convey the unreliability of what is being reported (see more details in Section 5.3); for simplicity, I will refer to these clauses as simulative in Tables 1–8, meaning both simulative and complement clauses;⁶
- epistemic contexts with the marker *možet* 'maybe';
- and contexts with the optative marker *xot'by* 'if only'.

The data on UIs extracted from the Newspaper corpus are presented in Table 1. There are three contexts in which *kto* as a UI is rare or absent: imperatives, subjunctive, and disjunctive clauses. In other contexts, the ratio of the UI is about 0.2 or higher, with the optative contexts having the highest ratio of about 0.5.

As a next step, I selected five contexts from my list, with which all further work continued. These are conditional, interrogative, imperative, subjunctive, and simulative contexts. Thus, three contexts, namely the optative *xot'by*, the disjunctive *ili* and the epistemic *možet*, were excluded from further work (however, both *xot'by* and *možet* are considered in Sections 4.2 and 5.3 with respect to the role of register). This was done to reduce the sample size but also for some more meaningful reasons. *Xot'by* and *možet* were excluded as contexts with lexical rather than grammatical markers, which, as discussed in Section 5.3, are not indicative of optative and epistemic contexts in general. Note also that the sample with *xot'by* is very small, and hence statistically not very useful. Disjunctive *ili*-contexts were excluded because in my sample with *ili*, many examples contain, along with *ili*, at least

⁵ The pronoun *čto* 'something/anything' was not considered simply to limit the sample size, which could not be too big since most of the data were filtered manually.

⁶ *Kak*, *kak budto*, *slovnno*, and *točno* may also be used as particles in a monoclausal construction, cf. (i). I considered such examples as simulative, since in that particular meaning in which they occur in my sample, they are semantically very close to simulative constructions.

(i) *Vdrug menja budto kto tolknul pod bok.*
suddenly I.ACC as.if someone.NOM push.PST.M.SG under side.ACC
'Suddenly, it was like someone hit me in the side.' [RNC, 2001]

Table 1

**Frequency of *kto* as a marked and unmarked indefinite
in the Newspaper corpus of the RNC⁷**

Context type	Marked	Unmarked	Ratio of unmarked
optative (<i>xot'by</i>)	19	22	0.537
yes/no-question (<i>li</i>)	483	163	0.350
conditional (<i>esli</i>)	3,470	994	0.223
epistemic (<i>možet</i>)	401	103	0.204
simulative (<i>kak</i> etc.)	96	22	0.186
disjunctive (<i>ili</i>)	1,067	56	0.050
subjunctive (<i>čtoby</i>)	917	7	0.008
imperative	109	0	0

one more feature associated with UIs (e.g., the question illocutionary force or an epistemic modal), as in (20).

- (20) *Sami* *vyrezali* *ili* *kto* *dal?*
 oneself.NOM.PL cut.PST.PL or someone.NOM give.PST.M.SG

‘Did you cut it yourself or did someone give it to you?’ [RNC, 2010]

In Table 2 (p. 28), the comparative frequency of the marked and unmarked *kto* in five selected contexts is given for the 20th century and the 18th–19th centuries.⁸ Note particularly that the frequency of *kto* in imperative and subjunctive clauses, in which *kto* is rare or absent in modern newspaper texts, drops consistently from the 18th–19th centuries to the 21st century. For both contexts, the difference between the Newspaper corpus and the 20th-century subcorpus, and between the Newspaper corpus and the 18th–19th-century subcorpus is statistically significant (χ^2 , df = 1, $p < 0.01$). The difference between the 18th–19th centuries and the 20th century is significant for subjunctive clauses (χ^2 , df = 1, $p < 0.01$), while for imperatives it is not. For clarity, the corresponding data are presented separately in Tables 3 and 4 (p. 28).

⁷ Here and below, all examples were filtered manually except for the contexts in which more than 1,000 examples were retrieved from the corpus; in this case, roughly the first thousand examples were filtered. Importantly, all samples over 1,000 examples contained few irrelevant examples (e.g., there were 12 irrelevant examples in the first thousand examples with *kto* within conditionals retrieved from the Newspaper corpus). The following types of queries were used to obtain the data in Table 1 as well as the similar data below: **если** на расстоянии 1 от **кто** -amark; **-вряд -»мало» -едва -»то»** на расстоянии 1 от **ли** на расстоянии 1 от **кто** -amark; **Vimper** на расстоянии 1 от **кто** -amark; **чтобы** на расстоянии 1 от **кто** -amark; **или** на расстоянии 2 от **кто** -amark; **будто | словно | «точно»** на расстоянии 1 от **кто** -amark; **«может»** на расстоянии 1 от **кто** на расстоянии 1 от **-угодно; «хоть»** на расстоянии 1 от **бы** -amark на расстоянии 1 от **кто** -amark.

⁸ As noted by an anonymous reviewer, considering the 18th and the 19th centuries as a single time period is undesirable. Although I fully agree with this, analyzing them as separate periods would be problematic, since the 18th-century subcorpus is substantially smaller than that of the 19th century, and the data for the 18th century alone would hardly be representative. The investigation of the differences in the use of UIs between the 18th and the 19th centuries are thus left for the future.

Table 2

**Frequency of *kto* as a marked and unmarked indefinite in 1701–1900 and 1901–2000
according to the Main corpus of the RNC**

Context type	1701–1900			1901–2000		
	Marked	Unmarked	Ratio of unmarked	Marked	Unmarked	Ratio of unmarked
conditional (<i>esli</i>)	212	831	0.804	1,222	1,231	0.502
yes/no-question (<i>li</i>)	160	469	0.746	418	550	0.568
imperative	82	13	0.137	256	23	0.082
subjunctive (<i>čtoby</i>)	253	77	0.233	766	42	0.052
simulative (<i>kak</i> etc.)	180	212	0.541	1,028	277	0.212

Table 3

**Frequency of *kto* as a marked and unmarked indefinite in imperative clauses
in different time periods**

Subcorpus	Marked	Unmarked	Ratio of unmarked
Newspaper corpus (1990s–2000s)	109	0	0
1901–2000	256	23	0.082
1701–1900	82	13	0.137

Table 4

**Frequency of *kto* as a marked and unmarked indefinite in subjunctive clauses
in different time periods**

Subcorpus	Marked	Unmarked	Ratio of unmarked
Newspaper corpus (1990s–2000s)	917	7	0.008
1901–2000	766	42	0.052
1701–1900	253	77	0.233

The data in Tables 1–2 show two more trends:

- There is a consistent drop in frequency of *kto* from the 18th–19th centuries to the 21st century in all contexts, imperatives and subjunctive clauses thus being part of a broader tendency.
- Imperatives and subjunctive clauses include *kto* the least frequently compared to other contexts not only in the Newspaper corpus, but in all time periods.

Crucially, the fact that imperatives and subjunctive clauses are almost out of use in present-day Russian can be seen as a logical consequence of these two trends. Indeed, these clauses included UIs rarer than other contexts already in the 18th century; that is why they were the first to stop using them.

As alluded to in the Introduction, the Newspaper corpus data were included in this research as I assumed them to represent the contemporary norm more adequately than the 20th-century subcorpus. This choice might be questioned since Russian UIs have been claimed to be limited to the spoken discourse, which is only sporadically present in the Newspaper corpus (e.g., in interviews). However, as mentioned in Section 2, the corpus study by Tret'yakova [2004] suggests that this is not the case. And even it were, there are at least two reasons to assume that considering the Newspaper corpus data is justified. First, the drop in frequency of UIs is a general tendency that can also be observed on the data of the Main corpus of the RNC alone

(although the latter reveal this tendency less consistently than the Newspaper corpus). Second, the frequency of UIs in the Newspaper corpus is different for different contexts: e.g., in conditionals UIs are more frequent than in imperatives. Importantly, the differences in frequency of UIs in different contexts are roughly the same in the Newspaper corpus as in the 20th and 18th–19th-century corpora. This suggests that the factors underlying the use of UIs are roughly the same in all three corpora.

3.2. *Kakoj*

The corpus data on *kakoj*, given in Table 5 for the Newspaper corpus, the 20th-century, and the 18th–19th-century corpora, show that in all time periods, *kakoj* as a UI is rarer than *kto*. But in other respects, the data on *kakoj* suggest roughly the same conclusions as the data on *kto*. Since the 18th century, the frequency of *kakoj* declined consistently in all five contexts. For the imperative and subjunctive contexts, it is also important that *kakoj* was attested both in the 18th–19th and 20th-century corpora but is absent in the Newspaper corpus. The difference between the Newspaper corpus and the 20th-century corpus is statistically significant both for imperatives and subjunctive clauses (2-tailed Fisher exact test, $P < 0.01$), as well as the difference between the Newspaper corpus and the 18th–19th-century corpus (2-tailed Fisher exact test, $P < 0.01$ for subjunctive clauses and $P < 0.05$ for imperatives). It is somewhat unexpected that *kakoj* is more frequent in imperatives in 1901–2000 than in 1701–1900. However, the difference is not statistically significant (2-tailed Fisher exact test, $P = 0.8$).

Table 5

Frequency of *kakoj* as a marked and unmarked indefinite in different subcorpora of the RNC

Context type	1701–1900			1901–2000			Newspaper corpus (1990s–2000s)		
	Marked	Unmarked	Ratio of unmarked	Marked	Unmarked	Ratio of unmarked	Marked	Unmarked	Ratio of unmarked
conditional (<i>esli</i>)	119	33	0.217	345	62	0.152	846	32	0.036
yes/no-question (<i>li</i>)	164	309	0.653	331	187	0.361	894	36	0.039
imperative	79	3	0.037	193	11	0.054	161	0	0
subjunctive (<i>čtoby</i>)	108	15	0.122	182	8	0.042	211	0 ⁹	0
simulative (<i>kak</i> etc.)	192	59	0.235	353	25	0.066	66	4	0.057

3.3. Preliminary conclusions

My analysis of the RNC data for the UIs *kto* and *kakoj* suggests that since the 18th century, the frequency of Russian UIs has dropped consistently in at least five contexts: conditionals, questions and imperatives, and subjunctive and simulative clauses. Two contexts, namely imperatives

⁹ One example within a subjunctive clause attested in the Newspaper corpus is a Bible citation and therefore was not considered here.

and subjunctive clauses, are of particular interest as they do not allow (or allow only very marginally) to use UIs in present-day Russian.

4. Further factors
that influence the use of Russian UIs

In this section, factors other than the type of the context are investigated: the ontological category of the pronoun (4.1), register (4.2), and emphasis (4.3).

4.1. Ontological category of the pronoun

As mentioned in Section 2, UIs associated with the position of adjuncts are assumed to be restricted in use, both in Russian and cross-linguistically. In this section, this view is verified for the pronouns *gde*, *kogda*, *počemu*, and *začem*. As interrogatives, they are used to ask for place, time, reason, and purpose, i.e. meaning ‘where’, ‘when’, ‘why’, and ‘what for’, respectively. As UIs, they may be expected to mean ‘somewhere’ (*gde*), ‘ever, some day’ (*kogda*), ‘for some reason’ (*počemu*) and ‘for some purpose’ (*začem*). The data in Table 6 show the comparative frequency of these UIs within conditional and simulative clauses in different time periods, suggesting the following conclusions:

- The pronouns *gde*, *kogda*, *počemu*, and *začem* are relatively rare as UIs in all periods; in particular, they are rarer than *kto* (cf. the comparative data on *kto*, *gde*, and *kogda* in Table 7, showing that the ratio of the uses of a pronoun in the function of a UI in conditionals from all its occurrences in the Newspaper corpus is the highest for *kto*).
- In all periods and contexts, the frequency declines according to the following hierarchy: *gde* > *kogda* > others.
- The frequency of all pronouns drops starting from the 18th century, which is in line with the general tendency observed for *kto* and *kakoj* in Section 3 (two examples with *kogda* in simulative clauses in 1901–2000s, which seem to contradict this generalization, are in fact not representative due to a very small sample size).

Table 6
Frequency of *gde*, *kogda*, *počemu*, and *začem* as UIs in different time periods according to the RNC

Pronoun	Context	1701–1900, 60,010,888 words		1901–2000, 158,277,632 words		Newspaper corpus (1990s–2000s), 228,521,421 words	
		N	ipm	N	ipm	N	ipm
<i>gde</i>	<i>esli</i>	116	1.93	65	0.41	48	0.21
	<i>slovno...</i>	7	0.12	4	0.03	0	0
<i>kogda</i>	<i>esli</i>	106	1.77	30	0.19	13	0.06
	<i>slovno...</i>	0	0	2	0.01	0	0
<i>počemu</i>	<i>esli</i>	1	0.02	0	0	0	0
	<i>slovno...</i>	0	0	0	0	0	0
<i>začem</i>	<i>esli</i>	0	0	0	0	0	0
	<i>slovno...</i>	0	0	0	0	0	0

Note that, although no uses of *začem* as a UI were attested in our sample, *začem* could be used as a UI in some earlier period. However, it denoted not purpose but cause, corresponding to the causal interrogative meaning of *začem*, which is now obsolete. Both the interrogative and indefinite causal meanings are listed for *začem* in [Barkhudarov et al. 1978: 338].

Table 7

Comparative frequency of *kto*, *gde*, and *kogda* in conditionals in the Newspaper corpus¹⁰

Pronoun	UIs	Total	Ratio of UIs
<i>kto</i>	994	236,527	0.004
<i>gde</i>	48	208,627	0.0002
<i>kogda</i>	13	335,637	0.00004

4.2. Register

As mentioned in Section 2, the corpus data presented by Tret'yakova [2004] dispute the assumption that UIs tend to be used in colloquial speech. In this section, this assumption is verified in a different way. Based on the corpus data for three synonymous epistemic modals *možet*, *možet byt'*, and *byt' možet*, all meaning 'may be', and the optative particle *xot' by*, I demonstrate that there is a correlation between their register characteristics (colloquial vs. formal) and the frequency of their occurrence with UIs: more colloquial items combine with UIs more frequently.

Intuitively, *možet* differs from *možet byt'* and especially *byt' možet* in that it is more colloquial. *Xot' by* also sounds colloquial. This intuition is confirmed by the RNC. In texts created in 1901–2016 (see Table 8), *možet* is 24 times more frequent and *xot' by* is 12 times more frequent in informal than in formal texts (the former include vernacular texts and electronic communication, while the latter consist of official business texts and academic writing). On the contrary, *možet byt'* is only 2.5 times more frequent in informal texts than in formal ones, while *byt' možet* is almost twice rarer in informal texts than in formal ones.

Table 8

Frequency of *možet*, *možet byt'*, *byt' možet*, and *xot' by* in formal and informal texts (1901–2016 subcorpus of the RNC)¹¹

	Informal 6,907,660 words		Formal 29,180,561 words	
	ipm	N	ipm	N
<i>možet</i>	195	1,347	8	244
<i>možet byt'</i>	250	1,730	100	2,915
<i>byt' možet</i>	22	154	38	1,098
<i>xot' by</i>	24	169	2	47

¹⁰ In Table 7, the data were collected differently for *kto*, on the one hand, and *gde* and *kogda*, on the other: the distance between *esli* 'if' and the pronoun was one word for *kto* (which means direct adjacency) and from one to four words for *gde* and *kogda*. This was due to the fact that the one-word distance would result in a too small sample for *gde* and *kogda*, while the distance of more than one word would result in a too large sample for *kto*. However, the conclusion on the higher frequency of *kto* is not compromised by this methodology: with the distance from one to four words, the frequency of *kto* would turn out to be even higher than in Table 7.

¹¹ The following queries were used to obtain the data in Table 8: “**может**” (acomma | first) bcomma; “**может**” (acomma|first) -bcomma на расстоянии 1 от “**быть**” bcomma; “**быть**” (acomma|first) -bcomma на расстоянии 1 от “**может**” bcomma; “**хоть**” на расстоянии 1 от бы -acomma.

According to the data in Section 3, the ratio of the UI *kto* in the context of *xot'by* is relatively high in the Newspaper corpus, namely 0.5. If we accept the hypothesis that Russian UIs tend to be used in colloquial texts, the high ratio of *kto* in the context of *xot'by* could be explained by the colloquial character of the latter. The corpus data on the frequency of *kto* in the context of *možet*, *možet byt'*, and *byt' možet*, given in Table 9, suggest a similar interpretation: in all periods, *kto* combines with *možet* more frequently than with *možet byt'* or *byt' možet*. The latter, being the “least colloquial”, was not attested at all in combination with *kto*.

Table 9

Frequency of *kto* with epistemic modals in different time periods

Subcorpus	<i>možet kto</i>	<i>možet byt' kto</i>	<i>byt' možet kto</i>	Ratio of <i>možet</i>
1701–1900	31	12	0	0.72
1901–2000	115	14	0	0.89
Newspaper corpus, 1990s–2000s	103	0	0	1

Similar frequency data are provided by Tret'yakova [2009a: 177] for *kto* in the context of *možet*, *možet byt'*, and *byt' možet*, although she refers to the introspection and not to a corpus analysis. Tret'yakova [Ibid.] also suggests an interpretation to this comparative frequency. She argues that what is relevant here is the information structure, namely the fact that UIs can never be focused. *Možet*, *možet byt'*, and *byt' možet* differ in these terms in that *možet* tends to defocalize the constituent with which it combines, and *byt' možet* introduces a focus, while *možet byt'* does not specify the information structure. Thus, *možet* conforms best and *byt' možet* conforms worst to the “non-focus” status of the UI. However, this difference does not seem to be well pronounced. The examples provided by Tret'yakova [2009a: 177] do not sound convincing to me: in (21), in which *byt' možet* is assumed to contribute to the importance of the information ‘something could have happened’, it can be substituted with *možet*. Cf. a corpus example with *možet* in (22).

- (21) *Byt' možet* (^{OK}*možet*), *čto-to slučilos'*.
be.INF may.PRS.3SG maybe what.NOM-INDF happen.PST.N.SG
‘Maybe, something happened.’ (based on [Tret'yakova 2009a: 177])

- (22) *Počemu my celuju nedelju emu ne zvonili?*
why we.NOM whole.ACC week.ACC he.DAT NEG call.PST.PL
Možet, čto-to slučilos'.
maybe what.NOM-INDF happen.PST.N.SG
‘Why haven’t we called him for a week? Maybe, something happened.’ [RNC, 2006]

To summarize, the hypothesis about the colloquial character of Russian UIs provides an account for the distribution of *kto* in the context of the epistemic modals *možet*, *možet byt'*, and *byt' možet*, and the optative particle *xot'by*. Note that this hypothesis does not necessarily contradict the results of the corpus study by Tret'yakova [2004], according to which Russian UIs are more frequent in written narrative discourse than in direct-speech contexts (see Section 2). Indeed, register is just one of the factors that influence the use of UIs in Russian (along with, for example, the semantic type of the context), and it is to be examined how exactly different factors interact (see more on this assumption in Section 5).

4.3. Emphasis

Emphasis, and in particular emphasis by the particle *i*, was mentioned as a factor that facilitates the use of Russian UIs (see Section 2). The RNC data confirm this view. The role

of emphasis is especially evident with respect to the pronouns that are associated with the adjunct position, for which the use as UIs is generally problematic (see Section 4.1). The UI *počemu* ‘for some / any reason’ occurs in the Main corpus only twice, and in both cases it is accompanied by the emphatic particle *i*:¹²

- (23) *A esli počemu F1 Challenge i budet*
 and if for.some.reason F1 Challenge PTCL AUX.3SG
ostavat'sja vne konkurencii kak imitator
 remain.INF out competition.GEN as imitator.NOM
Formuly-1, tak èto potomu, čto ne
 Formula 1.GEN then this because because NEG
budet ničego novogo.
 be.FUT.3SG nothing.GEN new.GEN

‘And if for some reason F1 Challenge will remain out of competition as an imitator of Formula 1, it is because there will be nothing new.’ [RNC, 2005]

The data in Table 10 suggest that the particle *i* contributes to the felicitousness of both *gde* ‘somewhere’ and *kogda* ‘ever, some day’ as UIs. In the Newspaper corpus, the ratio of examples with the particle *i* in conditionals is higher in combination with *gde* or *kogda* than with *kto* (the differences between *kto* and *gde* and between *kto* and *kogda* are both statistically significant, 2-tailed Fisher’s exact test, $P < 0.05$).

Table 10

Frequency of *kto*, *gde* and *kogda* with and without the emphatic particle *i* in conditionals in the Newspaper corpus of the RNC¹³

Pronoun	With <i>i</i>	Without <i>i</i>	Ratio of examples with <i>i</i>
<i>kto</i>	166	394	0.3
<i>gde</i>	10	9	0.5
<i>kogda</i>	4	0	1

5. Discussion

In this section, the discussion of the Russian data presented so far is divided into three parts. In 5.1, I argue that the boundaries of the Russian UIs on the semantic map of indefinite pronouns are different from those proposed in earlier works. In 5.2, previous accounts for the distribution of Russian UIs across contexts are discussed. In 5.3, my own analysis of such a distribution is developed.

¹² Only one of these two examples, namely the example from the 18th century, was part of my sample with *počemu* in Section 4.1. Example (23), retrieved from the Main corpus, was not considered there due to its date (2005).

¹³ Since the particle *i* is commonly placed before the finite verb, the queries were used in which *i* immediately precedes the verb (for the sample with emphasis) or is absent from this position (for the sample without emphasis): **если** на расстоянии 1 от **кто** -amark на расстоянии от 1 до 4 от **и** -amark на расстоянии 1 от **v indic** -amark; **если** на расстоянии 1 от **кто** -amark на расстоянии от 1 до 4 от **-и** -amark на расстоянии 1 от **v indic** -amark.

5.1. Russian UIs on the semantic map of indefinite pronouns

According to Tret'yakova [2004; 2009a; b], Russian UIs span the 'irrealis non-specific', 'question', and 'conditional' functions on the semantic map of indefinite pronouns proposed by Haspelmath [1997; 2003: 222], as in Fig. 1, repeated below in Fig. 3 for convenience.

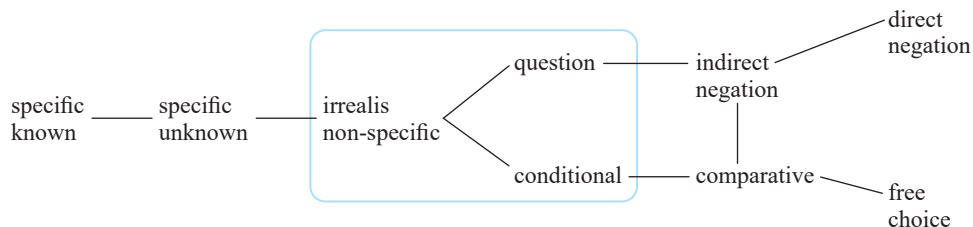


Figure 3. Boundaries of the Russian UIs according to Tret'yakova [2004; 2009a; b], on the semantic map by Haspelmath [1997]

However, the RNC data discussed above suggest that the presentation in Fig. 1 is imprecise. In present-day Russian, UIs are almost absent from the 'irrealis non-specific' contexts. Tret'yakova [2009b: 146] qualifies the following contexts as belonging to the 'irrealis non-specific' function, suggesting that Russian UIs can be used in all of them, excluding future contexts:

- future contexts;
- imperative contexts;
- optative contexts;
- disjunctive contexts;
- simulative contexts;
- contexts with epistemic modals;
- purpose contexts.

But in fact, as the data from the Newspaper corpus show (see Section 3), UIs are extremely rare or absent from imperative, purpose and disjunctive contexts in modern Russian. The problem with the epistemic and optative contexts is that UIs within them are restricted to certain lexical markers, and hence do not pertain to the epistemic or optative contexts in general. Indeed, the only optative marker Tret'yakova [2004; 2009a; b] considers is *xot' by*; some other optative constructions, as in (24) with the marker *čtob*, seem to be less compatible with UIs. No instances of *kto* or *kakoj* as UIs have been attested in the RNC within an optative construction with *čtob*. Epistemic modals that do not combine with UIs in modern Russian include *dolžno byt'* 'it is likely' [Yanovich 2005: 322] and *navernjaka* 'for sure'.

- (24) *Čtob mne kto-nibud' (čtob) tak platil!*
 OPT I.DAT who.NOM-INDF someone.NOM so pay.SBJV.M.SG
 'I wish someone would pay me so [much]!' [RNC, 2001]

- (25) *Navernjaka kogo-nibud' (čtob) vstretiš'.*
 for.sure who.ACC-INDF someone.ACC meet.FUT.2SG
 'You will definitely meet someone.' [RNC, 2003]

Thus, the only type of contexts from the above list whose combinability with UIs is not shaky are simulative contexts. But exactly these contexts, in my view, are not in fact 'irrealis non-specific' (pace Tret'yakova [2004; 2009a; b]) — they share important traits with the 'specific unknown' function. Consider the example of a UI in the simulative context in (26). The semantics of the embedded clause introduced by the subordinator *budto* can be split into two meaning

components: 1) ‘The speaker assumes that in the actual world, nobody shot the wheel’ (see also [Letuchiy 2017: 180]), 2) ‘the speaker assumes that in a possible world, different from the actual, there exists someone who shot the wheel’. The first component is rendered explicit in such environments as (27).¹⁴

- (26) *Koleso spustilo na rovnom meste,*
wheel.NOM flatten.PST.N.SG on even.LOC place.LOC

budto kto vystrelil v nego.
as.if someone.NOM shoot.PST.M.SG in it.ACC

‘The wheel flattened out of the blue, as if someone had shot it.’ [RNC, 2001]

- (27) *Zamečatel’no prosto! Kak budto leto! — Počemu “kak budto”?*
great just as if summer why as if

Normal’noe leto! (RNC, 2012)
normal summer

‘It’s just great! It’s as if it were summer! — Why [are you saying] “as if”? It’s really summer!’ [RNC, 2012]

Although the first meaning component is in line with identifying simulative clauses as ‘irrealis’, the second introduces the speaker’s commitment to the existence of an individual referred to by the pronoun (albeit in a possible world), which is crucial for the ‘specific unknown’ function (see [Haspelmath 1997: 38; Gärtner 2009: 7]). Now, it seems that at least for modern Russian, it is the second, and not the first component, that determines to which function simulative contexts correspond. That they correspond to the ‘specific unknown’ function is suggested by the distribution of the *to*- and *nibud*’-series. On the one hand, only *to*-series can be used in the ‘specific unknown’ function according to the semantic map proposed by Haspelmath [1997: 65] (see Figure 2). On the other hand, in simulative clauses the *to*-series is substantially preferred over the *nibud*’-series. The corpus data in Table 11 show that this is the case for all time periods under scrutiny, with the Newspaper corpus manifesting this trend most clearly.

Table 11

Frequency of *kto-to* and *kto-nibud*’ in simulative clauses in different time periods

Subcorpus	-to	-nibud’	Ratio of -to
1701–1900	138	42	0.77
1901–2000	1,012	16	0.98
Newspaper corpus (1990s–2000s)	96	0	1.0

One more piece of evidence in favor of the assumption that simulative clauses are ‘specific unknown’ comes from the pronoun *nekij* ‘certain’. *Nekij* may be used both in past declarative

¹⁴ Note that simulative markers do not convey a firm speaker’s commitment to the falsity of the proposition in question but only the speaker’s conjecture that the proposition is false. This is evidenced by the example in (i), where using *kak budto* does not prevent the speaker from having doubts.

- (i) *Vdrug ottuda stali donosit’sja otčëtlivye šoroxi,*
suddenly from.there begin.PST.PL be.heard.INF distinct.NOM.PL rustle.NOM.PL
kak budto kto-to xodit. I ja
as like who.NOM-INDF walk.PRS.3SG and I.NOM
rešila udostoverit’sja, čto postoronnix ljudej
decide.PST.F.SG make.sure.INF COMP foreign.GEN.PL people.GEN.PL
v zdanii net. Malo li čto?
in building.LOC not.be.PRS little Q what

‘Suddenly, distinct rustles began to be heard from there, as if someone was walking. And I decided to make sure that there were no strangers in the building. You never know.’ [RNC, 2013]

clauses (28), where indefinites must be specific, and in simulative clauses (29). At the same time, *nekij* is banned from ‘irrealis non-specific’ contexts, such as imperatives (30) (see more about *nekij* in [Shmelev 2002: 119 ff.]).

- (28) *Našli nekoego Blinova, togda rabotavšego*
 find.PST.PL certain.ACC Blinov.ACC then work.ACC
vo Dvorce sporta «Spartak».
 in palace.LOC sport.GEN Spartak.NOM
 ‘[They] found some Blinov, who was then working at the Spartak Sports Palace.’ [RNC, 1984–2001]

- (29) *A den' vsjo dlitsja i dlitsja,*
 and day.NOM still go.on.PRS.3SG and go.on.PRS.3SG
kak budto nekim Zanebesnym Dekretom otmenili tečenie vremeni.
 as.if as.if certain.INS celestial.INS decree.INS cancel.PST.PL passage.ACC time.GEN
 ‘And the day goes on and on, as if the passage of time was canceled by some kind of Celestial Decree.’ [RNC, 1999]

- (30) *Sročno najdi kakogo-nibud' (°nekoego) zemljaka.*
 urgently find.IMV.2SG which.ACC-INDF certain.ACC countryman.ACC
 ‘Urgently find some fellow countryman.’ [RNC, 2010]

Note that the possibility to use UIs in simulative clauses does not depend on a specific lexical marker: all complementizers that introduce simulative clauses (*kak*, *kak budto*, *slovno*, and *točno* seem to roughly exhaust the list) can combine with UIs. This is in contrast with optative or epistemic contexts, in which, as mentioned above, only certain markers are compatible with UIs.

The assumption that Russian UIs are not excluded from the ‘specific unknown’ function is further supported by the fact that UIs can be used in questions in which the speaker expects a positive answer, cf. (5). According to Haspelmath [1997: 82], it is cross-linguistically common that the indefinite series chosen for positive expectations is the one that has more specific functions, while the series chosen for neutral or negative expectations is the one that has more non-specific functions. In Russian, the *to-series* is used in questions accompanied by the speaker’s positive expectations, while the *nibud’-series* is generally excluded from such questions (31) [Kobozeva 1981: 170–171; Haspelmath 1997: 84]. Thus, UIs behave in this respect like *to*-pronouns, associated with the ‘specific unknown’ function.

- (31) *Stop! — perebil princ ministra i samogo sebja.*
 stop interrupt.PST.M.SG prince.NOM minister.ACC and oneself.ACC oneself.ACC
 — *Čixnul kto-to (°kto-nibud')? Ili mne poslyšalos'?*
 sneeze.PST.M.SG who.NOM-INDF who.NOM-INDF or I.DAT hear.PST.N.SG
 ‘“Stop!” — the prince interrupted the minister and himself. “Did someone sneeze? Or did I hear it?”’ [RNC, 1988]

To summarize, the boundaries of UIs in present-day Russian correspond to the semantic map in Figure 4 rather than to the one in Figure 3.¹⁵

In the Russian language of the 18th–19th centuries, as follows from the data presented in Section 3, UIs could be used in all contexts in which they are used today but could also combine

¹⁵ According to Fig. 4, Russian UIs do not occupy a contiguous area on the semantic map, thus being problematic for one of the essential principles, namely the principle of contiguity, lying behind the idea of semantic maps [Haspelmath 2003: 217]. This does not mean, I suppose, that there is some problem with the semantic map of indefinite pronouns itself. Rather, there is something special with UIs that makes them differ from the expected pattern. My hypothesis is that it is their dependency on context, in the sense of what is proposed in Section 5.3, that makes UIs so special. I will briefly come back to this proposal in the Conclusion.

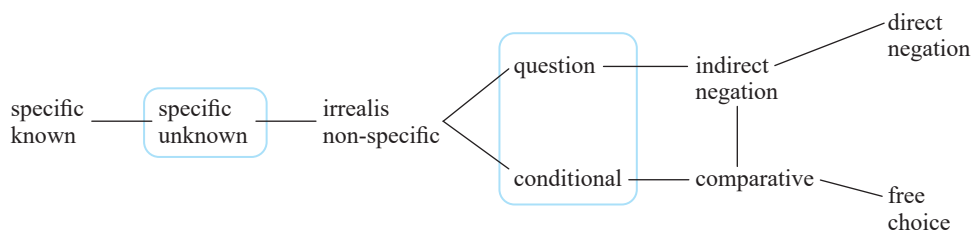


Figure 4. Boundaries of Russian UIs in present-day Russian according to the RNC data

with some ‘irrealis non-specific’ contexts. UIs thus occupied a contiguous area on the semantic map, as in Fig. 5, while the evolution of UIs since the 18th century can be described as a movement from the map in Fig. 5 to the map in Fig. 4.

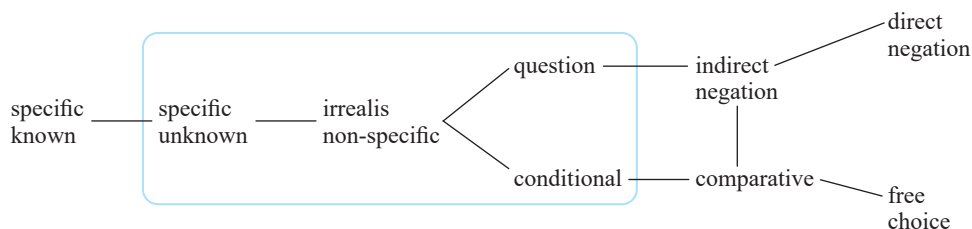


Figure 5. Boundaries of Russian UIs in the 18th–19th centuries according to the RNC data

5.2. Previous accounts

I am aware of two lines of reasoning that have been suggested to account for the distribution of Russian UIs across contexts. Both, however, leave unexplained several important facts.

One account goes back to Yanovich [2005] (see also [Belyaev, Haug 2020: 888]), who, following the analysis of Japanese bare indefinite roots by Kratzer and Shimoyama [2002], treats Russian UIs as Hamblin sets. The guiding idea is that UIs, exactly as interrogative pronouns, refer to a set of alternatives. To get an interpretation within a declarative sentence, a UI needs a **li-censer**, i.e., must be bound by a higher operator. This immediately explains why the sentence in (32), if taken on its own, can denote a question ‘who came?’ but not a statement ‘somebody came’. Indeed, questions are commonly assumed to yield non-singleton sets of propositions (cf. {*a* came, *b* came, *c* came ...} for (32)), while a declarative sentence would refer to a singleton proposition. In (33), on the contrary, there is a conditional operator that serves to yield a quantificational structure ‘for all *x*, if *x* comes, then call me’, compatible with the declarative clause type.

- (32) *Kto prišël?*
 who.NOM come.PST.M.SG
 ‘Who came?’

- (33) *Esli kto pridët, zovi menja.*
 if someone.NOM come.FUT.3SG call.IMV.2SG I.ACC
 ‘If anyone comes, call me.’

This approach sheds light on the cross-linguistic tendency, noted by Haspelmath [1997: 173], according to which UIs are “generally excluded from past or current present affirmative

declarative clauses, where indefinites must be specific”. However, at least the following two facts remain a mystery (for further critique, see [Tret'yakova 2009a: 58]):

— There are also languages in which UIs can be used in unembedded declarative sentences, as in the German example (34). How does the UI do without a licenser in this case?

- (34) *Gestern ist wer gekommen.*
 yesterday AUX someone come.PTCP

‘Someone arrived yesterday.’ [Tret'yakova 2009a: 133]

— As observed by Yanovich [2005] himself, there is cross-linguistic variation in the type of contexts compatible with UIs. Specifically, in some languages the list of such contexts may seem inconsistent. In Russian, e.g., UIs in optative and epistemic contexts are allowed with some operators and are banned with others (see Section 5.1).

Another line of reasoning was suggested by Tret'yakova [2004], who explains the distribution of Russian UIs across contexts by the following principle: UIs tend to be used in contexts that express indefiniteness on their own. These are, in her opinion, conditionals, questions and different types of ‘irrealis non-specific’ contexts, i.e. exactly those types of contexts that allow UIs according to her analysis.

The idea that due to the absence of a specific marker of indefiniteness, UIs highly rely on the context in their distribution and use, seems to me a very insightful and promising one. In fact, my own analysis, presented in Section 5.3, has been inspired by this idea. However, if put in this form, it does not seem to explain much. Why are some irrealis contexts (e.g., future contexts) incompatible with UIs in Russian? Why do some contexts combine with UIs more frequently than others (note especially imperatives and subjunctives, from which, according to the corpus data presented in Section 3, UIs are almost excluded in present-day Russian)? I try to offer answers to these questions in the following section.

5.3. Analysis

Based on the results of the corpus study presented in Sections 3 and 4, I assume that the distribution of UIs in Russian is determined by several factors, which include the type of the context, register (colloquial vs. formal), grammatical role of the pronoun (argument vs. adjunct), and emphasis on the pronoun. In what follows, I will propose several hypotheses as to how exactly these factors interact and what their origins are.

My guiding hypothesis is that the distribution of Russian UIs across contexts, both synchronically and diachronically, is determined by some sort of **economy principle**:

- (35) UIs tend to occur in contexts that are biased toward a certain interpretation of indefinite pronouns in terms of specificity (specific known vs. specific unknown vs. non-specific). While indefinite pronouns with a marker denote the type of specificity by means of the marker, UIs have no choice but to rely on the context.

This hypothesis offers an account for several facts observed hitherto.

First, it sheds light on the fact that both in modern Russian and the Russian language of the 18th–19th centuries, conditionals and questions are contexts in which UIs occur rather frequently (see Section 3). Indeed, conditionals and questions are strongly biased toward the non-specific interpretation of an indefinite. On the one hand, as mentioned in Section 5.1, the specific functions are associated with the commitment of the speaker to the existence of an individual or an item referred to by the pronoun [Haspelmath 1997: 38; Gärtner 2009: 7]. On the other hand, both in (real) conditionals and pragmatically neutral (i.e., non-biased and non-modalized) questions the speaker is unaware of the truth-value of the respective proposition (see,

e.g., [Liu et al. 2021: 1370–1371]). Consequently, he is most probably also unaware of whether what is denoted by an indefinite noun phrase within a conditional or a question exists. Neither in the conditional *I will be happy if you bring me a book*, nor in the question *Will you bring me a book?*, does the speaker assume anything as to whether a book that could be brought to him by the addressee exists.

Second, the principle in (35) provides an explanation for the fact that Russian UIs avoid the ‘irrealis non-specific’ function: they were rare in this function in the 18th–19th centuries and evolved toward the elimination of the ‘irrealis non-specific’ contexts from their usage in present-day Russian (see Section 3). Most contexts that span the ‘irrealis non-specific’ function are less clearly than conditionals or questions biased toward a certain interpretation of the indefinite in terms of specificity. The crucial difference between ‘irrealis non-specific’ contexts, on the one hand, and conditionals and questions, on the other, is that the former either imply the speaker’s assumption (albeit not the speaker’s commitment) that the participants of the situation in question **do exist** (which neither conditionals nor questions do), or they imply the speaker’s assumption that the participants of the respective situation **may exist**, doing so in a more explicit way than conditionals and questions. In both cases, ‘irrealis non-specific’ contexts turn out to represent a step toward the specific interpretation of an indefinite compared to conditionals and questions.

To illustrate this idea, let me consider different types of ‘irrealis non-specific’ contexts (future, imperative, subjunctive, disjunctive, optative and epistemic) one by one. If one makes a positive future statement *I will bring you a book*, or a positive imperative one *Bring me a book*, the speaker’s assumption that a book that could be brought exists is implied. This is also the case for positive purpose clauses, which are a subtype of subjunctive clauses in my terms. The statement *I went to a bookstore in order to bring you a book* implies that, according to the speaker’s assumption, there existed a book in the bookstore that could be brought to the addressee. This is in contrast both with the conditional *I will be happy if you bring me a book* and the question *Will you bring me a book?* In unmarked pragmatic conditions, as suggested above, neither of them implies the speaker’s assumption that a book that the addressee could bring them exists.

Both negative imperatives and negative purpose clauses are associated with a weaker implication: what they imply is not the speaker’s assumption that the object referred to by an indefinite noun phrase exists but rather that it may exist. Cf. the imperative *Don’t trip over something*, which implies that the speaker assumes there may be something to trip over. Similarly, the purpose statement *To prevent anyone from overhearing, the conversation took place in a separate room*, which contains an implicit negation due to the verb *prevent*, implies the speaker’s assumption that there could be someone who would overhear. This difference between positive and negative imperative and subjunctive contexts could be the reason for which in present-day Russian, UIs are more felicitous both in negative imperatives and negative subjunctive clauses. In my samples with UIs retrieved from the Newspaper corpus, no positive imperatives or subjunctive clauses have been attested.¹⁶

In the case of negative utterances, the difference between imperative and purpose clauses, on the one hand, and conditionals and questions, on the other, is less evident, since conditionals and questions also imply that the proposition within the scope of the conditional or the question operator may be true, hence, the participants of this proposition may exist. When asking *Will you bring me a book?*, the speaker indeed proceeds from the assumption that the proposition ‘you will bring me a book’ may *a priori* be true and a book that could be brought may *a priori* exist. Otherwise, i.e. in the case that the speaker firmly knows that this proposition is false, no questions

¹⁶ For imperatives, this difference between negative and positive clauses appears to have been absent in the language of the 18th–19th centuries. In my sample with the UI *kto* from the 1701–1900 corpus, there are 10 positive imperative examples and only three negative ones. For subjunctive clauses, however, the situation seems to have been similar to the present one: among 77 subjunctive examples with *kto*, only four contain neither explicit nor implicit negation.

are posed [Giannakidou, Mari 2019] or conditionals uttered. Still, what creates a difference between negative imperative and purpose clauses from conditionals and questions seems to be the **degree of explicitness** of the speaker's assumption that the participants of the proposition in question may exist. For conditionals and questions, this assumption is **presupposition-like**: it does not make sense to pose a question or utter a conditional if the respective assumption is false. For negative imperative and purpose clauses, the assumption is **assertion-like**: it is manifested by the speaker in that his utterance (or a situation described in it) is **motivated** by this assumption. Indeed, estimating the participants of the respective situation as probably existing is part of the motivation for an imperative utterance (e.g., assuming that there may be something to trip over motivates one to say *Don't trip over something*) or for having an intention expressed in the purpose clause (e.g., assuming that someone may overhear motivates one to set the goal 'to prevent anyone from overhearing'). For conditionals and questions, this is not the case. Both questions and conditionals are rendered meaningful by the assumption under discussion but cannot be said to be motivated by it. In this sense, conditionals and questions are a step further from the 'specific' zone than negative imperative and purpose clauses.

A similar reasoning is applicable to other 'irrealis non-specific' contexts. The disjunctive statement *They brought a book or something sweet as a gift* implies that the speaker assumes one of the disjunctive propositions (*he brought a book* or *he brought something sweet*) to hold. In other words, for each proposition, the speaker **asserts** the possibility that it may be true. Hence, he asserts the possibility that a book exists that was brought as a gift. The possible existence of a book is also directly asserted by the epistemic statement *Perhaps he brought a book*. In the optative statement *If only they would bring me a book!* the situation is slightly different in that logically speaking, a desire does not imply an assumption that the desired situation can become true. However, pragmatically speaking, if someone expresses their desire for something, they usually do not deem that thing as *a priori* unobtainable. In this case, what renders the respective assumption more explicit compared to conditionals and questions is that this assumption is at the same time the subject of the speaker's desires: the speaker puts this assumption into words by saying that this is what he wants (for the above example, the speaker wants the proposition 'they brought me a book' to be true).

As is clear from the above discussion, the interpretation of an indefinite pronoun in terms of specificity is determined by what the speaker may assume about the truth-value of the proposition that contains this pronoun. Thus, the idea expressed in (35) may also be posed in terms of the state of the speaker's knowledge: the acceptability of UIs in Russian seems to decline according to the hierarchy in (36).

- (36) The speaker is unaware of, and makes no assumptions about, the truth-value of the proposition containing the UI > the speaker knows or assumes the proposition containing the UI to be false or unfeasible > the speaker knows or assumes the proposition containing the UI to be true or feasible.

The link between (35) and (36) is as follows: if the speaker is unaware of the truth-value of the proposition containing the UI, he has no reason to assume that the item referred to by the pronoun exists, which is a strong bias toward the non-specific interpretation, hence, against the specific interpretation; if, on the contrary, the speaker has some information about the truth-value status, this gives him a reason to assume something on the existence of the item in question. Still, the assumption that the respective proposition is false is a more solid reason to assume the non-existence, hence, the non-specificity, than the assumption that the proposition is true. In the latter case, a strong bias toward specificity arises; however, in the absence of a special marker of indefiniteness, there is no way to distinguish 'specific known' and 'specific unknown' meanings, which, according to the principle in (35), is problematic for the use of UIs.

Conditionals and questions represent the upper point of the hierarchy (36), which accounts for the relative acceptability of UIs in these contexts across different time periods. Note that conditionals and questions could also be associated with the lower point of (36), since they

express the speaker's assumption that the proposition containing an UI is feasible, as discussed above. However, this is not what conditionals and questions convey in the first place, according to my above suggestions. Furthermore, simulative clauses are associated with the intermediate point of (36): they imply that the speaker assumes the respective proposition to be false (see the semantic analysis of simulative clauses in Section 5.1). This explains why simulative clauses, which are 'specific unknown' according to my assumptions, combine with UIs more frequently than most of the 'irrealis non-specific' contexts (imperatives, subjunctives, disjunctives, etc.) in all time periods, the latter being associated with the lower point of the hierarchy (36).

As mentioned in Section 3.1, the markers *kak*, *kak budto*, *slovno*, and *točno* most commonly introduce simulative clauses but may also introduce complement clauses, in which case they convey that the speaker or the subject of the main clause doubts what is being reported in the embedded clause. Cf. a simulative clause in (37) and a complement clause in (38).

- (37) *On — v dver', a ona ne poddaetsja,*
 he.NOM in door.ACC and she.NOM NEG give.in.PRS.3SG
budto kto eë podpër snaruži.
 as.if someone.NOM she.ACC prop.PST.M.SG outside
 'He [pushes] against the door, but it does not give in, as if someone propped it outside.'
 [RNC, 2000]

- (38) *Snačala žena uslyšala, kak budto kto butylku razbil,*
 at.first wife.NOM hear.PST.F.SG as.if someone.NOM bottle.ACC break.PST.M.SG
a potom polyxnulo.
 and then blaze.PST.N.SG
 'At first, the wife heard that someone seemed to have broken the bottle, and then it blazed.'
 [RNC, 2008]

Both in the Newspaper corpus and in the corpus of 18th–19th centuries, simulative clauses with the UI *kto* are more frequent than complement clauses (the ratio is 20:1 in the Newspaper corpus and 204:8 in the 1701–1900 corpus). This is generally in line with the hierarchy (36). Indeed, in simulative clauses, the speaker assumes that the proposition containing the UI is false, while in complement clauses the speaker only doubts it to be true. Although strictly speaking the hierarchy in (36) does not make a difference between these two options, the former option is somewhat stronger than the latter and could be associated with the better acceptability of UIs if (36) is on the right track. Note, however, that the subordinators *kak*, *kak budto*, *slovno*, and *točno* seem to be generally more frequent as markers of simulative clauses than as complementizers in complement clauses, which could also be the reason why they occur with UIs in simulative clauses more frequently. Still, the fact that the difference in frequency corresponds to what is predicted by (36) and that it holds for different time periods is noteworthy.

Not only *kak budto*- [*budto*- etc.] clauses but also subjunctive clauses, introduced by the subordinator *čtoby*, include different semantic-syntactic types. Besides purpose clauses, *čtoby* can also introduce complement clauses, which, in turn, may be of two types: purpose-like complement clauses, which are close to purpose clauses both semantically and syntactically; and the so-called epistemic complement clauses, which convey that the speaker considers the likelihood of the situation expressed in the embedded clause to be low [Dobrushina 2012]. Cf. a purpose clause in (39), a purpose-like complement clause in (40), and an epistemic complement clause in (41):

- (39) *K domu ego prigonjal,*
 to house.DAT he.ACC drive.PST.M.SG
čtoby kto kakuju zapčast' ne snjal.
 so.that someone.NOM some.ACC spare.part.ACC NEG remove.SBJV.M.SG
 'He drove it to the house so that someone would not remove some spare part.' [RNC, 2007]

- (40) *Vsjo vremja ona bojalas, čtoby kto ne nabrël*
 all time she.NOM be.afraid.PST.F.SG COMP someone.NOM NEG come.across.PST.M.SG
na neë v kustarnike, ne uvidel.
 on she.ACC in bush.LOC NEG see.SBJV.M.SG

‘All the time she was afraid that someone would come across her in the bushes, would see [her].’ [RNC, 1982]

- (41) *Navrjad li, čtoby kto poveril ...*
 unlikely PTCL COMP someone.NOM believe.SBJV.M.SG

‘It is unlikely that anyone would believe [this].’ [RNC, 1966]

The hierarchy in (36) predicts UIs to be more frequent in epistemic clauses than in purpose or purpose-like clauses, as the former are associated with the intermediate point in (36), while the latter, according to the above assumptions, represent the lower point. Unfortunately, my sample of subjunctive clauses is too small to assess whether this prediction is borne out. But importantly, my data are not in contradiction to it. Among seven examples in the sample from the Newspaper corpus, three are epistemic, two are purpose and two more are purpose-like. Note also that generally, epistemic clauses are the rarest type of *čtoby*-clauses in modern Russian [Dobrushina 2012: 155].

There is, however, one more prediction suggested by (36) that can be verified. It follows from (36) that counterfactual conditionals should combine with UIs less frequently than real conditionals, since the former but not the latter imply that the proposition contained in the conditional clause is false. This prediction is borne out. The data in Table 12 show that in conditionals, the UI *kto* occurs with the subjunctive particle *by* (or its variant *b*), which marks counterfactual conditionals in Russian, less frequently than marked indefinite pronouns (the difference is statistically significant, χ^2 , $df = 1$, $p < 0.01$).

Table 12

Frequency of marked and unmarked indefinites with *kto*
 in subjunctive and indicative conditional clauses
 in the Newspaper corpus of the RNC¹⁷

	With <i>by/b</i>	Without <i>by/b</i>	Ratio of examples with <i>by/b</i>
<i>Kto</i> unmarked	27	994	0.03
<i>Kto</i> marked	262	3,470	0.07

So far, I discussed the facts that are in line with the predictions suggested by (35) and (36). However, there are also several facts that contradict (35) and (36). These are the data concerning the distribution of UIs in epistemic and optative contexts, and in the context of questions for which the speaker expects a positive answer (see Section 5.1). Indeed, according to the data of the Newspaper corpus (see Section 3), UIs combine rather frequently with the epistemic marker *možet* and the optative marker *xot'by*. Both are associated with the ‘irrealis non-specific’ function, and both, as I assumed earlier, correspond to the lower point of the hierarchy in (36) (*možet* probably more so than *xot'by*). Questions with the speaker’s positive expectations also fit the lower point of (36).

¹⁷ The particle *by/b* most typically follows the subordinator *esli* [Dobrushina 2009: 284–285]. Therefore, the data in the second column of Table 12 (examples without *by/b*) coincide with the data in Table 1, where the frequency of marked and unmarked indefinites was calculated for structures with *kto* immediately following *esli*. To retrieve the data in the first column of Table 14, the following queries were used: *если на расстоянии 1 от бы | б* -amark на расстоянии 1 от *кто-нибудь* -amark; *если на расстоянии 1 от бы | б* -amark на расстоянии 1 от *кто-то* -amark; *если на расстоянии 1 от бы | б* -amark на расстоянии 1 от *кто* -amark.

My hypothesis is that in all three cases, a factor different from the principle (35) comes into play. This factor is the **pronouncedly colloquial** character of all three constructions. Importantly, this idea allows one to explain the selectivity UIs demonstrate when combining with epistemic and optative markers. Recall that, as shown in Section 5.1, UIs can be used not with all epistemic and optative markers but just with some of them. In what follows, I will expand on my hypothesis separately for each of the three cases.

Epistemic markers *vrjad li* ‘hardly’, *možet / možet byt’ / byt’ možet* ‘perhaps’, *navernoe* ‘probably’, *navernjaka* ‘for sure’, and *dolžno byt’* ‘apparently’ behave differently with respect to combinability with UIs in present-day Russian. Only three markers can combine with UIs — *vrjad li* (42) (see also [Tret’yakova 2004]), *možet*, and *možet byt’*, with *možet* doing so more often than *možet byt’* (cf. the corpus data in Section 4.2).

- (42) *V ètoj oblasti vrjad li kto imeet vlijaniya*
 in this.LOC area.LOC hardly PTCL someone.NOM have.PRS.3SG influence.GEN
bol’she menja.
 more I.GEN

‘In this area, hardly anyone has more influence than me.’ [RNC, 2014]

Four other markers do not combine with UIs at all in modern Russian: no examples from either the 20th or 21st century have been attested in the Main corpus of the RNC.¹⁸ Crucially, the markers in question also differ in their epistemic status. *Vrjad li* implies that the speaker assumes the proposition with which it combines to be false. *Navernjaka* conveys an opposite implication, namely that the speaker is sure that the respective proposition is true. Another five markers also imply positive expectations of the speaker, albeit much weaker ones than *navernjaka*. Therefore, the hierarchy in (36) predicts *vrjad li* to be more felicitous with UIs than other markers, which is indeed borne out with the only exception being the marker *možet*. As shown in Section 4.2 based on the RNC data, *možet* also differs from *možet* or *byt’ možet* in that it tends to be used in informal texts, i.e. is more colloquial. In this respect, *navernoe*, *navernjaka* and *dolžno byt’* are more similar to *možet byt’* than to *možet* (I did not conduct a corpus study to verify this assumption, but based on my intuition as a native speaker, this seems to be undeniable). *Možet* can thus be assumed to violate the hierarchy in (36) because it is colloquial.

Like *možet*, the optative marker *xot’ by* also tends to be used in informal texts according to the data in Section 4.2, which could be the reason for the relatively high ratio of UIs in the context of *xot’ by*. Note, however, that to verify this assumption, other optative markers should be evaluated as to whether they are used in the colloquial language and whether and how often they combine with UIs. This is a separate big question, which I leave open for now.

Finally, questions with the positive speaker’s expectations, as in (43), are also colloquial. Indeed, they may only occur as direct questions, hence, in a dialog. Russian embedded questions necessarily contain the question particle *li* (with the only exception being alternative questions, see [Shvedova 1980: 480]), which is associated with epistemically neutral questions than do not convey the speaker’s positive or negative expectations [Kobozeva 2000: 304]. *Li* can also be used

¹⁸ One example with *navernoe* (i) was attested, in which a “multiple partitive construction” [Haspelmath 1997: 177] is used. The bare interrogative in this construction occurs together with one or more instances of the same interrogative pronoun in a parallel coordinate clause.

(i) *Mužčin sredi nix bylo malo vidno — navernoe, kogo*
 man.GEN.PL among they.GEN COP.PST few visible probably someone.ACC
soslali, a kto podalsja na novostrojki v goroda.
 exile.PST.PL and someone.NOM move.PST.M.SG to construction.site.ACC.PL in city.ACC.PL
 ‘There were few men among them — probably, someone was exiled, and someone moved to construction sites in the cities.’ [RNC, 1995]

This seems to be a construction on its own, not a UI in the strict sense, since it occurs in languages (e.g., in Italian or French) that otherwise do not use bare interrogatives as indefinites.

in direct questions, but even in this case the question is epistemically neutral. For this reason, *li* is infelicitous in (44), where the speaker's positive expectations are suggested by the context.

- (43) *Molites' za upokoj duši. — Umer kto? —*
 pray.IMV.2PL for rest.ACC soul.GEN die.PST.M.SG someone.NOM
sprosila niščenka, vzdrognuv.
 ask.PST.F.SG beggar.woman.NOM start.CNV

“Pray for the rest of the soul.” “Did someone die?” asked the beggar woman with a start.¹⁹
 [RNC, 1997]

- (44) *Molites' za upokoj duši. — ??Umer li kto? —*
 pray.IMV.2PL for rest.ACC soul.GEN die.PST.M.SG Q someone.NOM
sprosila niščenka, vzdrognuv.
 ask.PST.F.SG beggar.woman.NOM start.CNV

Intended meaning: ‘Pray for the rest of the soul.’ “Did someone die?” asked the beggar woman with a start.’

To sum up, I argue that the epistemic *možet* and the optative *xot'by*, as well as questions with the speaker's positive expectations, although they are not expected to combine with UIs according to (35) and (36), do so because they are colloquial. The role of register for the use of UIs can thus be generalized as follows. Contrary to what is sometimes assumed (see Section 2), UIs are not strictly colloquial: they can be used, e.g., in newspaper texts. This is in line with the corpus data provided by Tret'yakova [2004] on the frequency of UIs in written and direct-speech contexts (see Section 2). For the contexts that are pronouncedly colloquial, however, the register factor may override other factors relevant to the use of Russian UIs, such as the principle (35) and the hierarchy (36).¹⁹

Two ingredients are still missing from my analysis of factors that determine the use of Russian UIs: emphasis and the grammatical status of the pronoun. In what follows, I discuss their role and the way they interact with the principle (35).

As shown in Section 4.3, emphasis by means of the particle *i* facilitates the use of UIs associated with the role of adjunct. The semantic contribution of *i* is similar to the concessive-conditional ‘even if’ meaning [Uryson 2011: 280], which is at the same time a contrastive meaning (on the affinity between ‘even’ and contrastiveness, see, e.g., [Yanko 2001: 56]). Contrastiveness generated by *i* can be of two types. In one type, illustrated by (45), contrastiveness arises in a cleft-like construction. In the second type, as in (46), there is contrastiveness but no cleft.

- (45) *Esli kogda i brala bol'ničnyj, to tol'ko iz-za detej.*
 if ever PTCL take.PST.F.SG sick.leave.ACC then only because.of child.GEN.PL
 ‘[Even] if she ever took sick leave, it was only because of the children.’ [RNC, 2010]

- (46) *Esli gde i obrzovalos' pjatno,*
 if somewhere PTCL form.PST.N.SG stain.NOM

¹⁹ An anonymous reviewer points out that UIs are felicitous in contexts introduced by the epistemic particle *vdrug*, as in (i). I suggest that this is also due to the colloquial nature of *vdrug*.

(i) *Sprosite tam, vdrug kto znaet.*
 ask.IMV.PL there PTCL someone.NOM know.PRS.3SG
 ‘Ask there, someone may know’. [RNC, 2015]

Another context mentioned by the reviewer as particularly felicitous with UIs is an interrogative context with the verb *byt'* ‘to be’, as in (ii). In this case, however, the interrogative pronoun is in fact not a UI but a relative pronoun that occurs in a free relative clause. Not surprisingly, it cannot be substituted with a *nibud'*-pronoun:

(ii) *U tebya est' kogo (*kogo-nibud') naznačit' direktorom?* [RNC, 1997]
 ‘Do you have someone to appoint as a director?’ (lit.: ‘Do you have whom to appoint as a director?’)

dostatočno poteret' ego obyčnym acetonom.
 enough rub.INF it.ACC ordinary.INS acetone.INS

'[Even] if a stain has formed somewhere, it is enough to rub it with ordinary acetone.'
 [RNC, 2004]

In both cases, the positive effect of *i* can be explained with reference to the principle (35). According to the common assumptions, contrastiveness makes the hearer generate a set of alternatives (cf., e.g., [Vallduví, Vilks 1998; Yanko 2001: 47]). For the sentences both with and without cleft, contrastiveness associated with the concessive-conditional semantics evokes the speaker to compare the positive alternative *P* with the negative one, $\neg P$ ('a stain has formed somewhere' vs. 'a stain has not formed anywhere' for (46)), with the former being less probable than the latter. This, on the one hand, emphasizes that the speaker is unaware of the truth-value of the conditional proposition. On the other hand, the low probability of the positive alternative is highlighted. Both points serve to strengthen what makes any conditional suitable for the use of UIs according to the principle (35): a bias toward the non-specific interpretation of an indefinite contained in the conditional clause.

Cleft constructions create particularly favorable (in the sense of (35)) conditions for the use of UIs. The point is that indefinites in such constructions can only have the non-specific interpretation, i.e. the construction itself predetermines the non-specific function of the pronoun (it seems that the German so-called reduced conditionals investigated by Schwarz [1998] display similar properties, see also [Bhatt, Pancheva 2006: 676]). To demonstrate this, it makes sense to compare conditionals with cleft and standard conditionals without cleft. Although, as assumed above, the latter are biased toward the non-specific interpretation of an indefinite, they are not strictly limited to such interpretation. This is evidenced by the fact that Russian indefinites of the *koe*-series, which are used when the referent is specific and the speaker knows its identity [Paducheva 1985: 90; Haspelmath 1997: 275], are not banned from conditionals, cf. (47).²⁰ However, the *koe*-pronoun is ungrammatical in cleft conditionals with the particle *i*. No examples with *koe-kto* like (48) have been attested in the RNC.

- (47) *Esli koe-kto iz nižeperečislennyx person vystupit*
 if INDF-who.NOM from following.GEN.PL person.GEN.PL perform.FUT.3SG
s Mariej duëtom, eë akcii navernjaka pojduť v goru.
 with Marija.INS duet.INS her share.NOM.PL surely go.FUT.3PL in mountain.ACC
 'If some of the following persons perform a duet with Maria, her shares will surely go up.' [RNC, 2007]

- (48) *Esli kto-to (??koe-kto) i pridët k nam na pomošč',*
 if who.NOM-INDF INDF-who.NOM PTCL come.FUT.3SG to we.DAT to aid.ACC
to ne Mark.
 then NEG Mark.NOM
 'If anyone comes to our aid, it won't be Mark.'

Contrary to emphasis (or contrastiveness), the role that the grammatical status of the pronoun has for the use of UIs does not seem to be directly connected to the principle (35). This role may be not specific to bare interrogatives but be common to indefinite pronouns in general (which, in turn, could be related to the fact that arguments seem to be generally more frequent than adjuncts). Haspelmath [1997: 31] mentions English and German as lacking causal marked indefinites, cf. **somewhy* and **irgendwarum*. However, this type of explanation does not give the whole story, since the comparative frequency of Russian UIs associated with adjuncts differs from that of marked indefinites. As shown in Section 4.1, *gde* 'somewhere' is more

²⁰ Still, they are rare: seven examples with *esli koe-kto* have been attested in the Main corpus and three examples in the Newspaper corpus of the RNC.

frequent than *kogda* ‘ever, some day’ in all time periods, while for the *nibud’*- and *libo*-series the ratio is reversed (in the Newspaper corpus, the ratio of *gde* and *kogda* is 3,225 : 6,284 for *nibud’* and 364 : 3,089 for *libo*). For the *to*-series, *gde* is somewhat more frequent than *kogda* (20,562 : 16,950), but this is most probably due to the fact that *gde-to*, along with the indefinite meaning, has also lexicalized to mean ‘approximately’.

Thus, a factor specific to UIs must also be at play here. I suggest that it is the greater dependence of UIs on context compared to marked indefinite pronouns that hinders the former from being used as adjuncts. Hence, in a sense, the rarity of UIs as adjuncts can be related to the principle in (35).

Indeed, since arguments are more tightly integrated with their heads both semantically and syntactically than adjuncts, their interpretation relies more on the context (e.g., on the meaning of the verbal head) than that of adjuncts. Consider the dialogues in (49) and (50). In (49), the indefinite pronoun *čto-to* ‘something’ is the direct object of the verb *poterjat’* ‘lose’. Due to the selective features of the verb, we know, for instance, that it is most probably a physical object that got lost. In (50), we know little, if anything, on the meaning of the indefinite *začem-to* ‘for some / any reason’ that would be due to the external context and not to the meaning of the pronoun itself. Expectedly enough, the marked indefinite can be substituted by the respective UI in (49) but not in (50).

- (49) — *Poterjal čto-to (OK čto)?*
lose.PST.M.SG what.NOM-INDF something.ACC

— *Da, pasport.*
yes passport.ACC

‘— Have you lost something? — Yes, my passport.’

- (50) — *Ne mogu najti pasport.*
NEG can.PRS.1SG find.INF passport.ACC
— *Možet, ty ego sprjatal začem-to (*začem)*
maybe you.NOM it.ACC hide.PST.M.SG for.what.reason-INDF for.some.reason
i zabył?
and forget.PST.M.SG

‘— I can’t find my passport. — Maybe, you hid it for some reason and forgot about it?’

This line of reasoning is corroborated by the fact that adjuncts of place and time are more felicitous as UIs than adjuncts of reason and purpose (see the corpus data in Section 4.1). The former are sometimes considered to be intermediate between true arguments and true adjuncts: like arguments, they may be subcategorized by the verb, and like adjuncts, they are usually not restricted to a particular morphosyntactic form (cf. the term “adjacts” (= adverbial arguments) introduced for such noun phrases by Feuillet [2006: 383], see also [Testelefs 2001: 189]). Thus, the more pronoun is an argument, the more felicitous it is as a UI, cf. (49) and (50) with (51)²¹.

- (51) — *Ne mogu najti pasport.*
NEG can.PRS.1SG find.INF passport.ACC
— *Možet, ty ego sprjatal gde-to (gde) i zabył?*
maybe you.NOM it.ACC hide.PST.M.SG where-INDF somewhere and forget.PST.M.SG

‘— I can’t find my passport. — Maybe you hid it somewhere and forgot about it?’

²¹ An anonymous reviewer suggests that the UIs’ being more frequent as arguments than as adjuncts could be due to the fact that arguments could prefer a less heavy marking of indefiniteness “simply because they are less autonomous”, cf. the tendency for object pronouns to be clitics. However, this sort of explanation does not seem to differ essentially from the one suggested above: being less autonomous is interrelated with being less predicted by the meaning of the verbal head. In a similar way, Haspelmath [2008] associates the length of a reflexive marker with the likelihood of an anaphoric pronoun being coreferential with the subject.

5.4. On the evolution of Russian UIs: preliminary conclusions

The analysis proposed in this section suggests that the evolution of Russian UIs can be generalized as follows. In the Russian language of the 18th–19th centuries as well as earlier (see [Krys'ko et al. 2020: 178] on the use of UIs in Old Russian), UIs were used more widely than in present-day Russian. As the marked indefinites of *nibud'*- and *to*-series were grammaticalized, the UIs were pushed to the periphery (cf. a similar view in [Malovitskii 1971: 69]). At least for *nibud'*, the process of grammaticalization presumably completed rather late—according to Pen'kova [2016: 281], not earlier than by the 18th century (for *to*-, the origin and the later development are unclear). What determined the choice of the contexts abandoned by UIs (imperatives, subjunctive clauses, etc.) and those in which they persisted (conditionals, questions, and simulative clauses first of all) were mainly the principle formulated in (35) and the register factor. The principle in (35) reflects the fact that the interpretation of a UI strongly relies on context due to the absence of a marker that would express the semantic type of indefiniteness. The role of register, i.e., the tendency to use UIs in the colloquial language, was convincingly interpreted by Malovitskii [1971: 66], who associates the fact that UIs are stylistically non-neutral with their peripheral status compared to *nibud'*- and *to*-pronouns.

6. Typology

The principle (35), which I suggested to be the leading factor in the evolution of Russian UIs in the previous section, allows us to make a few typological predictions.

First, if (35) really holds true, we can expect that conditionals and questions are the environments that consistently allow UIs in languages that make use of bare interrogatives as indefinites. This indeed seems to be the case (see [Haspelmath 1997: 172; Belyaev, Haug 2020: 888]).

Second, specific contexts, such as past or present declarative clauses, are predicted by (35) to be cross-linguistically problematic for UIs since they are not biased toward a certain interpretation of an indefinite pronoun, being equally compatible both with the 'specific known' and 'specific unknown' readings. This prediction is also borne out at least for western Indo-European [Haspelmath 1997: 173].

However, there are also languages that make extensive use of bare interrogatives as indefinites both in specific and non-specific contexts. These include, but are not limited to, languages (see the list thereof in [Tret'yakova 2009a: 88–90]) that have no series with a marker, with bare interrogatives being the only type of indefinite pronouns. To take such languages into account, a hypothesis more general than the principle (35) may be put forward:

- (52) The more series of indefinite pronouns with a marker, spanning the functions from 'specific-known' to 'conditional' on the semantic map of indefinite pronouns, there are in a language, the less likely the language is to use UIs in specific contexts. In other words, in a language that has two or more series with a marker that share the functions from 'specific-known' to 'conditional', bare interrogatives are not expected to be used as indefinites in specific functions. It may play a role, however, whether, in addition to series that share the specific and non-specific functions, there is also a series with a marker that can be used in both types of functions: this could facilitate the extensive use of UIs, see the discussion for German below.

The logic that lies behind the assumption in (52) is as follows. In case there are at least two different series with a marker that share specific and non-specific functions in a language, a semantic split within the zone of specificity turns out to be grammatically drawn, hence, salient for the speakers of this language (see, e.g., [Jakobson 1971: 481] on how grammars define as to what

“aspects of experience must be expressed”). Therefore, a need arises to differentiate between different types of specificity also in case a marker is absent. It is generally in this case that the principle (35) comes into force. While conditional or interrogative contexts can serve to indicate the non-specific interpretation of an indefinite, the contexts corresponding to the specific functions cannot help to distinguish between ‘specific known’ and ‘specific unknown’, hence, cannot individuate the function on their own.

To illustrate this idea, a small sample of eight languages (Classical Greek, Chinese, Korean, Serbian, Polish, Lithuanian, German, and Old Russian) is considered below. Three of them (Classical Greek, Chinese, and Korean) do use UIs in specific functions and have no or only one series with a marker for functions from ‘specific-known’ to ‘conditional’, which is in line with (52). The other five languages, on the contrary, do not allow UIs in specific functions (with German being an exception to be discussed below) and possess at least two series with a marker for the zone from ‘specific known’ to ‘conditional’. This is also in accordance with (52).

In Classical Greek, UIs (cf. the indefinite pronoun *τις* ‘someone/anyone’ and the corresponding interrogative pronoun *τις* ‘who’) can be used both in ‘specific known’, ‘specific unknown’, ‘irrealis non-specific’, ‘conditional’, and ‘question’ functions [Boas et al. 2019: 356]. Importantly, no series with a marker are allowed in these functions [Schwyzer, Debrunner 1950: 213].

Chinese, according to Haspelmath [1997: 307], restricts the use of bare interrogatives as indefinites to non-specific functions. However, Gärtner [2010: 13, fn. 21] notices that Haspelmath himself [1997: 171] gives an example that can be interpreted as a ‘specific unknown’ use of the bare interrogative *shenme* ‘what’. Furthermore, Tret’yakova [2009a: 112] reports that bare interrogatives can be used in Chinese in the ‘specific known’ function. Generic nouns can also be used in specific functions; there is, however, no series with a marker and, in general, no alternative to UIs for the ‘irrealis non-specific’, ‘conditional’, and ‘question’ functions [Haspelmath 1997: 307]. The resulting picture is thus in line with (52).

Korean makes extensive use of UIs, which are allowed both in specific and non-specific functions with the only exceptions being ‘comparative’, ‘direct negation’, and ‘free-choice’ (i.e., the functions on the right of the semantic map) according to Haspelmath [1997: 314] and ‘direct negation’ and ‘free-choice’ according to Tret’yakova [2009a: 148]. Crucially, Korean possesses only one series with a marker that spans all the functions from ‘specific known’ to ‘conditional’ [Haspelmath 1997: 314].

In Serbian, UIs are not allowed in specific functions [Tret’yakova 2009b: 145]. As for the series with markers, their functional distribution is interpreted differently by Haspelmath [1997: 270] and Tret’yakova [2009b: 146]. But according to both descriptions, Serbian has at least two series with a marker that share the zone from ‘specific known’ to ‘conditional’. This is again congruent with what is predicted by (52).

In Polish, the situation is very similar to that in Serbian. UIs are limited to non-specific contexts [Tret’yakova 2009b: 144], while the functions from ‘specific known’ to ‘conditional’ are shared by two series with a marker [Haspelmath 1997: 271; Tret’yakova 2009b: 144].

In Lithuanian, there are three series with a marker that can be used in the functions from ‘specific known’ to ‘conditional’ [Haspelmath 1997: 275; Kozhanov 2011: 104], while UIs are limited to non-specific functions [Kozhanov 2011: 90–91]. Interestingly, Kozhanov [2015: 469] reports that bare interrogatives may be used as indefinites with specific functions in dialects, but in this case they tend to replace the series with a marker used for ‘specific’ and ‘non-specific’ functions, i.e., the latter distinction is likely to disappear.

In German, there are several series with a marker that can be used in specific and non-specific functions [Haspelmath 1997: 244]. At the same time, German makes more extensive use of bare interrogatives than Russian — it allows bare interrogatives in affirmative declarative clauses in the ‘specific-unknown’ function [Tret’yakova 2009a: 132]. Cf. the contrast in (53).

(53) a. GERMAN: *Da kommt wer.*
 there come.PRS.3SG someone.NOM

b. RUSSIAN: **Idët kto.*
 come.PRS.3SG someone.NOM

Intended meaning: ‘Someone is coming.’

This seems to contradict the hypothesis in (52). However, the use of UIs may be facilitated by the fact that the pronoun *etwas* ‘something/anything’ can be used in both specific and non-specific functions including ‘specific known’ [Haspelmath 1997: 245], thus smoothing out the distinctions between different types of indefiniteness for native speakers of German.

Finally, in Old Russian (Indo-European, East Slavic), UIs were used much more extensively than in Modern Russian [Krys’ko et al. 2020: 178]. Still, they were limited to non-specific functions [Penkova, Rabus 2021: 245; Malovitskii 1971: 65]. At the same time, there were two major series with markers: one, with the marker *ne-*, had both specific and non-specific uses [Krys’ko et al. 2020: 178], while the other, with the marker *ljubo*, to which the modern *-libo* goes back, could only be used in non-specific functions [Penkova, Rabus 2021: 241, 245]. All in all, the situation in Old Russian fits the assumption in (52).

The eight languages considered above thus provide a piece of evidence for (52). Of course, much more typological work is needed to verify this hypothesis.

7. Conclusion

In this paper, I have shown that UIs have become less frequent in Russian compared to the 18th–19th centuries, while their present-day distribution differs from the one assumed in previous studies. Contrary to expectations, in modern Russian UIs are almost absent from ‘irrealis non-specific’ contexts such as imperatives or subjunctive clauses, in which they could be used as early as the 19th century, but are allowed in at least two ‘specific unknown’ contexts — simulative ‘as if’ clauses and questions associated with the speaker’s positive expectations. Although the latter contexts have been listed as compatible with UIs in earlier works, so far they have been treated as ‘irrealis non-specific’. I suggest, however, that both display important traits of the ‘specific unknown’ function. This is confirmed in particular by the fact that both in simulative clauses and questions associated with the speaker’s positive expectations, the indefinite *to*-series, which can be used in the ‘specific unknown’ function, is highly preferred over the *ni-bud*’-series, excluded from the ‘specific unknown’ function.

I argue that the distribution of Russian UIs, both synchronically and diachronically, can be accounted for by some sort of **economy principle**, according to which UIs tend to be used in contexts that are strongly biased toward a certain interpretation of an indefinite in terms of specificity (specific known vs. specific unknown vs. non-specific). This is in contrast with marked indefinites, which do not depend on the type of context in the same way since they convey the semantic type of indefiniteness by means of the marker. This principle, as I suggest, allows one to explain a few facts revealed by my corpus study, including the following:

- UIs are relatively frequent in conditionals, questions and simulative ‘as if’ clauses compared to other contexts both in modern Russian and the Russian language of the 18th–19th centuries.
- UIs were relatively rare in imperatives and subjunctive clauses in the 18th–19th centuries and are almost absent from them in present-day Russian.

The reasoning that underlies this analysis is that conditionals, questions or simulative clauses are more strongly biased toward a certain, namely non-specific, interpretation of an indefinite than imperatives and subjunctive clauses, as well as most of other ‘irrealis non-specific’ contexts.

This, in turn, is connected to the truth-value of the proposition in question and ultimately to what the speaker may or may not assume on the existence of the situation or item referred to by the indefinite pronoun, with existence being strongly associated with specific and non-existence with non-specific uses.

It is this kind of context dependency, I suppose, that underlies the fact that synchronically, Russian UIs do not occupy a contiguous area on the semantic map (see Fig. 4 in Section 5.1). As assumed by Haspelmath [1997: 119–120], the semantic map is arranged in such a way that all functions that share the same value of several relevant parameters, the specific vs. non-specific parameter being one of them, form a contiguous area on the map. While this provides a contiguous arrangement on the map for marked indefinites, for UIs a separate parameter, namely the extent to which a function **predicts** a certain type of specificity of the pronoun, appears to be the most crucial one. This parameter produces a different configuration of functions, resulting in a non-contiguous arrangement of UIs on the map. Simulative clauses are an instance of a context in which the latter parameter goes contrary to other parameters. On the one hand, simulative contexts are specific in that they are associated with the speaker's commitment to the existence of an individual or an item referred to by the pronoun in a possible world (see Section 5.1). Being specific, they are not adjacent to conditionals and questions, two functions in which UIs are frequent, on the map. On the other hand, simulative clauses are irrealis. As such, they strongly predict an indefinite within them to be non-specific, hence, are favorable contexts for UIs like conditionals and questions. Essentially the same point may also be stated from the diachronic point of view. In the 18th–19th centuries, Russian UIs occupied a contiguous area on the semantic map (see Fig. 5 in Section 5.1). As their use declined, what fell out first were 'irrealis non-specific' contexts, which are less clearly biased toward the non-specificity of an indefinite than conditionals, questions, and simulative clauses. This gave rise to a gap in the semantic map.

I argue, furthermore, that the suggested economy principle may be overridden by one more factor, namely register: optative constructions with *xot' by* and epistemic constructions with *možet*, which are 'irrealis non-specific' contexts and as such are predicted to be problematic for UIs by the above principle, combine with UIs rather frequently because they are colloquial. Importantly, this offers an account for the following fact characterized as mysterious in the previous literature: neither optative nor epistemic contexts are consistent in their combinability with UIs, i.e. while some epistemic and optative markers do combine with UIs (e.g., *xot' by* and *možet*), others do not.

Two typological tendencies are predicted to hold by my analysis. On the one hand, conditionals and questions are expected to allow UIs in languages that use bare interrogatives as indefinites. On the other hand, past or present declarative clauses, which are equally compatible with 'specific known' and 'specific unknown' readings of the indefinites, are expected to avoid UIs cross-linguistically. Both tendencies have indeed been reported in the literature.

However, there are also languages in which UIs may be used in all kinds of specific functions. I tentatively argue that in these languages, a more general principle comes into force: the more different markers of indefiniteness, spanning the functions from 'specific known' to 'conditional', there are in a language, the more salient the contrast is between different types of specificity for native speakers, and hence the stronger the dependency of UIs on context and the less likely their occurrence in specific contexts. This hypothesis has been preliminarily tested and confirmed based on the data of eight languages (Classical Greek, Chinese, Korean, Lithuanian, Serbian, Polish, German, and Old Russian).

My study also shows the benefits of micro-diachronic analysis in grammatical studies. In previous works on UIs such as [Tret'yakova 2009a] and [Zevakhina, Oskol'skaya 2013], data from the 19th and early 20th centuries were considered as indicative for present-day Russian. As a result, the shift that occurred in the use of UIs since the 19th century could not be registered, and the parallels with other processes in which UIs are cross-linguistically involved could not be analyzed.

ABBREVIATIONS

1, 2, 3 — 1st, 2nd, 3rd person
 ACC — accusative
 AUX — auxiliary
 COMP — complementizer
 COP — copula
 DAT — dative
 F — feminine
 FUT — future
 GEN — genitive
 IMV — imperative
 INDF — indefinite
 INF — infinitive
 INS — instrumental
 LOC — locative

M — masculine
 N — neuter
 NEG — negation
 NOM — nominative
 OPT — optative
 PASS — passive
 PL — plural
 PRS — present
 PST — past
 PTCL — particle
 PTCP — participle
 SG — singular
 SBJV — subjunctive
 Q — question particle

REFERENCES

- Barkhudarov et al. 1978 — Barkhudarov S. G. et al. (eds.). *Slovar' russkogo yazyka XI–XVII vekov* [Dictionary of the Russian language of the 11th–17th centuries]. No. 5. Moscow: Nauka, 1978.
- Belyaev, Haug 2014 — Belyaev O., Haug D. The genesis of *wh*-based correlatives: From indefiniteness to relativization. Handout presented at the conf. Sinn und Bedeutung 19 (Formal Theories of Meaning Change). Göttingen, 15–17 September 2014.
- Belyaev, Haug 2020 — Belyaev O., Haug D. The genesis and typology of correlatives. *Language*, 2020, 96(4): 874–907.
- Bhatt, Pancheva 2006 — Bhatt R., Pancheva R. Conditionals. *The Blackwell companion to syntax*. Vol. 1. Everaert M., van Riemsdijk H. (eds.). Oxford: Blackwell, 2006, 638–687.
- Boas et al. 2019 — van Emde Boas E., de Bakker M., Rijksbaron A., Huitink L. *Cambridge grammar of Classical Greek*. Cambridge: Cambridge Univ. Press, 2019.
- Dobrushina 2009 — Dobrushina N. R. Semantika chastits *by i b* [Semantics of the particles *by* and *b*]. *Korpusnye issledovaniya po russkoi grammatike* [Corpus studies in Russian grammar]. Kiseleva Ks. L., Plungian V. A., Rakhilina E. V., Tatevosov S. G. (eds.). Moscow: Probel-2000, 2009, 281–313.
- Dobrushina 2012 — Dobrushina N. R. Subjunctive complement clauses in Russian. *Russian Linguistics*, 2012, 36(2): 121–156.
- Feuillet 2006 — Feuillet J. *Introduction à la typologie linguistique*. Paris: Honoré Champion, 2006.
- Gärtner 2009 — Gärtner H. More on the indefinite-interrogative affinity: The view from embedded non-finite interrogatives. *Linguistic Typology*, 2009, 13: 1–37.
- Giannakidou, Mari 2019 — Giannakidou A., Mari A. *Modalization and bias in questions*. Ms., 2019. <http://home.uchicago.edu/~giannaki/pubs/modalquestions.pdf>.
- Haspelmath 1997 — Haspelmath M. *Indefinite pronouns*. Oxford: Clarendon Press, 1997.
- Haspelmath 2003 — Haspelmath M. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. *The new psychology of language*. Vol. 2. Tomasello M. (ed.). New York: Erlbaum, 2003, 211–243.
- Haspelmath 2008 — Haspelmath M. A frequentist explanation of some universals of reflexive marking. *Linguistic Discovery*, 2008, 6(1): 40–63.
- Jakobson 1971 — Jakobson R. *Selected writings*. Vol. II: *Word and language*. The Hague: Mouton, 1971.
- Kobozeva 1981 — Kobozeva I. M. Opyt pragmaticheskogo analiza *to-* i *nibud'*-mestoimenii. [Towards a pragmatic analysis of the *to-* and *nibud'*-pronouns]. *Izvestiya AN SSSR. Seriya literatury i yazyka*, 1981, 40(2): 165–172.
- Kobozeva 2000 — Kobozeva I. M. *Lingvisticheskaya semantika* [Linguistic semantics]. Moscow: Librokom, 2000.
- Kozhanov 2011 — Kozhanov K. A. Notes on the use of Lithuanian indefinite pronouns. *Baltic Linguistics*, 2011, 2: 79–110.

- Kozhanov 2015 — Kozhanov K. A. Lithuanian indefinite pronouns in contact. *Contemporary approach to Baltic linguistics*. Arkadiev P., Holvoet A., Wiemer B. (eds.). Berlin: De Gruyter, 2015, 465–490.
- Kratzer, Shimoyama 2002 — Kratzer A., Shimoyama J. Indeterminate pronouns: The view from Japanese. *The Proc. of the 3rd Tokyo Conf. on Psycholinguistics*. Otsu Y. (ed.). Tokyo: Hituzi Syobo, 2002, 1–25.
- Krys'ko et al. 2020 — Krys'ko V. B., Kuznetsov A. M., Pen'kova Y. A. Mestoimeniya [Pronouns]. *Istoricheskaya grammatika russkogo yazyka: Entsiklopedicheskii slovar'* [Historical grammar of Russian: Encyclopaedic dictionary]. Krys'ko V. B. (ed.). Moscow: Azbukovnik, 2020, 172–189.
- Kuz'mina 1989 — Kuz'mina S. M. Semantika i stilistika neopredelennykh mestoimenii [Semantics and stylistics of indefinite pronouns]. *Grammaticheskie issledovaniya. Funktsional'no-stilisticheskii aspekt. Supersegmentnaya fonetika. Morfologicheskaya semantika* [Grammatical studies. Functional-stylistic aspect. Suprasegmental phonetics. Morphological semantics]. Shmelev D. N. (ed.). Moscow: Nauka, 1989, 178–231.
- Letuchiy 2017 — Letuchiy A. B. Sravnitel'nye konstruktii [Comparative constructions]. *Materialy k korpusnoi grammatike russkogo yazyka* [Materials for a corpus-based grammar of Russian]. No. 2: *Sintaksicheskie konstruktii i grammaticheskie kategorii* [Syntactic constructions and grammatical categories]. Plungian V. A., Stoinova N. M. (eds.). St. Petersburg: Nestor-Istoriya, 2017, 132–204.
- Liu et al. 2021 — Liu M., Rotter S., Giannakidou A. Bias and modality in conditionals: Experimental evidence and theoretical implications. *Journal of Psycholinguistic Research*, 2021, 50: 1369–1399. <https://doi.org/10.1007/s10936-021-09813-z>.
- Malovitskii 1971 — Malovitskii L. J. Voprosy istorii predmetno-lichnykh mestoimenii (*kto, chto*) [Issues of history of substantival pronouns (*kto, chto*)]. *Uchenye zapiski Leningradskogo gosudarstvennogo pedagogicheskogo instituta im. Gertsena*, 1971, 517: 3–130.
- Oskol'skaya, Zevakhina 2013 — Oskol'skaya S. A., Zevakhina N. A. Neopredelennye mestoimeniya bez markera neopredelennosti v russkom yazyke [Indefinite pronouns without a marker of indefiniteness in Russian]. Handout presented at the international conf. “Constructional and Lexical Semantic Approaches to Russian” (St. Petersburg, 2013).
- Paducheva 1985 — Paducheva E. V. *Vyskazyvanie i ego sootnesennost' s deistvitel'nost'yu* [Utterance and its relationship to reality]. Moscow: Nauka, 1985.
- Pen'kova 2016 — Pen'kova Ya. A. Neopredelennye mestoimeniya s formantami *nibud'* i *libo/lyubo* v istorii russkogo yazyka XV–XVII vekov: Korpusnoe issledovanie [Indefinite pronouns of the *nibud'*- and *libo/lyubo*-series in the history of Russian of the 15th–17th centuries: A corpus study]. *El'Manuscript–2016. Rašytinis palikimas ir informacinės technologijos. VI tarptautinė mokslinė konferencija Pranešimai ir tezės* (Vilnius, 2016). Baranov V. A., Timchenko T. A. (eds.). Vilnius; Iževskas: Loreta Uzdraite, 2016, 280–284.
- Penkova, Rabus 2021 — Penkova Y., Rabus A. East Slavic indefinite pronouns: A corpus-based approach. *Russian Linguistics*, 2021, 45: 227–252. <https://doi.org/10.1007/s11185-021-09247-0>.
- Rakhilina et al. 2016 — Rakhilina E. V., Borodina M. A., Reznikova T. I. “Taman' segodnya”: korpusnoe issledovanie russkogo yazyka XIX veka [“Taman today”: A corpus study of 19th-century Russian]. *Proc. of the V. V. Vinogradov Russian Language Institute*, 2016, 10: 242–255.
- RNC — *Russian National Corpus*. <https://ruscorpora.ru>.
- Shmelev 2002 — Shmelev A. D. *Russkii yazyk i vneyazykovaya deistvitel'nost'* [Russian language and extralinguistic reality]. Moscow: Yazyki Slavyanskoi Kul'tury, 2002.
- Schwarz 1998 — Schwarz B. Reduced conditionals in German: Event quantification and definiteness. *Natural Language Semantics*, 1998(6): 271–301.
- Schwyzler, Debrunner 1950 — Schwyzler E., Debrunner A. *Griechische Grammatik: auf der Grundlage von Karl Brugmanns griechischer Grammatik*. 2. Bd.: *Syntax und syntaktische Stilistik*. München: C. H. Beck'sche Verlagsbuchhandlung, 1950.
- Shvedova 1980 — Shvedova N. Yu. (ed.). *Russkaya grammatika* [Russian grammar]. Vol. 2: *Sintaksis* [Syntax]. Moscow: Nauka, 1980.
- Tatevosov 2002 — Tatevosov S. G. *Semantika sostavlyayushchikh imennoi gruppy: kvantornye slova* [The semantics of constituents of a noun phrase: Quantifier words]. Moscow: Gorky Institute of World Literature, 2002.
- Testeleits 2001 — Testeleits Ya. G. *Vvedenie v obshchii sintaksis* [Introduction to general syntax]. Moscow: Russian State Univ. for the Humanities, 2001.
- Tret'yakova 2004 — Tret'yakova O. D. Seriya neopredelennykh mestoimenii s nulevym markerom neopredelennosti v russkom yazyke [A series of indefinite pronouns with a zero marker of indefiniteness

- in Russian]. *Computational linguistics and intellectual technologies. Proc. of the international conf. "Dialog-2004"*. Kobozeva I. M., Narin'yan A. S., Selegei V. P. (eds.). Moscow, 2004, 602–607.
- Tret'yakova 2009a — Tret'yakova O. D. *Neopredelennye mestoimeniya, lishennye markera neopredelennosti, v tipologicheskoi perspektive* [Indefinite pronouns lacking overt markers of indefiniteness in a typological perspective]. Candidate diss., Lomonosov Moscow State Univ., 2009.
- Tret'yakova 2009b — Tret'yakova O. D. Funktsii neopredelennykh mestoimenii bez pokazatelya neopredelennosti (na materiale slavyanskikh yazykov) [Functions of indefinite pronouns lacking overt markers of indefiniteness in Slavic]. *Vestnik Moskovskogo universiteta, seriya 9: Filologiya*, 2009, 1: 134–147.
- Uryson 2011 — Uryson E. V. *Opyt opisaniya semantiki soyuzov* [An attempt of description of conjunctions semantics]. Moscow: Yazyki Slavyanskikh Kul'tur, 2011.
- Vallduví, Vilksuna 1998 — Vallduví E., Vilksuna M. On rheme and kontrast. *Syntax and Semantics*, 1998, 29 (*Syntax at the limits*. Culicover P., McNally L. (eds.)): 79–108.
- Yanko 2001 — Yanko T. E. *Kommunikativnye strategii russkoi rechi* [Communicative strategies of Russian speech]. Moscow: Yazyki Slavyanskoi Kul'tury, 2001.
- Yanko-Trinit'skaya 1977 — Yanko-Trinit'skaya N. A. Mestoimennyye slova so znacheniem neopredelennosti [Pronominal words with the meaning of uncertainty]. *Russkii yazyk v shkole*, 1977, 1: 71–73.
- Yanovich 2005 — Yanovich I. Choice-functional series of indefinites and Hamblin semantics. *Proc. of Semantics and Linguistic Theory (SALT) 15*. Georgala E., Howell J. (eds.). Ithaca (NY): Cornell Linguistics Publications, 2005, 309–326.

Получено / received 14.09.2022

Принято / accepted 27.09.2022