

Народная речь в Национальном корпусе русского языка

Ирина Борисовна Качинская, Московский государственный университет им. М. В. Ломоносова
(Россия, Москва), kacza@yandex.ru

Анна Викторовна Малышева, Институт русского языка им. В. В. Виноградова (Россия, Москва),
annamaly@mail.ru

DOI: 10.31857/S013161170005368-3

АННОТАЦИЯ: Количество информационных ресурсов по русской диалектологии, доступных онлайн, ежегодно увеличивается. В сети Интернет представлены диалектные словари, интерактивные диалектологические карты, звучащая речь. Статья посвящена корпусу диалектных текстов Национального корпуса русского языка (НКРЯ). Этот корпус, созданный в 2005 г., постоянно пополняется и предполагает включение любых диалектных текстов на русском языке, записанных как на территории исконного проживания русского населения, так и на территориях раннего и позднего заселения, миграций. В него входят полевые записи, аудио- и видеорасшифровки, тексты из хрестоматий и сборников. Сейчас в открытом доступе на постоянно действующем сайте <http://www.ruscorpora.ru/search-dialect.html> выставлены тексты, записанные в двадцати пяти областях России. В статье приводятся примеры текстов, входящих в корпус диалектных текстов НКРЯ, с указанием «паспорта» записи, рассказывается о метатекстовой и грамматической разметке и поиске в корпусе. Пополнение диалектного подкорпуса даст возможность лингвистам, филологам, историкам, культурологам, этнографам и всем, кто любит русское слово, работать с образцами народной речи в самых разных целях.

КЛЮЧЕВЫЕ СЛОВА: русская диалектология, корпусная лингвистика, Национальный корпус русского языка, корпус диалектных текстов

для цитирования: Качинская И. Б., Мальшева А. В. Народная речь в Национальном корпусе русского языка // Русская речь. 2019. № 4. С. 103–118. DOI: 10.31857/S013161170005368-3.

БЛАГОДАРНОСТИ: Работа осуществляется благодаря поддержке Российского фонда фундаментальных исследований, проект № 18-012-00557 «Полное дополнение корпуса диалектных текстов Национального корпуса русского языка».

Russian National Corpus

Folk Speech in Russian National Corpus

Irina B. Kachinskaya, Moscow State University by Lomonosov (Russia, Moscow), kacza@yandex.ru
Anna V. Malysheva, Vinogradov Russian Language Institute (Russian Academy of Sciences)
(Russia, Moscow), annamalys@mail.ru

АБСТРАКТ: The number of online information resources on Russian dialectology increases every year. There are dialect dictionaries, interactive dialect maps, speech recordings available on the Internet. This article examines the Corpus of Dialect Texts, a part of the of the Russian National Corpus. Launched in 2005, the Corpus of Dialect Text continually expands to include any Russian dialect texts recorded either in the original area of residence of Russian population or in territories that where settled and migrated to at a later point in time. The corpus includes field recordings, audio and video transcripts, texts from anthologies and compilations, and its collection continues to grow. Presently, the web-site <http://www.ruscorpora.ru/search-dialect.html> provides free access to texts recorded in 25 areas of Russia. The article presents samples from the Corpus of Dialect Texts specifying the “passport” data of each record, provides details on methatextual and grammatical annotation of texts and describes search functions of the corpus.

One of the primary tasks of the dialect subcorpus has to do with collection of new records and further growth of the Bank of dialect texts. The V. V. Vinogradov Russian Language Institute eagerly seeks to enrich the dialect subcorpus with texts from diverse sources from various regions of Russia and other countries in order to create a representative database readily available to anyone who appreciate the folk word — both linguists and non-linguists.

KEYWORDS: Russian dialectology, corpus linguistics, Russian National Corpus, Russian Dialects' Corpus

FOR CITATION: Kachinskaya I. B., Malysheva A. V. Folk Speech in Russian National Corpus. Russian Speech = Russkaya Rech'. 2019. No. 4. Pp. 103–118. DOI: 10.31857/S013161170005368-3.

ACKNOWLEDGEMENTS: This research is supported by a grant from Russian Science Foundation, project No. 18-012-00557 “Corpus of dialect texts within the Russian National Corpus: Additions”.

1. Электронные ресурсы по русской диалектологии. В последние десятилетия диалектологи активно осваивают пространство Интернета. На постоянно действующих сайтах размещаются *диалектные словари*: сводный «Словарь русских народных говоров», «Архангельский областной словарь», «Псковский областной словарь с историческими данными» и др. Частично оцифрованы и опубликованы в Интернете *лингвистические карты* «Общеславянского диалектологического атласа» [ОЛА] и «Диалектологического атласа русского языка» [ДАРЯ]. Доступны исследования по русской диалектологии [Русская диалектология], адресованные не только специалистам-диалектологам, но и широкому кругу читателей. Например, «Школьный диалектологический атлас: Язык русской деревни» на портале ГРАМОТА.RU представляет собой альбом тщательно отобранных лингвистических карт с комментариями и заданиями для самопроверки [Язык русской деревни]. *Звучащую речь* русских деревень из различных областей можно прослушать и скачать на странице «Электронная библиотека русских народных говоров» [Казанский университет]. На сайтах университетов создаются страницы с материалами студенческих диалектологических экспедиций [Шуя] и др.

Но основная масса аудиозаписей диалектной речи, расшифрованных текстов, хрестоматий, созданных на их основе, не выложена в Интернет и по-прежнему остается недоступной для широкого пользователя.

В этих условиях особое значение приобретают корпуса и базы данных, в которых есть возможность работать с диалектными текстами, а часто и прослушивать их. Это, например, «Корпус бассейна реки Устья» с материалами из Устьянского района Архангельской области [Даниэль, Добрушина и др. 2013–2018], «Корпус говора села Малинино» Липецкой области [Ter-Avanesova et al., 2019], «Корпус села Спиридонова Буда» с текстами из Брянской области [Garder et al., 2018] и др. Все тексты имеют аудиосопровождение, расшифровки сделаны в стандартизированной орфографии. Тексты в фонетической транскрипции и с аудиосопровождением, записанные в деревнях Шатурского района Московской области и Харовского района Вологодской области, можно найти на сайте «Вавилонская башня» // «Электронные базы по русским народным говорам» [Крылов, Тер-Аванесова 2011, 2012]. Диалектный корпус этнолингвистических материалов Кубанских говоров находится на сайте лаборатории компьютерной и региональной лексикографии «Региональная этнолингвистика» [Кубань]. Поиск в корпусах может осуществляться по разным критериям, т. к. в них по-разному решаются проблемы отражения фонетики, грамматики, лексики, семантики. Однако ко многим корпусам еще нет доступа через Интернет.

2. Корпус диалектных текстов. Корпус диалектных текстов в составе Национального корпуса русского языка [НКРЯ] появился уже в 2005 году и с тех пор продолжает пополняться новыми материалами. Диалектный подкорпус предполагает включение любых диалектных текстов на русском языке, записанных на территории исконного проживания русского населения (европейская часть России), раннего заселения (Русский Север), позднего заселения (Сибирь, Дальний Восток, Дон, Нижнее Поволжье), а также миграций (говоры старообрядцев и протестантов Прибалтики, Азербайджана, Армении, Болгарии, Румынии, Австралии, Канады, Америки и т. д.). Сейчас в него входят тексты из 25 областей России, все они находятся в открытом доступе на постоянно действующем сайте [Диалектный подкорпус НКРЯ].

В 2018 году в диалектный подкорпус было передано почти 300 текстов общим объемом около 100 тыс. слов из Архангельской, Вятской, Тверской, Пермской, Тюменской областей, с территории Среднего Поволжья и Башкирии, в том числе 36 текстов с аудиосопровождением и 3 текста с видеосопровождением.

В диалектном подкорпусе пользователь может работать не только с фрагментами текстов (как в других подкорпусах НКРЯ), но и с целыми текстами.

Приведем пример текста с сайта НКРЯ с указанием «паспорта» записи.

(1) Ставропольский край, Левокумский район, пос. Новокумский. Информант: Н. Т. Пушечкина, 1945 года рождения, род. в пос. Коджакель

(Турция), в 1962 г. в 18 лет приехала в Россию, образование: 3 класса в турецкой школе и курсы по обучению письменной русской речи в России; конфессия: старообрядцы, казаки-некрасовцы. Записала в 2008 г. В. М. Грязнова, текст предоставлен В. М. Грязновой, д. ф. н., проф. кафедры русского языка Гуманитарного института Северо-Кавказского федерального университета; не публиковался. Отрывок из текста «Сказка про Иссея»:

...[Сколько вам было лет, когда вы приехали сюда?] *Васимнацать была. Ехали с радостью, мячтали в Расию. Пайдём, будим платя нгдивать, на кубулках будим хадить, на канфетнай фабрики будим работать. Веритья, мы када жили ф Турсии, мама нам рассказывала пра Иссея скаску, как Иссе́й, вон жа морим плыл, искал эта самая и приплыл вон к остраву, а на этам острави, мама нам рассказывала, людаёв жыл, однаулазый. И там тока вон авец нас, ну эта тут. Эта ана нам в децтви енти скаски рассказывала. Мы приижжаим в Расию, и паказывають на тилявизару этат фильм. Мине мёньшая сястра, мёньша ана мине, звóнить: «Аксинь, Аксинь, ты видиш ента, ты смотриш этат фильм? Мамину скаску». Я у́вварю: «Да». А мы так, вот ана нам рассказывала эту скаску...*

3. Обработка материалов для корпуса диалектных текстов. Грамматическая и метатекстовая разметка осуществляется в среде «Рабочее место диалектолога» (автор программы Т. А. Архангельский). Для разметки текст требуется предварительно подготовить: отделить речь информанта от речи собирателя, разделить словоформы для их корректного распознавания грамматическим анализатором, перевести поданную держателями текстов транскрипцию в единый шрифт юникода и т. д. При открытии подготовленного текста в «Рабочем месте» автоматически срабатывают детранскрипторы: сначала транскрипция, в которой подан текст, переводится в унифицированный вид, а потом — в орфографический «подстрочник», необходимый для работы стандартного грамматического анализатора. «Подстрочник» вручную доводится до уровня орфографической записи. После такой обработки осуществляется грамматическая разметка текста, в ходе которой устраняется грамматическая омонимия (как в Основном корпусе), а также отмечаются диалектные грамматические особенности.

Для каждой изменяемой части речи предусмотрена возможность указывать диалектные особенности на многих уровнях. Для этого в «Рабочее место» внедрены грамматические таблицы по каждой из пяти изменяемых частей речи (глагол, существительное, прилагательное, местоимение, числительное). Если необходимо указать особые грамматические характеристики, то в специальной таблице «Диалектные особенности» нужно выбрать соответствующую грамматическую категорию и пометить ее как «диалектную».

Например, диалектные особенности глагола могут быть связаны с категорией вида, переходностью, возвратностью, наклонением, типом спряжения. Глаголы в русских народных говорах могут иметь специфические окончания в личных формах 2 и 3 л. ед. и мн. ч. (*идём* вместо *идём*, он *растёт*, они *растут*, он *идё*, они *смóтрют*), особые основы (*пахáл* — *пахáют*, *сял* — *сядут*, *прóбал* — *прóбают*), иные формы инфинитива (*игрáти*, *подмéсть*, *прийтíть*, *печí*, *пекчí*, *пектí* = печь), императива (*бежí*, *отгóнь*, *подвéзь*, *курь*), причастий (*у пса убéжено* = пес убежал, *вскóпáтый* = вскопанный, *взýденый* = взятый), деепричастий (*устáвши*, *устáмиши*, *устáтши*, *устáлиши*, *насмотрéвши*), употребляться в конструкциях, связанных с иным временем (перфект, плюсквамперфект). Во всплывающей таблице «диалектные особенности» для глагола могут быть помечены диалектные основа, суффикс, флексия, форма, вид, переходность, возвратность, время.

В литературном языке начальной формой всех адъективов является им. п. ед. ч. муж. рода с безударными флексиями *-ий/-ый* (*новый*, *синий*, *всякий*) и *-ой* в случаях ударной флексии (*больной*, *большой*, *такой*). Это уже сложившаяся традиция, хоть она и противоречит фонемному принципу, определяемому по сильной позиции. В таком случае диалектной флексией для адъективов должны считаться безударные окончания *-ой/-ей*, встречающиеся в северных говорах с полным оканьем (*стáрой*, *сíней*), или ударные окончания *-ый* и *-ей*, характерные для западной зоны (*золотэй*, *золотый*, *какéй*). Диалектный суффикс или диалектная основа указываются при наличии нестандартной сравнительной степени прилагательных / наречий (*скоря́й*, *скоря́е*, *толицéй*, *молодéй*, *тяжéльше*, *ра́ньшей*) и т. д.

Помимо грамматической разметки, для каждого диалектного текста осуществляется метаразметка, которая содержит три уровня.

1) «Паспортные» данные текста: кем, когда, где, от кого записан текст, публиковался ли, где хранится и проч.

2) Фонетический уровень: отмечаются некоторые особенности фонетики говора в области вокализма и консонантизма. При этом нужно учитывать, что многие тексты, переданные для выставления на сайт Национального корпуса, представлены не в виде транскрипции, а в орфографической записи. Указание на то, орфографизирована ли запись, содержится в «паспорте» каждого текста.

3) Текстовая метаразметка содержит три подуровня: жанр (тип), тематика текста, место и время описываемых событий. Жанр делится на четыре категории: устные и письменные нефольклорные тексты, устные и письменные фольклорные тексты. Предпочтение отдается устным нефольклорным текстам, хотя и в них могут содержаться элементы фольклора: колыбельные песни, частушки, пословицы, поговорки, загадки. В банке диалектных текстов имеется значительное количество письменных текстов,

как фольклорных (сборники песен, заговоры, записанные самими носителями), так и нефольклорных (письма, мемуары, дневники и проч.).

4. Поиск в корпусе диалектных текстов. Таким образом, поиск в диалектном подкорпусе осуществляется по разным параметрам. В пределах *метаразметки* тексты можно искать по региону записи, времени записи, дате рождения информанта, месту и времени описываемых событий, по жанру и тематике, фонетическим особенностям. Для этого на странице лексико-грамматического поиска [Диалектный подкорпус НКРЯ] нужно нажать на кнопку «Задать подкорпус» в правом верхнем углу (см. рис. 1).

Рис. 1. Создание подкорпуса | Fig. 1. Creation of a subcorpus

Появится окно выбора подкорпуса. Кнопка **«Регион записи выбрать»** используется для создания подкорпуса по «географическому фильтру», если необходимо получить тексты только из какой-то определенной области. Можно выбрать несколько регионов записи и сформировать подкорпус текстов, записанных на территории распространения определенных говоров — например, только севернорусских или только южнорусских и т. д. (см. рис. 2).

Рис. 2. Поиск по региону записи | Fig. 2. Search by region of recording

Приведем пример поиска **по тематике текста**, которая очень подробно дифференцирована: «Быт. Жизнь», «Война», «Досуг. Развлечения. Игры», «Духовная культура. Религия», «Народный этикет», «Здоровье и медицина.

Народная медицина», «Материальная культура», «Природа», «Транспортные средства» и др. (рис. 3).

The screenshot shows a search form titled "Жанр и тематика" (Genre and topic). It contains several input fields with dropdown menus. The first field is for "Жанр текста" (Text genre) with radio buttons for "любой" (any), "устный нефольклорный" (non-folklore oral), and "устный фольклорный" (folklore oral). Below it is a field for "Жанр: устные нефольклорные тексты" (Genre: non-folklore oral texts) with a "выбрать" (select) button. The next field is for "Жанр: устные фольклорные тексты" (Genre: folklore oral texts) with a "выбрать" button. The field "Тематика текста" (Text topic) is highlighted with a red box and has a "выбрать" button. The final field is for "Место и время описываемых событий" (Place and time of events) with a "выбрать" button.

Рис. 3. Поиск по тематике текста | Fig. 3. Search by text topics

Задав подкорпус по тематике текста «Война», получим 32 документа из Курской, Волгоградской, Тверской и Тамбовской областей. Прямо здесь, на странице подкорпуса, можно посмотреть полный текст, нажав на значок $\leftarrow \dots \rightarrow$ (рис. 4).

The screenshot shows the search results page titled "Результаты поиска в диалектном корпусе" (Search results in the dialectal corpus). It displays statistics: "Объем всего корпуса: 627 документов, 15 873 предложения, 285 281 слово." (Total corpus volume: 627 documents, 15,873 sentences, 285,281 words). Below, it says "Найдено 32 документа общим объемом 2 261 предложение, 18 248 слов." (Found 32 documents with a total volume of 2,261 sentences, 18,248 words). A link "Сравнить подкорпус и перейти к страницам поиска" (Compare subcorpus and go to search pages) is visible. Under "Страницы: 1" (Pages: 1), there are three search results. The first result is highlighted: "1. Бутылки с горючей смесью (Викторова, Суджановский район, Курская область, 1990) [показать текст]". The "показать текст" button is highlighted with a red box. The other two results are partially visible: "2. Военная служба у казаков (хутор Никровка, Фроловский район, Волгоградская область, 1984) [показать текст]" and "3. Военная служба у казаков. Приготовление пряника (хутор Краснодонецкий, Илевский район, Волгоградская область, 1987) [показать текст]".

Рис. 4. Просмотр полного текста | Fig. 4. Viewing a full text

Тексты предоставляются в том виде, в котором они были поданы в корпус — в орфографической записи (2) или в транскрипции (3):

(2) *Кашу их ели. У них бутылки были с горючей смесью, противотанковые. Части тут стояли, орудия у них противотанковые и бутылки. Карандаш вставляешь и прямо бросаешь. Камень не камень, кирпич не кирпич — всё плавится, что по танкам бросать. Я хожу-хожу: «Дядька, ну дай бутылочку». «Да, — говорит, — командир ругаться будеть». «Да я, — говорю, — тихонько». Ящички такие, как из-под водки, отдельно стояли, каждая бутылка из-под водки. Да бутылки разные были: из-под вина, из-под молока — жидкость наливали. А то кирпичину несу, камень; он мне даёт — раз! Удовольствие! Кирпич плавится, горит белым — радость невозможная. (Бутылки с горючей смесью; Викторова, Суджановский район, Курская область, 1990);*

(3) ...ну и патом черясь нескълька время / оой / немцы едутъ / немцы едутъ / ну праўда в нас аны ничаво ня дельги / тока рас / как-та / ўот та нетиль [тёлку] забрали / патом (...) ў анной жэньшчыны свинья была пѣрасиўша / двацѣть пять пѣрасятык / свинью забрали / а пѣрасятки так асталися и згибли / мѣлака не былъ // ну / и пѣрасятѣк вбирали яны / пѣрасят / патом (...) нетиль атабрали жэньшыну / три лошыди / фура така ўот длинна / да сих пор / ўот така длиннѣя была / три (...) три лошыди запрежын было // а ну (...) матка приходитъ и гварить / маш дѣктим / Толя / бяжы / дяди Люхи скажы / маш их дѣктим и гонѣти / грязных аны свиней брать ня будуть! (Немцы в деревне; Галибицы, Торопецкий район, Тверская область, 2011). Нажав на название текста, получаем информацию о том, что запись сделана А. И. Рыко и М. А. Мельниковой от Зинаиды Ивановны Барановой 1932 г. р. и что текст сопровождается аудиофайлом.

Подкорпус по тематике «Народный календарь и календарная обрядность» на сегодняшний день включает 10 документов из разных районов Тамбовской области: «Вербоклѣст, бей до слѣз» (с. Ярославка, Никифоровский район, Тамбовская область, 2002), «На Рождество дети ходили, служили» (с. Хоботово, Первомайский район, Тамбовская область, 2002), «Пасха. Пели песни пасхальные» (Беломестная Криуша, Тамбовский район, Тамбовская область, 2003), «Рождество — гадания и обряды» (с. Кужное, Мордовский район, Тамбовская область, 2002) и др., например:

(4) [Под Вербное вербочку ходили ломать?] Да как... Нет, не ходили. Я не ходила никогда. Нет, нет, я не ходила. [И не хранишь её у иконки?] Пойдѣшь, бывало, в церкву, купишь тама вербу, принесѣшь домой, на икону положишь. [И она весь год хранится?] Да. [Правда ли, что первый раз скотину выгоняют вербочкой?] Да, да, нужно. Было дело. Я сама выгоняла, была молодая, корову имела. Выгоняла с вербушкой. Корову я вела в стадо. [А зачем вербушкой выгоняют?] Ну, дочк, эт я не знаю, так положено уж. [Не приговаривали ли у вас: «Вербохлѣст, бей до слѣз»?] Это, бывало, матеря, кто спитъ, они и начинаютъ: Вставай, вербохлѣст, бей до слѣз, Вставай ране, бяжи дале [Это тех ребят, которые спят?] Ну, это кто бѣёт, а кто нет. Было дела такая, было. (Выгоняла с вербушкой; с. Озерки, Никифоровский район, Тамбовская область, 2002).

С помощью **фонетических фильтров** можно искать тексты с определенным типом безударного вокализма (аканье или оканье и типы аканья или оканья), с особенностями произношения **ĕ* (старого «ятя») под ударением (*песня* или *писня*), смычным или фрикативным характером звонкой заднеязычной фонемы /г/ и некоторыми другими диалектными

фонетическими особенностями, например «цоканьем». В русском литературном языке две аффрикаты: твердая [ц] ([ц]апля) и мягкая [ч'] ([ч']ай). В диалектах аффрикаты могут реализовываться разными звуками, например на месте мягкой аффрикаты [ч'] может произноситься твердый звук [ч]: [ч]ай, звук с утратой затвора [ш']: [ш']ай, мягкий [ц']: [ц']ай или палатальные звуки [ц''], [ч'']. Помимо систем с двумя аффрикатами в говорах существуют системы с неразличением аффрикат и совпадением их в одной, так называемое цоканье. Самый древний вид такого неразличения — мягкое цоканье, то есть совпадение *ц* и *ч* в мягком [ц']: [ц']ай и [ц']апля. Найдем тексты с совпадением аффрикат. В строке «Наблюдается ли в говоре совпадение аффрикат» нажимаем «выбрать» и в появившемся окне выбираем «да» (рис. 5).

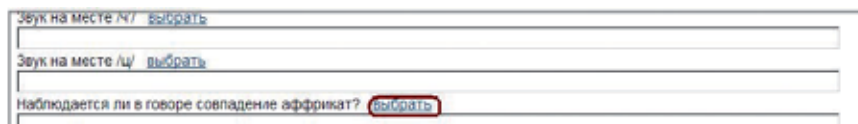


Рис. 5. Выбор подкорпуса по одному из фонетических параметров

Fig. 5. Selection of a subcorpus based on a phonetic parameter

В результате получаем 10 текстов, в которых при разметке было отмечено неразличение аффрикат. Примеры совпадения аффрикат в звуке [ц'] из текста «Как вылечили бельмо» (д. Лукинцы, Слободской район, Кировская область, 1959): *в больницу, она испекотыця* (испечется), *ф пець* (в печь), *ис пецц, цистой* и др.

Можно задать подкорпус, в который будут входить только тексты, поданные в транскрипции или только в орфографической записи (в орфографизированном виде).

На странице *лексико-грамматического поиска* есть возможность, как и в Основном корпусе, искать **по точной форме** (по форме слова или фразе) или **слову** (во всех его формах). Задав запрос на слово *толока* 'взаимопомощь, коллективная работа', получим следующие контексты, записанные в Калужской области:

(5) *Ну вот было пятница подойдет, собирает народу прясть. На толоку пойдём ну. Помогать друг другу.* (Рассказ Родиной о потере слуха; Калужская область, 1988);

(6) *И песни играли / по толокам ходили.* (Рассказ Родиной о потере слуха; Калужская область, 1988);

(7) *Ну а кто побогаче, молотилку нанимали. Собирали толоку / много народу. Толоку кормили хорошо, резали скотину али кислу делали, квас,*

стюдень. Солодуху варили, пироги пекли с квасолями. (Рассказ Солодовой о семье и быте; Калужская область, 1973);

(8) *На толоку люди шли охотно: и друг на друга посмотреть, и поест хорошо.* (Рассказ Солодовой о семье и быте; Калужская область, 1973).

Можно задать **грамматические параметры**: для имен существительных — род, склонение, одушевленность, падеж, число, для глаголов — вид, переходность, спряжение, наклонение, залог, для прилагательных — степень, полную или краткую форму и др. Кроме того, возможен поиск по диалектным морфологическим особенностям (рис. 6).

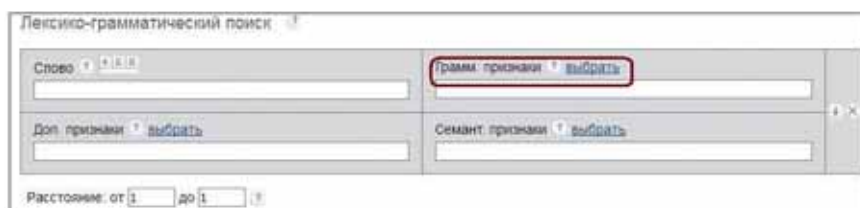


Рис. 6. Поиск по грамматическим признакам | Fig. 6. Search by grammatical properties

Известно, что в говорах преимущественно западных областей отмечается окончание *-и (-ы)* в дат. и предл. падежах (*к воды, в воды*). Как пример поиска определенной морфологической формы сформируем запрос «дат. п. ед. ч. существительных 1 скл.», при этом ограничим поиск только диалектными формами, задав параметр «диалектная флексия» (рис. 7).

Часть речи <input checked="" type="checkbox"/> существительное <input type="checkbox"/> прилагательное <input type="checkbox"/> числительное <input type="checkbox"/> числ-прил <input type="checkbox"/> глагол <input type="checkbox"/> наречие <input type="checkbox"/> предикатив <input type="checkbox"/> вводное слово <input type="checkbox"/> место-сущ <input type="checkbox"/> место-прил <input type="checkbox"/> место-предикатив <input type="checkbox"/> местоименное наречие <input type="checkbox"/> предлог <input type="checkbox"/> союз <input type="checkbox"/> частица <input type="checkbox"/> междометие	Падеж <input type="checkbox"/> именительный <input type="checkbox"/> родительный <input type="checkbox"/> родительный 2 <input checked="" type="checkbox"/> дательный <input type="checkbox"/> винительный <input type="checkbox"/> винительный 2 <input type="checkbox"/> творительный <input type="checkbox"/> предложный <input type="checkbox"/> предложный 2 <input type="checkbox"/> обиходная форма <input type="checkbox"/> диалектная особенность	Наклонение / Форма <input type="checkbox"/> изъявительное <input type="checkbox"/> повелительное <input type="checkbox"/> повелительное 2 <input type="checkbox"/> инфинитив <input type="checkbox"/> причастие <input type="checkbox"/> деепричастие	Степень / Краткость <input type="checkbox"/> сравнительная <input type="checkbox"/> сравнительная 2 <input type="checkbox"/> диалектная сравнительная <input type="checkbox"/> превосходная <input type="checkbox"/> полная форма <input type="checkbox"/> диалектная полная <input type="checkbox"/> краткая форма <input type="checkbox"/> диалектная отглагольная	Диалектная морфология <input checked="" type="checkbox"/> диалектная флексия <input type="checkbox"/> диалектная основа <input type="checkbox"/> диалектный суффикс <input type="checkbox"/> диалектная форма
Антропонимы <input type="checkbox"/> фамилия <input type="checkbox"/> имя	Число <input checked="" type="checkbox"/> единственное <input type="checkbox"/> множественное <input type="checkbox"/> диалектная двойственность	Время <input type="checkbox"/> настоящее <input type="checkbox"/> будущее <input type="checkbox"/> прошедшее <input type="checkbox"/> диалектная особенность	Переходность <input type="checkbox"/> переходный <input type="checkbox"/> непереходный <input type="checkbox"/> диалектная особенность	Прочее <input type="checkbox"/> цифровая запись <input type="checkbox"/> аномальная форма <input type="checkbox"/> мультислоговая форма
	Род <input type="checkbox"/> мужской <input type="checkbox"/> женской	Лицо <input type="checkbox"/> первое <input type="checkbox"/> второе <input type="checkbox"/> третье		Типы склонения <input checked="" type="checkbox"/> 1 склонение <input type="checkbox"/> 2 склонение
	Залог <input type="checkbox"/> действительный <input type="checkbox"/> страдательный			

Рис. 7. Поиск диалектной морфологической формы | Fig. 7. Search of a dialect morphological form

Получаем следующие примеры диалектных окончаний дат. падежа ед. ч. 1 скл. из Торопецкого района Тверской области:

(9) *хаді́ли / пѣ зямлі́ // ѣт то́жа ві́дила / ѣд бо́льшы тако́ва я ни ви́дала* (Домовой. Русалки; Уварово, Торопецкий район, Тверская область, 2011);

(10) *я ўо́т такі́х ви́дала / о́т такі́х ба́льшы́х / о́й! / и такі́м зигза́гэм / на сьпі́ны / чо́рный нзприме́р са́м / а зигза́к э́тым [...] жо́лтый* (Змеи; Княжово, Торопецкий район, Тверская область, 2011);

(11) *а во́т я хаді́ла к то́й ба́бы / ана́ мне́ вѣц сказа́ла то́ и то́* (Ведьма; Уварово, Торопецкий район, Тверская область, 2011) и др.

и из Красночикийского района Забайкальского края:

(12) *И ён ушó шо к нявэ́сты пришо́л к сваёй* (Попытка примирения; с. Урлук, Красночикийский район, Забайкальский край, 2002–2004) и др.

Корпус предоставляет также возможности для изучения диалектных синтаксических особенностей, в частности особенностей глагольного управления. Известно, что в южнорусских говорах валентность адресата при глаголах речи реализуется не только формой дат. п.: *говорить отцу*, но и предложно-падежным сочетанием *на + вин. п.*: *говорить на отца*. Найдем примеры такого управления в корпусе (рис. 8). Для этого зададим на странице лексико-грамматического поиска слово «говорить», слово «на» (на расстоянии от 0 до 10), *сущ./мест.-сущ.* в форме *вин. падежа* (на расстоянии от 0 до 1).

Получим некоторое количество контекстов, из которых вручную легко выбирается интересующая нас конструкция:

(13) *Пришла в штаб, там уже пожилой был начальник наш, наш. «Ну, — говорит на солдата, — иди, пусть девчат кормит, что ж они голодные?» Ну, пришел, полез к ней на потолок, поскидал мясо, тогда немцы убивали скот, уводили.* (Военные годы; Ширково, Львовский район, Курская область, 2000);

(14) *Да. Говорим на тракториста: Сломай его...* (На Троицу вили, а на Семик развивали; село Кужное, Мордовский район, Тамбовская область, 2002).

Еще один запрос, построенный так, чтобы глагол оказался в постпозиции: «на» + *сущ./мест. в вин. п.* + «говорить», — дает нам следующие контексты:

(15) *А на маму говорить: «Ты, щас будешь нам готовить».* (Детство, замужество, семейная жизнь; Большое Солдатское, Львовский район, Курская область, 2000);

Лексико-грамматический поиск

Слово Грамм. признаки

Доп. признаки Семант. признаки

Расстояние: от до

Слово Грамм. признаки

Доп. признаки Семант. признаки

Расстояние: от до

Слово Грамм. признаки

Доп. признаки Семант. признаки

Рис. 8. Комбинированный поиск по слову и грамматическим признакам

Fig. 8. Combined search by word and grammatical properties

(16) Ты чё эт на минё насупóрилси? Хлю́станка то́лько на жéнщин уаваря́ть — ну, а́на урязню́чая... (Погода. Характеристика человека; станица Усть-Хопёрская, Серафимовичский район, Волгоградская область, 2001).

Таким образом, в корпусе можно искать тексты различной тематики, подбирать материалы для изучения региональной и диалектной лексики, фонетики и грамматики.

5. Пополнение диалектного подкорпуса НКРЯ. Одной из основных задач подкорпуса является сбор и накопление Банка диалектных текстов. В настоящее время многие диалектологи заняты составлением диалектных словарей. Эти словари опираются на значительные по объему картотеки, часто содержащие не только иллюстрации диалектного слова, но и развернутые контексты. Во многих университетах собраны большие фонотеки, расшифрованные лишь в малой части. Диалектологическая и краеведческая работа часто осуществляется в школах, где тоже накоплены интереснейшие данные. Многие люди, связанные с деревней, далекие от филологии, но искренне преданные своей малой родине с ее неповторимыми особенностями, в том числе языковыми, записывали речь своих старших родственников на магнитофоны и диктофоны. В социальных

сетях собираются группы земляков, вспоминая деревенский быт, детские игры, диалектные слова. Силами местных жителей-энтузиастов создаются «наивные» словари, т. е. словари, составленные непрофессионалами, например замечательный «Устьянский народный словарь», содержащий не только словарные статьи от А до Я, но и «живую речь Устьи»: истории, пословицы, устойчивые выражения [Истомин и др. 2013].

Все эти материалы могут быть использованы для пополнения диалектного подкорпуса, который даст возможность лингвистам, филологам, историкам, культурологам, этнографам и всем, кто любит русское слово, работать с самыми разными образцами народной речи. Примеры из текстов и сами тексты могут выступать в качестве справочного материала, материала для научной и педагогической работы, демонстрации этнографических, этнокультурных традиций, особенностей русского менталитета. Мы надеемся, что со временем этот корпус станет репрезентативным собранием диалектных текстов и будет одним из самых любимых и востребованных пользователями.

Источники

Даниэль М. А., Добрушина Н. Р., фон Вальденфельс Р. Говор бассейна Устьи. Корпус севернорусской диалектной речи. 2013–2018 [Электронный ресурс]. URL: www.parasolcorpus.org/Pushkino (дата обращения: 30.03.2019).

Диалектный подкорпус — Корпус диалектных текстов в составе Национального корпуса русского языка [Электронный ресурс]. URL: www.ruscorpora.ru/search-dialect.html (дата обращения: 30.03.2019).

Казанский университет — Электронная библиотека русских народных говоров [Электронный ресурс]. URL: <http://dialekt.rx5.ru/index.html> (дата обращения: 30.03.2019).

Крылов С. А., Тер-Аванесова А. В. Электронные базы данных по русским народным говорам. 2011, 2012 [Электронный ресурс]. URL: <http://starling.rinet.ru/cgi-bin/main.cgi?root=ruscorpora&encoding=utf-rus> (дата обращения: 30.03.2019).

Кубань — Региональная этнолингвистика [Электронный ресурс]. URL: <http://www.ethnolex.ru/> (дата обращения: 30.03.2019).

Национальный корпус русского языка [Электронный ресурс]. URL: www.ruscorpora.ru (дата обращения: 30.03.2019).

Русская диалектология — Информационный центр «Русская диалектология» [Электронный ресурс]. URL: http://www.ruslang.ru/rus_dialectology (дата обращения: 30.03.2019).

Шуя — «Межкафедральный кабинет лингвокраеведения» Шуйского филиала Ивановского государственного университета [Электронный

ресурс]. URL: <http://sspu.ru/pages/subfaculties/rya/lingvo/index.html> (дата обращения: 30.03.2019).

Garder et al., 2018 — M. O. Garder, N. S. Petrova, A. B. Moroz, A. B. Panova, N. R. Dobrushina. Corpus of Spiridonova Buda dialect. 2018. Moscow: Linguistic Convergence Laboratory, HSE [Электронный ресурс]. URL: <http://linghub.ru/spiridonovabuda/> (дата обращения: 30.03.2019).

Ter-Avanesova et al., 2019 — A. V. Ter-Avanesova, F. A. Balabin, S. V. Dyachenko, A. V. Malysheva, V. A. Morozova. Corpus of the Malinino dialect. 2019. Moscow: Linguistic Convergence Laboratory, NRU HSE; Vinogradov Russian Language Institute of the Russian Academy of Sciences [Электронный ресурс]. URL: <https://linghub.ru/malinino/> (дата обращения: 30.03.2019).

Литература

Архангельский областной словарь [Электронный ресурс]. URL:

<http://www.philol.msu.ru/~dialectology/dictionary/> (дата обращения: 30.03.2019).

ДАРЯ — В. Соловьев, П. Салимов, А. Пилюгин, И. Исаев, В. Байрашева и др. Проект КФУ и ИРЯ РАН «Создание электронной базы данных русских говоров» [Электронный ресурс]. URL: <http://dialecty-maps.kpfu.ru/dialecty/> (дата обращения: 30.03.2019).

Истомин А. А. и др. (ред.). Устьянский словарь. Муниципальное бюджетное учреждение культуры «Устьянский краеведческий музей». П. Октябрьский [Архангельская область], 2013. Изд. 2-е, исправленное. 496 с.

ОЛА — Общеславянский лингвистический атлас [Электронный ресурс]. URL:

www.ruscorgo.ru <http://www.slavatlas.org> (дата обращения: 30.03.2019).

Псковский областной словарь с историческими данными [Электронный ресурс]. URL: <http://pos.spbu.ru/> (дата обращения: 30.03.2019).

Словарь русских народных говоров [Электронный ресурс]. URL:

<http://iling.spb.ru/vocabula/srng/srng.html> (дата обращения: 30.03.2019).

Язык русской деревни — Школьный диалектологический атлас: Язык русской деревни [Электронный ресурс]. URL: <http://gramota.ru/book/village> (дата обращения: 30.03.2019).

References

Arkhangel'skii oblastnoi slovar' [Archangelsk Region Dialect Dictionary]. Available at: <http://www.philol.msu.ru/~dialectology/dictionary/> (accessed 30.03.2019).

ДАРЯ — V. Solov'ev, P. Salimov, A. Pilyugin, I. Isaev, V. Bairasheva. *Proekt KFU i IRYa RAN "Sozdanie elektronnoi bazy dannykh russkikh govorov"* [The project "Development of electronic

database of Russian dialects”]. Available at: <http://dialecty-maps.kpfu.ru/dialecty/> (accessed 30.03.2019).

Kazanskii universitet. Elektronnaya biblioteka russkikh narodnykh govorov [Electronic library of Russian dialects]. Available at: <http://dialekt.rx5.ru/index.html> (accessed 30.03.2019).

Istomin A.A. et al. (ed.). *Ust'yanskii narodnyi slovar'* [Ustyansky national dictionary]. Municipal budgetary institution of culture “Ustyansky Museum of local lore”. The Village Of Oktyabrsky Arkhangelsk region, 2013. Ed 2-e, corrected. 496 p.

OLA – *Obshcheshlavyanskii lingvisticheskii atlas* [The Slavic linguistic Atlas].

Available at: www.ruscorpora.ru <http://www.slavatlas.org> (accessed 30.03.2019).

POS – *Pskovskii oblastnoi slovar' s istoricheskimi dannymi*. [Pskov regional dictionary with historical data]. Available at: <http://pos.spbu.ru/> (accessed 30.03.2019).

Russkaya dialektologiya – Informatsionnyi tsentr “Russkaya dialektologiya” [Information center Russian dialectology]. Available at: http://www.ruslang.ru/rus_dialectology (accessed 30.03.2019).

Slovar' russkikh narodnykh govorov [Dictionary of Russian dialects]. Available at: <http://iling.spb.ru/vocabula/srng/srng.html> (accessed 30.03.2019).

Yazyk russkoi derevni. Shkol'nyi dialektologicheskii atlas [School dialectological Atlas “Language of Russian village”]. Available at: <http://gramota.ru/book/village> (accessed 30.03.2019).