

Кюршунова Ирина Алексеевна*
Гусева Елена Рафхатовна**

Петрозаводский государственный университет
Институт филологии

*Доцент кафедры русского языка как иностранного

Доктор филологических наук

Ведущий научный сотрудник сектора языкоznания Института языка, литературы и истории
Карельский Научный Центр РАН

**старший преподаватель кафедры русского языка

Кандидат филологических наук

185910, г. Петрозаводск, пр. Ленина, 33.

Особенности создания диалектной текстовой базы данных (по материалам записей Карельского Поморья)¹

Аннотация. Статья посвящена актуальной проблеме диалектологии — целенаправленной фиксации диалектных текстов, хранящихся в Лаборатории лингвистического краеведения и языковой экологии Гуманитарного парка ПетрГУ, содержащих не только языковые данные, но и информацию, полезную исследователям различных гуманитарных направлений. В качестве образцов выбраны рукописные тексты и аудиозаписи, относящиеся к Карельскому Поморью, региону, который реже других русских территорий Карелии (Заонежья, Прионежья, Пудожья) становился объектом полевых исследований.

Особое внимание удалено особенностям представления диалектного материала в медиатеке «Диалектная лексика и ономастика Карелии». Разработаны правила учёта материалов, их описания, предложен алгоритм обработки диалектных сведений: затрагиваются фонетический (особенности транскрибирования текстов), лексический, лексикографический, грамматический аспекты. Дан образец оформления материалов, ускоряющий, по мнению авторов, процесс введения диалектных текстов разной представленности в научный оборот.

Ключевые слова: диалектный корпус; диалектный текст; Карельское Поморье; медиатека; русские говоры Карелии.

Активное использование компьютерных технологий в научном процессе даёт возможность обратиться к разработке различных прикладных вопросов любой лингвистической области исследования. Не является исключением диалектологическое направление: созданы словарные картотеки, издаются различные типы региональных словарей, использующих материалы картотек, проводятся монографические исследования и т.д. В этой работе вряд ли когда-то будет поставлена точка. Несомненно, она будет иметь больший эффект, если

¹ Статья подготовлена при финансовой поддержке РФФИ. Проект № 18-012-00810

LINGuae ANALITIO / ЛИНГВОКРАЕВЕДЕНИЕ

к стратификации материалов и особенно их сохранению будут применяться электронные способы обработки информации, которые позволят ввести в исследовательский оборот огромную, практически неисчислимую, фактологическую базу, сделают её достоянием научного сообщества. Таким образом, появляется возможность оперативно проводить сопоставление материалов, и возможно, создавать новые приёмы и методики изучения диалектного контента с последующим включением их в широкие изыскательские программы.

Именно поэтому необходимо поддержать концепцию по созданию корпуса диалектных текстов, которая особенно активно стала развиваться, начиная с 10-х годов XXI в., хотя сама идея была обозначена ещё в 80-90-е гг. XX века А. С. Гердом [2], В. Е. Гольдиным [3], [4]. Успешным следует признать опыт разных научных коллективов, занимающихся изучением говоров. Так, саратовские диалектологи апробировали электронный корпус диалектной коммуникации и показали, как организован поисковый механизм внутри корпуса [1, 24–27]. Томские исследователи совместили электронный вариант текстового архива говоров Среднего Приобья с компьютерным словарём этих же говоров [19]. Подобная работа проводится в лаборатории региональной и компьютерной лексикографии Кубани и в других диалектологических и лингвокраеведческих центрах России (см. далёко не полный список электронных диалектных корпусов [20], [21], [22], [23] после списка литературы).

Пожалуй, самым значимым этапом такой работы следует признать создание в 2005 году корпуса диалектных текстов на сайте Национального корпуса русского языка [20]. В нём учитывается комплекс различных текстовых продуктов на электронных и бумажных носителях, представляющих разные территории. В исследованиях В. Е. Гольдина [3], И. Б. Качинской [7], [8], [9], [10], [11], [12], О. Ю. Крючковой [13], [14], А. Б. Летучего [15], [16] и др. определены цели, задачи, концепция, принципы создания электронной базы диалектных текстов, а также описан механизм отбора материала, формы представления текстов, особенности их расшифровки, параметры метаразметки и др. стороны этой прикладной работы. Однако даже учитывая всплеск интереса к проблеме, следует согласиться с разработчиками Диалектного корпуса Национального корпуса русского языка, что в корпусе не так много текстов [11]. Более того, остаётся много нерешённых прикладных и теоретических вопросов важной во всех отношениях работы.

Настоящая статья — это своего рода отклик на призыв подключиться к общему для всех диалектологов делу, такому же важному, как всероссийская работа над Лексическим атласом русских народных говоров.

В Лаборатории лингвистического краеведения и языковой экологии Гуманитарного парка ПетрГУ хранятся диалектные нарративы, территориально связанные с Прионежьем, Заонежьем, Пудожьем и Беломорьем. Несмотря на то что диалектологическая работа активно велась преподавателями и студентами Карельского государственного педагогического института (университета, академии) и Петрозаводского государственного университета в течение 50 лет, до сих пор нет хрестоматии с записями речи русских жителей Карелии. Текстовый аспект оказался вообще не представленным при достаточно тщательном лексикографическом описании русской диалектной системы, серии статей, посвящённых различным лексическим явлениям русских говоров Карелии. Существенным аргументом обращения к заявленной теме является тот факт, что Карелия — уникальный регион, периферийная часть русского мира, в языковой системе которой, по словам Н. А. Мещерского, в большей степени, чем на территориях, в прошлом приближенных к Москве, а ранее к Новгороду и Пскову, сохранились архаические черты [17, 113]. Эти черты присущи не только лексике, но и грамматике, явления последней предпочтительнее рассматривать и наблюдать именно в тексте.



В настоящий момент в рамках проекта «Карельское Поморье: лексика и ономастика (XVI—XXI вв.)» ведётся работа по созданию медиатеки, составной частью которой является база данных диалектных записей. Первые экспедиционные записи сделаны в 1968 году, последнее в 2018 году в русских деревнях Беломорского, Кемского, Лоухского районов Карелии, определяемых традиционно как Карельское Поморье¹.

Отметим, что сейчас составляется каталог диалектных записей по Карельскому Поморью, проводится оцифровка рукописных и аудиотекстов, их расшифровка и комментирование. Каталог, как и все материалы медиатеки, имеет открытый вход, чтобы в него при необходимости оперативно включались новые материалы. Сами материалы в разных формах подачи будут представлены в виде серии публикаций текстов с комментариями к ним на страницах научного электронного журнала «CARELICA», а также в информационной системе (медиатеке) «Диалектная лексика и ономастика Карелии» (<https://lexis.petsru.ru/>) [6].

Опишем этапы работы и особенности представленного в них материала.

I этап. Составление каталога диалектных текстов

Каталог включает описания текстов. Это условный рабочий паспорт каждой диалектной записи, включающий следующие атрибуты: различные временные и географические параметры, данные о диалектоносителе, тематика бесед, сведения о способах представления записи, их хранении, информация о тех, кто вёл сбор и обработку материалов и примечания. Каталог представлен в системе Microsoft Excel и выглядит следующим образом:

Фрагмент каталога диалектных записей по Карельскому Поморью.

Год записи	Нас. пункт	Район	Диалектоноситель	Г.р.	Место рожд.	Темы беседы	№ тетради	Стр.	Записал	Руководитель	Сведения об		Примечания	
											оцифровке	расшифровке		
А	В	С	Д	Е	Ф	Г	Н	І	Ј	К	Л	М	Н	О
1969	Гридино	Кем.	Мехнина Ирина Ярионовна	1914	Гридино	Меведь-разбойник	15	23-24	Кукшиева Нина		Рукоп.	Оцифрована Кюршуновой И. А.	Расшифрована Баженовой А. И.	Упрощ. транскр., необход. доп. проверка
1983 ? 1984	Нюхча	Белом.	Воронина Анна Евдокимовна (в записях Воронова)	1918, 65 лет	Нюхча	Бытов. темы	18	25-29	Кравцова Н.		Рукоп.			Упрощ. транскр., дигл. явл. выдены непоследовательно, дата рождения и точная фамилия установлены по переписке с Семёновой Н. С.
1986	Шиженя	Белом.	Каманина Александра Николаевна	1905		Бытов. темы	5	2-4	?	Михайлова Л. П.	Рукоп.	Оцифровано — Бермус Е., Дьяконова	Расшифровка — Бермус Е., Дьяконова	Имя, отчество, место записи текста установлены по тетради Масло А. А. (тетрадь № 4, 1986).
1996	Сумпосад	Белом.	Пашкова Евдокия Михайловна	1926		Части дома, устройство русской печки, бани, дом. утварь	3	31-33	Ткалич Н. А.	Мухина Е. А.	Рукоп.			Упрощ. транскр., опред. тематики бесед выполн. Гусевой Е. Р., Павловой А. В.
2014	Нюхча	Белом.	Кокотова Мария Фёдоровна (а. ф. Докучаева)	1933	Нюхча	Бытов. темы, рождение, о семье, топонимы		21 стр.	Лазаренко И., Пинигина Кс.	Кюршунова И. А.	Word + Аудио-	Лазаренко И., Кюршунова И. А.	Лазаренко И., Кюршунова И. А.	Упрощ. транскр., дигл. явления выделены непоследовательно

Как видим, в каталоге есть лакуны. Они маркированы: красным шрифтом помечена информация, которая или была изменена в процессе обработки материалов, или на неё в дальнейшем следует обратить внимание. Ячейки, выделенные цветовым маркером, — показатель

¹ Временную границу по записи диалектной речи можно отодвинуть до 1939 года, если привлечь публикации образцов русской речи, представленных в известном всем диалектологам пособии Н. П. Гринковой и В. И. Чагишевой «Практические занятия по диалектологии» (1957).

отсутствия сведений. Таким образом, естественно обозначается задача предстоящих полевых изысканий в исследуемом регионе — обратить внимание на слабые звенья, по возможности добыть недостающую информацию.

Количество полей каталога может быть увеличено, а их наполнение изменено. Например, можно добавить ономастическую рубрику, отметить наличие фольклорных материалов, а также разных фактов об образовании, профессии, месте работы опрашиваемых и т.д.

II этап. Работа с диалектными записями

Следующий этап предвидит решение ряда вопросов, касающихся насущных теоретических и прикладных аспектов диалектологии. Представим их кратко, поскольку им может быть посвящено отдельное исследование.

- ⌘ Представляется проблемным. Сомнений не вызывают тексты, к которым приложена аудиозапись, поскольку есть возможность сопоставить речь и её «перевод» на бумажный и электронный носители. Однако большая часть текстов является рукописной. Они выполнены студентами до появления технических средств, и даже, казалось бы, при тщательной проверке содержат ошибки. *Имеет ли право такой текст на место в базе данных?*

Думается, что следует учитывать все материалы, даже выполненные неопытной рукой и записанные плохим почерком, без выделения характерных особенностей говора, с различными информационными пробелами. Даже в ущербных записях иногда содержатся исключительно важные сведения.

Вопросы о том, что считать текстом и как соотносится этот термин с принятой в лингвистике теорией текста, остаётся открытым в диалектологии (и кажется, в самой теории текста). Если считать основными свойствами текста связность и целостность, то отметим: идеальных текстов немного (даже когда расшифровке подлежат аудиозаписи). Так, в процессе беседы, иногда длящейся 2-3 часа и более, диалектоноситель «поставляет» значительный по объёму материал, который чаще всего членится на отдельные темы. Возникает вопрос о соотношении темы и текста. По ходу беседы возникают дополнительные вопросы (так называемые пресуппозиции), на них информант должен отреагировать комментарием. Кроме того, собеседника могут отвлечь и появляются немотивированные вопросами собирателя вкрапления, или реплики. *Можно ли считать эти минимальные структуры (реплики, комментарии, отступления от темы) текстом?*

Вероятно, с технической точки зрения в базе данных следует разграничить текст и запись. Одна запись может содержать разные типы текстов, зафиксированные от одного информанта. В них могут быть представлены разные темы. В этом случае решается проблема, когда начинающему диалектологу не удается записать связный текст на определенную тему, а только отдельные фрагменты беседы, речения.

Перейдём к описанию процедуры электронной обработки текстов.

1) Оцифровка материалов

- a) перевод в цифровой вид и запись на электронные носители аудиозаписей;
 - b) перевод в цифровой вид и запись на электронные носители рукописных текстов в формате pdf. Текст сканируется в оригиналe (со всеми замечаниями и исправлениями, пометами и выделениями, представленными в рукописном образце).

2) *Расшифровка материалов*, представленных в рукописном виде и в аудиозаписи, переводится в печатный вид (формат Word). При этом расшифровка рукописных материалов производится с сохранением постраничной разбивки.

Затем рукописные тексты *переводятся в орфографическую запись*, поскольку оригинал записи оцифрован. Аудиозаписи сначала транскрибируются. Подробная транскрипция используется для записей, в которых чётко прослеживаются фонетические диалектные черты: оканье, цоканье, шоканье, произношение *–ого*, *–его* в окончаниях прилагательных и проч. К упрощенной транскрипции обращаемся, когда фонетический строй речи совпадает в большей степени с литературным произношением. Если диалектоноситель отмечает характерные черты речи и аутентично представляет это во время беседы, то при расшифровке следует дать пояснение о переключении на подробную транскрипцию и обратно. Приведём пример, демонстрирующий подобное переключение речевых регистров:

очень цокали [далее подробная транскрипция] м'йша / а м'йша / зб'ёгај-ка / на дál'н'иј колоð'еу / пр'ин'ес'й-ка / холоðној вод'иц'к'и / и'ајн'иц'ек согр'ёт' / и'ајоц'ку нап'иц'а / гор'ац'его и'ају хоц'еу а-л'и / н'ёт-л'и [далее упрощенная транскрипция] (Демьянчук Александра Константиновна, 1938 г.р., Нюхча, Белом., 2014).

Как видим, помимо мягкого цоканья речи нюхчан было свойственно и полное оканье, но информант не замечает его, поскольку является носителем этого признака.

Перевод текстов в орфографический формат имеет определённую цель. К корпусу будут обращаться не только лингвисты (диалектологи, фонетисты, историки языка и проч.), но и более широкий круг возможных пользователей: историки, социологи, культурологи, краеведы, учителя, учащиеся, жители Беломорья, для которых тонкости диалектологической работы с текстом не важны, а важна содержательная сторона материалов, имеющих историко-культурное, нравственно-воспитательное значение и возможность быть представленными в другой научной парадигме знания. Более того, одна из задач этой работы — вернуть в обработанном виде материалы людям, чьи родственники на протяжении полувека являлись нашими информантами, даря не только языковой материал и знания о своей нелёгкой жизни, но и душевное тепло общения.

Именно текст в орфографическом оформлении становится полем обработки материала. В нём маркируются диалектные слова, а также лексемы, поиск которых осуществляется при работе по вопроснику Лексического атласа русских народных говоров [18], выделяется ономастическая лексика, грамматические особенности, выявленные в речи диалектоносителей.

III этап. Эти текстовые процедуры открывают следующий этап — этап обработки материалов.

А. Описание диалектных записей.

Ко всему представленному тексту, записанному во время беседы от одного информанта, составляется *справка*, включающая по возможности самые точные данные о диалектоносителе, сведения о месте и времени записи текста, далее следует характеристика записи: номер тетради, в которой представлен текст, с указанием страниц; кем выполнена запись; руководитель полевых работ и обработки материала; особенности почерка; выделения / невыделения в нём диалектных лексических единиц, грамматических форм); при представленной транскрипции текста следует отметить последовательность / непоследовательность отражения в записи фонетических явлений; необходимо также отметить наличие иллюстраций, схем, карт, фото-, видео-, аудиоматериалов); дать сведения об особенностях обра-



ботки текста (оцифровке, орфографической расшифровке); кратко представляется тематика бесед (см. пример 1, 2, 3).

Пример 1.

Каманина Александра Николаевна, 1885 г.р., род. в Сумпосаде, Белом. (год рождения установлен по тексту, в 1968 г. А. Н. Каманиной было 83 года).

Записано в Сумпосаде в 1968 году, тетрадь № 28 (98 стр.); собиратели — Симакина, Войтенкова; используется упрощённая транскрипция, лексические и грамматические диалектизмы не выделены; запись отражает полное оканье, цоканье мягкое (представлено непоследовательно); почерки разные, читаемые.

Тетрадь оцифрована в 2018 г.: файл pdf: 1968_Сумпосад_Белом.pdf

Темы: семейная жизнь, описание Сумпосада, занятия населения, рыбная ловля, церковные праздники, сбор ягод, история Сумпосада, история Марфы Борецкой, погодные явления (названия ветров), торговля с Мурманском, Норвегией, Архангельском, особенности общения с лопарями, купцы, устройство дома, свадьба, князь (царь) Алексей Михайлович, староверы, истории из жизни, о Пётре Первом, фольклорные тексты (свадебные притчания, песни — «Утушна песня», «Полынецка», сказка «Сестрица Алёнушка и братец Иванушка, предания), сватовство.

Пример 2

Дементьева Надежда Петровна, 1905 г.р., Шижня, Белом.

Записано в Шижне в 1986 году, тетрадь № 4 (с. 1—14); собиратель — Масло Л. Л.; используется упрощённая транскрипция, лексические и грамматические диалектизмы не выделены; запись отражает полное оканье, цоканье не отражено; почерк читаемый, но смешиваются «ч» и «г».

Вероятно, велась параллельная запись руководителем — Л. П. Михайловой (тетрадь № 5; 1986), поэтому текст представлен связно.

Тетрадь оцифрована, файл pdf: 1986_Шижня_Белом_2.pdf

Темы: огород, юбилей, семейная жизнь, приготовление пищи, рыболовные и торговые суда, рыболовство, кухонная утварь, приготовление масла, приговорки.

Пример 3

Пайкачева Нина Григорьевна, 1920 г.р., Колежма, Белом. (данные о месте рождения, девичьей фамилии, образовании, месте работы отсутствуют; год рождения установлен по записям 1998 г., выполненными М. О. Грызиной, О. Аникиной, тетрадь без номера).

Записано в Колежме Белом. р-на в 1998 году, тетрадь без номера (с. 2—24); собиратели — М. О. Грызина, О. Аникина; рук. И. А. Кюршунова. Используется упрощённая транскрипция, выделены лексические и грамматические диалектизмы; возможно, неточно представлено членение речи на потоки; исходя из записи речи диалектносителя характерно полное оканье; цоканье не фиксируется; почерк читаемый, но смешиваются «ш»~«м», «л»~«и»; имеются иллюстрации поморского сарафана, сорочки, повойник и их деталей, схема крепления жемчуга и бисера на повойнике.

Возможно, расшифровка текста велась по аудиозаписи, поскольку текст представляет не отрывочные фразы и отдельные речения.

Компьютерный набор с сохранением упрощённой транскрипции осуществлён в 2002 г. Стасевич Светланой Васильевной, рук. И. А. Кюршунова; файл: 1998_Колежма_Белом.doc.

Орфографическая запись и выделение диалектных особенностей текста выполнено А. И. Баженовой в 2018–2019 гг.

Тетрадь оцифрована, файл pdf: 1998_Колежма_Белом.pdf.

Темы: особенности отделки дома, части дома, сенокос, уборка сена на хранение, русская печь, обувь, рукавицы, одежда, женская одежда (сарафан, его детали, сорочка, поварник, платки), одежда рыбаков.

Эти записи свидетельствуют, как менялось отношение к оформлению диалектных материалов в полевых условиях, как со временем осознавалась важность этого вида работы. Конечно, особую роль сыграли появившиеся вопросы и программы сбора материалов (особенно программа ЛАРНГ [18]), методические рекомендации по работе с диалектоносителями, но в проведении полевых изысканий имеются временные лакуны, связанные с наступившим перестроичным и особенно постперестроичным этапом в стране и последовавшим недофинансированием, не позволявшим полноценно проводить плановые поездки в русские населённые пункты Белого моря.

Б. Работа с лексическим материалом.

Определяется значение слов по контексту, сопоставляются материалы с данными, представленными в диалектных словарях. Безусловно, большая часть лексического материала зафиксирована в известных лексикографических источниках. Наблюдения показали, что наполнение текстов диалектными лексемами меняется: они постепенно уходят из органично вплетённого в речь привычного лексикона и становятся фоном, свидетельствуя о нивелировании говоров.

В. Работа с грамматическим материалом.

Подобным образом исчезают яркие грамматические особенности речи. Грамматическая разметка текстов, в том виде как её предлагают составители корпусов диалектной речи, нами пока не проводится по разным объективным причинам, но по ходу обработки материалов обязательно обращается внимание на такие явления, как:

на полосы, у ларьку, до мосту, огурцей, ремонту наделал, угощу мятыма ягодамы, поле в концы деревни, на няши, церковь было на горы, у купцей было, промышляли навагамы, палтасым, проздравленье с блинамы дай сковороды; своима, егова, еёный, одна си, он ю бил; востра, круты, продажна; малтат, окучиват, роботат, берегчи, пекёт, берегёт; сходить бани можно затопить

Представим образцы обработки материала и его подачи в текстовой базе данных (см. образец 1, 2, 3).

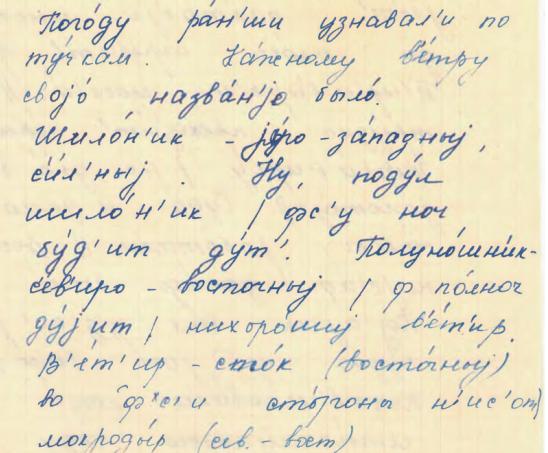
Первая колонка содержит сканы отрывков записей бесед, во второй колонке дана орфографизированная запись, в которой выделены лексические и грамматические явления, указана тема беседы, а также список литературы, представляющий исследованность темы (можно указать только фамилии исследователей, занимавшихся определенной темой); третья колонка — лексикографическая обработка материалов: сначала указывается слово



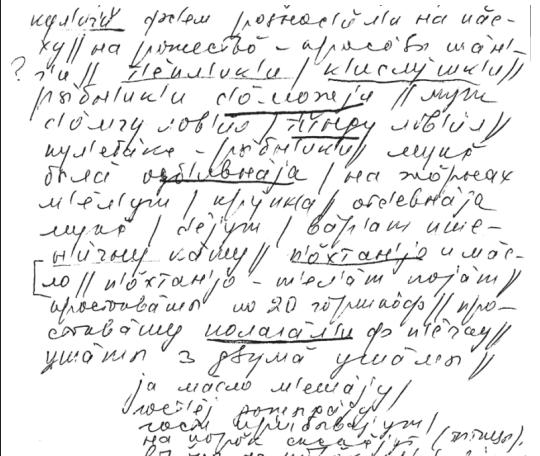
и значение, определённое по контексту; затем приводятся сопоставительные материалы, отмеченные в диалектных словарях. При необходимости даётся комментарий о фиксации слова, ареала его бытования в исследуемом по тексту говоре.

Материалы первой колонки актуальны для диалектологов, текстологов, вторая колонка — орфографическая запись, пред назначенная для широкого круга исследователей, в третьей колонке представлен обрабатываемый лексический материал. Эта часть имеет открытый вход, её можно дополнять новыми сведениями, разного рода комментариями.

Образец 1

Отрывок записи беседы с Каманиной Александрой Николаевной, 1885 г.р., Сумпосад, Белом. Запись 1968 г. С. 13	Орфографизированная запись Тема Литература по теме	Лексикографическая обработка
	<p>Погоду раньше узнавали по тучкам. Каждому ветру своё название было. Шилоник — юго-западный, сильный. Ну, подул шилоник, всю ночь будет дуть. Полуночник — северо-восточный, в полночь дует, нехороший ветер. Ветер сток (восточный) во все стороны несёт, мокродыр (северо-восточный).</p>	<p>Шелонник — юго-западный ветер. Слово представлено во всех словарях (см. СРГК, 6, с. 856).</p>
	<p>Тема: погода, названия ветров</p> <p>Арзуманова Н. Г. Названия ветров в русских говорах Карелии и сопредельных областей // Лексический атлас русских народных говоров. Материалы и исследования 1995. — СПб., 1998. — С. 102—111.</p> <p>Березович Е.А. Об одной группе севернорусских названий ветров // Язык и прошлое народа. — Екатеринбург, 1993. — С. 150—157.</p> <p>Демидова В. Е. Наименование ветров в северных говорах // Лексический атлас русских народных говоров. Материалы и исследования 1998. — СПб., 2001. — С. 74—78.</p> <p>Карамышева Л.М. Словосочетания — названия ветров в русских говорах Карелии // Севернорусские говоры. Вып. 9. — СПб., 2008. — С. 40—45.</p> <p>Новикова Е.И. Лингвогеографическое описание ветров в русских говорах Карелии // Лексический атлас русских народных говоров. Материалы и исследования 1998. СПб., 2001. С. 63—74.</p>	<p>Полуночник — северо-восточный ветер.</p> <ul style="list-style-type: none"> ♦ СРГК: значение — северный ветер в Канд., Тер. говорах (5, с. 65); ♦ Дуров: значение — северо-восточный, Повс. (с. 322); ♦ СРНГ: значение — северо-восточный, Помор. Арх., Волог., Медв., Белом., Кем. Арх., Мурм. и др. (29, с. 155—156). <p>Сток — восточный ветер.</p> <ul style="list-style-type: none"> ♦ СРГК: значение совпадает, но нет указания на фиксацию слова в Белом. говорах (указаны Прион., Сег., Канд., Кем., Кондоп., Онеж., Пуд., Тер.); у Дурова помета Повс. (с. 391); ♦ СРНГ: также нет указания на Белом. говоры, отмечена фиксация в Олон., КАССР без точной географии (41, с. 192). <p>Мокродыр — северо-восточный ветер.</p> <ul style="list-style-type: none"> ♦ СРНГ: значение и говор совпадают (5, с. 246); ♦ Дуров: отсутствует; ♦ СРНГ: отсутствует.



Отрывок записи беседы с Дементьевой Надеждой Петровной, 1905 г.р., Шижня, Белом. Запись Л. П. Михайловой, 1986 г.	Орфографизированная запись Тема Литература по теме	Лексикографическая обработка
	<p>Куличи всем разносили на Пасху. На Рождество — просовы шаньги. Теплики, кислушки. Рыбники сёможки. Муж сёмгу ловил, тинду ловил. Кулебяка, рыбники. Мука была отбивная. На жёрнах мелют, крупка. Отсевная мука, сеют. Варят пшеничную кашу. Пёхтаньё и масло. Пёхтанье — телят поят. Простокваша по 20 горшков. Простоквашу полагали в печку. Ушаты с двумя ушами. Я масло мешаю, Гостей ожидаю. Гости прибывают, На порог скакают (птицы) Вороны на порог прибывают.</p>	<p>Шаньга — ватрушка с кашей из проса. ◆ СРГК: род ватрушки или лепёшки, смазанной маслом, сметаной и т.п. География широкая, включая Белом. говоры (6, с. 832); Дуров: шанёжка. Повс. (с. 442), шаньга не фиксируется.</p>
		<p>Теплик — выпечка — ? ◆ СРГК: пирог круглой формы с какой-н. начинкой или без начинки. Кириш., Тихв., Кем., Выт., Новг. — <u>Белом. говоры не отмечены</u> (6, с. 451); ◆ Дуров: теплик — пышка из ржаного теста, лепёшка в виде маленькой булочки, испечённая на противне или сковороде. Сор. (с. 404).</p>
		<p>Тема: пища</p>
	<p>Публикаций много. Широко представлены разные аспекты этой темы в говорах. В наиболее полном виде собрана библиография в диссертации Пьяковой К. С. Тема по говорам Карелии не описана.</p>	<p>Кислушка — ? возможно, что-то испечённое из кислого теста или имеющее кислую начинку, ◆ СРГК: кисленник — пирог с начинкой из щавеля; кислое — что-то испечённое из кислого теста; киломолочник — ватрушка с творогом; кислопряжа — булочка из дрожжевого теста; а также кислопряженец, кислопряженник, кислопряжка, кислопряжник, кисляк (2, с. 350—353); ◆ Дуров: кисла шаньга (с. 168)</p>
		<p>Тинда — рыба ◆ СРГК — небольшая рыба, наподобие семги. Канд., Кем. (6, с. 460). <u>Белом. говоры не представлены</u>.</p>
		<p>◆ Дуров — летняя мелкая семга. Повс. (с. 405)</p>
		<p>Пёхтаньё — отходы от взбивания масла;</p>
		<p>◆ СРГК: то же, Онеж., Кем., Тер., Белом. (+), Канд., Карг., Плес., Пуд. (4, с. 496);</p>
		<p>Дуров: пёхтаньё — сыворотка, остающаяся после взбивания сметаны или сливок на масле (с. 297).</p>
		<p>Полагать — класть</p>
		<p>◆ СРГК — класть, помещать куда-н. Пуд., Тихв., Лоух., Прион.; <u>Белом. говоры не фиксируются</u>.</p>
		<p>Дожидать — ждать</p>

Образец 3

Отрывок записи беседы с Пайкачевой Ниной Григорьевной, 1920 г. р.; Колежма, Белом. Запись 1998 г. С. 14.	Орфографизированная запись Тема Литература по теме	Лексикографическая обработка
<p>Уна́е ^укóну вéт' не ^украсище // Кóнчи бýкоши / как тéсна // Тéсну и нé носи- ши // <u>Опти́сшина</u> иши / зна́хчи бýрчи таш / кóнмо норкошиш / шéтши / тóт дý- рет носи́т / а тóдёх нову соши́т //</p> <p>к кóлбéе кóрчиш / от сапóк отрéшут //</p> <p><u>Опти́сши</u> звáиши // Голáкши то отрé- шут // кóрчиш / как ја кóнчи / таш <u>у</u>ши и от сапóк отрéшут этиш // Тáкше и тараб / Уна́е нé вши волши саптéх</p>	<p>У нас кожу ведь не красили. Ноги выросли, да тесна. Тесну и не носили. Обтеснилась мне, значит, бери там, кто подрастает, меньшее, тот будет носить, а тебе нову сошьют. Опорками звали. Голяшки-то отрежут. Ходить, как я хожу, там у меня от сапог отрезано эти. Так же и тогда. У нас не было вообще лаптей.</p>	<p>Обтесниться — стать тесным. <u>Не зафиксировано в словарях.</u> СРНГ: обтиснуться — сжаться от мороза (о капусте). Пск., Твер. (22, с. 244).</p>
	<p>Тема: Обувь</p>	<p>Опорка — обувь с обрезанным голенищем. СРГК: широкая география, Белом. (+) (4, с. 219); Дуров: информация отсутствует. СРГН: отмечена широкая география, <u>Белом. говоры не представлены</u> (23, с. 280).</p>

Предварительный анализ показывает, что в исследовании говоров Карельского Поморья, в изучении языковых материалов ещё рано ставить точку. Беломорская лексика в меньшей степени, нежели та, которая бытует в других русских говорах Карелии, отображается в лексикографических источниках. Так, в каждом небольшом отрывке, представленном выше, присутствуют материалы, не имеющие пометы Белом. (в таблице они подчеркнуты), или отмечено слово, не зафиксированное в диалектных словарях (*кислушки, обтесниться*).

Описанный выше опыт презентации диалектных текстов Карельского Поморья является первым шагом по созданию текстовой диалектной базы данных Карелии, мобильным способом представления материалов специалистам разных гуманитарных направлений, а главное — возможностью присоединиться к общему для всех диалектологов России делу.

Литература

1. *Батраева И. А. Поисковый механизм Саратовского диалектного корпуса / И. А. Батраева, В. Е. Гольдин, О. Ю. Крючкова // Компьютерные науки и информационные технологии: матер. междунар. науч. конф. — Саратов: СГУ, 2009. — С. 24—27.*
 2. *Герд А. С. Типы русских текстов и организация Машинного фонда русского языка [/ А. С. Герд // Машинный фонд русского языка : идеи и суждения. — Москва : Наука, 1986. — С. 67—75.*
 3. *Гольдин В. Е. К проекту текстового диалектологического подфонда Машинного фонда русского языка / В. Е. Гольдин // Доклады Третьей Всероссийской конференции по созданию Машинного фонда русского языка. — Москва, 1990. — С. 92—103.*
 4. *Гольдин В. Е. Машиннообрабатываемые корпусы диалектных текстов и проблемы типологии русской речи / В. Е. Гольдин // Русистика сегодня. — Москва 1995. — № 3. — С. 72—84.*
 5. *Гринкова Н. П. Практические занятия по диалектологии : Пособие для пед. ин-тов / Н. П. Гринкова, В. И. Чагишева. — Ленинград : Учпедгиз. Ленинград. отд-ние, 1957. — 206 с.*
 6. *Гусева Е. Р. Информационная система «Диалектная лексика и ономастика Карелии» как динамический медиаресурс по краеведению / Е. Р. Гусева, А. Г. Марахтанов // Учен. зап. Петрозав. гос. ун-та. Сер. Обществ. и гуманитар. науки. — Петрозаводск, 2019. — № 3. — С. 89—93.*
 7. *Качинская И. Б. Диалектный Подкорпус НКРЯ. Новый стандарт подачи. Новое рабочее место / И. Б. Качинская, Е. В. Моисеева // Русская устная речь : материалы междунар. науч. конф. «Баранниковские чтения. Устная речь: русская диалектная и разговорно-просторечная культура*



общения». Межвузовское совещание «Проблемы создания и использования диалектных корпусов», Саратов, 15—17 ноября 2010 г. / ред. О. Ю. Крючкова, А. И. Буранова, В. Е. Гольдин, Л. В. Балашова. — Саратов, 2011. — С. 245—255.

8. Качинская И. Б. Диалектный подкорпус сегодня / И. Б. Качинская, Д. В. Сичинава // Тр. Института русского языка им. В. В. Виноградова. — Москва, 2015. — Вып. 6. — С. 142—162.
9. Качинская И. Б. Корпус диалектных текстов в Национальном корпусе русского языка: состояние и перспективы / И. Б. Качинская // Лексический атлас русских народных говоров : Материалы и исследования. 2009. — Санкт-Петербург : Наука : 2009. — С. 57—68.
10. Качинская И. Б. Корпус диалектных текстов в национальном корпусе русского языка: сегодняшнее состояние и перспективы / И. Б. Качинская, Д. В. Сичинава // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Междунар. конф. «Диалог», Бекасово, 4—8 июня 2014 г. — Москва, 2014. — Вып. 13 (20). — С. 593—600.
11. Качинская И. Б. О корпусе диалектных текстов в национальном корпусе русского языка [Текст] / И. Б. Качинская, Д. В. Сичинава // Вопросы лексикографии. — 2017. — № 11. — С.71—85
12. Качинская И. Б. О корпусе диалектных текстов в национальном корпусе русского языка / И. А. Качинская, А. В. Малышева // Актуальные проблемы русской диалектологии. Материалы международной конференции 26—29 октября 2018 г. — Москва : Институт русского языка им. В. В. Виноградова РАН, 2018. — С. 118—120.
13. Крючкова О. Ю. Корпус русской диалектной речи: концепция и параметры оценки / О. Ю. Крючкова, В. Е. Гольдин // Компьютерная лингвистика и интеллектуальные технологии («Диалог-2011»): тр. междунар. конф. — Москва : РГГУ, 2011. — Вып. 10 (17). — С. 359—367.
14. Крючкова О. Ю. Электронный корпус русской диалектной речи и принципы его разметки / О. Ю. Крючкова // Изв. Сарат. ун-та. Нов. сер. 2007. Т. 7. Сер. Филология. Журналистика, вып. 1. — Саратов: Изд-во Сарат. ун-та. — С. 30 — 34.
15. Летучий А. Б. Диалектный корпус: состав и особенности разметки / А. Б. Летучий // Национальный корпус русского языка: 2006—2008. Новые результаты и перспективы / отв. ред. В. А. Плунгян. — Санкт-Петербург : Об-во с ограниченной ответственностью «Нестор-История», 2009. — С. 114—129.
16. Летучий А. Б. Корпус диалектных текстов: задачи и проблемы // Национальный корпус русского языка: 2003—2005. Результаты и перспективы. — Москва: Индрик, 2005. — С. 215—232.
17. Мещерский Н. А. К изучению русских народных говоров на территории Карельской АССР / Н. А. Мещерский // Учен. зап. Карел. пед. ин-та. — Петрозаводск, 1963. — Т. XIII: Гуманитар. науки. — С. 112—130.
18. Программа собирания сведений для лексического атласа русских народных говоров: [В 2 ч.] / Рос. АН, Ин-т лингв. исслед. — Санкт-Петербург : ИЛИ : Науч.-изд. центр «НеГа», 1994.
19. Юрина Е. А. Томский диалектный корпус: в начале пути / Е. А. Юрина // Вестник Томского государственного университета. Филология. — 2011. — № 2. — С. 58—63.

Электронные ресурсы:

20. Диалектный корпус Национального корпуса русского языка — URL: <http://www.ruscorpora.ru/search-dialect.html>
21. Информационный центр «Русская диалектология». — URL: http://www.ruslang.ru/agens.php?id=rus_dialectology, (Институт Русского языка им. В. В. Виноградова РАН (20.02.2019).
22. Электронная библиотека русских народных говоров. — URL: <http://dialekt.rx5.ru/index.html> (Казанский (Приволжский) федеральный университет (02.2019).
23. Электронные базы данных по русским народным говорам. — URL: <http://starling.rinet.ru/cgi-bin/main.cgi?root=ruscorpora&encoding=utf-rus> (тексты, записанные в деревнях Харовского района Вологодской обл. и Шатурского района Московской обл., С. А. Крылов и А. В. Тер-Аванесова (20.02.2019).

Словари:

Дуров — Дуров И. М. Словарь живого поморского языка в его бытовом и этнографическом применении [Текст] / И. М. Дуров ; Карельский науч. центр Российской акад. наук, Ин-т яз., лит. и истории. — Петрозаводск : Карельский науч. центр РАН : Ин-т яз., лит. и истории КарНЦ РАН, 2011. — 453 с.

СГРС — Словарь говоров Русского Севера / Ур. гос. ун-т им. А.М. Горького ; [авт.-сост. : Ю. В. Алабу-
гина и др.] ; под ред. А. К. Матвеева. Т. 3: Г-Ж / [авт.-сост.: Е. Л. Березович и др.]. — Екатеринбург
: Изд-во Ур. ун-та, 2005. — 387 с.

СРГК — Словарь русских говоров Карелии и сопредельных областей / С.-Петербург. гос. ун-т ;
Гл. ред. А. С. Герд. — Санкт-Петербург : Изд-во С.-Петербург. ун-та, 1994—2005.

СРНГ — Словарь русских народных говоров / Академия наук СССР, Институт русского языка, Словарный сектор. — Москва : [б. и.] ; Ленинград : Наука, Ленинградское издание, 1965 — .

Вып. 22: Обивень-Одалбливать / Сост. Л. И. Балахонова и др.; гл. ред. Ф.П. Филин. — 1987. — 368 с.

Вып. 23: Одале-Осеть / [Составители Н. И. Андреева-Васина и др.]. — Л. : Наука : Ленингр. изд-ние, 1987. — 375 с.;

Вып. 29: Покорочеть-Попрятчиться / [Сост. Ю. Ф. Денисенко и др.]. — Санкт-Петербург : Наука : Ленингр. отд-ние, 1995. — 346 с.

Список сокращений географических районов дан по словарям:

- 〃 Дуров И. М. Словарь живого поморского языка в его бытовом и этнографическом применении (с. 17—18);
 - 〃 Словарь говоров Русского Севера (т. 1, с. 9—10);
 - 〃 Словарь русских говоров Карелии и сопредельных областей (т. 1, с. 15);
 - 〃 Словарь русских народных говоров (т. 1, с. 162—183).

Irina A. Kyurshunova*

Elena R. Guseva**

*Petrozavodsk State University
Institute of Philology

****Associate Professor of the Department of Russian as a Foreign Language
Doctor of Philology;
Leading Researcher of the Linguistics Sector at the Institute of Language, Literature and History
Karelian Scientific Centre of Russian Academy of Sciences**

** Senior Lecturer of the Department of Russian Language
PhD in Philology

Features of creating a dialect text database (based on records of Karelian Pomorye)

Annotation: the article is devoted to the actual problem of dialectology - the purposeful fixation of dialect texts stored in the Laboratory of Linguistic Local History and Language Ecology, which contain not only linguistic data, but also information useful for researchers of various humanitarian fields. Handwritten texts and audio recordings related to the Karelian Pomerania, a region that is less common in other Russian territories of Karelia (Zaonezhie, Prionezhie, Pudozhya) became the subject of field research, were selected as samples.

Particular attention is paid to the peculiarities of the presentation of dialect material in the media library «Dialect vocabulary and onomastics of Karelia». The rules for accounting of materials, their descriptions were developed, an algorithm for processing dialect data was proposed: the phonetic (transcribing features), lexical, lexicographical, grammatical aspects are affected. A sample of the design of the materials is given, which, in the authors' opinion, speeds up the process of introducing dialect texts of different representation in scientific circulation.

Key words: dialect case; dialect text; Karelian Pomerania Pomorye; media library; Russian dialects of Karelia.