REVIEWS AND REPORTS
Larisa V. Nenasheva, Lyudmila S. Shurykina. Electronic Dictionary of Arkhangelsk Dialects

201

Electronic Dictionary of Arkhangelsk Dialects

Larisa V. Nenasheva [1✉], Dr. Sci. (Phil.), Associate Professor, Professor
Lyudmila S. Shurykina [2], Assistant

[1, 2] Northern (Arctic) Federal University named after M.V. Lomonosov, Naberezhnaya Severnoy Dviny, 17, Arkhangelsk, Russia
[1] l.nenasheva@narfu.ru ✉, ORCID: https://orcid.org/0000-0001-8580-5980
[2] l.shurykina@narfu.ru, ORCID: https://orcid.org/0009-0004-8547-1967

Abstract. The research group of the Northern (Arctic) Federal University is implementing the project "Thematic dictionary of Arkhangelsk dialects with electronic support" supported by the Russian Science Foundation. The project aims to publish three issues of the dictionary, which will include the vocabulary reflecting the traditional world picture of the Arkhangelsk peasant: the first issue contains the vocabulary related to the names of clothes, shoes, hats; the second issue contains the names of residential and non-residential buildings; the third issue contains the vocabulary naming the food and drinks of the Arkhangelsk inhabitant. The second task of the project is to create an electronic dialect corpus devoted to the Arkhangelsk dialects, which is currently not available in universities of the North-West region. The article presents the experience of the project, which is at the intersection of dialectology, ethnography, cultural studies and corpus linguistics. The results of this research and the electronic database can be used in dialectological, folklore and ethnographic expeditions, in teaching Russian literature at universities and schools, in organization of local history work, in educational projects aimed at promoting northern spiritual culture, in preparation of cultural and scientific events devoted to the language and culture of the Russian North.
Keywords: *Arkhangelsk dialects, corpus linguistics, dialect dictionary*

Acknowledgments and funding

Introduction

The Department of Russian Language and Speech Culture of the Higher School of Social Sciences, Humanities and International Communication of the Northern (Arctic) Federal University named after M.V. Lomonosov keeps a card index of Arkhangelsk dialects recorded during dialectological expeditions in the Arkhangelsk Oblast, starting from 1960s. This interesting, extensive material is a great source for studying the living northern vocabulary. Currently, it is necessary to systematize this rich material and make it accessible to researchers and specialists dealing with issues of dialectology, folklore, ethnography, and culture of the Russian North.

REVIEWS AND REPORTS
Larisa V. Nenasheva, Lyudmila S. Shurykina. Electronic Dictionary of Arkhangelsk Dialects

202

Along with the National Corpus of the Russian Language [1], where a collection of texts in Russian is presented, there is a subcorpus of historical and dialect texts. In recent years, scientists from various universities have been creating electronic dialectal corpuses, which are posted on university web pages. Such electronic databases are available at the Volgograd State Socio-Pedagogical University, Dostoevsky Omsk State University, Tomsk State University, etc.

Currently, the collected dialect materials of the card index of the Department of Russian Language and Speech Culture are kept in handwritten form on paper media, and access to them is difficult for researchers and specialists. Therefore, there is a need to create an electronic corpus and information system in order to duplicate the traditional method of storage and provide convenient work with data. The Department of Russian Language and Speech Culture of NArFU named after M.V. Lomonosov is developing a web application and a mobile application to provide convenient access to dialectal materials. The mobile application will also allow working with the materials in the field and will facilitate systematization and classification of the collected materials on site.

Main part

In 1980s, A.S. Gerd proposed to compile a digital fund of historical and dialect texts. His idea was supported by V.E. Goldin [1]. Today there is a wide range of corpuses, both foreign and Russian, that store dialect texts and demonstrate elements of dialect speech, for example, foreign corpuses — The Nordic Dialect Corpus [2], The Freiburg English Dialect Corpus [3], Helsinki Corpus of British English Dialects [4]), Russian ones — dialect subcorpus within the National Corpus of the Russian Language (NCRL) [5], Saratov dialect corpus [6], Tomsk dialect corpus [7], Kuban dialect corpus [8], Dialect corpus of linguistic culture of the Northern Angara region [9], Corpus of folk speech of the Middle Irtysh region [10], Volgograd lexical atlas [11].

The dialects of the Russian North, in particular the Arkhangelsk dialects, have attracted the attention of linguists, philologists, and ethnographers for decades due to their remoteness from

---

[1] National Corpus of the Russian Language. URL: https://ruscorpora.ru (accessed 25 July 2023).

[2] Nordic Dialect Corpus. URL: http://www.tekstlab.uio.no/nota/scandiasyn/ (accessed 14 July 2023).

[3] Freiburg English Dialect Corpus. URL: https://fred.ub.uni-freiburg.de/ (accessed 14 July 2023).

[4] Helsinki Corpus of British English Dialects. URL: https://varieng.helsinki.fi/CoRD/corpora/Dialects/ (accessed 14 July 2023).

[5] Corpus of dialect texts: National Corpus of the Russian Language. URL: https://ruscorpora.ru/corpus/dialect (accessed 24 July 2023).

[6] Kryuchkova O.Yu., Goldin V.E. Mul'timediynyy dialektologicheskiy korpus — vazhneyshiy resurs sokhraneniya i izucheniya narodno-rechevoy kul'tury [Multimedia dialectological corpus is the most important resource for preserving and studying folk speech culture]. URL: http://sarteorlingv.narod.ru/dialekt/kru4kova-goldin.html (accessed 24 July 2023).

[7] Tomsk dialect corpus. URL: https://losl.tsu.ru/?q=corpus (accessed 24 July 2023).

[8] Kuban dialect corpus. URL: https://ethnolex.ru/kubdk/ (accessed 20 July 2023).

[9] Electronic text corpus of linguistic culture of the Northern Angara region. URL: http://angara.sfu-kras.ru (accessed 20 July 2023).

[10] Dictionary of constants of folk speech of the Middle Irtysh region. URL: http://dict.univer.omsk.su/ (accessed 20 July 2023).

[11] Word on the map. Volgograd lexical atlas. URL: http://dialekt.vspu.ru/?q=node/2 (accessed 20 July 2023).

REVIEWS AND REPORTS
Larisa V. Nenasheva, Lyudmila S. Shurykina. Electronic Dictionary of Arkhangelsk Dialects

203

the center and archaic nature. In recent decades, domestic scientists have been seriously mastering the vocabulary of the Russian North; research has resulted in the publication of such dictionaries as the "Arkhangelsk Regional Dictionary", published by Moscow State University named after M.V. Lomonosov, edited by O.G. Getsova and E.A. Nefedova, "Dictionary of dialects of the Russian North", edited by A.K. Matveev, published by Ural University, "Dictionary of Russian dialects of Karelia and adjacent regions", edited by A.S. Gerd, published by St. Petersburg State University, "Dictionary of Pinega dialects" by A.N. Levichkin and S.A. Myznikov (the first issue of the dictionary with sample articles was published in 2014).

Currently, the work on creating thematic dictionaries has also intensified, such as, for example, "Thematic Dictionary of dialects of the Tver Oblast" (2002–2006) [2], "Lovetskoe slovo: Dictionary of Volga-Caspian fishermen" by E.V. Kopylova (1984) [3], dictionary by Kostroma local historian A.V. Gromov "Vocabulary of flax growing, spinning and weaving in Kostroma dialects along the Unzha River" (1992) [4], "Dictionary of geographical terminology in Russian speech of the Perm Krai" by E.N. Polyakova (2007) [5].

The rich dialect material stored at the Department of Russian Language and Speech Culture makes it possible to create a thematic dictionary in which words are grouped according to a thematic principle. The dictionary contains vocabulary for one or another sphere of people's life (clothing, food, buildings, nature, flora and fauna, etc.) within the Arkhangelsk Oblast. The dictionary materials can serve as a good source for reconstructing the traditional picture of the world of the Arkhangelsk peasant. Due to the development of computer technology, there is a need to create an electronic version of the "Thematic Dictionary of Arkhangelsk Dialects".

Within the framework of the grant by the Russian Science Foundation, it is planned to prepare three issues of a thematic dictionary of Arkhangelsk dialects, including vocabulary associated with the names of clothing, residential and non-residential buildings, food and drinks, i.e. vocabulary reflecting the traditional picture of the world of the Arkhangelsk peasant.

The first issue of the thematic dictionary "Clothes, shoes, headwear, accessories, fabrics" includes such subgroups as "general name of clothing", "qualitative name of clothing", "women's outerwear", "men's outerwear", "sundresses, dresses, women's sweaters, shirts, skirts", "men's shirts, trousers", "materials used to make clothes", "mittens, gloves", "hats, female and male", "shoes, female and male", "socks, stockings", "children's clothing", "vocabulary of wearing clothes and shoes", "vocabulary of manufacturing and repairing clothes and shoes", "accessories".

At the moment, the dialectal material recorded by teachers and students of the Pedagogical Institute (now — Northern Arctic Federal University) during dialectological expeditions is stored in the form of field recordings on paper (in notebooks and on cards) and on tape cassettes. This material is very valuable because it preserves the original, unique, drawling, melodious northern speech. The dialect material in notebooks indicates personal data of informants: last name, first name, patronymic; year of birth (age); education; it is also noted whether the informant is an indigenous person — this information is especially important as the authentic northern speech is

REVIEWS AND REPORTS
Larisa V. Nenasheva, Lyudmila S. Shurykina. Electronic Dictionary of Arkhangelsk Dialects

204

preserved among indigenous people. Dialect words in notebooks are underlined and accented. The word is written in context, based on which the meaning can be formulated. One can also find drawings of everyday objects in notebooks, which illustrate how the spoken object looks like. Fig. 1, 2 show samples of field notes recorded during dialectological expeditions in the village of Ukhta, Kargopol region, in June—July 1991 (Fig. 1) and in Oshevensk, Kargopol region, in June—July 1993 (Fig. 2).
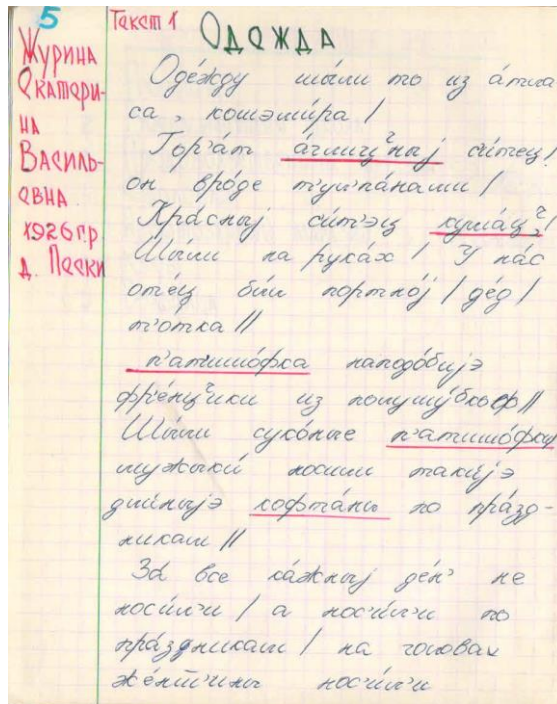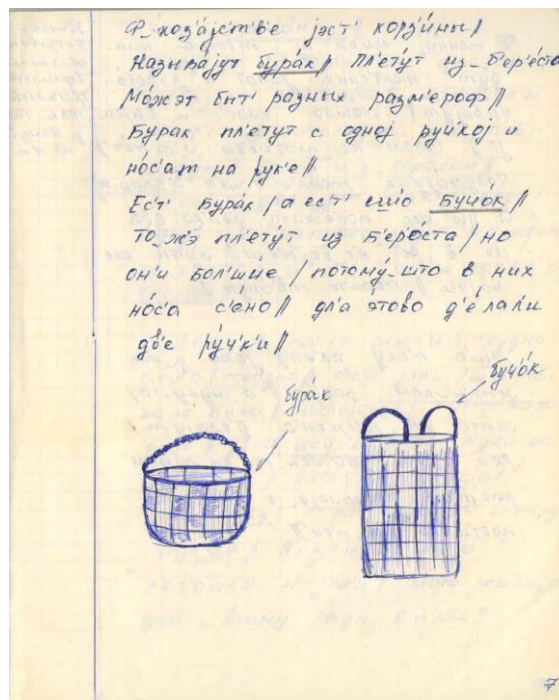

Fig. 1. Dialect record.

Fig. 2. Dialect record with illustrations.

From the collected records, dialect words were written down on cards in their original form: noun — in the nominative case, singular or plural; adjective — in the nominative case, mas-

REVIEWS AND REPORTS
Larisa V. Nenasheva, Lyudmila S. Shurykina. Electronic Dictionary of Arkhangelsk Dialects

205

culine, singular; verb —in the infinitive. The interpretation of the word (definition of the meaning of the word) was formulated, and the context revealing the meaning of the word (illustrative material) was recorded on the card. The locality and district of the Arkhangelsk Oblast where the word was recorded, as well as the initials of the informant, were also noted down.

Fig. 3. The word "Kazachina" [long fur coat].

At the moment, the first issue of the thematic dictionary "Clothes, shoes, headwear, accessories, fabrics" has been prepared for publication. The same material has been loaded into the electronic dialect corpus. The formed cards are entered into the database of dialect words with the help of a specially developed application "Word Box".

"Word Box" allows forming lists of dialect words. For this purpose, information about each word is entered into a special electronic card. Entries can be edited or deleted if necessary. The resulting list of words can be exported in .docx format as a set of dictionary entries or in .csv format as a dataset suitable for analysis.

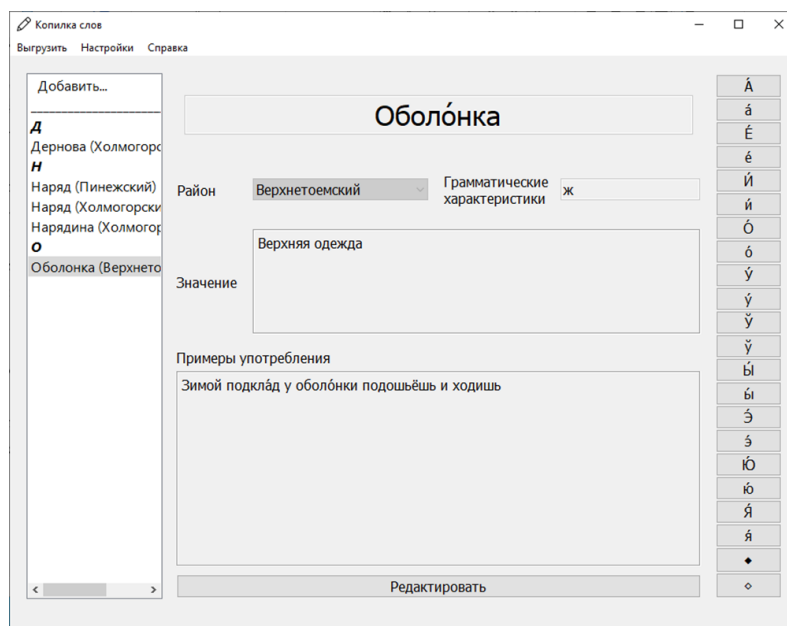The interface of the application is shown in Fig. 4.

Fig. 4. Application interface.

The menu bar allows accessing the program's features. In the "Download" submenu there are two lines: 1) "csv", which allows downloading the collected vocabulary cards in .csv format for machine analysis, and 2) "word", which allows downloading the collected vocabulary cards in .docx format as an alphabetically sorted dictionary. The "Settings" submenu contains the line "Change font size", by clicking on which the dialog box for setting the font of text elements is called up. The "Help" submenu contains a corresponding line that opens a window with brief information about the application and its functionality and with the developer's contacts.

In the center there is a vocabulary card, in the fields of which the relevant information from paper cards is entered.

To the left of the vocabulary card, there is a word organizer. Words are displayed without accents to avoid unnecessary visual noise, but with an indication of the areas in which these words are used. The words in the organizer are sorted alphabetically and divided into sections according to the first letter; each section is preceded by an initial — the corresponding capital letter.

Clicking on the "Add…" allows switching to the mode of adding a new word. By clicking on the selected word, the corresponding vocabulary card is shown. Clicking on the "Edit" allows switching the vocabulary card to editing mode. In order to continue working with the list of words, it is necessary to save or cancel changes in the current card, and at least the first field of the card must be filled in.

To the right of the vocabulary card area, there is a panel of additional symbols. They are used to place accents (Á), indicate brevity (Ў), and add other necessary information (♦, ◇). You can use them by placing the mouse cursor in the "word", "word meaning" or "examples of use" field and then click on the desired symbol with the left mouse button. The symbol will be inserted at the cursor position. You can also use apostrophes to set accents: after editing the field, all vowel letters with apostrophes will be replaced with vowel letters with accents.

As part of testing, the application was offered to a group of eight students of the Philology department as one of the tasks for practical training. As a result of testing, it was found that the application works correctly and performs the stated functions. The files generated as a result of processing in .csv format (datasets) are suitable for analysis.

A dataset was collected and analyzed on the basis of the materials included in the first issue of the thematic dictionary "Clothes, shoes, headwear, accessories, fabrics". Thus, Fig. 5 shows a word cloud that reflects the most frequent words. The more often a word is used, the larger it appears in the picture.

REVIEWS AND REPORTS
Larisa V. Nenasheva, Lyudmila S. Shurykina. Electronic Dictionary of Arkhangelsk Dialects

207

Fig. 5. Word cloud.

Figure 6 shows a bar graph showing the distribution of the most frequent words in this dataset by region. The figure shows the relative frequencies — the shares occupied by each of the frequency words in each specific area. The diagram demonstrates the uneven use of words: for example, the word "ранний" ["early"] is found in the Vinogradovskiy district more often than other frequency words and occupies a leading position among other districts, and the words "сарафан" ["sundress"] or "шить" ["sew"] are not reflected in the dictionary for this region at all. The diagram also shows that the word "носить" ["wear"] is the absolute leader both among all frequency words and among all regions.
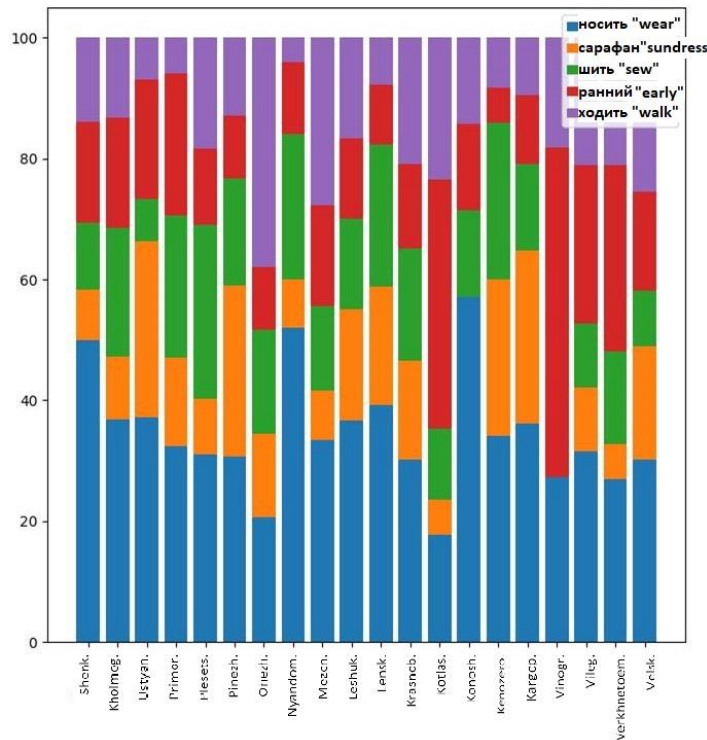

Fig. 6. Distribution of the most frequent words by region

REVIEWS AND REPORTS
Larisa V. Nenasheva, Lyudmila S. Shurykina. Electronic Dictionary of Arkhangelsk Dialects

208

Conclusion

Thus, on the basis of the Northern (Arctic) Federal University named after M.V. Lomonosov, a multi-volume thematic dictionary of dialects of the Arkhangelsk Oblast is being prepared for publication, and for the first time, an independent corpus dedicated to Arkhangelsk dialects is being created. The materials of the thematic dictionary of Arkhangelsk dialects will become available to a wide range of readers, to all who love living words, and the dialect corpus filled with these materials will help specialists in researching Russian folklore, in works on ethnography, social history, folk costume and life, as well as in organizing museum and ethnographic activities. The results of this study can be used in teaching Russian literature at universities and schools; in organizing local history work; in educational projects aimed at popularizing northern spiritual culture; in preparation of cultural events dedicated to the language of the Russian North; in organizing museum and ethnographic activities.

References

1. Goldin V.E. Dialektologicheskiy tekstovyy mashinnyy fond govora i issledovanie dialektnykh izmeneniy [Dialectological Text-Machine Fund of the Vernacular and the Study of Dialectal Changes]. In: *Sovremennye protsessy v russkikh narodnykh govorakh* [Modern Processes in Russian Folk Dialects]. Saratov, 1991, pp. 17–28. (In Russ.)
2. Kirillova T.V., Novikova L.N. *Tematicheskiy slovar' govorov Tverskoy oblasti. Vyp. 1* [Thematic Dictionary of Dialects of the Tver Region. Issue 1]. Tver, TvGU Publ., 2002, 184 p. (In Russ.)
3. Kirillova T.V., Novikova L.N. *Tematicheskiy slovar' govorov Tverskoy oblasti. Vyp. 2* [Thematic Dictionary of Dialects of the Tver Region. Issue 2]. Tver, TvGU Publ., 2003, 240 p. (In Russ.)
4. Kirillova T.V., Novikova L.N. *Tematicheskiy slovar' govorov Tverskoy oblasti. Vyp. 3* [Thematic Dictionary of Dialects of the Tver Region. Issue 3]. Tver, TvGU Publ., 2004, 228 p. (In Russ.)
5. Kirillova T.V., Novikova L.N. *Tematicheskiy slovar' govorov Tverskoy oblasti. Vyp. 4* [Thematic Dictionary of Dialects of the Tver Region. Issue 4]. Tver, TvGU Publ., 2005, 190 p. (In Russ.)
6. Kirillova T.V., Novikova L.N. *Tematicheskiy slovar' govorov Tverskoy oblasti. Vyp. 5* [Thematic Dictionary of Dialects of the Tver Region. Issue 5]. Tver, TvGU Publ., 2006, 150 p. (In Russ.)
7. Kopylova E.V. *Lovetskoe slovo: Slovar' rybakov Volgo-Kaspiya* [Lovetskoe Slovo: Dictionary of Volgo-Caspian Fishermen]. Volgograd, Nizhne-Volzhskoe knizhnoe izdatel'stvo Publ., 1984, 128 p. (In Russ.)
8. Gromov A.V. *Leksika l'novodstva, pryadeniya i tkachestva v kostromskikh govorakh po reke Unzhe: slovar': uchebnoe posobie* [The Vocabulary of Flax Growing, Spinning and Weaving in Kostroma Dialects along the Unzha River]. Yaroslavl, Yaroslavl State Pedagogical University named after K.D. Ushinsky Publ., 1992, 118 p. (In Russ.)
9. Polyakova E.N. *Slovar' geograficheskikh terminov v russkoy rechi Permskogo kraya* [Dictionary of Geographical Terms in the Russian Speech of the Perm Region]. Perm, 2007, 423 p. (In Russ.)