

СЛОВАРНЫЕ ПРОЕКТЫ И ТРУДЫ

УДК 81'374(038)=811.161.1'28; 004.65

DOI: 10.17223/22274200/18/5

С.С. Земичева

ОТ АБАРМА ДО ЯЩИЧИШКА: РАЗРАБОТКА ЛЕКСИКОГРАФИЧЕСКОГО КОМПОНЕНТА ТОМСКОГО ДИАЛЕКТНОГО КОРПУСА¹

Предлагается проект интеграции опубликованного диалектного словаря с электронным корпусом текстов одного региона, не имеющий аналогов в российской научной практике. Описаны этапы работы, обоснованы принципы отбора единиц в словник и подходы к представлению материала с учётом цели лексикографического компонента корпуса и особенностей автоматической обработки слова. Приведены образцы словарных статей планируемого электронного ресурса.

Ключевые слова: русские говоры Сибири, диалектный корпус, электронный словарь, цифровая диалектология.

Введение

Одной из важнейших тенденций современной диалектологической науки является создание электронных ресурсов, на что уже неоднократно обращали внимание исследователи [1–4]. В электронную форму переводятся диалектные картотеки [5], фонотеки [6, 7], диалектологические атласы [8, 9]. Существуют электронные версии многих диалектных словарей. Часть изданий при этом представлена в свободном доступе в формате скан-версий печатных текстов: «Словарь русских народных говоров» [10], «Архангельский областной словарь» [11] и др. Коллекция оцифрованных диалектных словарей, включающая в том числе малотиражные, редкие лексикографические труды, представлена на сайте Института русского языка им. В.В. Виноградова РАН [12]. На материале наиболее известных, крупных словарей

¹ Исследование выполнено за счёт гранта Российского научного фонда «Разработка электронных ресурсов для исследования народно-речевой культуры Среднего Приобья» (проект № 19-78-10015).

создаются также электронные базы данных. Так, размещена в интернете электронная версия словаря В.И. Даля [13], ведётся работа над созданием базы данных «Псковского областного словаря» [14]. Существуют мультимедийные учебные пособия по русской диалектологии, размещённые в сети Интернет: школьный диалектологический атлас «Язык русской деревни» [15], вузовский курс «Фонетика русских диалектов» [16], массовый открытый онлайн-курс «Русские диалекты: взгляд из Сибири» [17] и др. Подобные тенденции характерны не только для российской, но и для мировой науки в целом [1, 3, 18]. Разработка новых электронных источников позволяет поставить вопрос о формировании цифровой диалектологии как особого научного направления в русле digital humanities (цифровых гуманитарных технологий). К диалектологическим ресурсам примыкают фольклорные базы данных, как общероссийские, например фундаментальная электронная библиотека «Русская литература и фольклор» [19], так и локальные – база данных псковского фольклора [20], база данных среднеобского фольклора [21] и др. Создание любых подобных ресурсов имеет важное значение, так как позволяет сохранить языковые элементы традиционной народной культуры, стремительно уходящей в прошлое.

В то же время необходимо отметить, что цифровые диалектологические источники весьма различны по своим техническим характеристикам. Думается, что их создание не должно сводиться к простой оцифровке имеющихся материалов. С этой точки зрения особое место среди электронных диалектологических ресурсов занимают базы данных и корпуса. Их число пока относительно невелико.

Так, на материале русских говоров Удмуртии создана лингвогеографическая система «Диалект», включающая возможности словного поиска и визуализации лингвистических данных на карте [22]. С 2005 г. создаётся сводный диалектный подкорпус НКРЯ (285 281 слово) [23], имеющий многоуровневую разметку (центральное положение занимает морфологическая характеристика единиц со специальным маркированием диалектных особенностей). Существуют диалектные корпуса отдельных регионов. Впервые идея такого ресурса выдвинута создателями Саратовского диалектологического корпуса [24]. В настоящее время ряд морфологически размеченных диалектных корпусов, снабжённых аудиозаписями, создан лингвистами Высшей школы экономики – корпус бассейна реки Устья с материалами из Устьянского района Архангельской области, насчитывающий более миллиона словоупотреблений

[25, 26], корпус говора села Малинино Липецкой области (166 639 словоупотреблений) [27], корпус села Роговатка Белгородской области (114 600 словоупотреблений) [28]. Кроме того, начата работа над тематически размеченными корпусами – диалектным корпусом лингвокультуры Северного Приангарья (170 813 словоупотреблений) [29], Кубанским диалектным корпусом (без сведений об объёме) [30] и др.

Краткий обзор подобных ресурсов позволяет сделать вывод, что большая часть из них находится на начальном этапе разработки, включает достаточно небольшое количество материала. При этом как российские, так и зарубежные исследователи указывают, что корпусное направление является одним из наиболее актуальных для современной диалектологии [2, 18, 31]. Добавим, что создание электронных корпусов и баз данных по русским говорам не только позволяет сохранить особый культурный пласт, но и вносит значимый вклад в развитие корпусной лингвистики в целом, так как падёт возможность восполнить лакуну, связанную с представлением устной речи в имеющихся корпусных проектах [32. С. 42].

Внедрение в корпус лексикографического модуля в целом является актуальной тенденцией. В основной части НКРЯ для каждого слова имеется ссылка на словарный агрегатор academic.ru, где можно посмотреть значение лексемы и получить некоторую другую дополнительную информацию (в данный момент эта опция доступна только в «старой версии» корпуса) [33]. Часто внедрение лексикографической составляющей сопряжено с переводом на другой язык. Так, в корпусах, создаваемых на материале различных языков народов России, нередко имеется перевод на русский и возможен поиск по русской словоформе. Таковы, в частности, корпус хакасского языка [34], корпус калмыцкого языка [35], корпус вепского языка [36], корпус удмуртского языка [37] и др.

В диалектных корпусах дополнительные словарные компоненты, как правило, отсутствуют. Одним из немногих исключений является болгарский диалектный корпус [38], где каждому слову даётся перевод на русский и английский языки. Учёные отмечали, что интеграция диалектного корпуса со словарём – задача достаточно отдалённого будущего, если рассматривать русский язык в целом [39. С. 78]. В то же время на материале одного говора или группы говоров одного региона подобная задача выполнима¹.

¹ В частности, планировалось включение лексикографического модуля в лингвогеографическую систему «Диалект» [40], однако неясно, был ли этот проект осуществлён.

Томский диалектный корпус (далее ТДК) начал создаваться в 2010 г., когда была предложена общая концепция ресурса [41] и началась оцифровка экспедиционных материалов. Сегодня ТДК насчитывает более 1 млн 700 тыс. словоупотреблений. Осуществлена разметка и доступен поиск по экстралингвистическим параметрам, теме, типу текста, отдельному слову. Демо-версия корпуса представлена в открытом доступе [42], полная – доступна для зарегистрированных пользователей.

Данный ресурс проектируется как универсальная информационно-поисковая система, включающая три модуля: 1) текстовый (доступ к целостным текстам и аудиозаписям, тематическая и жанровая разметка); 2) собственно корпусный (грамматическая разметка); 3) лексикографический. Настоящая статья посвящена описанию последнего.

Идея интеграции ТДК с диалектным словарём впервые была предложена Е.В. Иванцовой [43]. Основная задача лексикографического компонента – выдача толкования собственно диалектных лексем, что обусловило обращение к дифференциальному толковому словарю для его интеграции в корпус. Объектом интеграции является «Словарь русских старожильческих говоров средней части бассейна р. Оби» под ред. В.В. Палагиной [44] и два дополнения к нему [45, 46] (далее для обозначения совокупности данных изданий используется аббревиатура СРСГ). Всего семитомник насчитывает 23 190 словарных статей. Ранее идея создания сводного словаря на базе СРСГ и других лексикографических трудов томских диалектологов (безотносительно к корпусу) выдвигалась О.И. Блиновой [47].

Реализация проекта интеграции корпуса и словаря предполагает следующие этапы:

- 1) сканирование, автоматическое распознавание бумажного словаря, выстраивание словарных статей в общем алфавитном порядке;
- 2) унификация подачи материалов: устранение неточностей автоматического распознавания текста, объединение словарных статей из разных томов;
- 3) создание компьютерной базы данных;
- 4) внедрение словарных статей в корпус;
- 5) добавление в корпус возможности поиска по словарным пометам.

Остановимся подробнее на втором этапе работы как центральном с филологической точки зрения (остальные задачи по большей части технические). Подготовку материалов к внедрению в корпус осу-

шестволяют старший научный сотрудник Лаборатории общей и сибирской лексикографии ТГУ С.С. Земичева и сотрудники кафедры русского языка Л.А. Иванова, М.М. Угрюмова.

Основные принципы и подходы к обработке словарных статей определяются, во-первых, целью электронного ресурса, во-вторых, техническими ограничениями и требованиями. Общая цель лексикографического компонента корпуса – представить толкования диалектных лексем. По итогам реализации проекта предполагается, что пользователь сможет получить словарную статью на интересующее его слово в тексте (в виде гиперссылки или всплывающего окна). При этом внутри системы будет сначала происходить морфологическая обработка (переход от формы слова в тексте к начальной форме, которая затем ищется в электронном словаре). Таким образом, лексикографический компонент корпуса не самостоятелен, он отличается от традиционных словарей, в частности, тем, что в нём отсутствует алфавитное упорядочивание единиц (хотя оно используется в служебных целях на подготовительном этапе работы). Этот принцип обозначается как автономное функционирование отдельной словарной статьи. Технические ограничения связаны с тем, что распознавание происходит на уровне отдельного слова, а не оборота, чем обусловлено исключение из электронной версии части фразеологических единиц.

1. **Состав словника.** В СРСГ были включены «слова с диалектными корнями, семантическими, словообразовательными, акцентологическими и лексикализованными фонетическими особенностями <...> фразеология, глаголы с особенностями управления и формообразовательными особенностями в исходной форме» [44. Т. 1. С. 4–5]. В дополнение к словарю были также включены (с некоторыми оговорками) заимствования из западноевропейских языков [45. Ч. 1. С. 3].

При обработке материалов словаря для внедрения в корпус возник ряд проблем. Проект лексикографического компонента в корпусе предполагает, что толкование слова будет выдаваться во всех текстах, включающих его. В связи с этим сложности вызвала, во-первых, задача неоднословных единиц. Из словника были исключены диалектные фразеологические и нефразеологические обороты, состоящие из общерусских слов (*камушки считать* ‘осматривать хозяйство жениха перед свадьбой’, *от ветру качать* ‘быть слабым’, *в голову кидаться* ‘пьянить’ и др.). В противном случае толкование выдавалось бы на каждое употребление соответствующего слова (*камушки, считать, от, ветер,*

качать и т.п.), что, на наш взгляд, излишне. Однако отметим, что исследователи диалектной фразеологии при необходимости смогут найти в корпусе примеры нужных им единиц, уточняя соответствующую семантику по опубликованным бумажным и электронным словарям.

Фразеологический оборот, включающий диалектную лексему, даётся при ней, например:

КОСМЫГНИ, мн. Неодобр. Взлохмаченные волосы. – Космы'ни вон какие (Кем. Лен.-Куз. Шаб.). / Кем. (Лен.-Куз.). ◇ ЗА КОСМЫГНИ ЗАВЕЯТЬ. Оттаскать за волосы. – Я хотел тебя за космы'ни завеять (Карг. Ил.).

При этом в ряде случаев словарные статьи на фразеологические и нефразеологические сочетания, данные в исходной версии словаря на общерусскую единицу, были перенесены на другие буквы алфавита. Например, словарная статья КЛЕПЁШНЫЙ в исходной версии словаря имеет вид отсылочной:

КЛЕПЁШНЫЙ, см. ДОСКА КЛЕПЁШНАЯ [45. Ч. 1. С. 197].

В итоговой версии она выглядит так:

КЛЕПЁШНЫЙ, прил. Относящийся к клепцам (капкану на мелкого зверя). ДОСКА' КЛЕПЁШНАЯ. Ловушка на медведя, состоящая из доски, в которую вбиты металлические стержни с зазубринами. – Доска клепёшна. Зверь идёт травой, все четыре ноги всодит и сидит (Пар. Нест.). / Том. (Пар.).

Во-вторых, специальной обработки потребовали слова, совпадающие с общерусскими по форме, но отличающиеся значением. Служебные единицы и глаголы, которые отличаются от литературного языка только сочетаемостью, исключались. Так, были исключены следующие словарные статьи:

КЛАСТЬ в сочетании с сущ. огонь. Разжигать костёр. – Огонь клади, я карасей пойду добуду (Крив. Ник.). / Том. (Крив.) [44. Т. 2. С. 82].

КАК, союз. Чем. – Траву же косить легче, как хлеб (В.-Кет. Кет.)¹ [44. Т. 2. С. 66] и др.

Отметим, что корпус, проект которого включает и морфологическую разметку, в будущем позволит выявлять некоторые из подобных значений, если они являются грамматически обусловленными. Так, например, использование союза *как* со сравнительной семантикой можно будет найти по запросу «прилагательное в сравнительной степени» + «как».

¹ Словарная статья приведена в сокращении.

В случае других единиц, совпадающих с общерусскими по форме и отличающихся семантически, решение о включении в лексикографический компонент корпуса принималось на основе критерия частотности. Хотя частотность является непостоянным параметром, опора на данный показатель может быть обоснована следующим образом. На сегодняшний день ТДК включает приблизительно 40% материалов, имеющих в архиве, при этом на долю текстов, записанных в 1940–1990-е гг., приходится более 60% от общего количества словоупотреблений, уже внесённых в корпус. Пополнять ресурс планируется как материалами старых, так и новых экспедиций, но в целом увеличиваться будет доля новых текстов, где диалектная семантика чаще всего оказывается невостребованной.

При проверке частотности возможны два варианта результатов. В первом случае более востребованным является общерусское значение, тогда единица исключается из словника. Так, были исключены такие слова, как *аккура'тный* 'обладающий красивым телосложением', которое употребляется в диалектном значении лишь 2 раза из 12; *ка'дочка* 'ручка цепа', насчитывающее 92 словоупотребления в корпусе в значении 'небольшая кадка' и 0 – в указанном диалектном значении. Во втором случае, если востребована только диалектная семантика, единицы сохранялись для корпусного представления словаря. Например, *американка* '1) сорт картофеля; 2) разновидность сохи' (а не 'жительница Америки'), *вяз* 'гибкий прут, связывающий противоположные столбики у саней' (а не 'дерево семейства ильмовых'), *кладь* 'продолговатый скирд хлеба, сена' (а не 'груз, поклажа') и т.п.

Третий тип единиц, которые были исключены из словника СРСГ, – лексемы, отличающиеся от литературного языка только грамматическими характеристиками (*яблока*, ж.). Семантических отличий в данном случае нет, толкование не требуется. Предполагается, что такие единицы будут обрабатываться с помощью модуля грамматической разметки, а не лексикографического компонента корпуса.

По итогам обработки материалов первой части словаря (буквы А–М) с учётом вышеобозначенных принципов из 12 450 словарных статей СРСГ в итоговой версии осталось 9 324, т.е. объём материала сократился примерно на четверть. При этом основная часть исключённых словарных статей – неоднословные единицы и единицы, совпадающие по форме с общерусскими.

Вопрос о пополнении словника остаётся на данном этапе открытым. Как в архивных записях, так и в материалах новых экспедиций обнаруживаются некоторые диалектные единицы, отсутствующие в словнике СРСГ (что связано, в частности, с отбором в словарь лексики только старожильческих говоров) [24. С. 18–19]. Сбор и систематизация подобных единиц требуют специальной работы, однако принцип автономного существования словарной статьи в лексикографическом модуле ТДК предполагает, что пополнение словника возможно на любом из этапов, в том числе после завершения работы над основной частью словаря (так как не требуется выстраивать статьи в общем алфавитном порядке).

2. Структура словарной статьи. Словарная статья включает заглавное слово, его частеречную характеристику, экспрессивные и лексико-грамматические пометы, толкование, контексты, локальные пометы. Часть информации, имеющейся в СРСГ, при подготовке материалов к внедрению в корпус была удалена. Так, в исходной версии словаря каждое слово снабжалось дополнительными данными о словоизменении (в частности, для существительных указывались окончания в родительном падеже, для глаголов – окончания в формах первого лица единственного числа и второго лица единственного числа). Устранение этой информации объясняется тем, что грамматические характеристики каждой единицы будут полно представлены в морфологическом модуле корпуса. В некоторых случаях указание окончаний заменялось на маркирование части речи (вместо «ая, ое» – «прил.»).

Кроме того, в томах 1–5 СРСГ в каждой словарной статье давались сведения о наличии диалектной единицы в словарных и несловарных источниках, известных к моменту публикации (словарь В.И. Даля, «Областной словарь колымского русского наречия» В.Г. Богораза, «Опыт областного великорусского словаря» и др.). Эти сведения опущены, так как более полная информация о бытовании слова на территории России на сегодняшний день дана в сводном «Словаре русских народных говоров», в связи с чем и редакторы СРСГ отказались от этой зоны словарной статьи в заключительных томах издания.

Имеющиеся пометы, данные перед толкованием слова («многократное», «уменьшительное», «уничижительное» и др.), а также локальные пометы, указывающие перечень населённых пунктов, где встречается то или иное слово, сохранены.

Далее опишем этапы работы над электронным словарём в соответствии с зонами словарной статьи: заглавное слово, толкование, контексты.

Заглавное слово. В соответствии со сложившейся лексикографической традицией сохранена подача заглавного слова прописными буквами, с указанием ударения. Такой подход удобен и для автоматической обработки, так как позволяет отделить зону заглавного слова от других зон словарной статьи. Варианты слова, как и в исходном издании, даются в разных словарных статьях (иначе их автоматическая обработка будет затруднена).

Определённую проблему представляет собой написание диалектных единиц. Общий подход, обозначенный в предисловии к СРСГ, заключается в том, что для слов с ясной этимологией даётся единственный вариант написания. «Если этимология неясна, на слово, кроме основной статьи, даётся ссылочная: *еман* см. *яман*, *бойдон* см. *байдон*» [44. Т. 1. С. 5]. Для подобных диалектных единиц в разрабатываемой электронной версии словаря даются дублирующие словарные статьи, чтобы пользователь мог получить толкование лексемы независимо от варианта написания.

Одной из центральных задач представления в корпусе многотомного словаря, издававшегося частями в течение достаточного длительного времени, стало упорядочивание словарных статей, данных на одно и то же слово в разных томах СРСГ. Совпадающие единицы давались в новой электронной версии либо как значения одного слова, либо как омонимы (если между ними нет семантической связи). Примером объединения разных значений полисеманта может служить слово *ко'ник*. В СРСГ включены две словарные статьи:

КО'НИК, а, м. Фольк. Ласкат. к конь. – Что, мой коник, сивогривый, вороной... (из песни) (Кож. Жур.). / Том. (Кож.). Даль. [45. Т. 1. С. 207].

КО'НИК, а, м. Скамья, имеющая вид лаля и используемая для сна и хранения одежды. – Коник – кровать-то, одежду куда класть. Мне всё родитель купил (Яшк. Полom.). Кроватей не было. Коники назывались. С одной стороны коник сделают до поперечнику, занавеску некуда повесить (Яшк. Полom.). [46. Ч. 1. С. 154].

Для представления в корпусе они объединены в одну:

КО'НИК, м. 1. Фольк. Ласкат. к конь. – Что, мой коник, сивогривый, вороной... [из песни] (Кож. Жур.). / Том. (Кож.). 2. Скамья, имеющая

вид ларя и используемая для спанья и хранения одежды. – Коник – кровать-то, одёжу куда класть. Мне всё родитель купил (Яшк. Полон.). Кроватей не было. Коники назывались. С одной стороны коник сделают до поперечнику, занавеску некуда повесить (Яшк. Полон.).

При формировании словарной статьи многозначной лексемы для определения порядка значений учитывалась, во-первых, семантика (от наиболее широкого значения – к частным, конкретным; переносные значения указываются после прямых), во-вторых, частотность (если два или более значений являются конкретными, неметафорическими, первым даётся более распространённое из них. Распространённость определяется как количество использований слова в том или ином значении в ТДК).

Если между значениями совпадающих слов не обнаруживалось семантической связи, они давались как омонимы. Например, словарные статьи на слово *косица* в СРСГ имели следующий вид:

КОСИ'ЦА, ы, ж. Висок. – Это брови, а это веки, норки, лоб, шшоки, санки передни и коренные, верхны и нижны, а это косица (В.-Кет. М. Яр). Сына у нас застрелили. В косицу прямо пуля попала (Кем. Лен.-Куз. Красн.). / Том. (В.-Кет.). Кем. (Лен.-Куз.). Опыт: Арханг. (Арханг., Мез.), Иркут., Перм., Якут. Кривошапкин. Молотиллов. Маляревский. Патканов и Зобнин. [44. Т. 2. С. 98].

КОСИ'ЦА, ы, ж. Женщина-косарь. – Я работала вно'го шибко. Летом сено косила, хлеб убирала. Шесть соток под руками выжать, кото'ры плохи' косицы, а нас человек семь хорошо косило (Юрг. Н.-Р.). [46. Ч. 1. С. 161].

Словарные статьи в корпусе даны как омонимы:

КОСИ'ЦА¹, ж. Висок. – Это брови, а это веки, норки, лоб, шшоки, санки передни и коренные, верхны и нижны, а это коси'ца (В.-Кет. М. Яр). Сына у нас застрелили. В коси'цу прямо пуля попала (Кем. Лен.-Куз. Красн.). / Том. (В.-Кет.). Кем. (Лен.-Куз.). КОСИ'ЦА², ж. Женщина-косарь. – Я работала вно'го шибко. Летом сено косила, хлеб убирала. Шесть соток под руками выжать, кото'ры плохи' коси'цы, а нас человек семь хорошо косило (Юрг. Н.-Р.).

На данный момент разграничения омонимичных значений не предусмотрено в связи с тем, что надёжные методики для решения подобного рода задач не разработаны. Таким образом, как и в случае с многозначными лексемами, пользователь будет видеть оба значения и иметь возможность самостоятельно определить нужное, опираясь

на контекст. В перспективе в части контекстов семантическая омонимия, как и грамматическая, возможно, будет снята вручную.

Толкования. Принцип автономного существования словарной статьи требует замены отсылочных толкований на содержательные. Так, в СРСГ словарная статья имеет вид:

БАЛМО'ШНЫЙ, ая, ое. См. ба'лмошный. – Соображенья не имеет, так балмошный (В.-Кет. Ат.) / Том. (В.-Кет.). [44. Т. 1. С. 25].

В корпусе с учётом толкования единицы БА'ЛМОШНЫЙ в [44] эта словарная статья будет представлена следующим образом:

БАЛМО'ШНЫЙ, прил. Взбалмошный. – Соображенья не имеет, так балмо'шный (В.-Кет. Ат.) / Том. (В.-Кет.).

В тех случаях, где толкование в СРСГ было дано в виде «Уменьшит.-ласкат. к ...», «Уничижит. к ...», «прил. к ...» и подобных в толковании приводилось и значение исходного слова. Так, в СРСГ имеется словарная статья, отсылающая к толкованию диалектного слова:

КАНДЕ'ПОЧКА, и, ж. Уменьшит.-ласкат. к канде'пка. – Дырочку доспешь, кандепочку приткнёшь, она примёрзнет (Крив. Ник.) / Том. (Крив.) [44. Т. 2. С. 68].

В этом же томе слово «канде'пка» толкуется как «Колышек для укрепления сети самолера и др. снарядов на льду» [44. Т. 2. С. 68].

В электронном словаре толкование восстанавливается, итоговая словарная статья выглядит так:

КАНДЕ'ПОЧКА, ж. Уменьшит.-ласкат. к канде'пка «колышек для укрепления сети самолера и др. снарядов на льду». – Дырочку доспешь, канде'почку приткнёшь, она примёрзнет (Крив. Ник.) / Том. (Крив.).

Контексты. Вопрос о том, приводить ли контексты из иллюстративной части СРСГ в словарных статьях планируемой электронной базы данных, был решён положительно. Разработчики руководствовались при этом двумя соображениями. Во-первых, такие иллюстрации помогают уточнить значение лексической единицы, особенно в ситуации наличия полисемии или омонимии. Во-вторых, некоторые из контекстов, вошедших в словарь, были утеряны и в основную часть корпуса не войдут. Пополнение словарных статей новыми контекстами не предусматривалось, так как выдача большого числа примеров использования слова является, с нашей точки зрения, задачей текстового модуля корпуса, а лексикографический компонент рассчитан на семантизацию диалектных единиц,

Таким образом, для добавления толкований лексем в диалектный корпус потребовалась выработка особого подхода к отбору материала и способам организации словарных статей. Принципы перевода материала в электронную форму для внедрения в корпус определяются тем, что лексикографический компонент рассматривается как часть универсальной электронной системы, при этом автоматическая обработка осуществляется на уровне отдельного слова, а каждая словарная статья функционирует как автономная. К настоящему моменту обработано около половины исходных материалов. Финальную версию электронного словаря как части Томского диалектного корпуса планируется представить к июню 2021 г. Внедрение лексикографического компонента в корпус значительно упростит и ускорит работу с диалектными текстами, а также сделает их понимание доступным широкому кругу лиц. К перспективам проделанной работы относятся, во-первых, расширение словника за счёт других опубликованных словарей и текстовых материалов [48]; во-вторых, внедрение в корпус поиска по единицам, имеющим те или иные словарные пометы («уменьшительно-ласкательное», «увеличительное», «женское» и др.). Представленные решения могут использоваться при разработке других диалектных корпусов.

Литература

1. *Жданова Е.А.* Проект корпуса русских говоров Удмуртии // Интеллектуальные системы в производстве. 2016. № 4 (31). С. 137–141.
2. *Крючкова О.Ю.* Научные парадигмы в диалектологии и диалектологическая традиция в Саратовском университете // И.И. Срезневский и русское историческое языкознание: опыт и перспективы. 205-летию со дня рождения И.И. Срезневского : сб. ст. Междунар. науч.-практ. конф. Рязань, 2017. С. 299–304.
3. *Земичева С.С., Иванцова Е.В.* Проект создания Томского диалектного корпуса в свете тенденций развития корпусной лингвистики // Сибирский филологический журнал. 2018. № 3. С. 192–205.
4. *Качинская И.Б., Малышева А.В.* Народная речь в Национальном корпусе русского языка // Русская речь. 2019. № 4. С. 103–118.
5. *Качинская И.Б., Крылов С.А.* Диалектная лексикография: электронная картотека «Архангельского областного словаря» // Диалог-2010. Компьютерная лингвистика и интеллектуальные технологии. М., 2010. Вып. 9 (16). С. 169–172. URL: <http://www.dialog-21.ru/media/1652/27.pdf>
6. *Кульшарипова Р.Э., Ибрагимов Т.И.* Электронная библиотека русских народных говоров Казанского университета: возможности применения, информа-

ционный потенциал // Международный журнал экспериментального образования. 2013. № 5. С. 95–96.

7. *Коконова А.Б.* Фонотека архангельских говоров // Вестник Российского гуманитарного научного фонда. 2015. № 2 (79). С. 231–236.

8. *Кузнецова Е.В.* Информационная система «Лексический атлас Волгоградской области»: научный материал в учебном процессе вуза // Лексический атлас русских народных говоров (материалы и исследования). СПб., 2013. С. 335–344.

9. *Создание* базы данных по русским диалектам и перспективы диалектометрических исследований / И.И. Исаев [и др.] // Вестник Российской академии наук. 2016. Т. 86, № 11. С. 972–977.

10. *Словарь* русских народных говоров. URL: <http://iling.spb.ru/vocabula/srng/srng.html> (дата обращения: 23.04.2020).

11. *Архангельский* областной словарь. URL: <http://www.philol.msu.ru/~dialectology/dictionary/> (дата обращения: 29.04.2020).

12. *Ссылки* на сканированные версии некоторых русских диалектных словарей и исследований по диалектологии // Институт русского языка им. В.В. Виноградова Российской академии наук. URL: http://www.ruslang.ru/dialectolog_centers_links (дата обращения: 10.05.2020).

13. *Даль В.И.* Толковый словарь живаго великорускаго языка. URL: <http://slovardalja.net/> (дата обращения: 29.04.2020).

14. *Кукушкина И.С.* Конвертация «Псковского областного словаря» в формат электронного словаря на базе DWS LINGVO CONTENT // Русский язык и литература в поликультурном коммуникативном пространстве : материалы Междунар. науч. конф. / отв. ред. Н.В. Большакова. Псков, 2012. С. 248–253.

15. *Школьный* диалектологический атлас: Язык русской деревни. URL: <http://gramota.ru/book/village> (дата обращения: 29.04.2020).

16. *Князев С., Моисеева Е., Шаульский Е.* Фонетика русских диалектов. URL: <http://dialect.philol.msu.ru/index.php> (дата обращения: 12.03.2020).

17. *Долганина А.А., Шевчик А.В.* Русский диалекты: взгляд из Сибири. URL: https://pushkininstitute.ru/external_courses/260 (дата обращения: 09.04.2020).

18. *Szmrecsanyi B.* Methods and Objectives in Contemporary Dialectology // Contemporary approaches to dialectology: The area of North, Northwest Russian and Belarusian vernaculars / eds. Ilja A. Seržant & Björn Wiemer. Bergen, 2014. Vol. 12. P. 81–92.

19. *Фундаментальная* электронная библиотека «Русская литература и фольклор». URL: <http://feb-web.ru/> (дата обращения: 12.03.2020).

20. *База* данных псковского фольклора // Научно-образовательная лаборатория региональных филологических исследований. URL: <http://nocpskoviana.pskgu.ru/colloquial.php> (дата обращения: 20.03.2020).

21. *База* данных Среднеобского фольклора // Томский межрегиональный институт общественных наук. URL: <http://mion.tsu.ru/srobannot> (дата обращения: 02.03.2020).

22. *Лингвогеографическая* система «Диалект». URL: <http://manuscripts.ru/dl/dialekt.main> (дата обращения: 02.03.2020).

23. *Диалектный корпус* Национального корпуса русского языка. URL: <http://www.ruscorpora.ru/search-dialect.html> (дата обращения: 12.03.2020).

24. *Крючкова О.Ю., Гольдин В.Е.* Корпус русской диалектной речи: концепция и параметры оценки // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог». Бекасово, 25–29 мая 2011 г. М., 2011. Вып. 10 (17). С. 359–367.

25. *Basic stats* // Даниэль М., Добрушина Н., Вальденфельс Р. *Говор бассейна Устья*. Корпус севернорусской диалектной речи. Берн ; Москва, 2013–2018. URL: <http://parasolcorpus.org/Pushkino/stats.php> (дата обращения: 19.05.2020).

26. *Waldenfels R., Daniel M., Dobrushina N.* Why Standard Orthography? Building the Ustyia River Basin Corpus, an online corpus of a Russian dialect // Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference “Dialogue”. Moscow, 2014. Is. 13. P. 270–278.

27. *Corpus statistics* // Malinino Corpus. URL: https://linghub.ru/malinino/#!/corpus_statistics (дата обращения: 20.05.2020).

28. *Corpus of Rogovatka dialect*. URL: <http://www.parasolcorpus.org/Rogovatka/> (дата обращения: 20.05.2020).

29. *Диалектный подкорпус* // Электронный текстовый корпус лингвокультуры Северного Приангарья. URL: <http://angara.sfu-kras.ru/?page=dialect#> (дата обращения: 12.04.2020).

30. *Диалектный корпус* // Региональная этнолингвистика. URL: <https://ethnolex.ru/kubdk/> (дата обращения: 20.05.2020).

31. *Демешкина Т.А.* Векторы развития современной русской диалектологии // Актуальные проблемы обучения русскому языку : материалы Междунар. науч. конф. Брно, Чехия, 05–07 мая 2014 г. Брно, 2014. С. 268–278.

32. *Резанова З.И.* Лингвистический корпус «Томский региональный текст»: типологически релевантные параметры сбалансированности и репрезентативности // Вестник Томского государственного университета. Филология. 2015. № 1 (33). С. 38–50.

33. *Национальный корпус русского языка*. URL: <http://ruscorpora.ru/old/> (дата обращения: 20.05.2020).

34. *Электронный корпус хакасского языка*. URL: <http://khakas.altai.ru> (дата обращения: 20.05.2020).

35. *Национальный корпус калмыцкого языка*. URL: <http://kalmcorpora.ru/dial> (дата обращения: 20.05.2020).

36. *Корпус вепсского языка*. URL: <http://vepsian.krc.karelia.ru/about/> (дата обращения: 20.05.2020).

37. *Корпус удмуртского языка*. URL: http://web-corpora.net/UdmurtCorpus/search/index.php?interface_language=ru (дата обращения: 25.04.2020).

38. *Bulgarian Dialectology as Living Tradition*. URL: <http://bulgariandialectology.org/> (access date: 05.03.2020).

39. *Качинская И.Б., Сичинава Д.В.* О Корпусе диалектных текстов в Национальном корпусе русского языка // Вопросы лексикографии. 2017. № 11. С. 71–85.

40. *Жданова Е.А.* Лексикографический модуль лингвогеографической информационной системы «Диалект» // Лексический атлас русских народных говоров (материалы и исследования). СПб., 2013. С. 185–192.

41. Юрина Е.А. Томский диалектный корпус: в начале пути // Вестник Томского государственного университета. Филология. 2011. № 2 (14). С. 58–63.
42. Томский диалектный корпус // Лаборатория общей и сибирской лексикографии НИ ТГУ. URL: <http://losl.tsu.ru/corpus> (дата обращения: 01.05.2020).
43. Иванцова Е.В. Томский диалектный корпус: обоснование концепции и перспективы развития // Вопросы лексикографии. 2017. № 11. С. 54–70.
44. *Словарь русских старожильческих говоров средней части бассейна р. Оби* / ред. В.В. Палагина. Томск : Изд-во Том. ун-та, 1964–1967. Т. 1–3.
45. *Словарь русских старожильческих говоров средней части бассейна р. Оби (Дополнение)* / ред. О.И. Блинова, В.В. Палагина. Томск : Изд-во Том. ун-та, 1975. Ч. 1–2.
46. *Среднеобский словарь: (Дополнение)* / ред. В.В. Палагина. Томск : Изд-во Том. ун-та, 1983–1986. Ч. 1–2.
47. Блинова О.И. Проект «Словаря русских старожильческих говоров Среднего Приобья» // Вестник Томского государственного университета. Филология. 2014. № 4 (30). С. 17–26.
48. Земичева С.С., Иванцова Е.В. Диалектный корпус как новый ресурс областной лексикографии // Вестник Томского государственного университета. 2019. № 446. С. 15–22.

From “*Abarmo*” to “*Yashchichishko*”: Creating the Lexicographic Component of the Tomsk Dialect Corpus

Voprosy leksikografii – Russian Journal of Lexicography, 2020, 18, pp. 98–116.

DOI: 10.17223/22274200/18/5

Svetlana S. Zemicheva, Tomsk State University (Tomsk, Russian Federation).

E-mail: optysmith@gmail.com

Keywords: Russian dialects of Siberia, dialect corpus, electronic dictionary, digital dialectology.

The study is supported by the Russian Science Foundation, Project No. 19-78-10015.

One of the most important trends in modern dialectological science is creating new electronic resources. The article gives an overview of Russian resources of this kind. Among them dialectal corpora hold a special place. The author of the article focuses on the Tomsk Dialect Corpus, which today includes more than 1,700,000 tokens. This resource is unparalleled in Russian scientific practice. It is designed as a universal information retrieval system which includes three modules: 1) textual, 2) grammatical, 3) lexicographic. The aim of the lexicographic component is to provide definitions of dialect lexemes. To do this, it is proposed to use the *Dictionary of Russian Old-Timers’ Dialects of the Middle Part of the River Ob Basin* (1964–1967) edited by V.V. Palagina and two supplements to it (1975, 1983–1986). The phases of the implementation of the lexicographic module into the Tomsk Dialect Corpus are described. The first phase was the automatic recognition of the above-mentioned paper dictionary. The second stage is editing the dictionary. The principles of editing the source material are determined by the fact that the lexicographic component is considered as part of a universal electronic

system. Two basic editing principles are: the possibility to process a word automatically and the autonomous functioning of each dictionary entry. In accordance with them, the vocabulary and the structure of the dictionary entry were formed. At the stage of forming the vocabulary, some dictionary entries (for example, two-word ones) were discarded. The structure of the dictionary entry contains the main areas: headword, definition and contexts. One of the main editing tasks is to combine dictionary entries from different volumes of the dictionary into one. These words are marked either as homonyms, or as the meanings of one word. Examples of dictionary entries before and after editing are presented in the article. By now, about a half of the original vocabulary has been processed (letters from A to M, 12,450 entries). The final version of the electronic dictionary as part of the Tomsk Dialect Corpus is planned to be presented on the website of the Laboratory of General and Siberian Lexicography (<http://losl.tsu.ru/>) by June 2021. The prospects of the project include, firstly, the expansion of the vocabulary, and secondly, the implementation of search by dictionary labels (diminutives, augmentative, etc.) into the corpus. The presented solutions can be used in the development of other dialect corpora.

References

1. Zhdanova, E.A. (2016) Science Project of the Linguistic Corpus of Russian Dialects of the Udmurt Republic. *Intellektual'nye sistemy v proizvodstve – Intelligent Systems in Manufacturing*. 4 (31). pp. 137–141. (In Russian).
2. Kryuchkova, O.Yu. (2017) [Scientific paradigms in dialectology and dialectological tradition at Saratov University]. *I.I. Sreznevskiy i russkoe istoricheskoe yazykoznanie: opyt i perspektivy. 205-letiyu so dnya rozhdeniya I.I. Sreznevskogo* [I.I. Sreznevsky and Russian Historical Linguistics: Experience and Prospects. To the 205th anniversary of the birth of I.I. Sreznevsky]. Proceedings of the International Conference. Ryazan: Ryazan State University. pp. 299–304. (In Russian).
3. Zemicheva, S.S. & Ivantsova, E.V. (2018) The project of Tomsk dialect corpus in keeping with trends of corpus linguistics development. *Sibirskiy filologicheskij zhurnal – Siberian Journal of Philology*. 3. pp. 192–205. (In Russian). DOI: 10.17223/18137083/64/18
4. Kachinskaya, I.B. & Malysheva, A.V. (2019) Folk Speech in Russian National Corpus. *Russkaya rech'*. 4. pp. 103–118. (In Russian). DOI: 10.31857/S013161170005368-3
5. Kachinskaya, I.B. & Krylov, S.A. (2010) [Dialect lexicography: an electronic corpus of “The Arkhangelsk region dialect dictionary”]. *Dialog-2010. Komp'yuternaya lingvistika i intellektual'nye tekhnologii* [Dialogue-2010. Computational linguistics and intelligent technologies]. Proceedings of the International Conference. Is. 9 (16). Moscow. pp. 169–172. [Online] Available from: <http://www.dialog-21.ru/media/1652/27.pdf>. (In Russian).
6. Kul'sharipova, R.E. & Ibragimov, T.I. (2013) Elektronnaya biblioteka russkikh narodnykh govorov Kazanskogo universiteta: vozmozhnosti primeneniya, informatsionnyy potentsial [Electronic library of Russian folk dialects of Kazan University: application possibilities, information potential]. *Mezhdunarodnyy zhurnal*

eksperimental'nogo obrazovaniya – International Journal of Experimental Education. 5. pp. 95–96.

7. Kokonova, A.B. (2015) Fonoteka arkhangel'skikh govorov [Sound library of Arkhangelsk dialects]. *Vestnik Rossiyskogo gumanitarnogo nauchnogo fonda – Vestnik RGNF*. 2 (79). pp. 231–236.

8. Kuznetsova, E.V. (2013) Informatsionnaya sistema “Leksicheskiy atlas Volgogradskoy oblasti”: nauchnyy material v uchebnom protsesse vuza [Information system “Lexical Atlas of Volgograd Oblast”: Research material in the educational process of the university]. In: Gerd, A.S. (ed.) *Leksicheskiy atlas russkikh narodnykh govorov (materialy i issledovaniya)* [Lexical atlas of Russian folk dialects (materials and studies)]. St. Petersburg: Nestor-Istoriya. pp. 335–344.

9. Isaev, I.I. et al. (2016) The Database of Russian Dialects Creation and Perspectives of Dialectometrical Research. *Vestnik Rossiyskoy akademii nauk – Herald of the Russian Academy of Sciences*. 86 (11). pp. 972–977.

10. *Dictionary of Russian Folk Dialects*. [Online] Available from: <http://iling.spb.ru/vocabula/srng/srng.html>. (Accessed: 23.04.2020). (In Russian).

11. *Arkhangelsk Regional Dictionary*. [Online] Available from: <http://www.philol.msu.ru/~dialectology/dictionary/>. (Accessed: 29.04.2020). (In Russian).

12. Vinogradov Institute of the Russian Language of the Russian Academy of Sciences. (2020) *Ssylki na skanirovannye versii nekotorykh russkikh dialektnykh slovarey i issledovaniy po dialektologii* [Links to scanned versions of some Russian dialect dictionaries and studies on dialectology]. [Online] Available from: http://www.ruslang.ru/dialectolog_centers_links. (Accessed: 10.05.2020).

13. Dal', V.I. (2020) *Tolkovyy slovar' zhivago velikoruskago yazyka* [Explanatory Dictionary of the Living Great Russian Language]. [Online] Available from: <http://slovardalja.net/>. (Accessed: 29.04.2020).

14. Kukushkina, I.S. (2012) [Conversion of the “Pskov Regional Dictionary” into the format of an electronic dictionary based on DWS LINGVO CONTENT]. *Russkiy yazyk i literatura v polikul'turnom kommunikativnom prostranstve* [Russian language and literature in the multicultural communicative space]. Proceedings of the International Conference. Pskov: Pskov State University. pp. 248–253.

15. Gramota.ru. (2020) *Shkol'nyy dialektologicheskyy atlas: Yazyk russkoy derevni* [School dialectological atlas: The language of the Russian village]. [Online] Available from: <http://gramota.ru/book/village>. (Accessed: 29.04.2020).

16. Knyazev, S., Moiseeva, E. & Shaul'skiy, E. (2020) *Fonetika russkikh dialektov* [Phonetics of Russian dialects]. [Online] Available from: <http://dialect.philol.msu.ru/index.php>. (Accessed: 12.03.2020).

17. Dolganina, A.A. & Shevchik, A.V. (2020) *Russkiy dialekty: vzglyad iz Sibiri* [Russian dialects: a view from Siberia]. [Online] Available from: https://pushkin-institute.ru/external_courses/260. (Accessed: 09.04.2020).

18. Szmrecsanyi, B. (2014) Methods and Objectives in Contemporary Dialectology. In: Seržant, I.A. & Wiemer, B. (eds) *Contemporary approaches to dialectology: The area of North, Northwest Russian and Belarusian vernaculars*. Vol. 12. Bergen: John Grieg AS. pp. 81–92.

19. *Russian Literature and Folklore. Fundamental Electronic Library*. [Online] Available from: <http://feb-web.ru/>. (Accessed: 12.03.2020). (In Russian).
20. Research and Educational Laboratory of Regional Philological Studies. (2020) *Baza dannykh pskovskogo fol'klora* [Database of Pskov folklore]. [Online] Available from: <http://nocpsko-viana.pskgu.ru/colloquial.php>. (Accessed: 20.03.2020).
21. Tomsk Interregional Institute of Social Sciences. (2020) *Baza dannykh Sredneobskogo fol'klora* [Database of the Middle Ob folklore]. [Online] Available from: <http://mion.tsu.ru/srobannot>. (Accessed: 02.03.2020).
22. “Dialect” *Linguo-Geographical System*. [Online] Available from: <http://manuscripts.ru/dl/dialekt.main>. (Accessed: 02.03.2020).
23. The Dialect Corpus of the Russian National Corpus. [Online] Available from: <http://www.ruscorpora.ru/search-dialect.html>. (Accessed: 12.03.2020). (In Russian).
24. Kryuchkova, O.Yu. & Gol'din, V.E. (2011) [Corpus of Russian dialectal speech: concept and parameters of assessment]. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii* [Computational linguistics and intelligent technologies]. Proceedings of the International Conference “Dialogue”. Bekasovo. 25–29 May 2011. Is. 10 (17). Moscow. pp. 359–367. (In Russian).
25. Daniel, M., Dobrushina, N. & Val'denfel's, R. (ed.) (2013–2018) *Govor basseyna Ust'i. Korpus severnorusskoy dialektnoy rechi* [Speech of the Ustyia Basin. Corpus of North Russian dialect speech]. Bern; Moscow. [Online] Available from: <http://parasolcorpus.org/Pushkino/stats.php>. (Accessed: 19.05.2020).
26. Waldenfels, R., Daniel, M. & Dobrushina, N. (2014) Why Standard Orthography? Building the Ustyia River Basin Corpus, an online corpus of a Russian dialect. *Computational Linguistics and Intellectual Technologies*. Papers from the Annual International Conference “Dialogue”. Moscow. Is. 13. pp. 270–278.
27. Malinino Corpus. (2020) *Corpus Statistics*. [Online] Available from: https://linghub.ru/malinino/#!/corpus_statistics. (Accessed: 20.05.2020).
28. *Corpus of Rogovatka Dialect*. [Online] Available from: <http://www.parasolcorpus.org/Rogovatka/>. (Accessed: 20.05.2020).
29. Electronic Text Corpus of Linguoculture of the Northern Angara Region. (2020) *Dialektnyy podkorpus* [Dialectal Subcorpus]. [Online] Available from: <http://angara.sfu-kras.ru/?page=dialect#>. (Accessed: 12.04.2020).
30. Regional Ethnolinguistics. (2020) *Dialektnyy korpus* [Dialectal Corpus]. [Online] Available from: <https://ethnolex.ru/kubdk/>. (Accessed: 20.05.2020).
31. Demeshkina, T.A. (2014) [Development vectors of modern Russian dialectology]. *Aktual'nye problemy obucheniya russkomu yazyku* [Topical issues of teaching the Russian language]. Proceedings of the International Conference. Brno, Czech Republic. 05–07 May 2014. Brno. pp. 268–278. (In Russian).
32. Rezanova, Z.I. (2015) Tomsk Regional Corpus: Typologically Relevant Parameters of Balance and Representativeness. *Vestnik Tomskogo gosudarstvennogo universiteta. Filologiya – Tomsk State University Journal of Philology*. 1 (33). pp. 38–50. (In Russian). DOI: 10.17223/19986645/33/4
33. *Russian National Corpus*. [Online] Available from: <http://ruscorpora.ru/old/>. (Accessed: 20.05.2020).

34. *Electronic Corpus of the Khakass Language*. [Online] Available from: <http://khakas.altai.ru>. (Accessed: 20.05.2020).
35. *National Corpus of the Kalmyk Language*. [Online] Available from: <http://kalmcorpura.ru/dial>. (Accessed: 20.05.2020).
36. *Corpus of the Vepsian Language*. [Online] Available from: <http://vepsian.krc.karelia.ru/about/>. (Accessed: 20.05.2020).
37. *Corpus of the Udmurt Language*. [Online] Available from: http://web-corpora.net/UdmurtCorpus/search/index.php?interface_language=ru. (Accessed: 25.04.2020).
38. *Bulgarian Dialectology as Living Tradition*. [Online] Available from: <http://bulgariandialectology.org/>. (Accessed: 05.03.2020).
39. Kachinskaya, I.B. & Sichinava, D.V. (2017) On the Corpus of Dialectal Texts in the Russian National Corpus. *Voprosy leksikografii – Russian Journal of Lexicography*. 11. pp. 71–85. (In Russian). DOI: 10.17223/22274200/11/5
40. Zhdanova, E.A. (2013) Leksikograficheskiy modul' lingvogeograficheskoy informatsionnoy sistemy "Dialekt" [Lexicographic module of the linguo-geographic information system "Dialect"]. In: Gerd, A.S. (ed.) *Leksicheskiy atlas russkikh narodnykh govorov (materialy i issledovaniya)* [Lexical atlas of Russian folk dialects (materials and studies)]. St. Petersburg: Nestor-Istoriya. pp. 185–192.
41. Yurina, E.A. (2011) Tomsk Dialectal Corpora: The Starting Point. *Vestnik Tomskogo gosudarstvennogo universiteta. Filologiya – Tomsk State University Journal of Philology*. 2 (14). pp. 58–63. (In Russian).
42. Laboratory of General and Siberian Lexicography, Tomsk State University. (2020) *Tomskiy dialektnyy korpus* [Tomsk Dialect Corpus]. [Online] Available from: <http://losl.tsu.ru/corpus>. (Accessed: 01.05.2020).
43. Ivantsova, E.V. (2017) Tomsk Dialect Corpus: substantiation of the concept and prospects of development. *Voprosy leksikografii – Russian Journal of Lexicography*. 11. pp. 54–70. (In Russian). DOI: 10.17223/22274200/11/4
44. Palagina, V.V. (ed.) (1964–1967) *Slovar' russkikh starozhil'cheskikh govorov sredney chasti basseyna r. Obi* [Dictionary of Russian Old-Timers' Dialects of the Middle Part of the River Ob Basin]. Vols 1–3. Tomsk: Tomsk State University.
45. Blinova, O.I. & Palagina, V.V. (eds) (1975) *Slovar' russkikh starozhil'cheskikh govorov sredney chasti basseyna r. Obi (Dopolnenie)* [Dictionary of Russian Old-Timers' Dialects of the Middle Part of the River Ob Basin (Supplement)]. Parts 1–2. Tomsk: Tomsk State University.
46. Palagina, V.V. (ed.) (1983–1986) *Sredneobskiy slovar': (Dopolnenie)* [Middle Ob Dictionary: (Supplement)]. Parts 1–2. Tomsk: Tomsk State University.
47. Blinova, O.I. (2014) The project of the Dictionary of Russian Old Timer Dialects of the Middle Ob. *Vestnik Tomskogo gosudarstvennogo universiteta. Filologiya – Tomsk State University Journal of Philology*. 4 (30). pp. 17–26.
48. Zemicheva, S.S. & Ivantsova, E.V. (2019) A Dialect Corpus as a New Resource of Regional Lexicography. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. 446. pp. 15–22. (In Russian). DOI: 10.17223/15617793/446/2