



Russian Learner Corpora in Teaching Russian as a Foreign Language: Covering the Domain of Indefiniteness

Olga Eremina¹

Abstract

Russian Learner Corpora includes texts by learners of Russian as a foreign language with marked and corrected errors. It is searchable both by lexical units and grammatical tags. The search can be general, level-specific, and language-specific (by speakers' native language). For a teacher of Russian as a foreign language it is a useful resource allowing for analyzing general trends and problems in acquisition of particular language phenomena and developing further recommendations and exercises for the students. In this paper, we use Russian Learner Corpora for evaluating students' problems with indefinite pronouns in Russian that include four sets marked by prefixes or suffixes (-to, -nibud', -libo, and koe-). The choice of a pronoun in a given situation depends on semantics, sentence type, and register of speech. For native speakers the choice of a pronoun is intuitive, L2 learners have to consider all the complexity of the factors. Often, their "calculations" are incorrect. The learner corpora help the instructor to find the "weak spots" in acquisition and develop student-oriented exercises that target specific difficulties.

1. Introduction

The acquisition of grammatical phenomena of a foreign language that have little or no equivalent in students' native language might prove challenging. If a phenomenon is complex and requires attention to multiple linguistic aspects, such as semantics, type of a sentence structure, register of speech etc., the difficulty of its acquisition increases. It is essential for a teacher to see which features of such complex language "puzzles" are acquired readily and which of them require extra class time.

Corpus resources provide a useful tool in studying the peculiarities of language acquisition and in developing further exercises in the problem areas. Russian Learner Corpora <http://www.web-corpora.net/RLC/> provides a rich supply of texts created by learners of Russian as a foreign language of different levels. All errors in the texts are marked by the type of error (Tense, Aspect, Case, etc.) and corrected – in a separate entry.

The corpus is searchable both by lexical units and grammatical tags. The search can be general, level-specific, and language-specific (in texts created only by native speakers of English, French, etc.).

This provides a useful resource for a teacher of Russian that helps analyze general trends and problems in acquisition of particular language phenomena (aspect, subject-verb agreement, pronouns, etc.), and subsequently to create practical exercises for students that are based on authentic language material and specifically target problem areas.

One of the areas, where using Russian Learner Corpora may give a very important feedback on how L2 students succeed in mastering a complex language phenomenon is the system of Russian indefinite pronouns.

2. Language background

The indefinite pronouns in Russian are marked by special prefixes or suffixes (-to, -nibud', -libo, and koe-). Referring to an indefinite entity of any sort (e.g. an indefinite person, equivalent to English *someone* / *anyone*; indefinite thing, *something* / *anything*; indefinite place, *somewhere* / *anywhere*; indefinite representative of a class of objects described by a noun, *some boy, book* / *any boy, book*, etc.) a speaker of Russian chooses from the four options. The choice depends on various characteristics: (i) familiarity of the object to the speaker or the addressee, or even existence of the object; (ii) information about the quantitative characteristics of the (plural) object (the whole group or just a part of the group); (iii) information about the situation as a whole that is provided by some grammatical "indicators" in the sentence (tense, markers of plurality, etc.), and (iv) register (formal or informal) of the speech. Features of each set of pronouns are based on a combination of all (or some) of the described characteristics.

¹ National Research University "Higher School of Economics", Russian Federation



Indefinites with the prefix *koë-* refer to an object that is known to the speaker, but unknown to the addressee (similar to English *a certain*).

Indefinites with the suffix *-to* are used in reference to specific objects that stay indefinite in the context, e.g. when the speaker has no information about the exact name, although he/she is or was familiar with the object itself.

Indefinites with the suffix *-nibud'* are usually described as non-specific (the named object is unknown and even its existence has not been confirmed). Another limitation is that these indefinites cannot be used in the situations where just any object will satisfy the speaker (i.e. as a universal like English *whoever*, or *just any*). In addition, these pronouns have a peculiar distribution (they can only be used in questions, conditional sentences, imperatives, future tense sentences, with modal operators, and in dictionary definitions).

The indefinites with the suffix *-libo* are very similar to *-nibud'* pronouns in their semantics and functions. They differ, first of all, in the register of speech (*-libo* indefinites are more formal than *-nibud'* indefinites). Secondly, these two sets of indefinites have a different degree of universality. *-Nibud'* indefinites can never receive a universal interpretation (e.g. be understood as *whoever*, *whatever*, *whenever*); *-libo* items, however, function as universals in some types of constructions, e.g. in comparisons.

The system also includes special negative pronouns which are used instead of indefinite pronouns in sentences with so-called "negative concord". As a result, area of usage for indefinite pronouns is not as wide as in English. In Russian, sentences like "He did not say anything to anyone" have to be rendered as "He did not say nothing to no one".

We expect that L2 learners of Russian may not only become confused when they have to choose one type of indefinites out of four, but also that they will overgenerate and erroneously use indefinites in the situations that require negative pronouns.

As we can see, the different sets of Russian indefinite pronouns present a continuum in which each of sets has its own semantics and functions and is dependent on the other parts of the system.

For native speakers the choice of a pronoun is intuitive; L2 learners have to acquire and take into consideration all the complexity of the abovementioned factors. We suggest that some of the key features may be learned easier, and some others may cause difficulties. We used the Russian Learner Corpora to test this hypothesis and tease out the "weak spots" of the acquisition process.

3. Data

In order to see how successfully L2 learners of Russian acquire the complex phenomenon of indefinite pronouns, we analyzed their usage in the sub-corpus of the speakers with a level of Russian of intermediate-mid and higher. We did not specify the native language of the learners. We excluded the texts of heritage learners of Russian.

We analyzed 50 randomly selected sentences where the students used *-to* pronouns; 50 randomly selected sentences with *-nibud'* pronouns; and the same number of sentences with *-libo* pronouns. This was the total number of *-libo* items in the sub-corpus (they are used much less frequently because of their stylistic limitations; they are characteristic of the formal register). The number of *-to* and *-nibud'* pronouns in the sub-corpus is much greater; therefore we limited the data for analysis to 50 items for each category simply to match the numbers of *-libo* items.

Unexpectedly, the search for the indefinites with the prefix *koë-* in writing samples of sub-corpus of speakers with a level of Russian of intermediate-mid and higher (the total number of sentences was 27653) gave zero results. Changing the conditions of the search (adding forms with deviations in spelling, and including L2 learners of all levels) did not change the result. Any textbook of Russian presents all four sets of indefinite pronouns. L2 students complete numerous exercises in which they have to choose between *-to*, *-nibud'*, and *koë-* pronouns. Our results show that despite all these efforts L2 learners do not use the latter set in their writing.

Thus, we will have to limit our further analysis to the other three sets of indefinite pronouns.

4. Results

4.1 Results for *-to* indefinites

Out of 50 sentences with *-to* indefinites, the choice of the pronoun was correct in 49 sentences (98%). The only problematic sentence with two instances of the indefinite pronouns used *-to* items in a non-specific



context, which normally only allows for *-nibud'* pronouns (or *-libo*, if the style permits it): *I hope that one day on a bus someone will speak with me or give me something to do.*

All other sentences, although they included some other problems with indefinite pronouns (e.g. spelling, case marking, etc.) used the correct lexical item.

4.2 Results for *-nibud'* indefinites

Out of 50 sentences with *-nibud'* indefinites, the choice of the pronoun was correct in 41 sentences (82%). The errors were caused by the following:

- (a) a non-specific pronoun is used in a context that requires a specific indefinite (i.e. *-to* pronoun): e.g. *Because of this disagreement, the group of supporters emerges in some natural way.* (4 errors)
- (b) an indefinite pronoun is used where a universal pronoun should be used (*-nibud'*, unlike *-libo*, never has a universal interpretation): e.g. *My experience allows me to approach just anyone and ask how I can help* (5 errors)

All other sentences, despite grammar and spelling problems, use the correct type of the indefinite.

4.3 Results for *-libo* indefinites

Out of 50 sentences with *-libo* indefinites, the choice of the pronoun was correct in 48 sentences (96%). The two errors appeared in the following situations:

- (a) an indefinite pronoun is where a negative pronoun is needed: *Many Americans mistakenly say "You simply cannot do anything".*
- (b) the pronouns with suffix *-libo* is used where a universal pronoun should be used (*-libo* items can be used as universals only in particular contexts, including comparisons): *Before a translator signs his work it is important that a different person checked the translation and corrected all errors.*

None of the samples contained a register error (all *-libo* items were, indeed, used in formal contexts). All other mistakes were instances of grammar and spelling errors.

5. Conclusions

Analysis of the data of the Russian Learner Corpora demonstrates that despite the complex nature of the relationship between different sets of Russian indefinite pronouns L2 learners have a good grasp on them. The students experience almost no difficulties with the specific *-to* indefinites; they are aware of the difference between those items and their non-specific counterparts with the suffixes *-nibud'* and *-libo*. However, the non-universal and partially-universal nature of *-nibud'* and *-libo* indefinites causes problems. The most striking finding is the fact that, although aware of the existence *koe-* indefinites learners do not produce them in their own writing.

Based on this information, we recommend instructors to pay more attention to the difference between the universals and indefinite pronouns with non-specific interpretation. Particular examples of correct and incorrect usage may be taken directly from the corpus. The students can be asked to look for these examples themselves and then explain the mistakes therein.

The data from the Russian Learner Corpora shows the importance of additional work on *koe-* indefinites. Working with the authentic data of the corpus instead of "artificial" assignments from textbooks creates an atmosphere of involvement in which students feel more "personal" about the language material and become active participants in the learning process through conscious analysis of errors instead of doing tedious and often inefficient drills.

References

- [1] Abusch, D. The scope of indefinites. *Natural Language Semantics*, 1994, 2, 83-135.
- [2] Cuto. Eugenija. Ruske neopredene zamjenice s cesticama. *Strani jezici* 33. 2004, 3-4: 323-329
- [3] Dahl, Osten. Some notes on indefinites. *Language*. 46, 1970, 33-41.
- [4] Kagan, Olga. Specificity as speaker identifiability. *Logic and Language* 9. August 24-26, 2006. Handout
- [5] Paducheva, E. B. Proposition and its relation to reality: referential aspects of pronouns semantics. 1985, Moscow, Science.
- [6] Pereltsvaig, Asia. "Russian *nibud'*-Series as Markers of Co-variation". In *West Coast Conference in Formal Linguistics Proceedings*. 2008, 370-378.



International Conference ICT for Language Learning



- [7] Seliverstova, O.N. *Pronoun in Language and Speech*. 1988, Moscow. Science.
- [8] Sheliakin, M.A. *Russian pronouns: meaning, grammatical forms and distribution*. 1986, Tartu.