Russian National Corpus as a Tool of Linguo-Didactic Innovation in Teaching Languages

Lyubov Dmitrievna Ponomareva, Lyubov Nikolaevna Churilina, Darya Sergeyevna Buzhinskaya, Elena Nikolayevna Derevskova, Oksana Vyacheslavovna Dorfman, Elena Petrovna Sokolova
Nosov Magnitogorsk State Technical University, RUSSIA.

ABSTRACT
Emphasis on universal learning activities of each student rather than acquisition of ready knowledge, as well as on how an individual masters a language necessitate the development and application of innovative technologies promoting functional-semantic and textual approaches. In the modern context, Russian language teachers, along with knowledge of the theory of language, methods of planning and conducting lessons of different types and forms, should be able to work with electronic resources. The Russian National Corpus, being a huge collection of texts, allows a teacher to draw on already existing practical teaching materials when planning his students’ route of learning Russian language. The article presents linguo-didactic examples of such work when studying vocabulary and grammar. Comprehensive analysis of the text is defined as “a complex multilevel process involving identification and methodological description of value, linguistic and language-generating levels” (Ponomareva, 2016). The analysis of the value level aims to identify cognitive, educational and aesthetic value of the text; the linguistic one - to estimate and analyze linguistic units as elements of the text. The language-generating level involves the encouragement, promotion and planning of such language-generating activities of students as interpretive (interpretation of the text), reproductive (semantic text compression), productive (creating their own texts of different styles and genres). The article presents a system of design and research tasks for university students to analyze the text “White Nights” by F.M. Dostoevsky, using the Russian National Corpus language as a linguo-didactic tool. Considering the findings of design and research activities of students, we can claim that the Russian National Corpus can be used as a universal linguo-didactic tool for the formation of a linguistic persona.

KEYWORDS
Russian National Corpus, Innovations, Information and Communication Strategy, Project and Research Work

ARTICLE HISTORY
Received 7 August 2016
Revised 2 November 2016
Accepted 13 December 2016

Introduction
Socio-cultural and political events of public life determine dynamic changes in the language system, updating of its units. Language teaching should not only reflect the synchrony and diachrony of the language, but also broaden the range of learning tools which require updating, implementation of innovative forms. Therefore, modern educational technologies, electronic resources should become a tool for the formation of competences of a linguistic persona.

CORRESPONDENCE Lyubov Ponomareva  L.Ponomareva@mail.ru
Innovation of language teaching contributes to the learning process becoming more student-oriented, for instance, it allows implementing project activities in accordance with the requirements of modern educational standards. “Using this approach, one can get interactive training, which enables to change the role of a teacher who in the classroom acts as an organizer, expert, tutor, and a facilitator” (Dorfman, 2016).

The potential of using the National Corpus in training should requires further promotion, despite the fact that the corpus linguistics began to form in the 1990s, and the Russian National Corpus (RNC) was formed in 2004. As a collection of diverse texts that represent a natural, dynamically developing language, the RNC increases students’ motivation to learn and promotes their research independence.

In the 21st century innovations in language teaching have been considered in a number of pedagogical, methodological, and linguistic studies. Regarding the issues covered in this paper, in accord with some researchers, we define innovation in education as “innovation, intended to solve the existing problem situation in order to improve the learning, its quality and to create favorable conditions for students mastering the material” (Basharina, 2009). The key parameter of the innovation at the micro level – in educational institutions – is the educational product used, updating of which is the key to innovation. That is why it is crucial to use information resources in education.

The dominance of a purely regulatory approach as the main feature of Russian language teaching in schools was highlighted in some papers: “We teach language as a system of rigid, dead rules that should be learnt and practiced, and there is practically no chance for any creative inspiration, directly connected with the subject studied. Schoolchildren have no need to speculate on some issue, since there is no need to think if there are set rules for everything” (Dobrushina, Levinzon, 2006). The researchers note that the skills of using the Internet resources in modern education (both in secondary school and university) are only beginning to form: students often do not distinguish between information search and analysis, and substitute understanding the material with finding it.

Talking about the potential of the RNC application in education, researchers noted that it is an instrument that compels a young Internet user to move away from the “dependent” position: the information that can be obtained with the help of the Corpus is extremely extensive, but one cannot use it without making some intellectual effort. The authors recommend using the RNC as the basis for creative tasks: the data obtained from it are not the answer to the question, as the information found should be analyzed. On the one hand, using the RNC it is easier to find examples necessary for presentation of the material; it is possible to quickly compare the linguistic data relating to different time periods; one can easily solve the problem of identifying the degree of importance of the phenomenon under study in the minds of native speakers. On the other hand, training turns into a “game”: even the first and the easiest stage of working with the RNC – finding examples – requires certain intellectual effort from a student. Finally, the information found in the RNC can form the basis for the creative search done by the teacher and the student together. Thus, the RNC provides the participants of the educational process with a real opportunity to develop new efficient ways of working with the language.
Moreover, the RNC can be used in the context of the DDL method – autonomous learning based on the data analysis, the concept developed by T. Johnes (1988), thus increasing students’ motivation. “A teacher, using the DDL, ceases to be the source of the ultimate truth: he is in charge of creating working conditions, giving clear instructions and initiating the thoughtful discussion of the outcomes. The teacher, therefore, passes to the student the most valuable information he possesses: the information on data analysis strategies” (Levinzon, 2015). The proposed method of working with the RNC seems to us highly successful when performing project tasks.

Some papers propose the following approaches to planning work with language learning using the RNC material: “As for the content, creation of the educational environment on the basis of the RNC makes it information-rich, diverse, and offers a huge didactic potential; as for its functionality, the designed environment creates conditions for mastering different types of activities (training, research, design, etc.) and organizational forms (individual, pair and group work, etc.)” (Fomicheva, 2015). The author outlines four areas of individual work: preparing and doing learning activities of various types, microanalysis, verification of dictionaries, textbooks, and a research project. “The ability to work with an electronic source of linguistic material indicates a high level of autonomy and implies application of a variety of particular skills: launching a project, planning individual or group work, project management, and presentation of results. Interacting with an information resource, students learn to design their own learning environment, which is one of the factors for the development of learning autonomy and a prerequisite for life-long learning” (Fomicheva, 2015).

The potential of the RNC regarding the application of the competence-based approach to teaching humanities is estimated as follows: “The RNC is seen primarily as a research tool, allowing one to quickly and effectively conduct linguistic research using a large text database. At the same time, it can be used effectively in devising training activities. The RNC makes it possible to conduct many scientific studies in school. Some research types associated with the collection of a large volume of data that used to take months and years, can now be performed in a matter of hours or even minutes. There is also a change in the focus of research activity planning. Creating an algorithm now turns to be the most difficult and time-consuming part of the study. This is what makes the RNC an effective tool that develops different linguistic competences of students and allows a teacher to use the activity approach not only for developing communication skills, but also for teaching linguistic theory” (Matyushkin, 2013).

Numerous articles, published in recent years and dwelling on the experience of practicing teachers and educators, prove the relevance of the RNC use in teaching languages (Slobodyan, 2010; Skovorodkina, 2013; Gerasimova, 2014).

Modern language teaching, actively using information resources such as the Russian National Corpus, aims to develop students’ cognitive abilities, promotes awareness of training, self-control and self-esteem. Teaching Russian within the framework of innovative technologies should contribute to the development of communicative and linguistic competencies of students, to increase their research skills. To this end, one should apply specific pedagogical management to the learning process, because it is the teacher/instructor who should create the environment that enable the development of these competencies.
Teaching Russian language through the RNC will be effective if the following pedagogical conditions are observed:
– students’ activity is motivated;
– the project method is used to create optimal organizational forms of students’ activity;
– there is an algorithm describing students’ activity which includes tasks for individual/group projects (case studies).

Methods

Using the RNC as an electronic collection of texts in the language studied equipped with a certain description a priori implies that corpus analysis methods are applied.

The markup options (morphological, syntactic, semantic, metalinguistic, etc.) provided by the RNC developers not only significantly simplify the procedure of searching linguistic units and structures with given semantic and grammatical properties that is part of any linguistic study, but also allow using methods of analysis traditional for linguistics (and philology in a wider sense) more efficiently: the method of oppositions, taxonomic method, contextological analysis and others.

The approach to working with a specific text, aimed at the analysis of its structural, grammatical and semantic features, implies the active use of techniques of qualitative and quantitative analysis: stylistic-statistical and quantitative methods, as well as the most current version of the latter — the method of content analysis.

If in the course of training a need arises to demonstrate the co-occurrence principle of language units, one can use different variants of distributional analysis, involving observation of typical or, on the contrary, unique context of the language unit. At the initial stage of working with the text we see it viable to use the method of concordances — identifying the nearest left and right context of the unit studied.

Analysis of the content aspect of a particular text, its thematic and semantic features implies the use of the method of component analysis, involving decomposition of the meaning into minimum elements. With a literary text, it is effective to use the keyword method, which along with component analysis becomes the key to understanding the meaning of the text as a whole by students.

Results

Analysis of lexical and grammatical units in the source text using the RNC

Working with the RNC, students use the following algorithm of actions:
2. Select “Search the Corpus” on the left of the page.
3. Click “Customize subcorpus” in the top right corner.
4. In the field “Title” specify “White Nights”, in the field “Author” type “Dostoevsky”. Click the “Next” button below.
5. Click “Save subcorpus and go to the search page”.

Analysis of the lexical structure of the source text using the RNC

Starting to work with the RNC, students are supposed to master the research methods of applied to the linguistic material and the analysis that follows it.

The task. Using the RNC markup, define the meaning of words презреть, нетривиальный, патетический, комиссия (to despise, non-trivial, fervent, a commission) in the text.

The next task includes not only finding a word in the Corpus and defining its meaning in the dictionary, but also analyzing the closest context influencing the meaning of the lexical unit. Doing this type of task, the student should determine the difference of lexical-semantic variants of a polysemantic word and to show the relationship of such semantic phenomena as polysemy and synonymy.

The task. Identify the context closest to the words “a sin” and “to feel” in the text. Analyze these contexts. Using markup, determine the difference in the meaning of each case of the word use, look up in special dictionaries of synonyms for each of the identified meanings.

It is generally accepted that the “lexical” worldview is directly connected with a “word-building” one, since the study of the semantics of derived words reveals systematic relationships between related units (those with the same root), thus making it possible to understand the mechanisms crucial for speech. This includes mastering the lexis, and the creation of nominative tools. Thus, the next task is aimed at studying the relationship of lexical and word-formation analysis. Conducting such research will enable one to identify different shades of meaning of derivatives.

The task. Type in word fragments чуд*, роб*/робк*, обид*, сме*, добр* (parts of the words *miracle, *shy/timid, *offense, *brave, *kind) in the field “Word”, click the hyperlink “Select” that goes after the heading “Grammatical Features”. In the box opened tick Noun, Adjective, Verb, Adverb and Predicative. Using the markup, analyze lexical meanings of all derivatives obtained that contain the analyzed roots. What are the additional components in the meanings of the derivatives?

Further tasks bring the students to the concept of the lexical structure of the text, which, as shown by research (Churilina, 2011; Derevskova, 2012) implies analyzing the text as a set of lexical units, paradigmatically and syntagmatically related. One of the first researchers developing the concept of the lexical structure of the text, V.V. Stepanova noted that the correlative relationships with the lexical system of the language “determine the specific features of paradigmatic and syntagmatic relations of words in the lexical structure of the text”. Paradigmatics of the lexical structure of the text is defined as “a set of lexical units, formed as a result of their selection, reflecting the features of these units as elements of the lexical system, and in the text – their ability to name the basic concepts of the text, to reveal their meaning” (Stepanova, 2006).

As the next step, students were to perform a task aimed at identifying the dominant meanings of the text. Here, it was viable to use the keyword method created in the framework of linguistics. This method has proven to be an efficient tool of text analysis due to the fact sufficient linguistic material can be used (Churilina, 2012a; Churilina, 2012b).
The task. Write down the key (main) words that reflect the central meaning of the text extract. Using the RNC materials, identify the common component in the meaning of each of the keywords. Make a topical glossary of this lexical group.

But how fine joy and happiness makes any one! How brimming over with love the heart is! One seems longing to pour out one's whole heart; one wants everything to be gay, everything to be laughing. And how infectious that joy is! There was such a softness in her words, such a kindly feeling in her heart towards me yesterday. . . . How solicitous and friendly she was; how tenderly she tried to give me courage! Oh, the coquetry of happiness!

The correlative relationships lay the basis which creates “the originality of syntagmatic relations of lexical units” (Stepanova, 2006), so the next task (the most difficult in terms of how to accomplish it) involves the application of several areas of analysis. Firstly, the onomasiological one, i.e. student should choose in their glossary a character that, from their point of view, can accurately convey the meaning of the text. Secondly, the semasiological one, i.e. student analyze the author’s language unit and carry out a comparative analysis of the meanings of these two words, immersing into the author’s original idea.

The task. Fill in the gaps with the words that, from your point of view, convey the meaning of the text. Compare the text obtained with the text by F.M. Dostoevsky. Using the RNC markup, analyze the lexical meanings of words; identify the difference between the words you chose and the author’s:

It was a ... night, such a night as is only possible when we are young, dear reader. The sky was so ... , so ... that, looking at it, one could not help asking oneself whether ... and ... people could live under such a sky. That is a ... question too, dear reader, very ..., but may the Lord put it more frequently into your heart! . . . Speaking of ... and ... people, I cannot help recalling my ... condition all that day. From early morning I had been oppressed by a ... despondency. It suddenly seemed to me that I was lonely, that every one was forsaking me and going away from me.

Performing this task, students approach the final stage of design and research of analyzing the lexical structure of the text – writing an essay.

The task. Using a set of key words, write about the contents of a text fragment.

Analysis of the morphological units of the original text using the RNC

The system devised by us for working with the grammatical structure of the studied text can solve fundamentally different tasks: when studying the text, the student can move both from the form to the meaning of the linguistic unit, and in the opposite direction – from the meaning to the ways of its expression. The first way is effective if we teach some grammar topic at the initial stage, the second – allows the learner to reach a higher level of understanding of the linguistic material. At the same time, we can vary the level of difficulty according to the stage of training and the level of students.

The task. Using the RNC search system, select proper nouns in the novel “White Nights” by F.M. Dostoevsky. What are the main differences between proper and common nouns? Using the RNC, identify the relationship between common and proper nouns and the category of number. Write a short essay upon the findings of this mini-study.
The task. Note which nouns and after what criteria are classified as animate/inanimate. The RNC search system gives examples of using nouns in the text. Define animation/inanimation of the nouns in the following extract.

The dreamer — if you want an exact definition — is not a human being, but a creature of an intermediate sort. For the most part he settles in some inaccessible corner, as though hiding from the light of day; once he slips into his corner, he grows to it like a snail, or, anyway, he is in that respect very much like that remarkable creature, which is an animal and a house both at once, and is called a tortoise.

The task. In the given subcorpus (the literary text), with the search system, select the examples of the usage of such words as: призрак, характер, лицо, мертвец (a ghost, a character, a face, a dead person). Derive the forms of the accusative plural from these nouns. Specify the form each case corresponds to.

Consider in which case the classification of the word as animate or inanimate may be influenced by one of the word meanings. According to the results of mini-research conducted, make a grammatical dictionary of homonymous forms, add two examples from the Corpus for each form, illustrating the variability of the animate/inanimate category.

The task. In the RNC search system, identify all cases of using such words in the analyzed text as сказать, говорить, думать, смеяться (say, talk, think, laugh). Write down personal forms of the verb in the context of the subject (for a two-member sentence) or with a dependent word (for a one-member sentence). Describe constant morphological characteristics of these verbs: aspect, transitive, reflexive, conjugation; variable morphological characteristics: imperative, tense, person, number, gender; syntactic role.

Analysis of transitive grammatical phenomena in the source text using the RNC

The greatest difficulty caused when studying Russian morphology is posed by so-called transitive phenomena, since “during the functional transition one and the same word can take the place of different parts of speech depending on the context” (Sokolova, 2014). Studying the text helps the students to compare their own language experience with the potential of the language system.

The task. Using the RNC search system, select all examples of the word один (one/alone) in the novel “White Nights” by F.M. Dostoevsky. Determine the part of speech this word takes in the text:

1) as a numeral denoting the number, amount;
2) as an adjective: a) “on one’s own, separately from the others, alone”; b) “no one else or nothing else; only”, “the only one available”, with words лишь, только (only, just); c) “the same, identical; similar”; with the words и тот же, и то же (the same, same); d) “integral, indivisible, single”; d) in combination with the word другой (another/other) when contrasting;
3) as a noun: a) “referring to a person”;
4) as a pronoun: a) “of some kind, some”; b) with a preposition из (from/out of); c) in combination with the word другой (another/other) when listing.

Represent the findings of the statistical study in a table indicating the number of occurrences of the word один (one/alone) as a numeral, adjective, noun and pronoun. Highlight features allowing classification of this word as each part
of speech. Consider why the author so often uses this word in the novel (29 occurrences/26 examples in the RNC) and why it is mainly used as a numeral or an adjective. How, in your opinion, is it connected with the ideas and the artistic content of the novel “White Nights” by F.M. Dostoevsky, as well as with the images of the main characters?

The task. With the RNC search system, find examples of using the word мимо (past/by) in the analyzed text. Determine whether this word is used as an adverb or a derivative preposition in each of the six considered statements. What are the features that allowed you to distinguish between the use of the word мимо (past/by) in the function of an adverb and as a preposition.

Similar to the studied examples, make two sentences with the words вокруг, около, посередине, напротив, впереди (around, near, in the middle, opposite, in front of), using each of them first as an adverb, and then – as a derivative preposition.

The task. In the given subcorpus, using the search system, choose examples of the word каждый (each/every). After analyzing all nine sentences provided by the RNC, specify the cases in which this word is used as an attributive pronoun, and in which it has the meaning of a noun.

Consider what changes take place when the analyzed word becomes another part of speech: how the semantic question and the general meaning alter (an indication of the characteristic of the subject or an object / person), syntactic function in the sentence. Identify the main criteria for differentiating between pronouns and homonymous forms of nouns.

Determine pronoun of which categories can act as nouns. Upon the findings of the conducted mini research, make a grammatical dictionary of homonymous forms, incorporating the words that can be used as pronouns and as nouns depending on the context. Add two of your own examples for every form to illustrate the phenomenon of changing the parts of speech.

The task. The RNC search system provided three examples how the word весело (lighthearted, gaily, gay) was used in the studied text:

1. Instantly I felt lighthearted, and I passed the barrier...
2. He is rich now with his own individual life; he has suddenly become rich, and it is not for nothing that the fading sunset sheds its farewell gleams so gaily before him...
3. One seems longing to pour out one’s whole heart; one wants everything to be gay, everything to be laughing.

Identify the part of speech of the selected word in each sentence, for which purpose analyze the syntactic structure of that part of the sentence which includes the word. Underline the clause, determine the syntactic function of the word весело (lighthearted, gaily, gay). Identify whether this part of the sentence is a one- or two-member one.

Summarizing the findings of the mini-study, conclude on the role of syntactic factors in determining the part of speech of the word. Make a table of correspondences, stating in the left column the part of speech, and in the right – the reason for this choice. The material for the first column: adjective, adverb, category of state; for the second: the predicate in a one-member impersonal
sentence; a verb in a two-member sentence; an adverbial dependent on the verb-predicate.

In a similar way, analyze homonymous grammatical forms: грустно, горько, прекрасно, тяжело, стыдно, холодно, страшно (sad, bitter, beautiful, hard, shameful, cold, scary) that are frequently used in the novel.

**The task.** In the given subcorpus find examples of the use of simple forms of comparative degree: лучше, меньше, больше, богаче, моложе (better, less, more, richer, younger), define their part of speech and function. Conclude on whether the criteria you devised and listed in the table can be applied to the forms of comparative degree.

Summarize the results of the study in a micro essay titled as “There is nothing in morphology which cannot first appear in the syntax...” (V.V. Vinogradov).

**The task.** With the RNC search system, identify all cases of using the word много (a lot) in the novel “White Nights” by FM Dostoevsky. Specify the syntactic function of this word in the examples found. For each sentence, write down phrases with simple structure with the word a lot, identify the main word and the type of subordination of the components (concord, government, adjournment).

Having analyzed and summarized the syntactic characteristics of the word a lot, indicate which part of speech it represents in each of the examples. Conclude on the role of syntactic features (agreement with a verb or a noun being the leading one), according to which the analyzed forms are classified as an adverb or a numeral.

In the given subcorpus of the RNC find examples of using the word столько (such a lot of). Similar to the micro study conducted, identify what part of speech it is.

**The task.** In the RNC search system, specify all cases of using of the word казалось (it seemed/it looked like) in the studied text. Analyze the syntactic structure of all seven sentences, and determine whether the word is a parenthetical expression or a part of a complex sentence.

Group the identified phrases after the syntactic function they perform. Describe the group of parenthetical expressions according to the plan: 1) the meaning of the parenthetical expression considering the relationship of the speaker to the reported information; 2) the position of the parenthetical expressions in a sentence (at the beginning, in the middle or at the end); 3) syntactic function, the presence or absence of grammatical relationships with other members of the sentence; 4) punctuation.

Carry out syntactical analysis of the sentences making up another group (the analyzed word is the part of the sentence): underline the main and secondary parts of the sentence, specify what parts of speech they are. Use arrows to display grammatical relations of the studied word. Note the way in which a member of the sentence represents the analyzed word and the type of clauses referring to it.

Based on the findings of the study, create an algorithm to distinguish between the use of homonymous grammatical forms as a parenthetical expression and a part of a sentence. Check that the algorithm is correct by analyzing the words верно, видно, наверное, возможно, кажется (right, probably, perhaps, maybe) that can be found in the given subcorpus of the RNC.
Conclusion

The modern teaching Russian language, which takes into account the current wide use of ICT and mass communication in all spheres of social life, requires a substantial expansion of the set of learning tools applied. The Russian National Corpus is one of the universal linguo-didactic tools.

The efficiency of the RNC in teaching Russian language stems from the following factors:

– taking into account specific features of how modern children and adolescents absorb information;
– reaching various educational objectives at lessons of various type and as part of extracurricular research work through application of various teaching technologies;
– using technology for illustrative purposes (multimedia);
– planning tasks for group and individual work of students;
– ensuring high speed and saving time required for selection and analysis of linguistic means;
– efficient planning and conducting various forms of testing to evaluate knowledge, abilities and skills of Russian language students, objective quality control assignments; measuring how well students have mastered the material.

The main risks associated with using the RNC as a linguo-didactic tool is that the Corpus contains different kinds of errors (the Corpus was not originally designed to strictly adhere to the norm); using the RNC too widely can hinder reaching educational objectives; a teacher may not be ready to combine various techniques, use the RNC in a flexible and variable way; a school may lack required facilities and equipment. Identification of risks allows the teacher to understand and timely deal with possible complications.

All considered ways of using the RNC as a linguo-didactic tool of innovation in language learning still require further development and study in papers on pedagogy and methodology. Design and experimental verification of linguo-didactic aids for Russian language teaching by means of the RNC will be the result of further cooperation of teachers, educators, linguists, teachers of Russian language and literature.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Lyubov Dmitrievna Ponomareva, D. Sc. in Pedagogics, Associate Professor, Department of Russian Language, General Linguistics and Mass Communication, Institute of Humanitarian Education, Nosov Magnitogorsk State Technical University (NMSTU)

Lyubov Nikolaevna Churilina, D. Sc. in Philology, Professor, Department of Russian Language, General Linguistics and Mass Communication, Institute of Humanitarian Education, Nosov Magnitogorsk State Technical University (NMSTU)

Darya Sergeyevna Buzhinskaya, Ph. D. in Philology, Department of Russian Language, General Linguistics and Mass Communication, Institute of Humanitarian Education, Nosov Magnitogorsk State Technical University (NMSTU)
Elena Nikolayevna Derevskova, Ph. D. in Philology, Associate Professor, Department of Russian Language, General Linguistics and Mass Communication, Institute of Humanitarian Education, Nosov Magnitogorsk State Technical University (NMSTU)

Oksana Vyacheslavovna Dorfman, Ph. D. in Pedagogics, Department of Russian Language, General Linguistics and Mass Communication, Institute of Humanitarian Education, Nosov Magnitogorsk State Technical University (NMSTU)

Elena Petrovna Sokolova, Ph. D. in Philology, Associate Professor, Department of Russian Language, General Linguistics an and Mass Communication, Institute of Humanitarian Education, Nosov Magnitogorsk State Technical University (NMSTU)

References


