

МИНИСТЕРСТВО НА ОБРАЗОВАНИЕТО И НАУКАТА НА РЕПУБЛИКА БЪЛГАРИЯ
КИРИЛО-МЕТОДИЕВСКИ НАУЧЕН ЦЕНТЪР ПРИ БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ
ИЖЕВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ИМ. М. Т. КАЛАШНИКОВА
НАУЧНОЕ СООБЩЕСТВО “ПИСЬМЕННОЕ НАСЛЕДИЕ”
DIGITAL MEDIEVALIST SCHOLARLY COMMUNITY
ФОНДАЦИЯ „УСТОЙЧИВО РАЗВИТИЕ НА БЪЛГАРИЯ“

**Писменото наследство
и информационните технологии**

El’Manuscript–2014

Материали от V международна научна конференция
Варна, 15–20 септември 2014 г.

София · Ижевск
2014

Сборникът е издаден с финансовата подкрепа на Министерството на образованието и науката на Република България по процедура за подкрепа на международни научни форуми.

Отговорни редактори: проф. д-р В. А. Баранов
 доц. д-р В. Желязкова
 д-р А. М. Лаврентъев

Редактори: Нели Ганчева, Веселка Желязкова (български текст)
 О. В. Зуга, В. А. Баранов (руски текст)
 Кевин Хокинс (Kevin Hawkins) (английски текст)

Писменото наследство и информационните технологии [Текст] : материали от V международна науч. конф. (Варна, 15–20 септември 2014 г.) / отг. ред. В. А. Баранов, В. Желязкова, А. М. Лаврентъев. — София ; Ижевск, 2014. — 448 с.

Сборникът съдържа материали от конференция, посветена на разработването и създаването на съвременни средства за съхраняване, описване, обработка, анализ и публикуване на ръкописни и старопечатни книжовни паметници и исторически извори, а също и на въпросите за подготвянето на електронни ресурси в областта на хуманитаристиката и тяхното използване в научните изследвания и преподаването.

© Кирило-Методиевски научен център — БАН, 2014
© Ижевский государственный технический университет
им. М. Т. Калашникова, 2014
© Авторски колектив, 2014
© Лилия Тошкова — графичен дизайн на корицата, 2014

ISBN 978–954–9787–25–2

Лингвогеографическая информационная система “Диалект” на фоне современной лингвистической географии

Е. А. Жданова, В. А. Баранов, М. В. Климов, Д. Б. Кожевников

Лингвистическая география, автоматизация картографирования и лексико-графирования, русские говоры Удмуртии

The Linguo-geographical Information System “Dialect” and Modern Areal Linguistics

Ekaterina Zhdanova, Victor Baranov, Mikhail Klimov, Dmitry Kozhevnikov

This paper presents a general description of the linguo-geographical system “Dialect”, which is designed for storing, editing, selecting and presenting lexical material of the Russian dialects between the Kama and Vyatka rivers, collected for the Lexical Atlas of Russian Folk Dialects, in comparison with another linguo-geographical projects presented online. Attention is drawn to the possibility of using a database for creating a dictionary representation of dialect material.

Изучение территориальной дифференциации языков лингвогеографическим методом началось более ста лет назад. Однако и в начале XXI века эту работу нельзя считать завершенной, несмотря на то, что уже созданы лингвистические атласы, отражающие распространение языковых явлений различных уровней, для большинства языков и диалектов Европы, начаты аналогичные исследования на территории Азии и Америки. В настоящее время продолжается работа над межъязыковыми и наднациональными атласами, целью создания которых является отражение результатов древнейших языковых процессов, установление хронологии формирования современных языков одной языковой группы (например, Общеславянский лингвистический атлас) или же демонстрация распространения и взаимодействия неродственных языков на определенной территории (например, Атлас языков межкультурной коммуникации в Тихоокеанском регионе, Азии и Америки). В то же время актуальным остается и создание диалектных атласов небольшой территории, задачей которых является, прежде всего, подробное исследование специфических языковых особенностей того или иного региона. Это позволяет, например, уточнить данные о заселении картографируемой местности, конкретизировать уже выявленные и обнаружить неизвестные частные изоглоссы, исследовать данные с учетом взаимодействий с неродственными языками. Отбор картографируемых явлений для таких атласов производится в соответствии с теми языковыми чертами, которые целесообразно изучить в говорах конкретной территории для решения поставленных актуальных сегодня задач.

Работа над региональным атласом ведется в настоящее время и в Удмуртии сотрудниками кафедры “Лингвистика” Ижевского государственного технического университета (ИжГТУ) им. М. Т. Калашникова. Говоры междуречья Камы и Вятки ранее не становились объектом лингвогеографического описания. Эта территория должна была входить в один из томов Диалектологического атласа русского языка, однако составителями ДАРЯ было принято решение ограничить картографирование центром Европейской части России (территорией более раннего заселения).

Сегодня в создаваемом Институтом лингвистических исследований РАН (Санкт-Петербург) Лексическом атласе русских народных говоров (ЛАРНГ) отражается и лексика русских говоров междуречья Камы и Вятки. Однако подобные масштабные проекты не позволяют детально изучить распространение диалектных явлений в отдельном регионе. Например, в сетку обследования ЛАРНГ вошел 21 из 25 районов Удмуртии, так как лишь в них значителен процент русского населения. Однако в районах, где проживает большое количество удмуртов, также есть русские деревни и села, говоры которых могут представлять интерес для исследователей. Кроме того, в ЛАРНГ для нанесения на карту используются данные не всех обследованных населенных пунктов района: в связи с тем, что единицей картографирования является административный район, в нем выбираются один-два наиболее характерных и показательных в языковом отношении пункта. Для лингвистического обследования данного региона актуальна также проблема распространения групп говоров русского языка, взаимодействия и взаимовлияния русских и удмуртских говоров, что особенно ярко отражается на лексическом уровне. Поэтому в начале работ по сбору материалов к ЛАРНГ учеными Удмуртского государственного университета было принято решение о том, чтобы параллельно с подготовкой данных для головной организации начать работы по подготовке регионального атласа. Для этого в 2005 году была создана Лингвогеографическая информационная система “Диалект” (ЛГИС “Диалект”) (руководитель проф. В. А. Баранов, программист И. С. Соломенников), которая впоследствии была значительно усовершенствована в ИжГТУ (руководитель доц. Е. А. Жданова, программист Д. Б. Кожевников).

В последнее десятилетие, в эпоху господства компьютерных технологий во многих областях науки, в том числе и в языкознании, начали появляться и электронные лингвистические атласы русского языка. Однако большая часть из них ориентирована на представление в Интернете готовых карт, опубликованных ранее или не изданных по ряду причин. Такой способ — создание интернет-версий печатных атласов — обеспечивает доступность ценной лингвистической информации, а также, при определенной доработке, удобство просмотра и анализа карт. В качестве примеров таких проектов можно назвать электронную публикацию карт и справочных материалов Общеславянского лингвистического атласа (<http://www.slavatlas.org/>), недавно открытый электронный атлас татарских народных говоров (<http://www.atlas.antat.ru/>), создаваемый электронный диалекто-

логический атлас русского языка (<http://it-claim.ru/Persons/Kononkov/WWWroot/klaster.html>).

Некоторые научные коллективы, приступающие к картографированию, идут иным путем: используют информационные технологии уже на этапе создания картотек в виде их копий в машиночитаемом структурированном виде или в виде баз данных, а также в процессе подготовки самих атласов, так как это значительно облегчает труд составителей карт, а в случае размещения их в Интернете делает результат доступным широкому кругу пользователей. Автоматизация части кропотливого труда, связанного с отбором и систематизацией материала и нанесением данных на карту, сокращает время работы над атласом, предоставляет широкие возможности для экспериментов при анализе данных и выборе средств их визуализации на карте, позволяет демонстрировать результаты как в электронном виде, так и — при необходимости — в привычной форме печатного издания. Таким образом действуют, например, создатели ЛАРНГ (гл. ред. Т. И. Вендина), которые используют для работы географическую информационную систему MapInfo, составители Диалектологического атласа удмуртского языка (гл. ред. Р. Ш. Насибуллин), авторы интернет-проекта “Лексический атлас Волгоградской области” (<http://dialekt.vspu.ru>).

В свою очередь и на основе переведенных в электронную форму диалектных материалов, собранных по вопросам ЛАРНГ в междуречье Камы и Вятки, создана база данных, разработаны интерфейсы ввода и редактирования материала, формы запросов и вывода на экран выборок (рабочий адрес ЛГИС “Диалект” — <http://lgis2.office.hlcompany.ru/>) [Баранов, Жданова, Соломенников 2009; Баранов, Жданова, Белых 2013].

Система имеет ряд принципиальных отличий от других систем других проектов, что существенно расширяет возможности ее использования. В первую очередь стоит отметить иной способ представления материала. База данных, содержащая не только ответы на вопросы программы в виде отдельных словоформ, словосочетаний и контекстов, но и сведения о населенном пункте, о информанте, о времени записи ответов и о собирателях, позволяет составить запрос, содержащий параметры, которым будет соответствовать представляемый на карте лингвистический материал: состав лексем, их группировку, временной интервал фиксации данных, характеристики информантов, вид значков и т. д.

ЛГИС “Диалект” отличается также разнообразием содержащегося в базе данных материала: это не только лексика, полученная в результате опроса по программе ЛАРНГ, но и сканкопии полевых материалов и аудиозаписи диалектной речи, сделанные в ходе диалектологических экспедиций студентами и сотрудниками УдГУ в 1970–2000-е гг. Система позволяет разметить графические и аудиофайлы, выделив фрагменты, установив связь с вопросом(ами) программы и подготовив транскрипцию, указать лемму, дефиницию и другие характеристики.

Существующие в разных формах лингвистические данные дают возможность пользователю ЛГИС “Диалект” хранить, анализировать и картографировать не

только лексические единицы, но и явления других языковых уровней. Например, помимо лексического, система снабжена фонетическим модулем, в основе которого лежит программа ДАРЯ.

Важной чертой, отличающей ЛГИС “Диалект” от других современных лингвогеографических электронных ресурсов, является неограниченность территории, которая может быть представлена на карте. Если большинство электронных проектов рассчитаны на лингвогеографическое представление определенного региона / страны / группы регионов или стран — территории, условно отграниченной составителями, — то в основе лингвогеографического модуля ЛГИС “Диалект” лежит система карт “Яндекс”, которая позволяет картографировать материал любой территории, собранный по программе. Это дает диалектологам возможность отражать лингвистическую информацию на постоянно обновляемой карте, данные которой соответствуют текущему положению границ регионов и других изменяющихся географических объектов, представление которых на карте может быть связано с лингвистическим наполнением. Более того, картографический сервис “Яндекс” дает возможность пользователю самому наносить на карту необходимые объекты, например, исчезнувшие населенные пункты, в которых было зафиксировано картографируемое языковое явление. Отметим также, что использование такого рода программного продукта, находящегося в открытом доступе, делает диалектолога независимым при расширении территорий картографирования или увеличении количества пунктов сбора данных.

Уникальным является и лексикографический модуль ЛГИС “Диалект”, который представляет собой инструмент для создания электронного диалектного словаря на основе материалов ЛАРНГ с привлечением примеров из текстовых и аудиофайлов. Словарь отвечает всем критериям, предъявляемым к подобного рода ресурсам — базируется на разнообразном и значительном по объему диалектном материале, имеет систему поиска и перекрестных ссылок. Большая часть имеющихся сегодня в Интернете диалектных лексикографических ресурсов является копией печатных словарей (в графическом или PDF-формате), которые предоставляют лишь традиционные возможности их использования. Разработка в рамках ЛГИС “Диалект” лексикографического модуля является примером применения материала, предназначенного для картографирования, для создания электронного словаря, статьи которого содержат сведения о значении лексемы, сведения о времени и месте фиксации и ссылки на омонимы и синонимы.

Таким образом, лингвогеографическая информационная система “Диалект” обеспечивает хранение, редактирование, разметку, отбор и вывод на экран диалектных данных в виде лингвистических карт или словарных статей и может быть использована научными коллективами для решения широкого круга задач по картографированию и лексикографированию лексического материала народных говоров разных регионов России.

Литература

- Баранов, Жданова, Соломенников 2009 — *Баранов В. А., Жданова Е. А., Соломенников И. С.* Лингвогеографическая информационная система “Диалект” как инструмент для представления (картографирования) диалектной лексики // Лексический атлас русских народных говоров (Материалы и исследования) 2009 / Ин-т лингв. исслед. СПб.: Наука, 2009. С. 96–101.
- Баранов, Жданова, Белых 2013 — *Баранов В. А., Жданова Е. А., Белых А. А.* Лингвогеографическая система “Диалект”: история создания, новые возможности, технологические решения, демонстрация данных // Интеллектуальные системы в производстве. Ижевск, 2013. 1 (21). С. 171–175.