Russian indefinite markers *to* and *nibud'*: a microdiachronic approach

Olga E. Pekelis[1,2]

Abstract

This paper deals with the evolution of the Russian indefinite series *to* and *nibud'* after the 18th century. The Russian National Corpus data show that in the 18th–19th centuries, *to* was almost strictly specific, while *nibud'* was slightly less non-specific compared to modern Russian. In the 20th century, *to* expanded to non-specific contexts but, crucially, did so inconsistently: it overtook *nibud'* in frequency in conditionals and questions, but not in imperatives. These facts raise the following questions: (i) What triggered the expansion of *to*? (ii) Why was this expansion inconsistent? The answers that I suggest are as follows: (i) *to* expanded because its relationship with *nibud'* was close to that of complementary distribution, which is a typologically rare, hence, unstable situation. (ii) *to* expanded to those non-specific contexts that helped to accommodate the originally specific semantics of *to* to the non-specific meaning of the context.

1 Introduction

In modern Russian, there are several series of indefinite pronouns. These include *koe-, to-, nibud'-, libo-, by to ni bylo-, ni-, ljuboj-* and *ugodno*-series; their distribution is presented in Fig. 1 according to the semantic map of Russian indefinite pronouns proposed by Haspelmath (1997, p. 65).[1]

Among these series, the *to-* and *nibud'*-series are the basic ones and are the subject of the present study. They cover several functions in the central (non-emphatic, in terms of Haspelmath, 1997, p. 125) part of the semantic map and, contrary to the multifunctional *libo*-series (Paducheva, 2015), are stylistically neutral.

---

[1] The map in Fig. 1 was further elaborated by Tatevosov (2002, p. 141). Although the map by Tatevosov differs from that by Haspelmath in several respects, in what follows I proceed from the latter map for two reasons. Firstly, the corrections introduced by Tatevosov are not relevant to the uses of indefinite pronouns considered in the present paper. Secondly, I take into account the typological data on the use of indefinite pronouns in other languages provided by Haspelmath (1997). It seems reasonable to use all the maps from one and the same source.

✉ O.E. Pekelis
   opekelis@gmail.com

1   Russian State University for the Humanities, Miusskaya sq. 6, Moscow, 125993, Russia

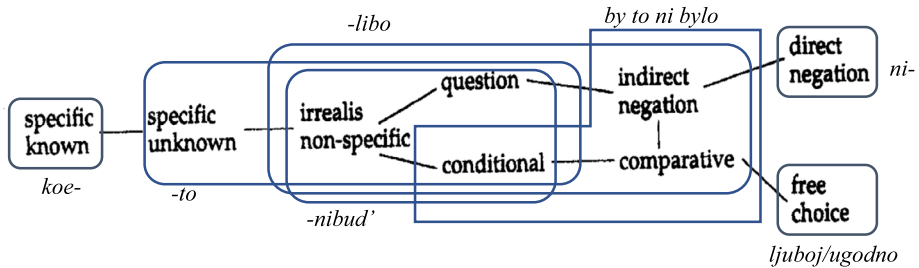2   HSE University, Myasnitskaya ulitsa 20, Moscow, 101000, Russia

Fig. 1 Boundaries of Russian indefinites according to Haspelmath (1997, p. 65)

Of the four functions that fall under the boundaries of *to* and *nibud'* in Fig. 1, in one, namely the 'specific unknown' function, indefinite pronouns are specific, while in the other three functions they tend to be non-specific. According to common assumptions, an important criterion of the specific use of an indefinite pronoun is the commitment of the speaker to the existence of the referent (Gärtner, 2009, p. 7, see also Haspelmath 1997, p. 38). This is the case in example (1) with an indefinite within a declarative past clause – a context that corresponds to the 'specific unknown' function. In line with Fig. 1, only the *to*-pronoun can be used here. A non-specific pronoun, on the contrary, is usually not associated with the speaker's commitment to the existence of the referent. The example in (2), which contains an indefinite within a conditional clause and is an instance of the 'conditional' function, is illustrative. In (2), both *to* and *nibud'* are admissible.[2]

(1)   *Vyjasnilos', čto kto-to* ([??]*kto-nibud'*) *priexal iz Moskvy.* (RNC, 1994–2003)
      'It turned out that someone came from Moscow.'

(2)   *Esli kto-to* ([OK]*kto-nibud'*) *priedet, to zavtra uvidimsja.* (RNC, 2012)
      'If someone comes, then we'll see each other tomorrow.'

It has been shown (see in particular Geist, 2008) that the Russian *to*- and *nibud'*-series also differ in terms of referential anchoring: with the *to*-marker, the commitment to the existence of the referent may be anchored not to the speaker but to some other discourse entity, most typically the subject of the matrix clause. In (3), it is a 68-year-old woman who believes that someone broke into her house. Although the speaker does not seem to share this view, only the *to*-pronoun is admissible in this case:

(3)   *68-letnjaja ženščina* […] *soobščila, čto kto-to* ([??]*kto-nibud'*) *pronik v ejo dom.* (RNC, 2017.08)
      'A 68-year-old woman reported that someone broke into her house.'

In what follows, I adopt a simplified approach to specificity. I do not differentiate point-of-view holders, assuming the presence of someone (be it the speaker or another discourse entity) who is committed to the existence of the referent to be criterial for specific uses. Further research is needed to discover the diachronic evolution of *to* and *nibud'* with respect to different anchors.

---

[2]Most Russian examples cited in this paper were sourced from the Russian National Corpus (RNC). The number next to the source of the example indicates the year or period the text was created. For the examples taken from the newspaper subcorpus of the RNC, both year and month of creation are reported. I have constructed the examples for which the source is not indicated.

The boundaries of *to* and *nibud'* shown in Fig. 1 are assumed to represent their use in modern Russian. However, there is a reason to think that as early as the 19th century these boundaries were different. In the middle of the 20th century, Fursenko (1969, p. 349) wrote about the expansion of *to*. This suggests that at the beginning of the 20th century and/or earlier the distribution of *to* was narrower than in Fig. 1.

In this paper, I check this hypothesis based on Russian National Corpus (RNC) data. The questions that I address are as follows:

(i)   What were the boundaries of *to* and *nibud'* in the 18th–19th centuries?
(ii) How did these boundaries evolve after the 19th century?
(iii) What were the triggers of this evolution?
(iv) What does this evolution mean from the typological point of view?

RNC data have shown that in the 18th–19th centuries, the distribution of *to* and *nibud'* indeed differed from the modern one: the relationship between *to* and *nibud'* was close to complementary distribution, as the *to*-series was almost strictly 'specific'. Based on the data on indefinite pronouns in 40 languages, elaborated by Haspelmath (1997), I argue that the relationship of complementary distribution between the specific and non-specific series of indefinites is a typological rarity, suggesting that this was a trigger for the further expansion of *to*.

However, I also show that *to* expanded to non-specific contexts inconsistently and *not* in accordance with typological expectations. In particular, its way of expansion is not predicted by the semantic map approach to the diachronic development of indefinite pronouns. As an alternative explanation, I suggest that *to* expanded to those non-specific contexts that helped to accommodate its original specific meaning to the non-specific meaning of the context.

The rest of the paper is structured as follows. In Sect. 2, I present the corpus data on the distribution of *to* and *nibud'* after the 18th century. I go on to analyse these data in Sect. 3, arguing that my analysis provides an explanation for the way *to* and *nibud'* evolved. In Sect. 4, the issue of complementary distribution between specific and non-specific indefinite markers is considered from the typological point of view. This, as I suggest, serves to shed light on what triggered the expansion of the Russian *to*. Section 5 concludes.

2 *To* and *nibud'* after the 18th century: the corpus data

This section presents corpus data and their statistical analysis. After an overview of the factors taken into account in the study in Sect. 2.1, I report the results of the study in Sects. 2.2 (descriptive part) and 2.3 (statistical part).

2.1 Factors under scrutiny

To form an indefinite pronoun, Russian indefinite markers *to* and *nibud'* attach to interrogative pronouns of different ontological categories, cf. *kto* 'who' vs. *kto-to* 'somebody', *čto* 'what' vs. *čto-nibud'* 'anything' etc. I concentrated on the ontological categories 'person', 'object', and 'property', i.e. on the pronouns *kto-to* 'someone', *kto-nibud'* 'anyone', *čto-to* 'something', *čto-nibud'* 'anything', *kakoj-to* 'some' and *kakoj-nibud'* 'any'. I also collected data on two pronouns of the *libo*-series – *kto-libo* and *kakoj-libo* – to make sure that the *libo*-series did not interfere significantly with the competition between the *nibud'*- and the

to-series in either of the periods under scrutiny (see the boundaries of *libo* in modern Russian in Fig. 1). I found that *libo* is used only sporadically in the contexts under consideration (see the list thereof below) in all periods, the two exceptions being questions and conditionals. In the latter contexts, *libo* is slightly behind *nibud'* in terms of frequency. This is in line with the assumption that in the contexts where the *nibud'*-series and the *libo*-series overlap, the difference between them is stylistic: the *nibud'*-series is more colloquial, while the *libo*-series is more formal (Haspelmath, 1997, p. 65, Paducheva, 2015). A similar view was suggested by Penkova (2016) for the Russian language of 15th–17th centuries, where the distribution of *nibud'* across contexts was broader compared to modern Russian and coincided with that of *libo*. I thus do not take *libo* into account in what follows.

I compared the data from three subcorpora of the RNC that represent three historical periods: the 18th–19th century subcorpus, the 20th century subcorpus (both being part of the main corpus) and the newspaper corpus containing media texts from the 1980–2000s. A shortcoming of this approach is that the texts to be compared turn out to be of different genres. However, the contrast is not that strong as the newspaper corpus, in turn, is heterogeneous in terms of register and genre – it contains both written and spoken texts, e.g. interviews. At the same time, media texts seem to be a more rigorous representation of the modern norm than fiction texts dominating in the main corpus, which I suggest to be even advantageous for the present study (see a more detailed discussion in Sect. 2.2). 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).

I collected the corpus data on the distribution of indefinite pronouns in six contexts: past declarative clauses (4); simulative 'as-if'-clauses introduced by the subordinators *slovno*, *budto*, *kak budto* and *točno* (5); imperative contexts with the second-person imperative forms (6); future declarative clauses (7); *yes*/*no*-questions with the question particle *li* (8); and conditional clauses introduced by *esli* 'if' (9).

(4)   *Kto-to priexal; sbegaj, uznaj.* (RNC, 1827–1832)
        'Someone has arrived; go and find out [who].'

(5)   *Iz kuxni slyšalsja gulkij zvuk, točno kto-to xlopal v ladoši.* (RNC, 1949–1956)
        'There was a booming sound from the kitchen, as if someone was clapping their hands.'

(6)   *Pozovite kogo-nibud' na pomošč.* (RNC, 2000.04)
        'Call someone for help.'

(7)   *Kakie-to sredstva otnesjem v bank.* (RNC, 2004.08)
        'We will take some money to the bank.'

(8)   *Izmenilos' li čto-nibud' posle gastrolej?* (RNC, 1989.09)
        'Has anything changed since the tour?'

(9)   *Esli kakie-to dannye menjalis', ix neobxodimo obnovit'.* (RNC, 2021.07)
        'If any data have changed, they must be updated.'

Note that the simulative markers *slovno*, *budto*, *kak budto* and *točno* can in fact introduce clauses of two different types, both of which were considered in my corpus study. One is an adverbial 'as if'-clause exemplified in (5). Another one is a complement clause (10) conveying that the speaker or the subject of the main clause doubts what is being reported in the embedded clause.

(10)   *Mne počudilos', budto kto-to stonet ili voet.* (RNC, 1950–1951)
         'I felt like someone was moaning or howling.'

These six contexts can be divided into groups in terms of functions of indefinite pronouns (cf. Fig. 1) and, more generally, in terms of specificity. Four contexts – imperative, future, question and conditional – belong to the non-specific reference type (Haspelmath, 1997, p. 120), with conditional and question representing functions of the same name, and imperative and future being instances of the 'irrealis non-specific' function on the semantic map of indefinite pronouns. The past context belongs to the specific reference type, more precisely, the 'specific unknown' function (in principle it is also compatible with the 'specific known' function, but no such examples occurred in my sample). Note, however, that a context with an indicative past verb is specific unless the verb is in the scope of a non-veridical operator such as a question or a conditional. Cf. (11), with the verb *otkliknulsja* in the scope of a question:

(11)   *Kto-nibud' otkliknulsja?* (RNC, 2003.08)
         'Has anyone responded?'

As for simulative contexts, both adverbial and complement, I want to suggest that they are also specific (see more on this proposal in Pekelis, 2023). This assumption is far from obvious and needs clarification. On the one hand, the situation introduced by a simulative subordinator is irrealis (Letučij, 2017, p. 180); in (5), no real clapping and in (10), no real moaning is intended. For this reason, simulative clauses are sometimes associated with the 'irrealis non-specific' function of indefinite pronouns (cf. Tretjakova, 2004). On the other hand, however, what the speaker assumes in (5) is that in a possible world, different from the actual, there existed someone who clapped. In the complement clause in (10) the degree of reality is even higher: the speaker believes that someone is moaning but is not sure about it. More generally, by using a simulative marker the speaker is carried away in thoughts in another world in which the referent of the pronoun exists. Now, as noted above, the speaker's commitment to the existence of the referent of an indefinite pronoun is assumed to be crucial for interpreting this pronoun as specific (Gärtner, 2009, p. 7, Haspelmath, 1997, p. 38). The Russian data confirm that Russian simulative clauses are indeed associated with the 'specific unknown' function. A piece of evidence is given by the distribution of the indefinite pronouns *nekto* 'someone' and *nekij* 'some'. On the one hand, both are specific indefinites that are allowed in 'specific known' (i.e., known to the speaker but not to the addressee, as in (12)) and 'specific unknown' (13) contexts but are banned from 'irrealis non-specific' contexts such as imperatives, cf. (14) (Paducheva, 1985, p. 214, Shmelëv, 2002, p. 120 ff.).

(12)   […] *zatejal ėto nekij Ika, tak ego nazyvali, odin iz zavsegdataev čerdaka.* (RNC, 2012)
         'It was started by a certain Ika, as he was called, one of the habitues of the attic.'

(13)   […] *nekto otkryl strel'bu na kampuse biznes-školy.* (RNC, 2011.11)
         'Someone opened fire on the campus of a business school.'

(14)   *Prinesi kakoj-nibud'* (??*nekij*) *paket.* (RNC, 1999.10)
         'Bring some package.'

On the other hand, both *nekto* and *nekij* are admissible in simulative clauses, adverbial (15) or complement (16).

Table 1 Comparative frequency of kto-to and kto-nibud' across contexts and periods (RNC)[a]

	Past		Simulative		Imperative		Future		Question		Conditional	
	to	*nib.*	*to*	*nib.*	*to*	*nib.*	*to*	*nib.*	*to*	*nib.*	*to*	*nib.*
1701–1900	421	10	116	45	0	100	1	11	0	192	10	247
1901–2000	2,621	43	995	17	5	240	81	69	28	416	562	722
Newspaper corpus, 1980–2000s	2,644	52	385	3	40	143	853	46	664	558	8,048	418

[a]All examples were filtered manually except for the contexts in which more than 1,000 examples were re-trieved from the corpus. In the latter case, roughly the first thousand examples were filtered. All samples over 1,000 examples contained few irrelevant examples (e.g., there were five irrelevant examples in the first thou-sand examples with *kto-to* within conditionals retrieved from the newspaper corpus). The following types of queries were used to obtain the data: кто-то, first v & indic & praet & sg & pf, -amark, на расстоянии от 1 до 1 от Слова 1 -бы & -б, на расстоянии от 1 до 1 от Слова 2; как, first | acomma | adash будто, -amark, на расстоянии от 1 до 1 от Слова 1 кто-то, -amark, на расстоянии от 1 до 1 от Слова 2; будто | словно | точно, first | acomma | adash кто-то, -amark, на расстоянии от 1 до 1 от Слова 1; V & imper & 2p кто-то, -amark, на расстоянии от 1 до 1 от Слова 1; кто-то, first v & indic & fut & sg & pf, -amark, на расстоянии от 1 до 1 от Слова 1; -едва & -"что" & -"то" & -вряд, "ли" на расстоянии от 1 до 1 от Слова 2, кто-то, -amark на расстоянии от 1 до 1 от Слова 3; если, кто-то, -amark на расстоянии от 1 до 1 от Слова 1.

Table 2 Comparative frequency of kakoj-to and kakoj-nibud' across contexts and periods (RNC)

	Past		simulative		Imperative		Future		Question		Conditional	
	to	*nib.*	*to*	*nib.*	*to*	*nib.*	*to*	*nib.*	*to*	*nib.*	*to*	*nib.*
1701–1900	169	6	129	41	2	85	1	8	1	194	3	125
1901–2000	697	9	330	26	9	179	10	11	49	255	98	239
Newspaper corpus, 1980–2000s	350	3	83	18	68	209	72	11	1,602	380	1,530	301

(15)   *Klaviši prygali sami! Budto nekij nevidimyj nažimal na nix!* (RNC, 2015)
        'The keys jumped by themselves! As if someone invisible pressed them!'

(16)   *On rasskazal, budto na nego napal nekij neizvestnyj emu čelovek.* (RNC, 2001.08)
        'He said that (lit.: as if) he was attacked by a man unknown to him.'

2.2 Descriptive statistics

The results of my corpus study are presented in Tables 1 (for the pronoun *kto*-), 2 (for *kakoj*-), and 3 (for *čto*-).[3] For simplicity and due to the scarcity of data for the 18th century, the 18th and the 19th centuries are considered as a single period (see below a more detailed analysis).

The data in Tables 1–3 suggest the following conclusions.

A. In four contexts, namely in *conditionals*, *questions*, *future* and *imperative* contexts, the *to*-series was only scarcely used in the 18th–19th centuries but became more frequent in the 21st century. Crucially, imperatives differ from the other three contexts in the following

---

[3]The sample database is accessible at the following link: all_data.xlsx. Note that the RNC does not allow downloading more than 5,000 thousand examples at a time. Due to this restriction, examples with *kto*-pronouns within conditional clauses, for which the sample exceeds 8,000 examples, could not be downloaded in full. All further statistical work was carried out with the downloaded part.

Table 3 Comparative frequency of čto-to and čto-nibud' across contexts and periods (RNC)

	Past		simulative		Imperative		Future		Question		Conditional	
	to	*nib.*	*to*	*nib.*	*to*	*nib.*	*to*	*nib.*	*to*	*nib.*	*to*	*nib.*
1701–1900	129	5	205	16	0	366	9	11	2	355	1	117
1901–2000	903	49	608	3	16	897	45	39	44	429	335	289
Newspaper corpus, 1980–2000s	380	17	174	0	201	640	65	18	692	286	2,816	109

CONDITIONAL
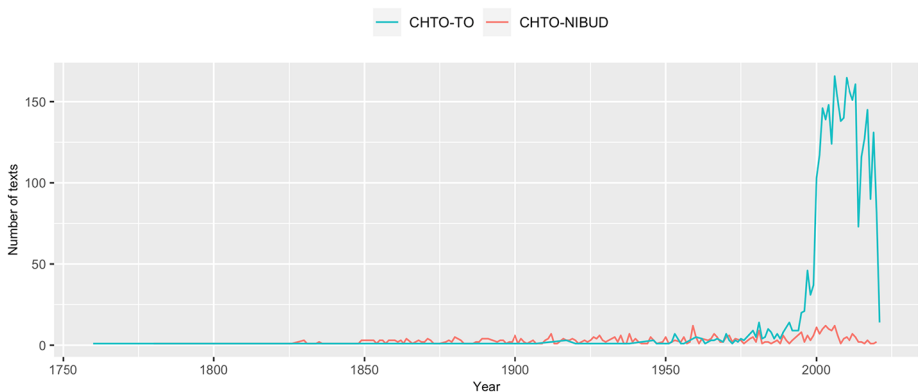
Fig. 2 Comparative frequency of *čto-to* and *čto-nibud'* in conditionals (RNC) (Color figure online)

way: in conditionals, questions, and future contexts the *to*-series has surpassed in frequency the *nibud'*-series, while in imperatives, this did not happen.[4] Figs. 2–5, for *čto*-pronouns, are illustrative.

B. In *simulative* contexts, *to* was more frequent than *nibud'* as early as the 18th–19th centuries. In modern texts, the dominance of *to* has become even more pronounced. This conclusion is distinctly supported by the data concerning *kto*- and *čto*-pronouns. With *kakoj*-pronouns the situation is less straightforward but at a closer examination, lends itself to the same interpretation. In many of the examples with *kakoj-nibud'* from the 18th–19th centuries and the 20th century and in *all* 18 examples with *kakoj-nibud'* from the newspaper corpus (cf. Table 2), *kakoj-nibud'* has an expressive – depreciative or appreciative – interpretation, i.e. it conveys the speaker's attitude toward the referent (Paducheva, 1985, p. 210, Nikolaeva, 2013, p. 275, a.o.). Cf. (17), in which *kakoj-nibud'* betrays the speaker's dismissive attitude toward turners and locksmiths.

(17)   *Genka emu kričit: «Tašči, Vasja!». Jašina nazvat' Vasej, slovno kakogo-nibud' tokarja ili slesarja.* (RNC, 2011.12)
    'Genka shouts to him: "Drag, Vasya!". [Just think –] to call Yashin Vasya, like some kind of turner or locksmith.'

---

[4]The dominance of *to* over *nibud'* in future contexts most probably regards not all future contexts but only that particular type of future context that prevailed in my sample (see details in Sect. 3.4).
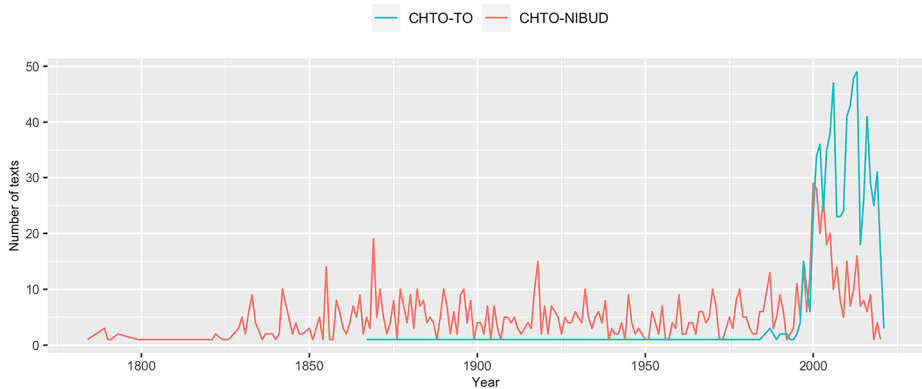
QUESTION

Fig. 3 Comparative frequency of *čto-to* and *čto-nibud'* in questions (RNC) (Color figure online)
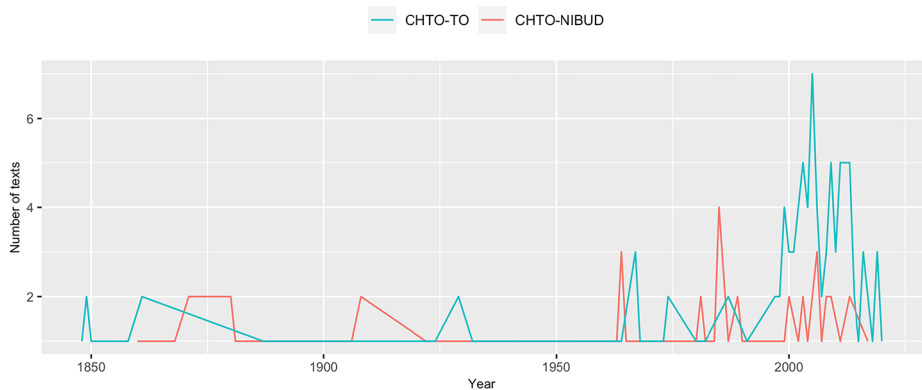
FUTURE

Fig. 4 Comparative frequency of *čto-to* and *čto-nibud'* in future contexts (RNC) (Color figure online)

As demonstrated by Bylinina (2010), the expressive *kakoj-to* differs in its distribution from the "ordinary" indefinite *kakoj-to*. This also seems to be the case for *kakoj-nibud'*. In particular, a depreciative *kakoj-nibud'* may occur in specific contexts, cf. (18), while the modern non-expressive *kakoj-nibud'* may not (see point C below). In (18), the speaker's negative attitude toward *agronomists* is conveyed by *kakoj-nibud'* and further supported by the context:

(18)   *«Kakoj-nibud' agronom privyk, čto v sovetskoe vremja ėtot tovar vsem prixodil be-*
         splatno, — ob"jasnjaet Birišev. — I emu očen' složno ponjat', kak možno ėto do-
         bro prodavat' po predoplate. Takoj čelovek ne budet rabotat' aktivno». Poėtomu v
         «AgroXimAl'janse» agronomov net. (RNC, 2000.02)
         ' "The agronomists (lit.: any agronomist) got used to the fact that in Soviet times
         this product [pesticides] came to everyone free of charge," explains Birišev. "And
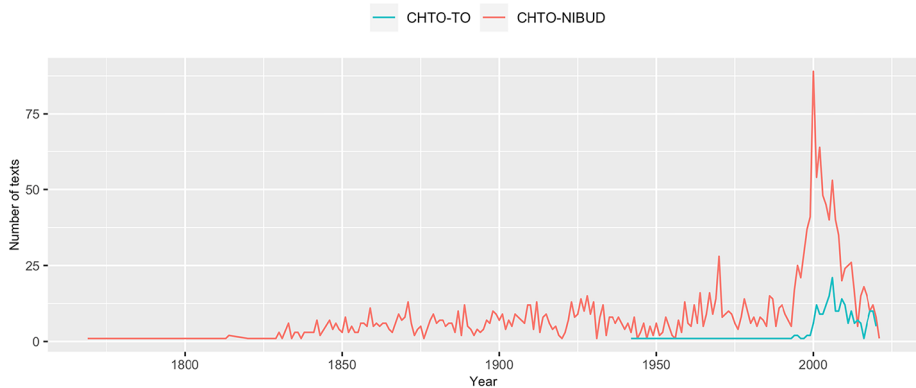
IMPERATIVE

Fig. 5 Comparative frequency of *čto-to* and *čto-nibud'* in imperatives (RNC) (Color figure online)

> it is very difficult for him to understand how it is possible to sell this good on an advance payment. Such a person will not work actively." Therefore, there are no agronomists in AgroKhimAlliance.'

Thus, the "expressive" examples with *kakoj-nibud'*, including all 18 examples attested in the newspaper corpus, do not in fact contain a genuine indefinite and therefore do not compromise the conclusion about the dominance of *to* as an indefinite marker in simulative clauses. Note also that the examples with *to* in simulative contexts, contrary to *nibud'*, are predominantly non-expressive:

(19) *Tol'ko tam, na pervom kurse, pedagogi raskryli menja kak artista — kak budto kakoj-to tumbler vključili.* (RNC, 2002.11)
'Only there, in my first year, the teachers revealed me as an artist – as if some kind of toggle switch was turned on.'

C. In *past* contexts, *to* dominated as early as the 18th–19th centuries. However, a closer look at the examples suggests that *nibud'* had specific uses in the 19th century that have been lost by today. There are a few examples with *nibud'* from the 18th–19th centuries and early 20th century in my sample in which the context is definitely specific, i.e., includes neither explicit nor implicit interrogative, epistemic, habitual or other operators that could render it non-specific. Cf. (20), where *kto-nibud'* refers to a perpetrator of a murder, hence to a person that definitely exists given that the murder has already happened.

(20) *— Ty kak dumaeš? — Ne znaju kto … — I ja ne znaju, konečno. Kto-nibud' ubil že!* (RNC, 1917)
'What do you think? – I don't know who… – Of course, I don't know either. [But] someone did kill!'

In my sample of the late 20th–21st centuries, no similar examples with *nibud'* have been attested. All modern examples with *nibud'* in past contexts are either rendered non-specific by some sort of an explicit or implicit operator, cf. a question in (21), or allow an expressive interpretation of the pronoun, as in (18). Specifically, among the 43 examples with *kto-nibud'*

from the 20th century, 20 examples are questions, and among the 52 examples with *kto-nibud'* from the newspaper corpus, *all* 52 are questions.

(21)   *Kto-nibud' ocenil ėto?* (RNC, 2008.02)
       'Has anyone appreciated this?'

In what follows (see Sect. 3), I focus on conditionals, questions, future and imperative contexts. Therefore, I assessed the evolution of *to* and *nibud'* in these contexts with greater accuracy based on the decision trees method (Therneau & Atkinson, 2022), taking year of creation, source corpus (main or newspaper) and context (conditional, question, imperative or future) as the predictors. I opted for this method as the data I collected is not linear, with a twist in distribution in the last quarter of the 20th century. Now, analyzing nonlinear data with such methods as a mixed-effects regression model is not the optimal way, due to the inherent assumptions of this approach. Most regression models fundamentally assume a linear relationship between the predictors and the dependent variable and as such, are not equipped to capture the twists of the type we found.

The decision trees (see Fig. 9–11 in Appendix) confirm the main conclusion drawn from the data in Tables 1–3: by the 21st century, *to* surpassed *nibud'* in frequency in all contexts under scrutiny but in imperatives. However, as in the decision trees the date is treated as a continuous variable, they provide more detailed information on the evolution of *to* and *nibud'* than Tables 1–3. They show, for example, that in questions *to* overtook *nibud'* in frequency later than in conditional and future contexts – approximately at the turn of the 20th and 21st centuries. In conditional and future contexts, this happened roughly in the third quarter of the 20th century. The decision trees also testify that the transition from *nibud'* to *to* is a consistent trend that breaks only at short time intervals for which there is little data.

As mentioned above, a drawback of my data is that the texts from the newspaper corpus and those from the main corpus are of different genres. The decision trees in Fig. 9–11 (see Appendix) show that the genre variable is indeed important (less so for *kto*-pronouns than for *kakoj* or *čto*; see also Sect. 2.3). I suggest, however, that this does not invalidate the conclusions and is even advantageous, since the media texts may be assumed to reflect the modern norm more accurately than the fiction texts dominating in the main corpus, the latter being more biased toward imitation of the previous norm. If this were true, the data from the newspaper corpus would be expected to generally emphasize the shifts outlined by the main corpus data, which indeed seems to be the case. On the one hand, while in conditionals and future contexts *to* starts to dominate over *nibud'* roughly in the second half of the 20th century according to the data from the main corpus, the newspaper corpus confirms this trend with better evidence. On the other hand, *to* does not surpass *nibud'* in imperatives either in the main or in the newspaper corpus. If, say, the high frequency of *to* in the newspaper corpus were due only to the newspaper genre and not to a consistent diachronic trend, one would not expect imperatives to differ from conditional and future contexts in the newspaper corpus in the same way they differ in the main corpus.

2.3 Statistical analysis

In addition to decision trees, I used random forest (Liaw & Wiener, 2002) for the analysis of the data. The Random Forest model was applied to a classification problem using the R programming language (RStudio Team, 2020, R Core Team, 2022). I took year of creation,

Table 4 Modeling data with Random Forest

	Čto	*Kto*	*Kakoj*
number of trees	100	50	20
OOB error rate	12.09%	13.21%	17.18%
Class error rate *to*	9%	5%	5%
Class error rate *nibud'*	14%	29%	37%

Table 5 Mean Decrease in Accuracy

Predictor	*Čto*	*Kto*	*Kakoj*
Year of creation	11.03	11.02296	5.636499
Context	17.42	11.84351	8.052270
Corpus	10.57	11.81705	5.492924

Table 6 Mean Decrease in Gini Index

Predictor	*Čto*	*Kto*	*Kakoj*
Year of creation	843.9990	954.8949	521.5547
Context	902.9295	551.7875	187.2339
Corpus	501.5686	631.9188	314.2280

source corpus (main or newspaper) and context (conditional, question, imperative or future) as the predictors. I built separate models for the distribution of *to* and *nibud'* for *čto, kto* and *kakoj*. Each forest had 100 trees for *čto*, 50 trees for *kto* and 20 trees for *kakoj*, as error rate stabilized after these numbers (see Table 4).

The Random Forest model also provided variable importance measures for each predictor in the model. These measures help identify the contribution of each variable to the model's predictive accuracy. The importance measures provided are Mean Decrease in Accuracy and Mean Decrease in Gini Index.

The Mean Decrease in Accuracy (cf. Table 5) is the average decrease in model accuracy that results when data for a particular variable is permuted across the out-of-bag observations. A higher value indicates a more important predictor variable.

The Mean Decrease in Gini Index (cf. Table 6) is a measure of how much a variable contributes to the homogeneity of the nodes and leaves in the Random Forest. A higher value indicates that the variable is better at splitting the data into pure nodes.

According to Table 5, we can tell that context is the most important predictor for the accuracy of the classification. This is expected as we already know that *to* and *nibud'* have different distribution in imperative. Considering the decrease in Gini index, year of creation contributes the most to node purity for *kto* and *kakoj*, which means that it helps to split data into nodes with much less internal variation than other predictors.

2.4 Preliminary conclusions

The RNC data considered in this section suggest that compared to modern Russian, the *to*-series was more strictly 'specific' in the 18th–19th centuries, while the *nibud'*-series was less strictly 'non-specific'. The evolution of their distribution across contexts after the 18th
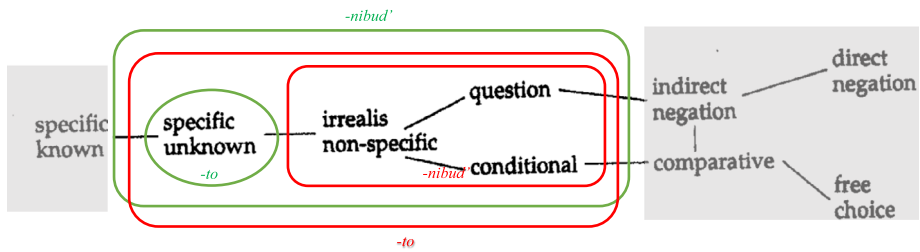
Fig. 6 Boundaries of *to* and *nibud'* in the 18th–19th centuries (green line; RNC data) and in modern Russian (red line; Haspelmath, 1997, p. 65 and RNC data) (Color figure online)

century is thus roughly as in Fig. 6. Note that functions other than 'specific unknown', 'irrealis non-specific', 'question' and 'conditional' have not been considered in my study; therefore, the map in Fig. 6 says nothing about the distribution of *to* or *nibud'* in these functions.

The assumption that *nibud'* narrowed its distribution after the 19th century is indirectly supported by the data on its usage before the 18th century. According to Penkova (2016), *nibud'* was being grammaticalized as an indefinite marker in the 15th–17th centuries and at that period it could be used not only in non-specific functions but also in the 'direct negation' and 'free choice' functions – the two rightmost functions on the semantic map that do not allow *nibud'* in modern Russian. As for *to*, it seems to be an even newer indefinite marker than *nibud'*. Galinskaja (2016, p. 294) cites an example from the second half of the 17th century. The fact of being new sheds light on why *to* had a rather narrow distribution in the 18th–19th centuries. The logic of its later evolution is, however, to be clarified. I will tackle this question in the next section.

3 Discussion

The way the *to*-series evolved after the 19th century runs contrary to the basic principle of the semantic map approach, according to which each marker must be extended incrementally, i.e., to contiguous functions at first (Haspelmath, 2003, p. 233). This principle predicts that the *to*-pronouns should have been extended to imperatives, which are an instance of the 'irrealis non-specific' function, before they extended to conditionals or questions, whereas in reality, according to the corpus data, the opposite happened.

As assumed by Haspelmath (1997, p. 119), the semantic map is arranged in such a way that all functions that share the same value of several relevant parameters (e.g., specific VS non-specific) form a contiguous area on the map. The 'specific unknown' function is opposed to both the 'irrealis non-specific' and 'conditional' functions in that the former is specific while the latter are not. Furthermore, conditionals differ from both 'specific unknown' and 'irrealis non-specific' contexts in that they are scale-reversing contexts (Haspelmath, 1997, p. 120). Consequently, 'specific unknown' and 'conditional' functions do not form a contiguous area on the semantic map (cf. Fig. 1). But then, why is *to* most frequent in exactly these two functions in modern texts? My hypothesis is that conditionals, as well as other non-specific contexts to which *to* has expanded from its original specific domain, are the contexts that best support its non-specific reading, or, in other words, best help to accommodate the specific meaning of *to* to the non-specific meaning of the context.

The rest of the section is structured as follows. I spell out the essentials of my proposal in Sect. 3.1. Then I move on to illustrate the proposal based on the corpus data about the distribution of *to* and *nibud'* in different contexts: conditionals (Sect. 3.2), questions (Sect. 3.3), future contexts (Sect. 3.4), and imperatives (Sect. 3.5).

3.1 Outline of the analysis

As the RNC data have shown, among the non-specific contexts *to* has extended the most to conditionals, questions and future contexts, while its extension to imperatives is slow (see Sect. 2). I suggest that this is because conditionals, questions and future contexts support the non-specific interpretation of *to* better than imperatives.

My data reveal two essential ways of how a context may "support" the non-specific interpretation of an indefinite. One way is exemplified by conditionals and questions in that they are strongly biased toward the non-specificity of an indefinite they contain (see more in Sects. 3.2 and 3.3). This non-specificity bias serves to erase the specific reading that is originally associated with the *to*-series. Another way is exemplified by future contexts (more precisely, by that particular type of future context that turned out to be frequent in my sample). In such contexts, the specific and non-specific interpretations are very close, so that it is difficult to draw a line between them (see details in Sect. 3.4). This eliminates the very need to switch from the specific interpretation to the non-specific one. However, neither the former nor the latter way of supporting the non-specific reading of *to* may apply to imperatives (see Sect. 3.5). This is why, as I suggest, the expansion of the *to*-series in imperatives is slow.

3.2 Conditionals

As mentioned above, the specific contexts are associated with the speaker's commitment to the existence of an individual or an item referred to by an indefinite (Haspelmath, 1997, p. 38, Gärtner, 2009, p. 7). But in (real) conditionals, the speaker is unaware of the truth-value of the protasis (Liu et al., 2021, pp. 1370–1371) and therefore is most probably also unaware of whether what is denoted by an indefinite within the protasis exists. When uttering *I will be happy if you bring me a book*, the speaker does not assume anything about whether a book that could be brought to him by the addressee exists. This is in contrast with imperatives (see Sect. 3.5) and, I argue, creates a bias toward the non-specific reading of an indefinite pronoun within a conditional.

This analysis predicts *to*-pronouns within conditionals to be *non-specific* just like *nibud'*-pronouns, which seems indeed to be the case. That *to* in conditionals is synonymous to *nibud'* was suggested by Sheljakin (1978, p. 17), see also (Kuz'mina, 1989, p. 210). This point of view is disputed by Paducheva (1985, p. 220), who claims that the sentences in (22a) and (22b) differ in that (22b) is only felicitous if the assumption 'he is waiting for someone' has already been mentioned in the pre-text:[5]

(22)   a.   *Esli on kogo-nibud' ždet …*
              'If he is waiting for someone…'
        b.   *Esli on kogo-to' ždet…*
              'If he is waiting for someone…'

---

[5]However, in a later work, Paducheva (2016) suggests that *to* is synonymous with *nibud'* when used in conditionals.

But this claim is not supported by the corpus data. In (23), the *esli*-clause is postposed to its main clause and introduces new information that has not been mentioned before. Not surprisingly, substituting *kto-to* with *kto-nibud'* does not seem to trigger any change in meaning:

(23)   *Otravljala žizn' ego čudoviščnaja revnost'.* […] *Emu ne nravilos', esli ja s kem-to (= s kem-nibud') pju kofe v kafe, komu-to (=komu-nibud') ulybajus', esli kto-to (=kto-nibud') pomogal mne donesti moi pokupki iz magazina v gostinicu.* (RNC, 2003.06)

'[My] life was poisoned by his monstrous jealousy. He did not like it if I drank coffee with someone in a cafe, smiled at someone, if someone helped me carry my purchases from the store to the hotel.'

The synonymy of *to* and *nibud'* in conditionals is further supported by the fact that they may cooccur in one and the same clause:

(24)   *Esli komu-nibud' čto-to izvestno ob ėtom užasnom proisšestvii, prosim pozvonit' v redakciju.* (RNC, 2009.01)

'If anyone knows something about this terrible incident, please call the editor.'

Note that *to*-pronouns cannot be substituted with *nibud'*-pronouns in factual conditionals, which are a particular type of conditional clause carrying the presupposition that someone (other than the speaker) believes the proposition expressed by the protasis to be true (Bhatt, Pancheva, 2006, p. 671). In (25), in which the protasis is introduced by the factual conditional subordinator *raz*, *kto-nibud'*, unlike *kto-to*, is infelicitous. However, this is fully predictable – factual conditionals, contrary to other types of conditionals, contribute to the specific interpretation of an indefinite within them. (25) implies that there exists a specific person who has been transferred to another position.

(25)   *Raz kogo-to* (^{??}*kogo-nibud'*) *pereveli na druguju dolžnost' — značit, bylo za čto.* (RNC, 2003.09)

'Since someone was transferred to another position, it means there was a reason [for this].'

To summarize, I suggest that the *to*-series expanded to conditionals because the semantics of conditionals help to erase the original specific meaning of *to*. This, in turn, is corroborated by the fact that *to* and *nibud'* in conditionals are synonymous and interchangeable (except for the special case of factual conditionals).

3.3 Questions

As in conditionals, in pragmatically neutral, i.e. non-biased and non-modalized, questions the speaker is unaware of the truth-value of the proposition being questioned. Consequently, he is most probably also unaware of, and does not assume anything about, whether the referent of an indefinite pronoun within a question exists. When uttering *Will you bring me a book?*, the speaker does not assume anything as to whether there is a (specific) book that could be brought to him by the addressee. Thus, specific indefinites being associated with the existence of the referent, pragmatically neutral questions can be assumed to be strongly biased toward the non-specific interpretation of indefinites, similarly to real conditionals.

As observed by Kobozeva (2000, p. 304), Russian questions with *li* are pragmatically neutral – they do not convey the speaker's positive or negative expectations. This predicts *to*-pronouns within questions with *li* to be *non-specific*, which seems to be borne out both for independent and embedded questions. In both (26) and (27), the *to*-pronoun can be substituted with the *nibud'*-pronoun without any clear shift in terms of specificity. In neither (26)

nor (27) does *čto-to* refer to something ('something left unsaid', 'something stolen from the supermarket') that the speaker assumes or knows to exist – otherwise the respective questions would not be posed.

(26) *Ostalos' li čto-to* (^{OK}*čto-nibud'*) *nevyskazannym v vašix knigax?* (RNC, 1997.05)
      'Is there anything left unsaid in your books?'

(27) *Bylo li čto-to* (^{OK}*čto-nibud'*) *poxiščeno iz supermarketa, ne soobščaetsja.* (RNC, 2003.09)
      'Whether anything was stolen from the supermarket is not reported.'

However, there is a subtle distinction between *čto-to* and *čto-nibud'* in (26) and (27) that is not directly linked to specificity but still can be traced back to the difference between the originally specific *to* and the non-specific *nibud'*. In a question with *to*, what seems to be emphasized is whether the referent of the pronoun *exists*, i.e. 'Does there exist anything left unsaid in your books?' in (26) and 'Does there exist anything stolen from the supermarket?' in (27). What is focused in a question with *nibud'* is the *situation* itself, i.e. 'Was anything left unsaid in your books?' in (26) and 'Was anything stolen from the supermarket?' in (27).

This assumption is corroborated by the following observation. *Čto-to* is infelicitous if the verb, followed by *li*, is preceded by the topicalization particle *a* (see about this function of *a* Zaliznjak & Mikaelian, 2018, p. 334), as in (28) and (29). The particle serves to emphasize that the focus of the question is the verb, hence, the situation as a whole, which creates unfavorable conditions for the use of the *to*-pronoun. Note that the variants with *to* and *nibud'* clearly differ in acceptability in (28) and (29) but not in (26) and (27), where the topicalization particle is absent.[6]

(28) *No vsjo èto bylo uže posle smerti xudožnika-samorodka. A pomog li kto-nibud'* ([?]*kto-to*) *emu pri žizni?* (RNC, 2006.01)
      'But all this was already after the death of the nugget artist. Did someone help him during his lifetime?'

(29) *Strukturu rasxodov činovniki sčitajut blestjašče. A posčital li kto-nibud'* ([?]*kto-to*) *strukturu urona ot vsex ètix novovvedenij?* (RNC, 2005.03)
      'Officials calculate the structure of expenses brilliantly. Has anyone calculated the damage structure from all these innovations?'

Thus, on the one hand, questions with *li* are strongly biased toward the non-specific interpretation of an indefinite and as such, contributed to the expansion of the *to*-series. On the other hand, however, there is a semantic distinction between *to* and *nibud'* in questions that, as I suggest, contributed to the retention of *nibud'*.

This could be the reason why with *kto*- and *čto*-pronouns, as evidenced by the data in Tables 1 and 3 (see Sect. 2), the frequency gap between *to* and *nibud'* in conditional contexts is noticeably larger than in interrogative ones in the newspaper corpus. The data are reproduced in Table 7 for convenience; the difference between conditionals and questions is statistically significant both for *kto*- and *čto*- (χ^2-test, $p < 0.01$).[7]

---

[6]Here and below, the acceptability of constructed examples was checked by six informants.

[7]In conditionals, a distinction between *to* and *nibud'* like the one observed for questions may also arise. In (i), *to* seems to emphasize the existence of the referent ('if there is someone who I don't like…'), while *nibud'* focuses on the situation as a whole ('if I don't like someone…'). However, contrary to questions this distinction does not seem to contribute to the retention of *nibud'*: I am not aware of an environment that would

Table 7 *To* and *nibud'* with *kto-* and *čto-*pronouns in the newspaper corpus of the RNC

	kto-			čto-		
	to	*nibud'*	ratio of *to*	*to*	*nibud'*	ratio of *to*
conditional	8,048	418	0.95	2,816	109	0.96
question	664	558	0.54	692	286	0.71

With *kakoj*-pronouns, however, the relative frequency of *to* and *nibud'* in conditionals and questions in the newspaper corpus is roughly the same: *to* is about five times more frequent than *nibud'* (see Table 2). A closer look at the data in my sample with *kakoj* suggests that this is due to a widespread type of environment in which the subtle semantic distinction between *to* and *nibud'*, assumed above, does not hold. In this environment, the head noun of *kakoj* is *implicitly non-specific*, i.e., may convey a non-specific reading even if it is used without an indefinite. The examples in (30) and (31) are illustrative. In both, *kakoj-to* can be omitted (or substituted with *kakoj-nibud'*) without a clear shift in meaning.

(30)   *Byli li* (*kakie-to*) *konsultacii s prezidentom na ėtu temu?* (RNC, 2001.11)
        'Were there any consultations with the president on this subject?'

(31)   *Nužna li* (*kakaja-to*) *pomošč v vosstanovitel'nyx rabotax?* (RNC, 2012.02)
        'Do you need any help with restoration work?'

The same is true for *kakoj-nibud'*: if the head noun is implicitly non-specific, *kakoj-nibud'* can be omitted (or substituted with *kakoj-to*) without any clear shift in meaning:

(32)   *Est' li* (*kakoj-nibud'*) *šans, čto situacija možet pomenjat'sja?* (RNC, 2015.08)
        'Is there any chance that the situation could change?'

(33)   *Slučajutsja li* (*kakie-nibud'*) *konflikty na ėtoj počve*? (RNC, 2015.02)
        'Are there any conflicts in this regard?'

Thus, *kakoj-to* and *kakoj-nibud'* in questions, at least when used with an implicitly non-specific head noun, are semantically even closer than *kto-to* and *kto-nibud'* or *čto-to* and *čto-nibud'*. This means that the factor that presumably contributes to the retention of *nibud'* with *kto-* and *čto-*, is not at play with *kakoj-*, which, in turn, accounts for the broader expansion of *kakoj-to* in questions compared to *kto-to* or *čto-to*.

To summarize, the hypothesis that *to* extended to contexts that help to cancel its specific meaning (see Sect. 3.1) provides an account for why *to* extended to questions. Combined with a few additional assumptions, this hypothesis also sheds light on why *to* overtook *nibud'* in questions less strongly than in conditionals, and why *kakoj*-pronouns differ in this respect from *kto-* and *čto-*pronouns.

3.4 Future contexts

Compared to conditionals and questions, future contexts are less biased toward the non-specific reading of an indefinite. When making a statement in the future tense, the speaker

serve to emphasize the latter reading to the prejudice of the former in conditionals in the same way the topic marker *a* in questions does.

(i)   *Esli kto-to* (*kto-nibud'*) *ne nravitsja, ja s nim prosto ne rabotaju*. (RNC, 2019.05)
        'If [I] don't like someone, I just don't work with them.'.

usually assumes that a referent of an indefinite exists. For example, when saying *I'll bring you a book*, the speaker assumes that there is a book that he could bring to the addressee. As suggested in Sects. 3.2 and 3.3, this is in contrast both to conditionals and questions.

Consequently, *to*-pronouns are not expected to be synonymous with *nibud'*-pronouns in terms of specificity in future contexts, and indeed, often they are not. While *nibud'*-pronouns are non-specific in future contexts, *to*-pronouns may be specific. In (34), a specific person is intended who will come from Moscow; hence, only *to* is felicitous here.

(34) *Ėkzameny v 8–9 klassax vyneseny «kak pokazatel'nye» na 2 ijulja. Kto-to* (??*kto-nibud'*) *priedet iz Moskvy.* (RNC, 1945)
      'Examinations in grades 8–9 have been scheduled as "demonstration" for July 2. Someone will come from Moscow.'

However, there is one particular type of the future context in which the meanings of *to* and *nibud'* converge. In this context, exemplified in (35) and (36), it is stated that there exists someone who will take the respective action, i.e., will say that I am a stupid person in (35) and will ask "Why such tricks?" in (36). As mentioned above, this is a prerequisite for the *specific* interpretation of the pronoun. At the same time, however, even if the indefinite is singular, it is usually not a single referent but a group of referents that is intended in such sentences. In (35), for example, the speaker wants to say that there will be *some people* who will say 'You are a stupid person'. In (36), some people may ask "Why such tricks?". The fact that no specific referent is intended is consistent with the *non-specific* reading of an indefinite. Thus, the two readings of an indefinite appear to be very close in this case – it is difficult to draw a line between them. Not surprisingly, both *to*- and *nibud'*-pronouns are felicitous in (35) and (36).

(35) *Kto-to* (OK*kto-nibud'*) *skažet, glupyj čelovek.* (RNC, 2020.09)
      'Someone will say [that I am] a stupid person.'

(36) *Kto-nibud'* (OK*kto-to*) *sprosit: «Dlja čego takie uxiščrenija?* (RNC, 2003.09)
      'Someone will ask: "Why such tricks?'

(37) and (38) are similar examples with *čto*- and *kakoj*-pronouns: in both, the speaker is confident of the existence of the referent, but no specific referent is intended. Here, too, *to* and *nibud'* are interchangeable:

(37) *Čto-to* (OK*čto-nibud'*) *proigraju, čto-to* (OK*čto-nibud'*) *vyigraju.* [lenta.ru, 2019.08]
      'There will be something that I will lose, there will be something that I will win'

(38) *Kakoj-nibud'* (OK*kakoj-to*) *priëm srabotaet.* (RNC, 2005.07)
      'At least one trick will work.'

This type of context appeared to be particularly frequent in my sample.[8] I want to suggest that this accounts for the frequency of *to*-pronouns in my sample of future contexts in modern texts. The affinity between the specific and non-specific meanings that characterizes this type of future contexts allows the speaker to use *to* as a synonym of *nibud'* despite the originally specific meaning of *to*.

Note that this mechanism is different from the one that I assumed to be responsible for the expansion of *to* in conditionals and questions. In the latter case, as I suggested, the con-

---

[8]This is because I looked for examples in which the indefinite pronoun is in the sentence-initial position (see the query in Sect. 2). Otherwise, the pronoun could have been preceded by such items as a conditional subordinator or a *wh*-word that would have changed the context type (e.g. from future to conditional or question).

text serves to erase the specific meaning of *to*. In the future contexts, the specific meaning associated with *to* stops being problematic because the context provides an affinity between the specific and non-specific readings of an indefinite.

3.5 Imperative contexts

The RNC data have shown that the expansion of *to* to imperatives is slow: *to* is less frequent than *nibud'* with both *kto-*, *čto-* and *kakoj'-* in the newspaper corpus (see Sect. 2). I suggest that this is because the mechanisms that facilitate the expansion of *to* in conditionals, questions and future contexts do not apply to imperatives.

Compared to conditionals and questions, imperatives are less biased toward the non-specific reading of an indefinite as they usually imply that the speaker assumes the referent of the pronoun to exist. For example, if one says *Bring me a book*, they most probably assume that a book that could be brought exists.

This seems to predict that like in future contexts, indefinites within imperatives may be both specific and non-specific. However, the situation involving imperatives is more complicated than in future contexts. Being less biased toward non-specificity than conditionals and questions, imperatives are at the same time less biased toward specificity than future contexts. Although both a future situation and an imperative situation are irrealis, only in the former case (cf. *I will bring you a book* vs. *bring me a book*) does the speaker express high confidence that the situation will take place.

For this reason, indefinites in imperative contexts are mostly non-specific. In (39), for example, no specific person is intended, i.e. a person who the speaker knows to exist. (Note that this does not change the fact that the speaker assumes there exists someone whom the addressee may notify about the trip, as suggested above.). Not surprisingly, *to* can be substituted with *nibud'* without any clear shift in specificity.

(39)　*Predupredite kogo-to* (^{OK}*kogo-nibud'*) *o poezdke*. (RNC, 2016.12)
　　　　'Notify someone about the trip.'

There are several environments in which the use of *to* seems to be supported by some context feature. Firstly, this happens when a *to*-pronoun combines with an elective construction, exemplified in (40). The elective construction sets a narrow circle of possible referents, each of which is known to exist, cf. *someone from the household* in (40), which seems to reconcile the originally specific semantics of *to* with the non-specific status of the imperative.

(40)　*Poprosite kogo-to iz domočadcev razbit' v ètot stakan svežee kurinoe jajco*. (RNC, 2017.04)
　　　　'Ask someone from the household to break a fresh chicken egg into this glass.'

The use of *kto-to* in (40) is not the specific use in the proper sense of the word since no specific person from the household is intended, but still, it is closer to the specific use than (39). It does not seem to be a coincidence that the elective construction is frequent in imperatives with *kto-to* – more frequent than, for example, in simulative contexts with *kto-to*. The data, retrieved from my sample with *kto-to* in the newspaper corpus (see Sect. 2), are presented in Table 8; the difference between imperatives and simulatives is statistically significant (2-tailed exact Fisher test, $p < 0.01$).

Secondly, the use of *čto-to* in imperatives seems to be supported when *čto-to* cooccurs with a bare adjective, as in (41):

(41)　*Narisujte čto-to prostoe*. [Труд-7, 2008.08]
　　　　'Draw something simple.'

Table 8 Frequency of *kto-to* with and without an elective construction in imperative and simulative contexts (newspaper corpus of the RNC)

	elective +	elective −	ratio of elective
imperative	10	30	0.25
simulative	14	371	0.04

Table 9 Frequency of *čto-to* with and without a bare adjective in imperative and simulative contexts (newspaper corpus of the RNC)

	adj +	adj −	ratio of adj
imperative	64	137	0.32
simulative	14	160	0.08

As the elective construction, a bare adjective narrows the circle of referents from which to choose, and this brings the use closer to the specific one. Examples in (42) and (43) illustrate this assumption. In (42), the pronoun *čto-to* is used without an adjective and sounds worse than *čto-to* in (43), where it combines with an adjective. Note that the pronoun *čto-nibud'* is felicitous in both cases.

(42)   [?]*Prinesi mne čto-to* ([OK]*čto-nibud'*) *poest'*.
        'Bring me something to eat.'

(43)   *Prinesi mne čto-to* ([OK]*čto-nibud'*) *vkusnoe*.
        'Bring me something tasty.'

As expected, *čto-to* with a bare adjective is more frequent in imperatives than, say, in simulative contexts. The data are presented in Table 9; the difference between imperatives and simulatives is statistically significant (2-tailed exact Fisher test, $p < 0.01$).

Thirdly, the use of the pronoun *kakoj-to* in imperatives seems to be facilitated by an implicitly non-specific head noun. Cf. (44), where the head noun *pomeščenie* 'room' refers to a non-specific room even if *kakoj-to* is absent. *Kakoj-to* in (44) may be omitted or substituted with *kakoj-nibud'*.

(44)   *Iščite* (*kakoe-to*) *pomeščenie s kondicionerom*. (RNC, 2013.06)
        'Look for some air-conditioned room.'

Fourthly, and finally, *kakoj-to* is the only option in imperatives when combined with the noun *vremja* 'time'. *Kakoe-to vremja* has lexicalized to denote an indeterminate period of time, the duration of which the speaker does not know (as in (45)) or does not want to specify (as in (46)).

(45)   *Poprobujte kakoe-to* ([??]*kakoe-nibud'*) *vremja ne solit' ovošči*. [Коммерсант, 2010.09]
        'Try not to salt the vegetables for a while.'

(46)   *Kakoe-to* ([??]*kakoe-nibud'*) *vremja nazad ja učastvovala v odnom koncerte v Voroneže*. (RNC, 2004.05)
        'Some time ago I participated in a concert in Voronezh.'

Since both with indicative (45) and imperative verb forms (46) the speaker intends a period of time that he assumes to exist, *kakoe-to vremja* satisfies the main criterion of specificity in these contexts. Diachronically, this sheds light on why *to*, and not *nibud'*, is used in this construction. Synchronically, however, it is the indeterminacy of the time period rather

than its existence that is foregrounded. Indeed, *kakoe-to vremja* is used in both specific (45) and non-specific contexts, be it imperatives, as in (46), or conditionals, as in (47), while *kakoe-nibud' vremja* sounds awkward in all these contexts. In my sample with *kakoj-to* and imperatives from the newspaper corpus, twelve of 68 examples contain *kakoe-to vremja*.

(47)   *No esli kakoe-to* (??*kakoe-nibud'*) *vremja lenjus', to potom starajus' zanimat'sja ser-*
         jëzno. (RNC, 2011.11)
         'But if I'm lazy for a while, then I try to get serious.'

The above contexts (with the exception of *kakoe-to vremja*, which is a case of lexical-ization), I suggest, combine frequently with *to*-pronouns because they help to mitigate the conflict that emerges due to the original specific semantics of *to* and the non-specific seman-tics of the imperative. However, these contexts are peripheral. Outside of such contexts, *to* in imperatives is rare and rather marginal. As a speaker, I find *to* in many such examples as inappropriate. Example (39), for instance, sounds better for me with *nibud'* than with *to*. It is noteworthy that Kobozeva (1981, p. 165) and Paducheva (2016) claim that *to*-pronouns cannot be used at all in imperative contexts.

The reason why the extension of *to* to imperatives is slow can thus be summarized as follows. In imperatives there is no consistent way to neutralize the original specific meaning of *to*. Imperatives are less biased toward the non-specific reading of indefinites than condi-tionals and questions; hence, imperatives do not cancel the specific meaning of *to* in the way that conditionals and questions do. There is also no way to neutralize the semantic distinction between 'specific' and 'non-specific' in the way future contexts do (see Sect. 3.4), except for few rather peripheral cases. At the same time, imperative contexts are more patently irrealis than future contexts. Therefore, they can tolerate the originally specific *to* with more diffi-culty than future contexts. The conflict that results from this is, in my view, the cause of the rarity of *to* in imperatives.

3.6 Preliminary conclusions

As an explanation for the evolution of the *to*- and *nibud'*-pronouns after the 18th century, I suggested that *to* extended most to those non-specific contexts that helped to reconcile the non-specific meaning of the context with the original specific semantics of *to*. The mech-anism of such "help" may be of two types. Firstly, a context itself may help to cancel the specific meaning of the pronoun; this is what happened in conditionals and questions. Sec-ondly, a context may converge the 'specific' and 'non-specific' readings, i.e., neutralize the distinction between them. This, as I suggest, occurred in the future context (more precisely, in one widespread subtype of the future context). In imperatives, the extension of *to* is slow since neither type of "help" applies, which results in a conflict between the original specific meaning of *to* and the non-specific meaning of the context.

This mechanism turns out to be stronger than the principles that underlie the distribution of semantic functions on the semantic map of indefinite pronouns (Haspelmath 1997): the latter erroneously predict that *to* should have extended to imperatives prior than to condi-tionals or questions.

4 Typology

In modern Russian, of the four functions of indefinite pronouns considered in my corpus study, the distribution of *to* and *nibud'* overlaps in three: 'irrealis non-specific', 'conditional',

Fig. 7 Latin (Haspelmath, 1997, p. 254)
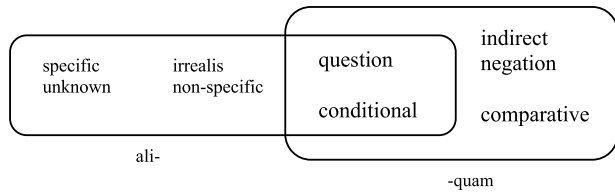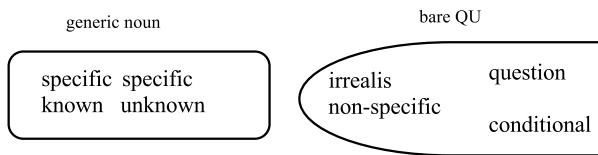

Fig. 8 Chinese (Haspelmath, 1997, p. 307)

and 'question' (see Fig. 9). In the Russian language of the 18th–19th centuries, however, their distribution overlapped only in one function, namely 'specific unknown', i.e., *to* and *nibud'* were in a relationship close to that of *complementary distribution*.

Typological data suggest that the complementary distribution between the 'specific' and 'non-specific' series of indefinite pronouns is a rarity cross-linguistically. As a source of such data, I used the semantic maps of indefinite pronouns of 40 languages, elaborated by Haspelmath (1997, p. 68 ff.). I considered whether there is an overlap in distribution in the functions 'specific unknown', 'irrealis non-specific', 'conditional' and 'question' between markers that have specific uses and those that have non-specific uses. In Fig. 7 and 8, fragments of two semantic maps are given for illustration, for Latin and Chinese. The specific and non-specific markers overlap in the former map and do not overlap in the latter.

In 16 of the 40 languages, according to the maps, there is no overlap in distribution of the specific and non-specific series, which at first glance seems like a great number. These are Swedish (Germanic, Indo-European), Serbian (Slavic, Indo-European), Latvian (Baltic, Indo-European), Irish (Celtic, Indo-European), Ossetic (Iranian, Indo-European), Yakut (Turkic), Lezgian (Nakh-Daghestanian), Nanay (Manchu-Tungusic), Maltese (Afro-Asiatic, Arabic), Hausa (Chadic, Afro-Asiatic), Georgian (Kartvelian), Kannada (Dravidian), Chinese (Sino-Tibetan), Ancash Quechua (Quechua), Japanese (Japonic), and Basque (Isolate). Upon closer examination, however, it turned out that in seven of these 16 languages (Swedish, Latvian, Lezgian, Nanay, Maltese, Hausa and Japanese), the absence of overlap is simply due to the fact that there is only one marker in the language used in these four functions, so the lack of overlap does not lead to complementary distribution. As for the remaining nine languages, the assumption that the 'specific' and 'non-specific' pronouns in them are in complementary distribution requires further verification. The data available to me show that at least for two languages, Chinese and Serbian, this assumption is not entirely correct.

In Chinese, according to Haspelmath (1997, p. 307), only generic nouns can be used in specific functions, while for the 'irrealis non-specific', 'question' and 'conditional' functions Chinese uses bare interrogatives (cf. Fig. 8). However, Gärtner (2009, p. 13, footnote 21) reports that Haspelmath himself (1997, p. 171) gives an example that can be interpreted as a 'specific unknown' use of the bare interrogative *shenme* 'what'. Furthermore, bare interrogatives can be used in Chinese in the 'specific known' function according to Tretjakova (2009, p. 112). This suggests that generic nouns and bare interrogatives are in fact *not* in complementary distribution.

In Serbian, according to Haspelmath (1997, p. 269–270), the zone from 'specific un-known' to 'conditional' is shared by two series of indefinite pronouns: the *ne*-series is used in 'specific' and 'irrealis non-specific' contexts, while the *i*-series is used in 'conditional' and 'question'. However, there are corpus examples with the *ne*-series within conditionals. Cf. (48):

(48)   *A šta ako se neko vozi bez suvozača*? (Intercorp v13 – Serbian)
        'А если кто-то едет без пассажира?'

The reason why languages avoid complementary distribution between specific and non-specific series could be the *subtlety of the boundary* between 'specific' and 'non-specific' interpretations. As Russian data have shown, there are many intermediate cases (cf. examples (35)–(38) and (40)).

Thus, it may be assumed that the expansion of the Russian *to* to the contexts in which *nibud'* was already used was triggered by the fact that the relationship between *to* and *nibud'* was close to complementary distribution, and therefore was unstable.

5 Conclusions

The main results and assumptions of this paper are the following:

- According to RNC data, in the 18th–19th centuries *to* was more strictly specific, while *nibud'* was slightly less non-specific compared to modern Russian. Of the four functions of indefinite pronouns considered in my corpus study ('specific unknown', 'irrealis non-specific', 'conditional', and 'question'), *to* and *nibud'* overlapped only in the 'specific unknown' function, i.e., had a relationship close to complementary distribution (Sect. 2).
- In the 20th century, *nibud'* stopped being used in specific contexts, while *to* expanded to several non-specific contexts, namely to future, conditional and interrogative ones (Sect. 2).
- It can be assumed that the expansion of *to* was triggered by its almost complementary distribution with *nibud'*, which is a typologically unstable situation (Sect. 4). The expansion affected *to* rather than *nibud'* because *to* was a newer marker with a narrower distribution (Sect. 2).
- Although *to* expanded to several non-specific contexts, it almost did not expand to imperatives (Sect. 2). This runs contrary to what is predicted by the semantic map approach to the evolution of indefinite pronouns (Sect. 3).
- I argue that *to* expanded to those non-specific contexts that helped to accommodate its original specific semantics to the non-specific semantics of the context. These are, as I suggest, conditionals, questions and future contexts, but not imperatives (Sect. 3).

Sources Intercorp v13 – Serbian – Intercorp (https://ucnk.ff.cuni.cz/InterCorp/), accessed via "Kontext" interface at kontext.korpus.cz.
    RNC – Russian National Corpus. URL: www.ruscorpora.ru.

Appendix: Decision trees

The decision trees for the pronouns *kto*, *čto* and *kakoj*, shown in Fig. 9–11, shall be read as follows: the upper node divides the data into two parts according to the condition specified
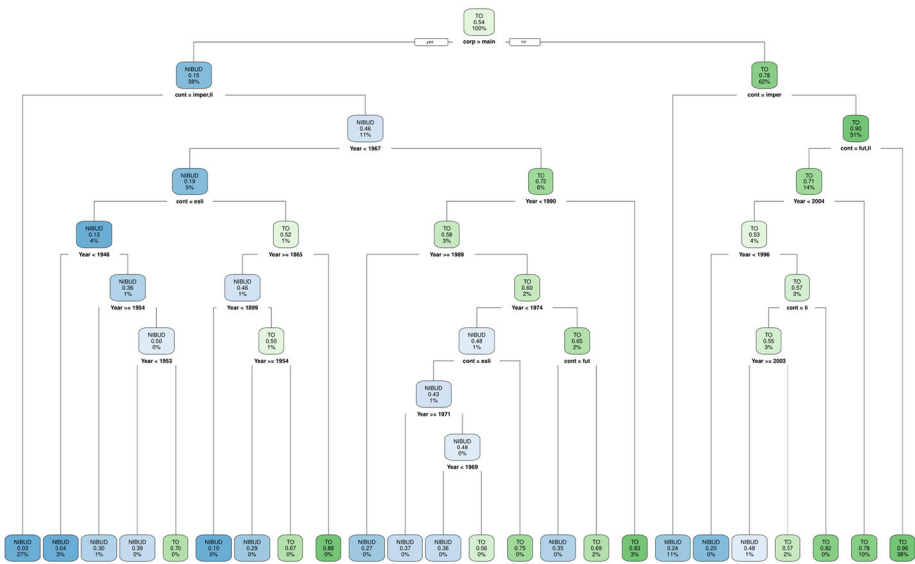
Fig. 9 Decision tree for *kto*-pronouns (Color figure online)

Fig. 10 Decision tree for *kakoj*-pronouns (Color figure online)

under the colored square (node). Data that meet the condition are on the left of the node, and those that do not meet the condition are on the right. Each subsequent node does the same with the data in its part of the division. The information on the node represents characteristics of this part of the data before the partition. The percentage on the node shows what part of the data is represented on this node before the partition – 100% for the upper node. The color of each node shows the marker that prevails in this part of the data - green for *to* and blue for *nibud'*. The decimal shows which part of this data is *to* (irrespectively to the color

Fig. 11 Decision tree for *čto*-pronouns (Color figure online)

of the node). Year of creation, source corpus (main or newspaper) and context (conditional, question, imperative or future) are used as the predictors.

Declarations

Competing Interests The author has no relevant financial or non-financial interests to disclose.

References

Bhatt, R., & Pancheva, R. (2006). Conditionals. In M. Everaert & H. van Riemsdijk (Eds.), *The Blackwell companion to syntax* (vol. 1, pp. 638–687). Oxford: Blackwell.

Bylinina, E. G. (2010). Depreciative indefinites: evidence from Russian. In G. Zybatow et al. (Eds.), *Formal studies in slavic linguistics: proceedings of formal description of slavic languages 7.5*, Frankfurt am Main (pp. 191–206).

Fursenko, D. I. (1969). O razgraničenii značenij i uslovijax upotreblenija neopredelënnyx mestoimenij v sovremennom russkom jazyke. In I. P. Palta (Ed.), *Materialy VIII u IX meždunarodnyx seminarov prepodavatelej russkogo jazyka stran socializma*, Moscow: Izdatel'stvo Moskovskogo universiteta (pp. 338–350).

Galinskaja, E. A. (2016). *Istoričeskaja grammatika russkogo jazyka*. Moscow: Lenand (2nd ed.).

Gärtner, H. (2009). More on the indefinite-interrogative affinity: the view from embedded non-finite interrogatives. *Linguistic Typology*, *13*, 1–37.

Geist, L. (2008). Specificity as referential anchoring: evidence from Russian. In A. Grønn (Ed.), *Proceedings of SuB12* (pp. 151–164). Oslo: ILOS.

Haspelmath, M. (1997). *Indefinite pronouns (Oxford studies in typology and linguistic theory)*. Oxford: Clarendon Press.

Haspelmath, M. (2003). The geometry of grammatical meaning: semantic maps and cross-linguistic comparison. In M. Tomasello (Ed.), *The new psychology of language*, New York: Erlbaum (Vol. 2, pp. 211–243).

Kobozeva, I. M. (1981). Opyt pragmatičeskogo analiza to- i nibud'-mestoimenij. Izvestija «AN SSSR». *Serija literatury i jazyka*, *40*(2), 165–172.

Kobozeva, I. M. (2000). *Lingvističeskaja semantika*. Moscow: Knižnyj dom "Librokom".

Kuz'mina, S. M. (1989). Semantika i stilistika neopredelënnyx mestoimenij. In D. N. Shmelëv (Ed.), *Grammatičeskie issledovanija. Funkcional'no-stilističeskij aspekt. Supersegmentnaja fonetika. Morfologičeskaja semantika* (pp. 178–231). Moscow: Nauka.

Letučij, A. B. (2017). Sravnitel'nye konstrukcii. In V. A. Plungjan & N. M. Stojnova (Eds.), *Materialy k korpusnoj grammatike russkogo jazyka. Vypusk II. Sintaksičeskie konstrukcii i grammatičeskie kategorii*, Saint-Petersburg: Nestor-Istorija (pp. 132–204).

Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R News*, *2*(3), 18–22.

Liu, M., Rotter, S., & Giannakidou, A. (2021). Bias and modality in conditionals: experimental evidence and theoretical implications. *Journal of Psycholinguistic Research*, *50*, 1369–1399. https://doi.org/10.1007/s10936-021-09813-z.

Nikolaeva, T. M. (2013). Funkcional'naja nagruzka neopredelennyx mestoimenij v russkom jazyke. In T. M. Nikolaeva (Ed.), *Lingvistika: izbrannoe*. (pp. 266–283). Moskva: Jazyki slavjanskoj kul'tury.

Paducheva, E. V. (1985). *Vyskazyvanie i ego sootnesënnost' s dejstvitel'nost'ju*. Moscow: Nauka.

Paducheva, E. V. (2015). Mestoimenija otricatel'noj poljarnosti. Materialy dlja proekta korpusnogo opisanija russkoj grammatiki, available at: http://rusgram.ru. Manuscript, Moscow.

Paducheva, E. V. (2016). Mestoimenija neizvestnosti. *Materialy dlja proekta korpusnogo opisanija russkoj grammatiki*, available at: http://rusgram.ru. Manuscript, Moscow.

Penkova, Ya. A. (2016). Neopredelennye mestoimeniya s formantami nibud' i libo/lyubo v istorii russkogo yazyka XV–XVII vekov: Korpusnoe issledovanie. In V. A. Baranov & T. A. Timchenko (Eds.), *El'Manuscript–2016. Rašytinis palikimas ir informacinės technologijos. VI tarptautinė mokslinė konferencija Pranešimai ir tezės*, Vilnius, 2016 (pp. 280–284). Vilnius, Iževskas: Loreta Uzdraitė.

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

Rakhilina, E. V., Borodina, M. A., & Reznikova, T. I. (2016). "Taman' segodnja": korpusnoe issledovanie russkogo jazyka XIX veka. In A. M. Moldovan (Ed.), *Trudy Instituta russkogo jazyka imeni V.V. Vinogradova*, Moscow (Vol. 10, pp. 242–255).

R Core Team (2022). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

RStudio Team (2020). *RStudio: integrated development for R. RStudio*. PBC, Boston, http://www.rstudio.com/.

Sheljakin, A. M. (1978). O semantike i upotreblenii neopredelennyx mestoimenij v russkom jazyke. *Učenye zapiski Tartuskogo universiteta*, *442*, 3–22.

Shmelëv, A. D. (2002). *Russkij jazyk i vnejazykovaja dejstvitel'nost'*. Moscow: Jazyki slavjanskoj kul'tury.

Tatevosov, S. G. (2002). *Semantika sostavljajuščix imennoj gruppy: kvantornye slova*. Moscow: IMLI RAN.

Therneau, T., & Atkinson, B. (2022). _rpart: Recursive Partitioning and Regression Trees_. R package version 4.1.16, <https://CRAN.R-project.org/package=rpart>.

Tretjakova, O. D. (2004). Serija neopredelënnyx mestoimenij s nulevym markerom neopredelënnosti v russkom jazyke. In I. M. Kobozeva, A. S. Narinjani, & V. P. Selegej (Eds.), *Proceedings of the international conference "Dialog-2004". Computational linguistics and intellectual technologies* (pp. 602–607). Moscow.

Tretjakova, O. D. (2009). Neopredelënnye mestoimenija, lišënnye markera neopredelënnosti, v tipologicheskoj perspektive. Moscow State University. Ph. D. Thesis.

Zaliznjak, A. A., & Mikaelian, I. (2018). Russkoe *a*: opyt integral'nogo analiza. *Russian Linguistics*, *42*(1), 321–344. https://doi.org/10.1007/s11185-018-9199-3.