УДК 81'33+81.374+811.161.1'282.2

С.С. Земичева, Е.В. Иванцова

ДИАЛЕКТНЫЙ КОРПУС КАК НОВЫЙ РЕСУРС ОБЛАСТНОЙ ЛЕКСИКОГРАФИИ

Исследование выполнено при поддержке гранта РФФИ № 19-012-00320 «Томский диалектный корпус как новый ресурс для изучения народно-речевой культуры».

Ставится вопрос об использовании электронного корпуса тестов в сфере диалектной лексикографии. Потенциал применения этого нового источника составления областных словарей проанализирован на примере работ томской диалектологической школы. Показано, какими информативными возможностями характеризуется Томский диалектный корпус при совершенствовании уже составленных ранее словарей среднеобских говоров и создании новых лексикографических трудов. Ключевые слова: диалектный корпус; диалектная лексикография; томская диалектологическая школа; среднеобские диалектные словари.

Лексикография, имеющая многовековую историю, и корпусная лингвистика, входящая в ряд самых молодых междисциплинарных областей науки, активно развиваются в наши дни, демонстрируя актуальность этих направлений научного поиска. Процессы интеграции названных дисциплин привели к формированию корпусной лексикографии. На Западе данное направление разрабатывается с 60-х гг. ХХ в. Уже в этот период «в США и Европе крупнейшие лексикографические дома обратились к созданию корпусов литературного английского языка <...> для дальнейшего составления и издания словарей» [1. С. 17]. На рубеже XX—XXI вв. корпусная лексикография начинает развиваться и в России.

В этой области формируются два подхода. Вопервых, в качестве источника для составления словарей может использоваться уже существующий универсальный ресурс. В российской словарной практике чаще всего лексикографы обращаются к Национальному корпусу русского языка. На его основе в Институте русского языка им. В.В. Виноградова РАН создан ряд экспериментальных словарей: «Грамматический словарь новых слов русского языка», «Новый частотный словарь русской лексики», «Словарь русской идиоматики. Сочетания слов со значением высокой степени», «Словарь глагольной сочетаемости непредметных имен русского языка [2]. Кроме того, иллюстрации из НКРЯ включаются в лексикографические труды разных типов: «Академический толковый словарь» [3], «Активный словарь русского языка» [4], «Словарь служебных слов» [5], «Словарь русской пищевой метафоры» [6], региональный ономастический словарь [7] и др. Вовторых, для лексикографических целей может собираться специальный корпус. Так, корпус текстов А.П. Чехова был создан для авторского частотного словаря [8]. Корпусная лексикография внедряется и в диалектологические исследования: электронная база данных, близкая к корпусу, создана составителями «Архангельского областного словаря» [9]; в проекте корпуса коми диалектов предполагается составление на его основе электронного диалектного словаря [10].

Привлечение новых электронных ресурсов в словаростроении на данном этапе опережает осмысление источниковедческих возможностей корпусов для совершенствования уже существующих и составления новых словарных трудов. Очевидно, эта задача значима как для лингвистики в целом, так и для ее отдельных областей.

В настоящей статье вопрос о возможностях использования диалектного корпуса как источника составления областных словарей рассматривается на примере работ томской диалектологической школы.

Специфика этой школы многогранна [11]; одним из направлений деятельности, развивавшимся на протяжении всех лет ее существования, является лексикографическое. С 1964 г. издано более 40 словарей (в том числе многотомных), относящихся к 28 словарным типам (обзор см.: [12]). Ведется работа по составлению новых лексиконов; задумано также создание исправленной и дополненной версии толкового дифференциального словаря среднеобских говоров (ССГ) – первого из лексикографических трудов томских диалектологов, отразившего лексико-фразеологический состав диалектов Среднего Приобья по данным полевого архива 1946-1981 гг. в русле традиций диалектной лексикографии своего времени. Словарь представляет собой семитомное издание [СРСГ, Д-1, Д-2].

С 2010 г. параллельно с традиционной лексикографической деятельностью был начат новый проект – Томский диалектный корпус (ТДК). Его первоначальная концепция развивается и совершенствуется [13]; с 2016 г. работа по наполнению и разметке компьютерной базы данных под руководством С.С. Земичевой приобрела систематический интенсивный характер.

Источником создания корпуса служит диалектный архив кафедры русского языка Томского государственного университета, состоящий из трёх частей. Первая часть включает записи «от руки» 40–70-х гг., вторая — блокнотные расшифровки аудиозаписей 80-х — начала 2000-х гг., третья — расшифровки материалов новых экспедиций 2008–2018 гг., снабжённые звуковыми файлами.

Общий объём ресурса на сегодняшний день – свыше 1 400 000 словоупотреблений, из них 40% всего введенного и размеченного контента приходится на последнее десятилетие. Включенные в корпус новые экспедиционные материалы, собранные с применением технических средств и в большей мере, чем прежде, ориентированные на отражение непринужденной речи информантов, расширяют возможности электронного ресурса для словарной работы. Самые ранние записи нередко представляют собой фиксацию несвязных фрагментов речи, в том числе отдельных слов.

В ТДК реализованы три вида разметки: паспортная (место и год записи, сведения об информанте – в том числе возраст, пол, образование), иерархическая тематическая разметка и разметка по типу текста (спонтанные тексты: диалог, полилог, ситуативные вкрапления, спровоцированные тексты с ответами по вопросникам); в дальнейшем предполагается также морфологическая и жанровая разметка. Кроме обозначенных параметров, возможен поиск слов по конкретной форме и по лемме. На данном этапе основная часть корпуса [ТДК] закрыта для широкого круга пользователей, в свободном доступе представлена его демоверсия [14].

На стадии апробирования этого ресурса изучаются в том числе его информативные возможности для использования в области словарной работы.

Было установлено, что корпус может служить для составителей диалектных словарей 1) источником нового текстового материала; 2) источником выявления новых единиц, подлежащих лексикографированию; 3) источником уточнения семантических, грамматических и функциональных особенностей слов и фразеологизмов, уже зафиксированных в среднеобских словарях.

Прежде всего для лексикографа представляют ценность отраженные в корпусе фрагменты диалектного дискурса, которые являются фундаментом любого областного словаря.

В течение длительного периода источниковая база диалектных словарей складывалась на основе ручных записей устной речи с широким привлечением сбора материала по опросникам. Такие особенности присущи и словарю среднеобских говоров, составители которого опирались на полевые данные 40–70-х гг. ХХ в.

Появление в экспедициях техники для аудиозаписи (в Томском государственном университете – с конца 70-х – начала 80-х гг. ХХ в.) существенно облегчило возможность точной фиксации речи диалектоносителей, в том числе связной монологической и диалогической речи, дало резкий прирост объема собранного архива текстов. Эти введенные в корпус материалы позволят дополнить описание лексикографируемых единиц.

В ССГ некоторые слова были представлены без примеров употребления в речи старожилов. Словарные статьи такого типа давались в тех случаях, когда в результате ответа на вопрос «Что это?» был получен однословный ответ [СРСГ. Т. 1. С. 6]. Без иллюстративной зоны, например, описана лексема бондарка:

БОНДАРКА, и, ж. Бондарня. В.-Кет. Б. Яр (Том. В.-Кет) 1 [Д-1. Ч. 1. С. 39].

Новые данные ТДК дают возможность заполнить эту лакуну. Ресурс содержит на сегодняшний день контексты, отражающие девять словоупотреблений лексической единицы: Бонда'рить — человек кадки делат, бочки, у меня вот муж был бонда'рь, так всё делал, в бондарке работает (Белый Яр, 1979); Бондарка — это бо'ндарь там, бочки делает. (Колпашево, 1982)² и др.

Корпус дает возможность дополнить словарные материалы примерами употребления слов в живой непринужденной речи. Так, словарная статья ЧЕЛЯДЬ в ССГ иллюстрируется только высказываниями старожилов, полученными в результате опроса:

ЧЕЛЯДЬ, и, ж. То же, что арда [дети³]. – Челядь – маленьких зовут. Челядь наберётся – это ребя'ты (Карг. Тым.). Всяко зовут: ребятишки и арда и челядь (Крив. Был.). Том. (В-Кет. Колп. Крив. Пар.) [СРСГ. Т. 3. С. 224].

ТДК позволяет дополнить иллюстративную зону фрагментами спонтанного дискурса: У отца — одиннадцать душ, у дяди — двенадцать душ. <...> Отец получал за их премии, похвальну грамоту давали. Челяди было много, много (Нарга, 1984).

Расширяют возможности представления спонтанной речи в словаре отраженные в корпусе дискурсивные фрагменты общения жителей села между собой. Словарные статьи ССГ репрезентируют только монологические высказывания — как, например, в статье на фразеологизм каку' (таку') беду:

БЕДА ◆ КАКУ (ТАКУ) БЕДУ. Очень, сильно. – Раньше ежели добром пойдёт (невеста), дак плачет каку беду. Плачет, дак спасу нет (Том. Н. Ишт.). Гуляли вон каку беду (Кем. Пром. Ваг.). Закуталась каку беду, насдевала на себя (Кем. Яшк. Мох.). Хотели нас сфотографировать, директор таку беду разошёлся (В. Кет. Мох.). Вот они с ней жить-поживать. Стал Иван Петрович с ней скучать, таку беду [из сказки] (В.-Кет. Кет.). Том. (В.-Кет. Том.). Кем. (Пром. Яшк.) [Д-1. Ч. 1. С. 25].

В корпусе можно найти иллюстративный материал с этим фразеологизмом в диалоге диалектоносителей: Г.Д. Чё-то, однако, хочет разгуляться, ветер-то. В.П. Чё? Г.Д. Дак чё, он каку' беду рвёт. В.П. Рвёт, ши'бко? Г.Д. Ну. В.П. Поди, уж плёнку сорвал там. Г.Д. Дак а чё, за'просто, щас таку' беду рвёт. (Вершинию, 1995). Пример демонстрирует также вариативность формы фразеологической единицы (каку'/таку' беду) в речи одного лица.

Актуальным для современной лексикографической практики представляется и более широкое включение в областные словари фрагментов дискурса, отражающих особенности материальной и духовной народной культуры. Корпус, включающий объемный массив текстов и охватывающий значительный временной диапазон, помогает лексикографу отобрать самые яркие, выразительные контексты, раскрывающие своеобразие народного быта, истории, традиций и обрядов, мировидения и миропонимания.

В словарной статье ЧИРОК, например, томскими лексикографами представлено ограниченное число

высказываний, позволяющих выявить семы «обувь с опушкой», «женская», «облегающая ступню»:

ЧИРОК, а, или у, м., чаще мн. ЧИРКИ. 1. См. чаро'к в 1-м знач. [кожаная женская обувь с опушкой и завязками]. – Чирки были с опушкой (Мол. В. Сор). Чирки – как глубоки калоши (Крив. Перш.). Ну, вот чирки были, как сейчас тапочки, их носили женшшыны (Кем. Яшк. Пача). Том. (повсеместно). Кем. (Мар. Яшк.). Даль. Доп.: Олон. (Каргоп.). Маляревский: черки. Молотилов. Осокин. Патканов и Зобнин: черки [СРСГ. Т. 3. С. 231].

Поиск по ТДК выдает контексты, в которых проявляются также семы «кожаный», «легкий», «кустарного производства», поясняется слово опушка как деталь чирков, описывается их изготовление: У нас пермяки приезжали, все в лаптях, а у нас-то всё чирки' были. Сами же кожу каку'-нибудь обдерут да и сошьют чирки'. А они лёгоньки, небось наденешь, дак только пятки говорят. И красивы они были, опушнят их, так и любо сглянуть-то. (Баткат, 1967); А на вечёрку холшовые чулки наденем, не было же ничё. Чирки с опушками. Русские чирочки. [А опишите, из чего они делались?] А они делались, скотину били, мы делали кожу, отец сам шил чирки. Вот чирок такой, это из кожи, тут опушка вот такая, сделам оборочки, сплетём. [А что за опушка?] А опушка, чтоб чирки затягывать, тут затянем, затягывали. (Парабель, 2012). Представляет интерес пример, отражающий восприятие информантами этой обуви как сибирской (в сопоставлении с лаптями) и исконно русской (русские чирочки), хотя, по сведениям словаря М. Фасмера, название является заимствованным («из тур., крым.-тат., кумык., кирг.») [Фасмер]. В первом из высказываний, кроме того, вербализуется прагматическая (лёгоньки; наденешь, дак только пятки говорят) и эстетическая оценка (красивы были; любо сглянуть) этого вида обуви.

Предельно лаконичная иллюстрация употребления слова *хрестник* в одном из томов ССГ может быть дополнена зафиксированным в ТДК фрагментом диалектного дискурса, в котором детально описывается церковный обряд крещения ребенка, а также воспроизводится народный обрядовый текст «приговаривания» над младенцем с прогнозированием его счастливой судьбы. Ср.:

ХРЕСТНИК, а, м. Крестник. – Сестрин сын хресник мне ешшо (Кем. Пром. Ваг.). Том. (Том.). Кем. (Пром. Яшк.) [Д-1. Ч. 2. С. 256].

ТДК: [Как отмечают крестины?] Коуда родится ребёнок. Берут в церкву крёстных мужчину и женщину. В церкви кум приготовит купель, хрёсная воды тёплой и выльет в купель. А тоуды' ложат ребёнка в купель. Из купели вынимают и ложат на руки хрёстному, если мальчик, ежели девочка, то ложат на руки хрёстной. Ребёнка постригут, помажут миром головку и тело. Волоски и кусочек от свечки и бросают в купель. Если утонет, то ребёнок помрёт. Мальчика носют за царские двери, а девочку нет. С церквы приедут, шубу вверх шерсту положат и приговаривают: «Мой хрестник как был под крестом, так чтоб стал под венцом. И как эта шуба космата, так

был бы мой хрестник богатый». Уж а потом долго гуляют (Каргала, 1967).

Примеры такого рода могут быть полезны для составления лингвокультурологических словарей, в том числе концептоориентированных.

Корпусные данные позволяют также полнее отразить в современных лексикографических трудах такую неотъемлемую черту народной речи, как ее образность, широкое употребление диалектоносителями разнообразных выразительных средств текста — сравнений, метафор, фразеологизмов, вкраплений малых жанров фольклора. Этот материал для пополнения томских областных словарей был обнаружен в ТДК не только среди записей последних лет, но и в архивах ранних экспедиций.

Так, в ССГ присутствует архаизирующееся слово *шулю'кан/шулюка'н*, обозначающее в народной культуре ряженого в период празднования святок:

ШУЛЮКАН, а, м. Устар. То же, что наряжунчик [ряженый]. – В новый год бегали шулюканы, ворожили (Крив. Был.). Наряжали, одевали маски, называли шулюкан (Шег. Карг.). В святки шулюкан оболокатся. Если в хорошую одежду, то ряженый, а если махор какой страшной, то шулюкан (Мол. В. Сор.). Том. (Кож. Колп. Крив. Мол. Том. Шег.). Ср. Даль: шуликун арх. Опыт: шуликун Арханг. (Арханг.). Молотилов: шуликан [СРСГ. Т. 3. С. 242].

ШУЛЮКАН, а, м. Устар. Ряженый. — В старо время шулюканом идевались. Там чё попало на себя идевали. Охотница я была до этого. Гармонисты приходят, садятся на лавку, а шулюканы пляшут (Кем. Ягун.) [Д-2. Ч. 2. С. 207].

Вместе с тем не только в нем, но и в аспектном «Словаре образных слов и выражений сибирского говора» [COCB] образные употребления от этой лексемы не представлены. Единственное сравнение как шулюкан дано в «Полном словаре диалектной языковой личности» [ПСДЯЛ. Т. 4. С. 341]. Корпус демонстрирует наличие в среднеобских говорах этого компаратива при обозначении человека, одетого в старую, рваную, грязную одежду (В.П. Давно'шный халатик у тебя. А.П. Тамарин, ауа. Наздева'юсь, как шулю'кан. В.П. Тепле'. А.П. Счас же Новый год. Хаха! Валя вот, у Коли там... грит: «Давай наряжаться, мы же в прошлом году-то наряжались!» Я грю: «Да мы и так наряжены, то'ко лицо вымазать сажей, да и всё» (Вершинино, 1995); Кото'ры в Святки наряжались. <...> Ты как шулюка'н грязный да махрявый. (Каргала, 1958) и языковой метафоры с близким значением, используемой как бранное слово (В святки-то наряжались, маскировались, наряженый. Шулю'кан было прозвище, когда одетый плохо (Трубачёво, 1958); Шулю'кан идетый плохо, всё отец говорил на ребятишек, на взро'слав так не говорили. (там же)). В электронном ресурсе отмечен также цитируемый диалектоносителем текст, который, вероятно, является прецедентным: Рожество' от четырнадиатого по-новому ворожили. В дом молодёжь наряжалась в шулюка'ны: кто махрявый, кто какой. «Эх, на хлеб на соль, на угольну лучину». Кто солому в рот наберёт, кто частокол наберёт. (Баткат, 1983).

В составе прецедентных текстов диалектного корпуса наиболее частотны пословицы и поговорки. Возможность обращения лексикографа к тексту, в котором обнаружена паремия, позволяет выявить ее семантику, типичные ситуации употребления, речевые трансформации, которые обычно не комментируются в словарях.

В ТДК, например, встречается пословица живёт кошка, живёт и собака...: Н.Н. Этих подсадных чучело'в-то [о чиновниках] там до фига' же сидит, в кабинетах. И каждому хоть по миллиону дай, и то сколь кабинетов по миллиону опять. Г.Д. Ну чё сде'лашь? Ничё не сде'лашь. В.П. Живёт, гыт, кошка, живёт собака – и мы так живём (Вершинино, 1995); Вижу всё как в тумане <...>. Я различить никого не различу уж. Живу. Живёт кошка, живёт собака – и я живу. Вот так вот, всяки живут. Чё я сделаешь, жить надо (Асино, 2017). Отраженные в корпусе фрагменты дискурса показывают, что народный афоризм вербализуется говорящим после констатации им или его собеседником какого-либо положения дел, которое является нарушением нормы, вызывает дискомфортное состояние. Паремия декларирует невозможность изменения этого положения, призывает к смириться с ним; ее смысл поддерживается репликами Чё сделаешь? Ничё не сделашь; Чё я сделаешь, жить надо. В паремиологической единице наблюдается вариативная мена форм множественного и единственного числа, возможность эллипсиса наречия так (мы так живём / я живу).

Такого рода контексты могут быть востребованы при составлении словарей прецедентных текстов.

Корпус позволяет выявлять новые диалектные единицы, оставшиеся за пределами областных словарей.

Среди них есть новые слова, лексикосемантические варианты (в том числе образные), формальные варианты, однокорневые лексемы к уже зафиксированным, а также новые фразеологические единицы. Ресурс, предоставляющий возможность ознакомления с широким контекстом, помогает также наиболее точно определить их значение.

Одним из слов, ранее не представленных ни в одном из словарей среднеобского региона, стало клёк, встретившееся в рассказах диалектоносителей о детских играх: Мы играли, в догоня'шки играли. [Угу.] В клёк. Ямку выкапывали, вот такая вот чурочка, и вот палками били, нету шевяка' этого, палочки, шевя'к, а-а-а, какие игрушки были у нас? (Большая Саровка, 2016); А.И. Ну игры у нас какая? Клёк. Клёк называлась. А.В. Клёк, ауа. А.И. По ногам не раз получал. [Смеётся] А.В. Ставят такую чурочку, маленькую, вот, делают там значит кружок, очерчивают, ставят эту чурочку вот такую небольшую, а здесь эти, также как у городков эти вот были, ну и вот, по порядку встают, ну пацаны то в основном в клёк играли, и вот кто собьёт. Вот, вот такая игра. <...> Вот с этого клёка и пошли городки. (Большая Саровка, 2016). У В.И. Даля это существительное дано с толкованием «игра городки или чурки» [Даль. Т. 2. С. 115]; в СРНГ это объединенное толкование разделено на два самостоятельных значения [СРНГ. Вып. 13. С. 275]. Примеры ТДК указывают на то, что клёк является модификацией городков; в такой игре на селе могли использоваться различные подсобные средства – от *чурочек* до *шевяков* (засохшего помета коров или лошадей).

В корпусе возможно обнаружение новых значений у диалектных лексем, уже включенных в созданные словари.

Так, в ССГ слово *жарница* отмечено при обозначении домашней утвари и посуды:

ЖАРНИЦА, ы, ж. То же, что жарник [совок для выгребания углей]. – Жарницей у нас зовётся, на котору уголь подгребаем (Кож. Елг.). Том. (Кож.) [Д-1. Ч. 1. С. 135].

ЖАРНИЦА, ы, ж. (Уд.?). Посуда для жаренья. – Жарница – жаровня, шумовка (Кож. Елг.) [Д-2. Ч. 1. С. 101].

Контекст из ТДК отражает новый ЛСВ «количество продукта, изжаренного за один раз»: Щас рыбачу помаленьку. Запрещают ведь ловить. Если быть членом Общества охотников и рыболовов, тогда одну сеть разрешается ставить, да и то место укажут, а там сетей полно, а рыбы-то и нет. Я вот чебачков только на жа'рницу одну ловлю. (Зырянское, 1982).

В числе выявленных встречаются диалектные ЛСВ, базирующиеся на метафорическом переносе от общерусских единиц. ТДК фиксирует, например, у слова горох значение «множество маленьких детей» в высказывании Ой-ой-ой, сколько. Десять нас, представляешь от. Горох. Ну и дисциплина была (Шестаково, 2014). С этой же семантикой лексема отмечена в «Архангельском областном словаре» [АОС. Т. 9. С. 363].

Корпус позволяет выявлять также формальные варианты уже лексикографированных слов.

В ТДК обнаружены, к примеру, контексты со словами *частогово'р / часто'говор*, находящимися в отношениях формального варьирования с отмеченным в ССГ существительным *частоговорка*:

ЧАСТОГОВ**О**РКА, и, ж. Человек, который быстро говорит. — Это затрешшала частоговорка (В.-Кет. М. Яр.). Том. (В.-Кет.). Ср. Даль: частоговорка, нвг., пен. Скороговорка. Опыт: частоговорка — скороговорка. Новгор. (Тихв.), Псков. (Опоч.). Доп.: частоговорка — красноречие. Псков., Твер. [Д-1. Ч. 2. С. 262].

ТДК: Бра'жкинцы нас звали частогово'рами. Быстро говорим. (Корнилово, 1965); Они нас звали часто'говоры. Говорили мы быстро (Там же).

Как и в случае с лексико-семантическими вариантами, в корпусе встречаются также новые формальные варианты от общерусских единиц. Например, в недифференциальный словарь диалектоносителя из Среднего Приобья включено слово купель «сосуд для крещения ребенка» [ПСДЯЛ. Т. 2. С. 111]. ТДК отражает формы купе'ля и купа'льня, совпадающие с лексемой купель по семантике, но отличающиеся от нее составом морфем и грамматическими характеристиками: Раньше это купеля такая была большая, как ванна большая. (Цыганово, 1982); Крестили в церкви, венчали как там — я это видел. Ну поп, он стоит на всю трибуну, он чита'т. Они, ну, те что же'нются, стоят. На их венки та-

ки' красивые. Он берёт их и вокруг купальни обведёт. (Парабель, 1985); В церкви кум, кума, вот ребёночка вот так на простыню кладут, несут кругом этой купальни (Молчаново, 1978).

По мере выявления лакун в составе гнезд однокоренных слов рассматриваемый электронный ресурс позволяет лексикографу отразить в областном словаре словообразовательные связи лексем, репрезентирующие диалектную систему:

РАСПРЕТИТЬ, чу, тишь, что, сов. Разрешить, отменить запрет. – На их охоту запретили, нонче должны распретить их много стало (Пар. Вял.). Том. (Пар.). [СРСГ. Т. 3. С. 103]. ТДК: Гоняли, не давали бить. А сколько шишки было, ужас! [А почему не давали?] А вот не давали, чтоб до распрета. А какой распрет ждать, кода есть охота (Парабель, 2012).

МЯКАТЬ, аю, аешь, несов. См. мявкать [мяукать]. – Кошка мякат (Колп. Чал.). А кот у нас не мякат (В.-Кет. Кет.). Том. (В.-Кет. Колп.). Кем. (Кем.) [СРСГ. Т. 2. С. 163]. ТДК: Кошка – мя'кат же, кошка, говорят, замя'кала, под ногами так вьётся, ей исть надо, вот и мурлычет, мя'кат она. (Баранаково, 1979).

Кроме представленных в примерах новых отглагольных существительных (распре'т) и глагольных приставочных дериватов (замя'кать), в корпусе часто обнаруживаются диминутивы к лексикографированным ранее единицам (годови'к «годовалый ребёнок» [Д-2. Ч. 1. С. 69] — годовичо'к), отыменные прилагательные (горница «чистая половина крестьянского дома» [Д-1. Ч. 1. С. 100] — горничный «расположенный в горнице») и др.

Корпус также дает возможность расширения круга лексикографируемых устойчивых словосочетаний и фразеологизмов, объединенных каким-либо опорным словом.

Наряду с представленным в ССГ фразеологизмом на голые зубы, ТДК выдал три словоупотребления выражения на голую кочку:

ГОЛЫЙ ◆ НА ГОЛЫЕ ЗУБЫ. Ничего не имея. – Приехавшие-то на голые зубы забрались и уехать нельзя. Много забрано; отработай и поезжай (Крив. Ник.) [Д-2. Ч. 1. С. 71].

ТДК: Спомню, как жили. Сестрёнка от голода заумирала. От голоду столь народу погибало. Отец хлебопашеством занимался. Выслали на голую кочку. (Тискино, 1971); В тридцать шестым году колхоз открывался, отец не пошёл в колхоз, раскулачили опеть. Мы опеть на голой кочке (Нарга, 1984); Родители на... с Курска. Мы курские. Вот. [А как они сюда переехали?] Они как переехали? Они переехали... ктото, кто-то, кто-то переехал. Приехали на голую кочку. Вот. Потом они, эти деды', вызвали се'мью свою, родню. Потом... дядьку нашего с тёткой они вызвали суда'. Что ты! Тут лес, тут земля, тут всё, угодья (Мельниково, 2018).

Это устойчивое словосочетание дано в «Словаре русских народных говоров» и «Фразеологическом словаре русских говоров Сибири» с дефинициями «приехать на пустое место» [СРНГ. Вып. 15. С. 133], «экспрессивн. На пустое место (переселиться)» [ФС. С. 97]. С учетом всех предложенных ранее толкова-

ний наиболее точным представляется определение «о переселении на новое место без всякого имущества».

Выборка не отраженных до настоящего времени в среднеобских диалектных словарях единиц с привлечением ТДК дала:

а) слова, совпадающие по форме и значению с представленными в некоторых русских областных словарях. К их числу относится, например, кувя'кушка: [А родители Ваши не были против?] Мама-то сильно не хотела, нет. Ну, говорит: «Вот я мучилась, вас ро'стила, об вас плакала, ты росла у меня, и вдруг, какой-то с Камы, с Волги приехал, приживёт тебе кувя'кушку, бросит тебя, и останешься ты, будешь мучиться» (Молчаново, 1978). Лексема есть в [СРНГ. Вып. 15. С. 393];

б) единицы, которые близки по форме и семантике отраженным в диалектных лексикографических трудах, но не тождественны им. Одна из них – то'лька в контексте из корпуса Тюлька – бревно коротенько. Ну тюлька! – Бестолковое бревно! (Белый Яр, 1979). Здесь одновременно представлены и прямое значение «толстый обрубок бревна», отмеченное в [СРНГ. Вып. 46. С. 22], и переносное «о глупом, бестолковом человеке». Метафорический ЛСВ в сводном «Словаре русских народных говоров» несколько иной – «о малообразованном человеке» [Там же. С. 23], хотя довольно близок к выявленному в ТДК;

в) слова, которые не фиксируются ни в известных нам диалектных словарях, ни в словарях литературного языка. В качестве примера можно привести глагол скондаши'ть: В доме – ни окон, ни дверей, без потолка и без полу. Избёнку скондаши'ли и сбили (Нарга, 1984). Контекст позволяет сформулировать толкование «сделать быстро, наспех». Возможно, глагол родственен просторечному выражению с кондачка или диалектным лексемам сгоношить «сберечь, собрать» [СРСГ. Т. 3. С. 131], гоноши'ть «готовить наскоро» [Д-1. Ч. 1. С. 98]. В СРНГ есть вокабулы сканды'рить «завести, приобрести что-л.» [СРНГ. Вып. 37. С. 385] и скондыря'ть «быстро сделать, смастерить что-л.» [СРНГ. Вып. 38. С. 76], отчасти напоминающие интересующую нас единицу и по форме, и по значению.

Квалификация подобных употреблений в качестве узуальных диалектных либо окказиональных требует привлечения как можно более широкого круга словарных и корпусных источников в сочетании с экспериментальными исследованиями. Эти материалы могут быть полезны для дополнения как среднеобских диалектных словарей, так и сводных областных словарей.

Диалектный корпус может служить источником дополнения и уточнения уже изданных лексикографических трудов.

Корпусные данные в некоторых случаях помогают лексикографу более полно и точно сформулировать толкование значения диалектных единиц.

Так, толкование названия растения *косарик* дано в ССГ предположительно, со знаком вопроса:

КОСАРИК, а или у, м. Одуванчик? – Косарики – маленькие, жёлтые, они отцветут – пухнатенькие, листочки маленькие, дунешь – разлетятся (Кем. Мар.

М. Ант.). Кем. (Мар.). Ср. Даль: косарики... раст. дикая герань [Д-1. Ч. 1. С. 213].

ТДК подтверждает это предположение: *Ало'й – до'хтур. Девичья краса – Ванька мокрый. Берёзка.* Стекля'ночка. Одуванчик – коса'рик (Колеул, 1960).

Автоматизированный поиск в корпусе по лемме в ряде случаев уточняет представление о заглавной форме слова и его грамматических характеристиках. Глагол *аннулировать*, в частности, был реконструирован составителями ССГ на основе двух примеров с причастиями. Оставался открытым вопрос и о переходности лексемы:

АННУЛИРОВАТЬ, рую, руешь, кого? что? сов., прич. АННУЛИРОВАН. Уничтожить, убрать. – Избушки аннулированы (Том. Губ.). Номер, он не застроены, отведена земля, а хозяева аннулированы куда-то (Том. Губ.). Том. (Том.) [Д-1. Ч. 1. С. 16].

Корпус подтвердил достоверность этой реконструкции наличием контекста Когда объединяли колхозы, их всех в одну согнали. Это тоже ни к чему, деревни аннулировали, каждая деревня по-своему жила (Вершинино, 1991). Пример указывает и на правомерность постановки пометы что при характеристике переходности.

Широко распространенным в говорах явлением можно считать вариантность форм имени при слово-изменении. ССГ использует в словарных статьях функционально-грамматическую помету «чаще...», но только при описании зоны заглавия (САБОГ, а или у, м. чаще мн. САБОГИ. Сапог [СРСГ. Т. 3. С. 118]; ОШУРОК, рка или рку, м., чаще мн. ОШУРКИ. Шкварки [Д-1. Ч. 2. С. 62] и др.). Корпус позволяет не только выявить вариантные словоформы, но и установить наиболее частотные из них, отразив эту информацию в грамматической зоне словарной статьи.

Продемонстрируем это на примере лексикографического представления слова *боярин*, для которого типична множественная вариативность словоформ Им. падежа мн. числа:

БОЯРИН, а, м, мн. бояра, бояры, бояре, боярины. – Устар. То же, что боляр [участник свадебного обряда, гость на свадьбе]. – Бояры – мушшыны, штук по пятнадцать, знакомы и родственники (Крив. Чаг.). Утром бояры садятся за стол, их в тот момент величают, бояров (В.-Кет. Мох.). Бояре кладут деньги (Том. Б.-Прот.). Боярин, тысяцкый – это хрёсный (Туг. Сем.). Дружка, тысячник и их товаришшы всем по цветочку втыкают. Это всё боярины (Кем. Мар. Тюм.). Бояра – все поезжане. Придут невесту брать – это всё бояра (Кем. Яшк. Кос.). Том. (повсеместно). Кем. (Кем. Мар. Юрг. Яшк.). Даль [СРСГ. Т. 1. С. 48].

Материалы ТДК указывают на преобладание в Им. мн. формы *бояры* (19 словоупотреблений); прочие низкочастотны. Данные такого рода важны не только для толковых дифференциальных словарей, но и для словарей полного типа.

Тексты корпуса в ряде случаев также помогают уточнить сведения о слове в зоне функциональных помет.

Для многих лексем, имеющих в ССГ помету «фольклорное», в ТДК обнаружены свидетельства их

употребления не только в малых жанрах устного народного творчества, но и в спонтанной бытовой речи. Ср.:

Д**О**ЧЕРЬ, и, ж. Фолькл. Дочь. – Дочерь моя родненькая, куда ты нас спокинула? (из песни) (Кож. Кож.) Она ходит, свою дочерь Любушку побуживает... (из песни) (Кем. Мар. М. Ант.) [Д-1. Ч. 1. С. 124].

ТДК: А дочерь счас за другим живёт <...> дочерь дом выкупила, зовёт к себе, сенки отдельно, колидорчик, на Черемошниках дом-то купила, огородик есь, содит немного так и картошки, да там и колодец пробитый. Хороший домишка (Баткат, 1983).

Аналогичные факты обнаружены в отношении слов *болю'чий*, *де'тушки*, *ра'денький* и др.

Паспортная разметка корпуса содержит также сведения об области, районе и населенном пункте, где записан тот или иной текст. Такая информация позволяет уточнить зону локальных характеристик слов с учетом новых экспедиционных материалов. Лексема кошева в значении «лёгкие глубокие сани с высоким задком, обшитые рогожей или кошмой», например, дана в СРСГ с локальными пометами семи районов Томской и Кемеровской области: Асиновского, Верхне-Кетского, Зырянского, Колпашевского, Кривошеинского, Парабельского и Юргинского [СРСГ. Т. 2. С. 101]. ТДК отразил также примеры употребления единицы в Кожевниковском, Молчановском, Томском, Крапивинском, Шегарском районах: Кошевы – это у которых ездют, маленьки кошёвочки (Кожевниково, 1963); Потом запрягли лошадей, кошева' – такие же полозья, как у саней, только у'же, с сиденьями, с кучером (Молчаново, 1975); У меня тоже кошева была. Новый год справляли – Дед Мороз, Снегурка нарядятся, из деревни в деревню ездют (Вершинино 1982); Была кошёвка, сани, полусани, кошева, пролётки, простая телега, на железном ходу, брички (Крапивино. 1968); Потом повезут к венцу в кошеве', наря'дют (Мельниково, 1984).

Итак, диалектный корпус, приходящий на смену рукописным архивам записей народной речи и традиционным картотекам, может быть взят на вооружение лексикографами в целях оптимизации словарной работы. Его материалы позволяют совершенствовать уже созданные областные словари, уточнять особенности семантики, грамматики, функционирования включенных в них лексико-фразеологических единиц. Корпус служит для составителей словаря источником выявления новых слов и выражений, которые обнаруживают многообразные виды связей с уже зафиксированными единицами, что дает возможность заполнения лакун в репрезентации диалектной лексической системы. Наконец, текстоориентированный электронный ресурс является бесценным кладезем фрагментов диалектного дискурса, сохраненных для наших современников и потомков. Наличие текстов, зафиксированных на протяжении многих десятилетий, в том числе при помощи современных технических средств, будет способствовать составлению областных словарей нового поколения с более широким отражением спонтанной речи, сведений о составляющих народной культуры, расширением круга помет и параметров описания диалектного слова, включением статических данных. К перспективам применения диалектного корпуса следует отнести создание лекси-кографических трудов новых типов: лингвокультуро-

логических словарей, словарей диалектной концептосферы, словарей речевых жанров, прецедентных текстов и др. Развитие корпусной лексикографии, несомненно, послужит совершенствованию теории и практики словарного дела.

УСЛОВНЫЕ СОКРАЩЕНИЯ СЛОВАРНЫХ ИСТОЧНИКОВ

- АОС Архангельский областной словарь / под ред. О.Г. Гецовой, Е.А. Нефедовой. М. : Изд-во МГУ ; Наука, 1980–2017. Т. 1–18; издание продолжается.
- Даль Даль В.И. Толковый словарь живого великорусского языка. М. : Рус. яз., 1989–1991. Т. I–IV.
- Д-1 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.
- Д-2 Среднеобский словарь: (Дополнение) / ред. В.В. Палагина. Томск: Изд-во Том. ун-та, 1983-1986. Ч. 1, 2.
- ПОС Псковский областной словарь с историческими данными. Л./СПб. : ЛГУ/СПбГУ, 1967–2016. Вып. 1–26; издание продолжается.
- ПСДЯЛ Полный словарь диалектной языковой личности / под ред. Е.В. Иванцовой. Томск : Изд-во Том. ун-та, 2006–2012. Т. 1-4.
- СОСВ Блинова О.И., Мартынова С.Э., Юрина Е.А. Словарь образных слов и выражений народного говора / под ред. О.И. Блиновой. 2-е изд., испр. и доп. Томск : Изд-во Том. ун-та, 2001. 308 с.
- $CPH\Gamma$ Словарь русских народных говоров / под ред. Ф.П. Филина, Ф.П. Сороколетова, С.А. Мызникова. М., Л., СПб. : Наука, 1965—2016. Вып. 1–49; издание продолжается.
- СРСГ Словарь русских старожильческих говоров средней части бассейна р. Оби / ред. В.В. Палагина. Томск : Изд-во Том. ун-та, 1964–1967. Т. 1–3.
- ССГ толковый дифференциальный словарь среднеобских говоров (выпуски СРСГ, Д-1 и Д-2 в совокупности).
- ТДК Томский диалектный корпус // Лаборатория общей и сибирской лексикографии ТГУ. URL: http://losl.tsu.ru/?q=corpus (дата обращения: 03.07.2019). Доступ для зарегистрированных пользователей.
- Фасмер Фасмер М. Этимологический словарь русского языка / пер. с нем. и доп. О.Н. Трубачева. Под ред. и с предисловием проф. Б.А. Ларина. 2-е изд., стереотип. М.: Прогресс, 1986. Т. I–IV. URL: http://etymolog.ruslang.ru/index.php?act=contents&book=vasmer (дата обращения: 03.07.2019).
- ФС Фразеологический словарь русских говоров Сибири / под ред. А.И. Федорова. Новосибирск : Наука, 1983. 232 с.

ПРИМЕЧАНИЯ

- ¹ Локальные пометы в словарных статьях указывают на районы и населенные пункты Томской и Кемеровской области, входящие в Среднее Приобье.
- ² При контекстах из корпуса приводятся название населенного пункта среднеобского региона и год записи.
- ³ После отсылочных толкований в квадратных скобках указано значение диалектного слова в словарной статье, к которой дана отсылка.
- ⁴ В заключительной части словарных статей СРСГ приводились сведения из словарей и этнографических очерков о Сибири XIX начала XX в., по которым составители проверяли лексикографируемые единицы; в дополнительных выпусках словаря (Д-1, Д-2) в связи с началом публикации СРНГ результаты проверки в издание уже не включались.

ЛИТЕРАТУРА

- 1. Маник С.А. Современная корпусная лексикография: особенности и перспективы развития // Вестник Ивановского государственного университета. Сер. Гуманитарные науки. 2016. № 3 (9). С. 17–20.
- 2. Словари, созданные на основе Национального корпуса русского языка. URL: http://dict.ruslang.ru/ (дата обращения: 01.07.2019).
- 3. Академический толковый словарь русского языка / отв. ред. Л.П. Крысин. М.: Издательский дом ЯСК, 2016. Т. 1, 2.
- 4. Активный словарь русского языка / отв. ред. Ю.Д. Апресян. М.: Языки славянской культуры, 2014. Т. 1. А-Б. 408 с.
- 5. Служебные слова в лексикографическом аспекте / Е.А. Стародумова [и др.]; отв. ред. Е.С. Шереметьева, Е.А. Стародумова, П.М. Тюрин. Электрон. дан. Владивосток: Дальневост. федерал. ун-т, 2017. 1 CD-ROM. Систем. требов.: процессор с частотой 1,3 ГГц (Intel, AMD); оперативная память от 256 МБ, Windows (XP; Vista; 7 и т.п.). Загл. с экрана.
- 6. Словарь русской пищевой метафоры / под ред. Е.А. Юриной. Томск: Изд-во Том. ун-та, 2015-2017. Т. 1, 2.
- 7. Бубнова Н.В. Возможности использования материалов Национального корпуса русского языка при составлении региональных ономастических словарей // Современные проблемы лексикографии: материалы конф. СПб.: Нестор-История, 2015. С. 26–27.
- 8. Потемкин С.Б. Авторский корпус и словарь языка Антона Чехова // Труды международной конференции «Корпусная лингвистика-2015». СПб.: Изд-во СПбГУ, 2015. С. 382–389.
- 9. 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.
- 10. Некрасова Г.А. Электронный диалектный корпус как ресурс сохранения и изучения коми диалектов // Финно-угорский мир. Саранск, 2010. № 1 (6). С. 13–16.
- 11. Томская диалектологическая школа: историографический очерк / под ред. О.И. Блиновой. Томск: Изд-во Том. ун-та, 2006. 392 с.
- 12. Иванцова Е.В. Тенденции развития русской диалектной лексикографии в зеркале локальной системы словарей // Известия РАН. Сер. Литературы и языка. 2018. № 2. С. 5–14.
- 13. 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.
- 14. Томский диалектный корпус демоверсия. URL: http://losl.tsu.ru/?q=corpus/demo (дата обращения: 20.06.2019).

Статья представлена научной редакцией «Филология» 15 мая 2019 г.

A Dialect Corpus as a New Resource of Regional Lexicography

Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal, 2019, 446, 15–22.

DOI: 10.17223/15617793/446/2

Svetlana S. Zemicheva, Tomsk State University (Tomsk, Russian Federation). E-mail: optysmith@gmail.com Ekaterina V. Ivantsova, Tomsk State University (Tomsk, Russian Federation). E-mail: ekivancova@yandex.ru Keywords: dialect corpus; dialect lexicography; Tomsk School of Dialectology; Middle Ob dialect dictionaries.

The article raises the problem of integrating the achievements of lexicography and corpus linguistics. The question of the possibilities of using a corpus as a source for compiling dictionaries is considered on the example of the works of the Tomsk School of Dialectology: a large series of dictionaries of dialects of the Middle Ob region and an electronic dialect corpus representing an archive of records made during dialectological expeditions in the same region (1946-2018; texts with 1,400,000 word usages have been entered and tagged in the corpus). It has been established that the corpus serves as a valuable source of new text materials for compilers of dialect dictionaries. The Central Ob dictionaries, created on the basis of hand-written field records of the 1940s-1970s, have a large proportion of concise examples and metatext illustrations. Fragments of dialect discourse fixed in the corpus (including using technical tools) give lexicographers the possibility of a wider reflection of the live spoken language of dialect speakers. The electronic resource provides illustrative material that reveals the originality of folk life, history, traditions and rites, world view and perception. The corpus data also make it possible to more fully reflect the figurativeness of folk speech, the dialect speakers' broad use of comparisons, metaphors, idioms, small genres of folklore in modern lexicographic works. The corpus is a source of identifying new units remaining outside the regional dictionaries. Among them, there are new words and idioms, as well as formal, lexicalsemantic variants and single-root derivatives of the already described words. Their inclusion in the dictionary fills the gaps in the lexicographic representation of the dialect lexical system. In addition, the corpus data can be used as a source for supplementing and specifying the already published lexicographical works. The content of the text-oriented dialect corpus helps the lexicographer to more fully and accurately formulate the interpretation of the meaning of dialect units, specify their grammatical characteristics (including taking into account the frequency of variant word forms), functional features and local prevalence, which allows refining the corresponding areas of dictionary entries. Thus, the dialect corpus, which replaces the handwritten archives of folk speech recordings and traditional card indexes, can be adopted by lexicographers in order to optimize their work on dictionaries. Its materials will help to improve the already created regional explanatory dictionaries and will become a support in the compilation of new generation dictionaries with a broader reflection of spontaneous discourse, with more information about folk linguistic culture, with more labels and description parameters of the dialect word, and with the inclusion of static data. The prospects for the use of the dialect corpus include the creation of new types of lexicographic works: linguoculturological dictionaries, dictionaries of the dialect conceptosphere, dictionaries of speech genres, precedent texts, etc.

REFERENCES

- 1. Manik, S.A. (2016) Contemporary corpus lexicography: peculiarities and perspectives of development. Vestnik Ivanovskogo gosudarstvennogo universiteta. Seriya "Gumanitarnye nauki". 3 (9). pp. 17–20. (In Russian).
- Russian National Corpus. (2008) Slovari, sozdannye na osnove Natsional' nogo korpusa russkogo yazyka [Dictionaries Created on the Basis of the Russian National Corpus]. [Online] Available from: http://dict.ruslang.ru/. (Accessed: 01.07.2019).
- 3. Krysin, L.P. (ed.) (2016) Akademicheskiy tolkovyy slovar' russkogo yazyka [Academic Dictionary of the Russian Language]. Vols 1, 2. Moscow: Izdatel'skiy dom YaSK.
- Apresyan, Yu.D. (ed.) (2014) Aktivnyy slovar' russkogo yazyka [Active Dictionary of the Russian Language]. Vol. 1. Moscow: Yazyki slavyanskoy kul'tury.
- Starodumova, E.A. et al. (eds) (2017) Sluzhebnye slova v leksikograficheskom aspekte [Functional Words in the Aspect of Lexicography]. [CD-ROM]. Vladivostok: Far Eastern Federal University.
- Yurina, E.A. (ed.) (2015–2017) Slovar' russkoy pishchevoy metafory [Dictionary of Russian Food Metaphor]. Vols 1, 2. Tomsk: Tomsk State University.
- Bubnova, N.V. (2015) [Possibilities of Using Materials of the Russian National Corpus when Compiling Regional Onomastic Dictionaries]. Sovremennye problemy leksikografii [Modern Issues of Lexicography]. Proceedings of the International Conference. St. Petersburg. 27–30 April 2015. St. Petersburg: Nestor-Istoriya. pp. 26–27. (In Russian).
- 8. Potemkin, S.B. (2015) [The Author's Corpus and Dictionary of the Language of Anton Chekhov]. *Korpusnaya lingvistika* 2015 [Corpus Linguistics 2015]. Proceedings of the International Conference. St. Petersburg. 22–26 June 2015. St. Petersburg: St. Petersburg State University. pp. 382–389. (In Russian).
- Kachinskaya, I.B. & Krylov, S.A. (2010) Dialektnaya leksikografiya: elektronnaya kartoteka "Arkhangel'skogo oblastnogo slovarya" [Dialect Lexicography: Electronic File Cabinet of the "Arkhangelsk Regional Dictionary"]. In: Dialog-2010. Komp'yuternaya lingvistika i intellektual'nye tekhnologii [Dialogue–2010. Computational Linguistics and Intellectual Technologies]. 9 (16). Moscow: Russian State University for the Humanities. pp. 169–172.
- 10. Nekrasova, G.A. (2010) Elektronnyy dialektnyy korpus kak resurs sokhraneniya i izucheniya komi dialektov [Electronic Dialect Corpus as a Resource for the Preservation and Study of Komi Dialects]. Finno-ugorskiy mir Finno-Ugric World. 1 (6). pp. 13–16.
- 11. Blinova, O.I. (2006) Tomskaya dialektologicheskaya shkola: istoriograficheskiy ocherk [Tomsk School of Dialectology: a Historiographical Essay]. Tomsk: Tomsk State University.
- 12. Ivantsova, E.V. (2018) Development Trends in the Russian Dialectal Lexicography in the Mirror of the Local System of Dictionaries. *Izvestiya RAN. Seriya Literatury i yazyka The Bulletin of the Russian Academy of Sciences: Studies in Literature and Language*. 2. pp. 5–14. (In Russian).
- 13. Zemicheva, S.S. & Ivantsova, E.V. (2018) The Project of Tomsk Dialect Corpus in Keeping with Trends of Corpus Linguistics Development. Sibirskiy filologicheskiy zhurnal – Siberian Journal of Philology. 3. pp. 192–205. (In Russian). DOI: 10.17223/18137083/64/18
- Tomskiy dialektnyy korpus demoversiya [Tomsk Dialect Corpus Demo Version]. [Online] Available from: http://losl.tsu.ru/?q=corpus/demo (Accessed: 20.06.2019).

Received: 15 May 2019