

УДК 81'374

doi 10.17072/2073-6681-2020-1-68-77

К ПРОБЛЕМЕ СОЗДАНИЯ ЭЛЕКТРОННОГО СВОДНОГО ДИАЛЕКТНОГО СЛОВАРЯ

Елена Ивановна Сьянова**к. филол. н., старший научный сотрудник****Институт лингвистических исследований Российской академии наук**199053, Россия, г. Санкт-Петербург, Тучков переулок, 9. syanovaei@mail.ru

SPIN-код: 2206-2275

ORCID: <http://orcid.org/0000-0003-1408-2943>*Статья поступила в редакцию 02.02.2020***Просьба ссылаться на эту статью в русскоязычных источниках следующим образом:**

Сьянова Е. И. К проблеме создания электронного сводного диалектного словаря // Вестник Пермского университета. Российская и зарубежная филология. 2020. Т. 12, вып. 1. С. 68–77. doi 10.17072/2073-6681-2020-1-68-77

Please cite this article in English as:

Syanova E. I. K probleme sozdaniya elektronnoego svodnogo dialektного slovarya [On the Problem of Creating an Electronic Consolidated Dialect Dictionary]. *Vestnik Permskogo universiteta. Rossiyskaya i zarubezhnaya filologiya* [Perm University Herald. Russian and Foreign Philology], 2020, vol. 12, issue 1, pp. 68–77. doi 10.17072/2073-6681-2020-1-68-77 (In Russ.)

Автор статьи обращается к вопросу создания электронного формата сводного диалектного словаря русских народных говоров. В контексте обсуждения будущего переиздания «Словаря русских народных говоров» подобная форма видится в качестве необходимой вехи на пути решения насущных проблем отечественной диалектной лексикографии. Электронный формат открывает новые возможности для лексикографических решений, в частности, для отражения функциональных особенностей живой диалектной речи, для представления в словарной статье сведений этнокультурного характера. Создание электронного сводного диалектного словаря предполагает разработку базы данных, отражающей диалектную лексику с учетом взаимодействия частных диалектных систем. База данных понимается как система хранения с широким спектром функциональных возможностей по обработке информации. Это будет способствовать решению вопросов прежде всего систематизации диалектного материала для последующей обработки в рамках будущего электронного сводного диалектного словаря, в том числе материалов, известных науке, но подчас разрозненных и по-разному интерпретируемых в региональных лексикографических проектах. Структура предполагаемой базы данных позволит в будущем создать электронную картотеку – как уже обработанных карточек, так и карточек на основе новых, оригинальных (собранных в полевых условиях) материалов. Географическая и временная детализация, предполагаемая в базе, послужит основой для создания лексических атласов, а также метахронных карт, позволяющих отображать лингвистический ландшафт определенной территории в течение того или иного временного отрезка. Первоочередными задачами признаются: 1) анализ структуры исходных данных, формирование требований к формату их представления; 2) разработка формата представления метаданных о пространственно-временных характеристиках диалектных единиц, отвечающего требованиям мультимодальности и интероперабельности; 3) разработка онтологии диалектных данных и метаданных для семантического аннотирования (в частности, при соотнесении с внешними онтологиями SKOS, Lemon, LexInfo) и т. п.

Ключевые слова: компьютерная лексикография; диалектная лексикография; электронный сводный диалектный словарь; база данных; русские народные говоры.

В начале 60-х гг. XX в. начинается подготовка, а затем и издание сводного «Словаря русских народных говоров» (СРНГ), который, безусловно, стал знаковым событием не только отечественной, но и мировой лексикографии и диалектологии. Данный научный проект сыграл важную роль прежде всего в развитии отечественной диалектной лексикологии. В настоящее время в Отделе диалектной лексикографии и лингвогеографии русского языка ИЛИ РАН (группа «Словаря русских народных говоров»: С. А. Мызников, И. В. Бакланова, Р. В. Гайдамашко, Ю. Ф. Денисенко, В. Б. Колосова, Е. В. Колосько, О. Н. Крылова, Е. И. Сьянова и др.) продолжается работа над СРНГ. В 2019 г. вышел 51-й том (Ход – хоюшки). Словарь стал одним из крупнейших компендиумов диалектной лексики русского языка [Вендина 2004: 3]. В нем обобщены лексические материалы, собранные в XIX–XXI вв. на всей территории распространения русского языка. В основе отбора диалектных слов лежит дифференциальный принцип. Диалектным считается слово, «имеющее локальное распространение и в то же время не входящее в словарный состав литературного языка (в любую его разновидность)» [Филин 1961: 22]. В качестве основного критерия диалектного слова признавалась его территориальная ограниченность. Диалектологи указывали на уязвимость данного критерия. О. Г. Гецова полагала, что «словарные изоглоссы не объективно данная величина, а представляет собою на самом деле не что иное, как искомую величину, такое неизвестное, которое нужно установить после того, как будут составлены диалектные словари» [Гецова 1964: 97]. Наличие у диалектного слова изоглоссы оставляло вне рамок диалектной лексикографии слова, распространенные во всех русских говорах, но не употребляемые в литературном языке (типа *девка*, *лихой* ‘злой’), а также слова диалектного происхождения, имеющие в говорах диалектные изоглоссы, но общепринятые в литературном языке (типа *нетель*, *терebить* (лен), *стерня*, *буран* и т. п.) [Нефедова, Качинская, Коконова 2013: 40]. С. А. Мызников, главный редактор «Словаря русских народных говоров» с 2014 г. (с 47-го т.), отмечает, что данное положение недостаточно раскрывает содержание таких понятий, как «диалектное слово», «диалектное значение» (подробнее о диалектном слове в лексико-системном аспекте см.: [Блинова 1975; Блинова 1984; Коготкова 1979; Оссовецкий 1982 и др.]), и указывает на тот факт, что «не всегда имеются достаточные основания для «разведения» фактов общенародных и территориально ограниченных,

диалектных» [Мызников 2013: 574–575]. Это позволяет расширять состав словника сводного диалектного словаря: включаются единицы, обозначающие характерные и специфические для конкретной местности названия предметов, действий и явлений, которые не имеют обозначения в общенародном языке, и архаичные (в том числе реликтовые) слова и выражения, сохранившиеся преимущественно в фольклоре и в речи старшего поколения носителей диалектов [там же: 575]. Более того, материалы «Словаря русских народных говоров» доказали, что так называемая «общенародная лексика» «по сути своей не является общенародной, так как часто используется лишь в некоторых областях России» [Вендина 2004: 3]. В 51-м т. (2019 г.) широко представлены слова общенародного языка. Главный редактор С. А. Мызников подчеркивает, что при подаче слов с большим семантическим объемом «дифференциальный принцип строго не соблюдался» (СРНГ 51: 4). Более широко в словник вводится лексика арго, что позволяет «проследить, как единица арго получает широкое распространение в русских говорах» (там же: 5). Дифференциальный подход достаточен при ретроспективной характеристике диалектного слова. В свою очередь, лексико-системный подход при исследовании диалектного слова обуславливает рассмотрение всего словарного состава диалекта как средства общения. В этом случае оказывается важной системная целостность наблюдаемого объекта.

Стоит указать на тот факт, что в последнее время источниковедческая база данного лексикографического проекта значительно увеличилась. Подаются материалы из отдельных региональных словарей, вышедших за последние два десятилетия: «Словарь вологодских говоров» (СВГ), «Словарь русских говоров Карелии и сопредельных областей» (СРГЛ), «Словарь русских говоров Одесщины» (СРГО), «Словарь смоленских говоров» (ССГ) и др. Однако книжная версия словаря не позволяет охватить в настоящее время всю диалектную номенклатуру, представленную в региональных словарях последних десятилетий. Очевидно, только при переиздании можно будет включить в состав словника СРНГ (на соответствующие буквы) материалы продолжающихся изданий: «Архангельский областной словарь» (АОС), «Псковский областной словарь с историческими данными» (ПОС), «Селигер: материалы по русской диалектологии» (Селигер), «Словарь говоров Русского Севера» (СГРС) и др. Не фиксируются составителями в настоящий момент материалы таких проектов, как «Живое ко-

стромское слово. Краткий костромской областной словарь» (КрКОС), «Областной словарь вятских говоров» (ОСВГ), «Словарь донских говоров Волгоградской области» (СДГВО) и мн. др.

Перед составителями встает насущный вопрос переиздания словаря, что определяет, без сомнения, новый взгляд на традиционные принципы его составления, на состав словника и соответственно его расширение, на структуру словарной статьи, способы представления системных связей слова, на детальную разработку географических помет и т. п. Очевидно, что составителям и редакторам стоит обратить внимание на отражение функциональных особенностей живой диалектной речи, на представление в словарной статье сведений этнокультурного характера. Электронная форма региональных словарей и соответственно сводного диалектного словаря служит решению данного вопроса.

Современные информационные технологии, безусловно, открывают новые возможности для лексикографических решений. К сожалению, одной из ключевых проблем отечественной диалектологии, в частности диалектной лексикографии, является низкая цифровизация.

Диалектные словари, в том числе сводный словарь русских народных говоров, достаточно консервативны в сравнении с другими лексикографическими проектами. Они создаются на основе устоявшихся концепций, достаточно традиционных. Большинство региональных словарей демонстрирует новые подходы к лексикографической разработке лексической составляющей народных говоров. Выражается это как в реализации идеи семантической подачи материала, так и в расширении объема словарной статьи и увеличении количества параметров, характеризующих диалектную единицу. К тому же анализ современного состояния проблемы позволяет сделать следующий вывод: диалектная лексика и фразеология нуждаются в новых подходах к их описанию и представлению. Исследование содержания и структуры региональных и сводных словарей показывает, каким образом традиционно заложенные основные характеристики формы репрезентации лексической единицы могут стать отправной точкой для нового методологического подхода к подаче диалектного материала. Безусловно, словарь нового поколения должен отвечать таким критериям, как: 1) электронная форма; 2) интерактивность; 3) мультимедийность; 4) многомерность; 5) визуализация. Чрезвычайно важно начать работу по разработке методологической парадигмы для создания электронной формы сводного словаря русских народных говоров.

Возникновение этого замысла во многом определяется и общим направлением развития зарубежной лексикографии. Например, в Институте исконных языков Финляндии (Kotimaisten kielten keskus – «Котус») создан «Словарь финских диалектов» (*Suomen murteiden sanakirja*), изначально включающий в себя примерно 300 000–350 000 слов. Словарь состоит из восьми томов, первый из которых был издан в 1985 г., последний вышел в 2008 г. Начиная с 2012 г. словарь представлен в электронной версии и постоянно пополняется новыми словами – около 6000 новых слов в год (<http://kaino.kotus.fi/sms/>).

Создание электронного сводного диалектного словаря предполагает разработку базы данных, отражающей диалектную лексику с учетом взаимодействия частных диалектных систем.

В современной лингвистике разработаны принципы построения текстовых корпусов (Британский национальный корпус, Национальный корпус русского языка и др.). На сегодняшний день известно два больших корпуса в отечественной лингвистике – Национальный корпус русского языка (НКРЯ) и эшелонированный корпус текстов всех жанров литературы XIX–XXI вв. объемом в 34 000 000 слововхождений, созданный в Секторе теоретической семантики русского языка им. В. В. Виноградова РАН. Пока же диалектный материал, имеющийся в различных диалектологических центрах, либо мало доступен широкому кругу исследователей, либо в том виде, в котором он обычно существует (без специального аннотирования), не может быть использован с максимальной пользой.

Еще в 1980-х гг. была высказана мысль о необходимости создания машинного фонда диалектных текстов [Машинный фонд русского языка 1986; Гольдин 1989]. В настоящее время существует ряд корпусов, репрезентирующих отдельные элементы диалектной речи: зарубежные корпуса диалектных текстов (например, Helsinki corpus of English dialects, Kirk's Northern Ireland Transcribed Corpus of Speech (NITCS), IViE (Intonational Variation in English) corpus, BBC Voices); диалектный подкорпус в составе Национального корпуса русского языка (НКРЯ); лексико-грамматическая база данных (ЛГБД) по говору с. Пустоша Шатурского р-на Московской обл., включающая тексты – образцы речи носителей говора [Тер-Аванесова, Крылов 2006]; «Текстовая репрезентация диалекта как культурно-коммуникативного образования» (Саратовский государственный университет им. Н. Г. Чернышевского); электронный корпус диалектной культуры Кубани на основе лингвокультуроло-

гической концепции репрезентации диалектного дискурса [Трегубова, Емельянова 2011], где объектом описания в корпусе является локальная традиция северо-западной части Кубани, формирующаяся на материнской основе украинской и южнорусской культур в непосредственном контакте с линейными казаками (восточная Кубань) в среде русскоязычного населения. Названные проекты различаются своими целями, принципами организации баз данных, методами обработки (разметки) текстов, включаемых в состав электронного корпуса.

В 1990-е гг. появляется много разработок и исследований в области компьютерной лексикографии. Каждые 2 года проходила конференция COMPLEX. Обзор развития компьютерной лексикографии представлен (см. работу: [Dictionaries 2013]). В Отделе грамматики и лексикологии Института русского языка им. В. В. Виноградова РАН идет работа над «Новым толковым словарем современного русского литературного языка с расширенными сведениями о слове (в книжной и углубленной электронной версиях)». В компьютерной лингвистике в настоящее время известны такие прогрессивные способы и методы технологий представления данных, в частности – лексикографических, как: открытые связанные данные; Linguistic Linked Open Data (LLOD) – облако Лингвистических открытых связанных данных, которое описывает методы создания, обмена и повторного использования языковых ресурсов в соответствии с принципами связанных данных; онтология SKOS, описывающая тезаурусы, таксономии и наборы понятий, связанных иерархическими отношениями; онтология Lemon, предназначенная для сложных лексических ресурсов, где базовыми единицами являются: лексикон, лексическая единица, форма лексической единицы, смысл лексической единицы и понятия из онтологии предметных областей; онтология LexInfo используется для описания языковых категорий: род, число, падеж, время, прямой объект, косвенный объект, синоним, антоним и т. д.; методы автоматической обработки естественного языка (в особенности при работе с диакритическими знаками) – методы, включающие в себя токенизацию, лемматизацию, морфологический и синтаксический анализ. Указанные технологии позволяют использовать преимущества экосистемы Semantic Web, которая включает в себя хранилища данных, логические системы вывода и различные приложения.

На основе представленных, известных в мире, способов и методов обработки лексического материала коллективом составителей СРНГ будет

разработана методологическая парадигма базы данных для репрезентации диалектного материала с целью создания в будущем электронной формы сводного словаря русских народных говоров.

Важным признается тот факт, что данные будут обладать возможностью расширения внешних и внутренних связей. Исходя из целей проекта, будут использованы модели представления пространственно-временных характеристик. Темпоральный подход позволит работать с данными с учетом временных свойств, а Геотеггинг – связать информацию с географическими метаданными. Целостная экосистема (в рамках терминологии компьютерной лингвистики) будет учитывать и последующий доступ к данным. Инструменты работы с технологиями семантического веба предусматривают использование языка запросов SPARQL, которое требует технических знаний. Однако будет разработан интерфейс взаимодействия с данными, не требующий таких знаний, а предлагающий интуитивно понятный блочный подход взамен языка запросов.

Таким образом, в состав задач подобного проекта входят:

- 1) анализ структуры исходных данных, формирование требований к формату их представления;
- 2) разработка формата представления метаданных о пространственно-временных характеристиках диалектных единиц, отвечающего требованиям мультимодальности и интероперабельности;
- 3) разработка онтологии диалектных данных и метаданных для семантического аннотирования (при соотнесении с внешними онтологиями SKOS, Lemon, LexInfo); развертка сервера с семантической базой знаний (RDF хранилищем) и платформой выполнения SPARQL-запросов; разработка шаблонных запросов на поиск и анализ данных;
- 4) проектирование и разработка интерфейса взаимодействия с RDF-хранилищем (поиск данных) в виде конструктора запросов, не требующего знаний языка SPARQL, в том числе с поддержкой запросов на естественном языке.

Проект по разработке базы данных как основы создания электронного сводного диалектного словаря и соответственно самого словаря (на основе информационных ресурсов Института лингвистических исследований РАН, Санкт-Петербург) будет являться первой попыткой исследования и разработки современной и доступной пользователю базы данных, охватывающей диалектный материал в его диахронической и синхронической динамике. Подобный проект

предполагает: определение и анализ новаций в сфере отечественной и зарубежной диалектной лексикографии, определение методологической базы для создания словаря нового поколения – электронного сводного диалектного словаря (на материале русских народных говоров); разработку общих принципов базы данных для электронного сводного диалектного словаря; разработку и описание формата представления метаданных о пространственно-временных характеристиках диалектных материалов (т. е. формата, отвечающего требованиям мультимодальности, интероперабельности и интерпретации), разработку онтологии диалектных данных для аннотирования и представления машиночитаемых данных (в частности с привлечением внешних онтологий – SKOS, Lemon, LexInfo).

Предполагается разработка и описание приемов и методов лингвистического обеспечения автоматизированной системы, решение задачи сохранения аутентичности и одновременно достоверности и полноты диалектных текстов, репрезентирующих речь диалектоносителя современных говоров русского языка. Одной из приоритетных при этом признается задача оперативного извлечения и обработки диалектных данных.

База данных рассматривается как система хранения с широким спектром функциональных возможностей по обработке информации. На ее основе предполагается в будущем создание электронного сводного диалектного словаря. Необходимой характеристикой базы данных в этом случае признается возможность пополнения ее новыми словарными статьями в любое время. При создании подобного рода базы данных важно соблюдать детальную паспортизацию репрезентируемого материала, что, в свою очередь, будет способствовать решению следующих вопросов:

1) обеспечит систематизацию диалектного материала для последующей обработки в рамках будущего электронного сводного диалектного словаря;

2) выступит в качестве основы электронного сводного диалектного словаря (на материале русских народных говоров);

3) позволит проследить динамику развития современных диалектов. Электронная форма даст возможность каждую репрезентируемую единицу сопровождать годом фиксации;

4) будет служить основой для систематизации известных науке, но подчас разрозненных и порозному интерпретируемых диалектных данных;

5) структура предполагаемой базы данных позволит в будущем создать электронную карто-

теку: как уже обработанных карточек, так и карточек на основе новых, оригинальных (собранных в полевых условиях) материалов. Пополнение базы может осуществляться после каждой диалектологической экспедиции и публикации нового регионального лексикографического проекта;

6) будет служить основой создания частотного словаря, что обеспечит возможность провести количественный анализ лексико-грамматической системы, выявить круг высокочастотных единиц, особенности их распределения. Будет проведен функциональный анализ внутрдиалектной и междиалектной вариативности;

7) географическая и временная детализация, предполагаемая в базе, послужит основой для создания лексических атласов, а также метахронных карт, позволяющих отображать лингвистический ландшафт определенной территории в течение того или иного временного отрезка;

8) использование разрабатываемой базы данных открывает большие возможности комплексного описания диалектных материалов с широким привлечением аудио- и видеоматериалов;

9) материалы разрабатываемой базы данных предоставят возможность по-новому посмотреть на отдельные вопросы эволюции диалектного материала, моделирования архаического ментального пространства; расширят спектр сопоставимых граней с материалами славянского континуума в целом. Тем самым углубится представление о соотношении универсального и локального, о характере взаимодействия языковых реалий в контексте системы диалекта.

База данных предоставит следующие возможности для пользователей системы: 1) хранение и обработку лексических данных в облачной системе (лексемы и их парадигмы); 2) обеспечение межсловарных связей. База данных будет строиться на основе технологий облачного хранения и обработки данных с возможностью дальнейшей масштабируемости. Предоставляется доступ через веб-интерфейс в браузере или при помощи локальной версии программы с возможностью простой синхронизации локально добавленных данных в центральную систему. Практически все данные в системе поддерживают версиюность изменений и отслеживание авторства правок. На первом этапе прототип серверной части системы использует СУБД PostgreSQL для хранения словарных данных и файловую систему для хранения объектных данных (аудио, разметки, видео), но в дальнейшем планируется миграция объектных данных на открытые объектные хранилища (например, Openstack Swift или Ceph) для масштабируемости по предоставлению до-

ступа к данным. Облачная платформа Openstack позволяет динамически создавать виртуальные окружения для пользователей с полным набором необходимых пользователю инструментов в виртуальных машинах на базе ОС Linux.

Актуальность подобного проекта обусловливается разработкой методики повышения производительности и эргономичности систем поиска данных диалектных исследований. Это потребует разработки и описания эффективных алгоритмов, ориентированных на запросы и работу с онтологиями, методов индексации данных и алгоритмов поддерживающих RDF-представление в области диалектной лексикографии.

Создание подобной базы данных и электронного сводного диалектного словаря в рамках отечественной диалектологии (проект) будет служить решению следующих вопросов:

1) сохранение диалектных данных, зафиксированных в XIX–XXI вв. на всей территории распространения языка. Русскими диалектологами были собраны богатейшие материалы по лексике и фразеологии. К сожалению, не весь существующий каркас лексических и фразеологических единиц нашел отражение в печатных источниках, не весь вошел в научный оборот. Работа составителей СРНГ (И. В. Баклановой, Е. В. Колосько, О. Н. Крыловой, Е. И. Сьяновой) с рукописными материалами по Ленинградской, Псковской, Новгородской областям (1936–1947 гг.), хранящимися в архиве Института лингвистических исследований РАН (Санкт-Петербург), показала, что лексические данные из рукописей могут быть использованы при подготовке переиздания СРНГ. Рукописи содержат слова, не зафиксированные в сводном словаре (*баран*, м. ‘одна из рукоятей сохи’, *бо́тать*, несов., перех. ‘бодать (о корове)’, *вя́клина*, ж., собир. ‘стебли гороха, свеклы, моркови’ и др.). Данные рукописей могут служить также для уточнения ареала функционирования той или иной единицы, а также ее значений. Например: *батог* ‘палка’ в (СРНГ 2: 144) имеет только следующие географические пометы – волог., олон., арх., новг., смол., тамб., курск., перм., урал., сиб.; *гоготать*, несов., неперех. ‘ржать (о лошади)’ в (СРНГ 6: 265) приводится с пометами челяб., оренб., куйбыш., ворон., курск., орл., тул., калуж., твер., новг.; слово *ва́чуги*, мн. (ед. *ва́чуга*, ж.) ‘холщовые рукавицы’, ‘холщовые рукавицы, надеваемые поверх шерстяных варежек’ в (СРНГ 4: 78) имеет оттенки значения ‘рукавицы’ (арх., олон.), ‘суконные или вязанные из шерсти рукавицы’ (арх., олон., астрах.), ‘холщовые рукавицы’ (новг., тихв.), ‘суконные или шерстяные рукавицы, обшитые

сверху кожей’ (арх.), ‘рукавицы, у которых сукно на ладони, а сверху кожа’ (арх.), ‘рабочие рукавицы из парусины, надеваемые поверх варежек’ (ленингр.); *рассказаться*, сов. ‘рассказать, поведать о ком-л., чём-л.’ в (СРНГ 34: 209–210) имеет помету север. Ср. *водонос*, м. ‘деревянная дуга с выемками или крючками на концах для ношения ведер на плечах; коромысло’ в приведенных материалах и ‘жердь с прицепом (привешенной палкой) для ношения ушатов с водой (носят два человека)’ (Слов. Акад. 1806. Моск., калуж., горьк., пенз., тул., орл., ряз., арх., петерб., яросл., калин., смол.) (СРНГ 4: 343). Без сомнения, современные фонды диалектных материалов требуют специальных форм хранения (базы данных, корпуса, электронные словари и т. п.);

2) разработка методологических основ создания электронной базы данных раскрывает большие возможности комплексного описания диалектных материалов с широким привлечением аудио- и видеоматериалов;

3) электронный ресурс будет способствовать решению многих задач традиционной диалектологии, в которой преобладает описательный подход: создание электронных словарей и электронных словарей; расширение возможностей диалектной лексикографии: создание словарной картотеки на основе базы данных, электронного построения словарных статей и автоматической обработки лексического материала и т. п. Реализация проекта требует создания комплексной цифровизации процессов, возникающих в деятельности научных исследователей, в частности, исследователей Отдела диалектной лексикографии и лингвогеографии русского языка Института лингвистических исследований РАН;

4) материалы базы будут служить выявлению хронологии появления и исчезновения отдельных значений, принципов действия семантических закономерностей и тенденций в лексике на разных этапах развития, причин появления семантических трансформаций и инноваций на материале диалектных систем;

5) разработка методологической парадигмы базы данных и соответственно электронного диалектного словаря сопряжена с решением ряда сложных теоретических проблем традиционной диалектной лексикографии:

а) принципы отбора слов. Как показывает анализ, диалектные словари далеко не полностью отражают план содержания словесных единиц, функционирующих в русских говорах. В начале 90-х гг. XX в. О. В. Загоровская указывала на необходимость создания диалектных словарей, с максимальной полнотой отражаю-

щих план содержания диалектных лексических единиц с учетом аккумулярованного в нем своеобразия духовной культуры народа [Загорская 1991: 4]. В последних томах АОС многозначные общерусские слова по возможности представлены полноструктурно, без отсечения значений, общих с литературными, уже независимо от того, обнаружены ли у них какие-нибудь отличия от литературного языка. Это обстоятельство позволяет рассматривать словарь как полный диалектный словарь. Как показывает анализ существующих региональных словарей, дифференциальные словари не охватывают полностью дифференциальную лексику, функционирующую в говорах. И словари полного типа также не отражают всей существующей в говоре лексики;

б) оформление заголовочного слова – одна из самых остро стоящих в диалектной лексикографии. Часто лексикографам приходится описывать материал, имеющий недостаточное количество фиксаций, с трудно восстанавливаемыми словообразовательными связями, непрозрачной внутренней формой, гипотетической этимологией. Особенно труден выбор заголовочной формы в случае варьирования толкуемой лексемы;

в) варьирование диалектных лексем. Сложным и до конца не решенным остается вопрос о представлении связи лексико-фонетического варианта с определенной территорией, вопрос о лексикализации фонетических явлений [Варбот 2018; Кузнецова 1994 и другие];

г) этимология, мотивированность, внутренняя форма, членимость диалектного слова;

д) разработка полисемии (предметная отнесенность слова, разная функциональная направленность денотата, сочетание этих аспектов при сохранении интегральных сем и семантических связей между разными значениями лексемы; этимологически исходный дифференциальный признак, переносные смысловые разновидности, их иерархическая зависимость) и др.

Решение подобных вопросов является, безусловно, важной вехой в развитии диалектной лексикографии в формате новых компьютерных технологий. В частности, можно будет наблюдать общерусское слово во всей совокупности его значений, в том числе и значений, совпадающих с литературным языком.

В целом предполагаемый электронный словарь должен содержать комплексную, по возможности максимально полную информацию, позволяющую представить слово как единицу языка и как единицу, функционирующую в широком культурно-историческом контексте. Эта задача предполагает описание слова в его системных языко-

вых связях: частеречных, парадигматических, синтаксических, деривационных, фразеологических, функциональных. В рамках решения данной задачи возможно также отображение обусловленности семантики лексической единицы экстралингвистическими факторами. Это отвечает запросам справочного и исследовательского характера. Регистрация всех случаев использования той или иной языковой единицы с паспортизацией позволяет не только отыскать необходимый фрагмент, но и проследить и проанализировать синтагматические связи, определить синтаксические характеристики и функции и т. п.

Список источников

АОС – *Архангельский областной словарь* / под ред. О. Г. Гецовой (вып. 1–12), О. Г. Гецовой, Е. А. Нефедовой (вып. 13–15), Е. А. Нефедовой (вып. 16–20). Вып. 1–20. М.: Изд-во Моск. ун-та, 1980–2019.

КрКОС – *Живое костромское слово*. Краткий костромской областной словарь / сост. Н. С. Ганцовская, Г. И. Маширова; отв. ред. Н. С. Ганцовская. Кострома: КГУ им. Н. А. Некрасова, 2006.

НКРЯ – *Национальный корпус русского языка*. URL: www.ruscorpora.ru.

ОСВГ – *Областной словарь вятских говоров*: в 12 вып. / отв. ред. В. Г. Долгушев, З. В. Сметанина. Киров, 1996–2018.

ПОС – *Псковский областной словарь с историческими данными* / ред. А. И. Лебедева, О. С. Мжельская, С. М. Глушкина, Л. А. Ивашко, А. И. Корнев, И. С. Лутовинова и др. Вып. 1–27. Л. / СПб.: Изд-во Ленингр. ун-та / Изд-во С.-Петербург. ун-та, 1967–2017.

СВГ – *Словарь вологодских говоров*: в 12 вып. / ред. Т. Г. Паникаровская (вып. 1–7); Т. Г. Паникаровская, Л. Ю. Зорина (вып. 8–12). Вологда: Вологодский гос. пед. ин-т / Вологод. гос. пед. ун-т, 1983–2007.

СГРС – *Словарь говоров Русского Севера* / под ред. А. К. Матвеева, М. Э. Рут. Т. 1–7. Екатеринбург: Изд-во Урал. ун-та, 2001–2018.

СДГВО – *Словарь донских говоров Волгоградской области*: в 6 т. / ред. Р. И. Кудряшова. Волгоград: Издатель, 2011.

Селигер – *Селигер*: материалы по русской диалектологии: Словарь. Вып. 1–7: А–У / гл. ред. А. С. Герд. СПб.; Тверь: Изд-во С.-Петерб. ун-та, 2003–2017.

СРГК – *Словарь русских говоров Карелии и сопредельных областей*: в 6 т. / гл. ред. А. С. Герд. СПб.: Изд-во С.-Петерб. ун-та, 1994–2005.

СРГО – *Словарь русских говоров Одесщины*. Т. 1–2. Одесса: Астро Принт, 2000–2001.

СРНГ – *Словарь русских народных говоров* / гл. ред. Ф. П. Филин (вып. 1–22), Ф. П. Сороколетов (вып. 23–46), С. А. Мызников (вып. 47–51). Т. 1–51. М.; Л.; СПб.: Наука, 1965–2019.

ССГ – *Словарь смоленских говоров*: в 11 вып. / под ред. А. И. Ивановой, Е. Н. Борисовой, Л. З. Бояриновой. Смоленск: СГПУ, 1974–2005.

Список литературы

Блинова О. И. Введение в современную региональную лексикологию: материалы для спецкурса. Томск: Изд-во Том. ун-та, 1975. 258 с.

Блинова О. И. Русская диалектология. Лексика. Томск: Изд-во Том. ун-та, 1984. 133 с.

Варбот Ж. Ж. Вариантность диалектной лексики и этимология // Исследования по славянской диалектологии. 19–20. Славянские диалекты в современной языковой ситуации. Диалектный словарь как способ исследования славянских диалектов. М.: Ин-т славяноведения РАН, 2018. С. 130–134.

Вендина Т. И. Лексический атлас русских народных говоров (пробный выпуск): предварительные итоги // Вопросы языкознания. 2004. № 2. С. 3–19.

Гецова О. Г. О характере областного (диалектного) словаря // Научные доклады высшей школы. Филологические науки. 1964. № 3. С. 96–105.

Гольдин В. Е. К проекту Диалектологического текстового подфонда Машинного фонда русского языка // Доклады Третьей всеросс. конф. по созданию Машинного фонда русского языка. Ч. 2. М., 1989. С. 3–5.

Загоровская О. В. Семантика диалектного слова и проблемы диалектной лексикографии: автореф. дис. ... д-ра филол. наук. Л., 1991. 32 с.

Коготкова Т. С. Русская диалектная лексикология (состояние и перспективы). М., 1979. 334 с.

Кузнецова О. Д. Слово в говорах русского языка. СПб.: ИЛИ РАН, 1994. 86 с.

Машинный фонд русского языка: идеи и суждения / АН СССР, Ин-т рус. яз., Науч. совет по лексикологии и лексикографии; отв. ред. Ю. Н. Караулов. М.: Наука, 1986. 239, [1] с.

Мызников С. А. Русская диалектная сводная и региональная лексикография (XIX–XXI вв.) // Славянская лексикография = Slavic lexicography: [международная коллективная монография] / Международный ком. славистов, Комис. по лексикологии и лексикографии, Нац. ком. славистов Российской Федерации, Ин-т русского языка им. В. В. Виноградова РАН; отв. ред. М. И. Чернышева. М.: Азбуковник, 2013. С. 563–578.

Нефедова Е. А., Качинская И. Б., Коконцова А. Б. «Архангельский областной словарь»:

прошлое и настоящее // Вестник Московского университета. Сер. 9. Филология. 2013. № 3. С. 39–60.

Оссовецкий И. А. Лексика современных русских народных говоров. М.: Наука, 1982. 198 с.

Тер-Аванесова А. В., Крылов С. А. Лексико-грамматические базы данных как инструмент диалектологического описания // Компьютерная лингвистика и интеллектуальные технологии: труды Междунар. конф. «Диалог 2006». М.: Изд-во РГГУ, 2006. С. 493–498.

Трегубова Е. Н., Емельянова М. В. Лингвокультурный дискурс и специфика его репрезентации в диалектном корпусе // Русская устная речь: материалы междунар. науч. конф. «Баранниковские чтения. Устная речь: русская диалектная и разговорно-просторечная культура общения» и межвуз. совещ. «Проблемы создания и использования диалектологических корпусов», Саратов, 15–17 ноября 2010 г. Саратов, 2011. С. 264–269.

Филин Ф. П. Проект «Словаря русских народных говоров». М.; Л.: АН СССР, 1961. 197 с.

Dictionaries An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography. Berlin: De Gruyter, 2013. 1592 p.

References

Blinova O. I. *Vvedenie v sovremennuyu regional'nuyu leksikologiyu: materialy dlya spetskursa* [Introduction to modern regional lexicology: materials for a special course]. Tomsk, Tomsk State University Press, 1975. 258 p. (In Russ.)

Blinova O. I. *Russkaya dialektologiya. Leksika* [Russian dialectology. Vocabulary]. Tomsk, Tomsk State University Press, 1984. 133 p. (In Russ.)

Varbot Zh. Zh. *Variantsnost' dialektnoy leksiki i etimologiya* [The variability of dialect vocabulary and etymology]. *Issledovaniya po slavyanskoy dialektologii. 19–20. Slavyanskie dialekty v sovremennoy yazykovoy situatsii. Dialektnyy slovar' kak sposob issledovaniya slavyanskikh dialektov* [Research on Slavic dialectology. 19–20. Slavic dialects in the modern language situation. The dialect dictionary as a way to study Slavic dialects]. Moscow, Institute of Slavic Studies of the Russian Academy of Sciences Publ., 2018, pp. 130–134. (In Russ.)

Vendina T. I. *Leksicheskiy atlas russkikh narodnykh govorov (probnyy vypusk): predvaritel'nye itogi* [Lexical Atlas of Russian folk dialects (a pilot issue): preliminary results]. *Voprosy yazykoznaniiya* [Topics in the Study of Language], 2004, issue 2, pp. 3–19. (In Russ.)

Getsova O. G. O kharaktere oblastnogo (dialekt-nogo) slovarya [On the nature of the regional (dialect) dictionary]. *Nauchnye doklady vysshey shkoly. Filologicheskie nauki* [Philological Sciences. Scientific Essays of Higher Education], 1964, issue 3, pp. 96–105. (In Russ.)

Gol'din V. E. K proektu dialektologicheskogo tekstovogo podfonda Mashinnogo fonda russkogo yazyka [To the project of the dialectological text subfund of the Russian language machine fund]. *Doklady tret'ey vserossiyskoy konferentsii po sozdaniyu mashinnogo fonda russkogo yazyka. Ch. 2* [Proceedings of the third all-Russian conference on the creation of the Russian language machine fund. Pt. 2]. Moscow, 1989, pp. 3–5. (In Russ.)

Zagorovskaya O. V. *Semantika dialektного слова i problemy dialektной лексикографии*: Avtoref. diss. ... d-ra filol. nauk [The semantics of a dialect word and issues of dialect lexicography: Abstract of Dr. philol. sci. diss.]. Leningrad, 1991. 32 p. (In Russ.)

Kogotkova T. S. *Russkaya dialektная лексикология (sostoyanie i perspektivy)* [Russian dialect lexicology (state and prospects)]. Moscow, 1979. 334 p. (In Russ.)

Kuznetsova O. D. *Slovo v govorakh russkogo yazyka* [The word in the dialects of the Russian language]. St. Petersburg, ILI RAN, 1994. 86 p. (In Russ.)

Mashinnyy fond russkogo yazyka: idei i suzhdeniya. AN SSSR, In-t rus. yaz., Nauch. sovet po leksikologii i leksikografii [Russian language machine fund: ideas and opinions (Academy of Sciences of the Soviet Union, Russian Language Institute, Scientific board on lexicology and lexicography)]. Ed. by Yu. N. Karaulov. Moscow, Nauka Publ., 1986. 239 p. (In Russ.)

Myznikov S. A. *Russkaya dialektная svodnaya i regional'naya leksikografiya (19–21 vv.)* [Russian dialect summary and regional lexicography (19th–21st centuries)]. *Slavyanskaya leksikografiya = Slavic lexicography: [mezhdunarodnaya kollektivnaya monografiya]*. *Mezhdunarodnyy kom. slavistov, Komis. po leksikologii i leksikografii, Nats. kom. Slavistov Rossiyskoy Federatsii, In-t russkogo yazyka im. V. V. Vinogradova RAN* [Slavic lexicography: International collective monograph (International Committee of Slavists, Commission on lexicology and lexicography, National Committee of Slavists of the Russian Federation, V. V. Vinogradov Russian Language Institute of the Russian Academy of Scien-

ces)]. Ed. by M. I. Chernysheva. Moscow, Azbukovnik Publ., 2013, pp. 563–578. (In Russ.)

Nefedova E. A., Kachinskaya I. B., Kokonova A. B. *'Arkhangel'skiy oblastnoy slovar'*: *proshloe i nastoyashchee* [Arkhangelsk regional dictionary: The past and the present]. [Moscow State University Bulletin. Series 9. Philology], 2013, issue 3, pp. 39–60. (In Russ.)

Ossovetskiy I. A. *Leksika sovremennykh russkikh narodnykh govorov* [Vocabulary of modern Russian folk dialects]. Moscow, Nauka Publ., 1982. 198 p. (In Russ.)

Ter-Avanesova A. V., Krylov S. A. *Leksikogrammaticheskie bazy dannykh kak instrument dialektologicheskogo opisaniya* [Lexical and grammatical databases as a tool for dialectological description]. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii: Trudy Mezhdunarodnoy konferentsii 'Dialog 2006'* [Computer linguistics and intellectual technologies: Proceedings of the international conference 'Dialogue 2006']. Moscow, Russian State University for the Humanities Press, 2006, pp. 493–498. (In Russ.)

Tregubova E. N., Emel'yanova M. V. *Lingvo-kul'turnyy diskurs i spetsifika ego reprezentatsii v dialektnom korpuse* [Linguistic and cultural discourse and the specifics of its representation in the dialect corpus]. *Russkaya ustnaya rech': materialy mezhdunar. nauch. konf. 'Barannikovskie chteniya. Ustnaya rech': russkaya dialektная i razgovorno-prostorechnaya kul'tura obshheniya' i mezhvuz. soveshchaniya 'Problemy sozdaniya i ispol'zovaniya dialektologicheskikh korpusov'* [Russian oral speech: Proceedings of the International Scientific Conference 'Barannikov's readings'. Oral speech: Russian dialect and colloquial-vernacular culture of communication and inter-university meetings 'Issues of creation and use of dialectological corpus']. Saratov, 2011, pp. 264–269. (In Russ.)

Filin F. P. *Proekt 'Slovary russkikh narodnykh govorov'* [The project of 'The dictionary of Russian folk dialects']. Moscow, Leningrad, Academy of Sciences of the Soviet Union Publ., 1961. 197 p. (In Russ.)

Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography. Berlin, De Gruyter, 2013. 1592 p. (In Eng.)

ON THE PROBLEM OF CREATING AN ELECTRONIC CONSOLIDATED DIALECT DICTIONARY

Elena I. Syanova

Senior Researcher

Institute for Linguistic Studies of the Russian Academy of Sciences

9, Tuchkov pereulok, St. Petersburg, 199053, Russian Federation. syanovaei@mail.ru

SPIN-code: 2206-2275

ORCID: <http://orcid.org/0000-0003-1408-2943>

Submitted 02.02.2020

The article addresses the issue of creating an electronic format for a consolidated dictionary of Russian dialects. In the context of discussing the forthcoming reissue of the *Dictionary of Russian Folk Dialects*, such a form is seen as a necessary milestone on the way to solving the pressing problems of Russian dialect lexicography. The electronic format opens up new possibilities for lexicographic decisions, in particular for reflecting the functional features of live dialect speech, for presenting ethnocultural information in a dictionary entry. The creation of an electronic consolidated dialect dictionary involves the development of a database that reflects dialect vocabulary taking into account the interaction of separate dialect systems. The database is understood as a storage system with a wide range of functions for processing information. This will facilitate the systematization of dialect material for its subsequent processing in the framework of the future electronic consolidated dialect dictionary, including materials known to science but disparate and interpreted differently in regional lexicographic projects. In the future, the structure of the proposed database will make it possible to create an electronic card index, containing both cards already processed, and cards based on new, original (collected in the field) materials. The geographic and temporal specification proposed in the database will serve as the basis for the creation of lexical atlases, as well as metachronal maps, allowing the linguistic landscape of a certain territory to be displayed over a given time interval. The priorities are: 1) analysis of the structure of the source data, formation of requirements for the format of their presentation; 2) development of a format for the presentation of metadata on the spatiotemporal characteristics of dialect units that would meet the requirements of multimodality and interoperability; 3) development of an ontology of dialect data and metadata for semantic annotation (in particular, when correlated with external ontologies SKOS, Lemon, LexInfo), etc.

Key words: computer lexicography; dialect lexicography; electronic consolidated dialect dictionary; database; Russian folk dialects.