

# Электронная библиотека русских говоров: идеи, проблемы, решения<sup>1</sup>

© Р.Э. Кульшарипова, Т.И. Ибрагимов, Ф.И. Салимов

Казанский (Приволжский) федеральный университет  
rkulshar@mail.ru, Tavzich.Ibragimov@ksu.ru, Farid.Salimov@ksu.ru

## Аннотация

Статья посвящена изучению возможностей электронной библиотеки русских говоров в теоретическом и прикладном аспектах. Архив электронной библиотеки русских говоров представляет интерес для разработки основ диалектной речевой культуры с применением компьютерных технологий и обогащения диалектного подкорпуса русского языка.

## 1 История проекта

В 1960 – 1990 гг. в Казанском государственном университете активно проводились диалектологические исследования русских говоров Европейской части России. Ежегодно организовывались диалектологические экспедиции в северо-западные районы России. Сбор материала проводился по особой программе, разработанной З.М. Альмухамедовой, для дальнейшего экспериментально-фонетического описания характерных звуковых структур. Программа сбора материала включала словники. Эти словники отражали характеристики тех фонетических позиций, которые крайне важны для выявления акустической специфики звука по параметрам длительности, интенсивности и тембра. Параллельно делались записи спонтанной речи от типичных представителей местного говора. Необходимость сбора такого материала определялась тем, что диалектная речь содержала фактический материал, который характеризовался ведущими лингвистами как образцы народной речевой культуры. Естественно, основное внимание уделялось речи носителей старшей возрастной группы, т. к. они были хранителями традиций русской языковой культуры в условиях ее трансформации. До сих пор в литературе уточняются понятия «диалектный язык», «параметры диалектного общения» и т. д.

Техника записи состояла в следующем: во-первых, по возможности обследовались все жители того или иного населенного пункта; во-вторых, выбиралась тематика бесед, которая включала освещение всех сторон повседневности, все, что оптимально могло передать ценности их жизни. Стереотипы диалектного речевого поведения требовали от ис-

следователей умения вжиться в их диалектное коммуникативное пространство, создавать с диалектными носителями ситуацию интеракциональной общности, не разрушая привычных для них основ коммуникации.

Лингвистическая теория того времени только еще приступила к детальному освещению теории текста и дискурса, вопросы онтологии диалектного языка не были предметом серьезных разработок. Сделанные диалектные записи интересовали исследователей только как фактический материал, содержащий типичные и нетипичные (относительно говоров) черты. Звуковой строй говора изучался в соответствии с принципами, разработанными С.С. Высотским. Особенности консонантизма уточнялись в соответствии с техниками слухового анализа, для выявления специфики диалектного звучания часть материала перезаписывалась на студийные магнитофоны; записи на больших скоростях были оптимально хороши для проведения спектрального анализа.

## 2 Задачи проекта

Создание электронной библиотеки предполагает перевод имеющегося материала в цифровой формат, лингвистическую обработку речи в различных аспектах и размещение результатов исследований в интернете. Основной задачей проекта было создание доступного электронного ресурса, позволяющего получить доступ к электронным записям широкому кругу исследователей. При этом предполагалось провести систематизацию имеющегося материала, лингвистическую обработку записей, составить словари, характеризующие фонетические и грамматические особенности говоров.

## 3 Формат библиотеки, построение сайта

Первой задачей, которую необходимо было решать в проекте, была задача оцифровки имеющихся фонограмм с соответствующей очисткой их от посторонних шумов. Задача оказалась достаточно сложной в силу ряда причин: магнитная пленка, на которой хранились фонограммы, оказалась в плохом состоянии, магнитный слой был частично разрушен, пленки при перезаписи разрывались, на многих пленках запись была достаточно сильно зашумлена. Кроме того, при записях использовалась разная техника, записи были сделаны на различных скоростях, условия хранения магнитных носителей не соответствовали требуемым нормам. Оцифровка потребова-

<sup>1</sup>Труды 12<sup>й</sup> Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2010, Казань, Россия, 2010

ла достаточно серьезной технической поддержки, был создан инструментарий для считывания записей в шаблонном режиме, разработаны цифровые фильтры для очистки речевого сигнала. Для хранения оцифрованного материала был создан специальный сервер, который впоследствии был использован в качестве хранилища при создании электронной библиотеки. На текущем этапе оцифровано множество записей на скоростях 19.0, 76.0, 9.5 м/сек, что составляет более 70 часов звучащей речи.

Оцифрованные записи были подвергнуты всестороннему анализу. Прежде всего, было необходимо определить время проведения записи, регион, в котором проводилась запись, разбить записи по дикторам (при этом в некоторых случаях приходилось решать задачу идентификации диктора) и пр. К сожалению, архив, идентифицирующий записи, был частично утерян, часто информация, приведенная в картотеке, не соответствовала фактическому положению дел. В результате всестороннего анализа записей, с привлечением воспоминаний участников экспедиций, информация была частично восстановлена и отражена в оглавлении библиотеки. Оглавление библиотеки содержит следующую информацию:

- регион записи, с указанием населенного пункта, в котором проводилась запись; названия областей и республик приводятся на момент времени записи; в случаях, когда не удалось идентифицировать регион, было указано, что регион записи не определен; справочник регионов построен по иерархическому принципу; имеется возможность фильтрации записей, относящихся к конкретному региону, посредством выбора интересующего региона из выпадающего списка записей;
- год записи – в случае, когда год записи идентифицировать не удалось, год записи не указан;
- сведения о дикторе – включают фамилию, имя и возраст информанта;
- тематика записи; записи фонограмм велись по определенным темам, впоследствии эта информация была утеряна; была проделана работа по восстановлению тематики имеющихся записей, составлен словарь затрагивавшихся тем, который представлен в оглавлении.

Кроме того, в оглавлении присутствует информация о типе говора, об особенностях записи и другие показатели.

Электронная библиотека оформлена как интернет-ресурс, расположенный по адресу <http://dialekt.rx5.ru/>. Базовое программное обеспечение, использованное для построения сайта, является стандартным: веб-сервер apache, язык программирования php, сервер баз данных mysql. Звуковые файлы специальным образом подготовлены, уменьшены частота дискретизации сигнала, разрядность, проведены нормализация по уровню громкости, сжатие в свободно распространяемый формат ogg. Это было сделано для облегчения воспроизведения фонограмм на компьютерах, оснащенных обычными звуковыми платами, а также для уменьшения объема файлов. При выборе формата сжатия учитывались его полная

открытость и свобода, а также качество воспроизведения декодированной информации. Для прослушивания фонограмм с использованием этого формата пользователю предлагается скачать соответствующие кодеки. Все файлы непосредственно хранятся на сервере в специально отведенном каталоге. Основной принцип, которому разработчики следовали при создании сайта, – обработка только необходимой информации, без излишней перегруженности и возможность работы на ограниченных ресурсах.

Сайт структурно разбит на три части, входная страница предлагает переход на следующие подразделы.

*Оглавление библиотеки.* Отображение списка записей с возможностью фильтрации записей по региону. В этом разделе дополнительно доступна такая информация, как прослушивание/загрузка звукового файла, просмотр текстового варианта записи (если таковая имеется), просмотр транскрипции записи (если транскрипция проведена), просмотр частотного словаря словоформ в облегченно-транскрибированном тексте.

*Фотоальбом:* здесь расположены некоторые архивные фотографии, созданные во время экспедиций.

*Служебная страница,* доступная только участникам проекта. В ней возможна синхронизация рабочей базы с информацией на сайте. Данная возможность позволяет поддерживать актуальность информации на сайте без участия программистов.

С целью регистрации возросшего объема оцифрованного материала был создан программный модуль, включающий в себя базы данных по обработанным записям, расширенные средства поиска и конвертации данных из одного способа представления в другой (при обработке текстовой информации использовались различные операционные системы и кодировки), статистические функции обработки транскрипций и текстовых аналогов звучащих файлов. В будущем с использованием данного модуля предполагается продолжить комплексную лингвостатистическую обработку оцифрованного материала.

#### **4 Лингвистическая обработка собранного материала**

Одна из современных научных парадигм – это дискурсивно-когнитивная. Диалектная речь в этом преломлении представляет немалый интерес и перспективу. Диалектный дискурс подходит под дефиницию Н.Д. Арутюновой «дискурс – речь, погруженная в жизнь». Поэтому архивы электронной библиотеки русских говоров позволяют успешно разрешать следующие проблемы теории дискурса. Диалектная речь, записанная от разных носителей, которые проживают на разных территориях, может быть изучена в рамках понятий коммуникативного события. Наши дикторы свободно кодировали выражаемую ими информацию, которая была зачастую вербализацией их прошлой языковой памяти. Они

охотно вспоминали и передавали в разнообразных приемах процесс концептуализации мира, в котором они живут. В одном повествовании встречались коммуникативные эпизоды разных временных срезов: молодость, военные годы, современная жизнь. Электронная библиотека позволяет уточнить специфику системы речевых событий, которые вербализуются диалектными носителями в актах общения с неносителями. Детализация явлений привычных сторон их традиционного коммуникативного пространства, зафиксированная в соответствующих разделах электронной библиотеки, показательна для решения задач этнолингвистического характера, т. к. сведения подобного рода попросту могут уже отсутствовать по естественным причинам.

Диалектная речь содержит параметры, которые важны для исследования природы содержательно-смысловой организации текста в русском языке с подключением факторов функциональной маркированности. Диалектное коммуникативное пространство, представленное многочисленными моделями актуализации смысла, помогает установить своеобразие также и дискурсивного мышления. Хотя принято считать, что диалектные носители по ранжиру коммуникативных приемов отстают от носителей литературного языка, но выставленные образцы оцифрованной диалектной речи, напротив, демонстрируют исключительную динамику тематического развертывания в актах коммуникации.

Речевые ценности, правила речевого поведения в каждом типе говора могут варьироваться, но в целом они указывают на универсальные начала ведения речи. В настоящее время многими исследователями-диалектологами принимается в качестве исходного положение о существовании «особого типа речевой культуры сельского общения».

Электронная библиотека русских говоров сохраняет накопленный материал и позволяет широкому кругу пользователей в реальности постичь как технику речевых действий диалектных носителей, так и специфику вербализации ментального лексикона, отражающего варианты их картины мира. Это наглядно прослеживается в актах речи, когда, повествуя об одном и том же концепте, наши дикторы вводят в семантическое пространство сведения разного рода и плана. Так, повествуя о своей жизни, наши информанты по ходу иллюстрируют навыки коммуникативно-прагматического характера; это могут быть сведения о жизни в юности вперемешку с описанием беседы у молодых; уборка хлеба с элементами разговора с бабушкой; служба в армии с событиями возвращения домой и т. д.

Конечно, лингвистов-исследователей интересуют также собственно метаязыковые понятия, поэтому в нашей электронной библиотеке содержатся и языковые данные по звуковым особенностям записанной речи, реальный словарь речи, в котором возможно уловить тонкости коммуникативно-прагматического эффекта (какие слова наиболее частотны при изложении тех или иных мотивов;

какова природа информационной базы носителей и т. д.).

Традиции речевой сельской культуры – не просто дань памяти и уважения к предшествующим поколениям; хотя, несомненно, этнокультурный фактор значим, но материалы электронной библиотеки представляют интерес в прикладном ракурсе. Изучение механизмов коммуникации, природы коммуникативных кодов в диалектном общении, систематизации речевых регистров способствует осмыслению многих кардинальных вопросов современного речеведения.

Функциональное изучение диалектной речи многое может объяснить и в системно-структурном образовании русского языка, его функциональных разновидностях. Звуковое представление диалектной речи также помогает уточнить характерные аспекты понятия «диалектная языковая личность», ее роль и статус в общем процессе организации речемыслительных действий на русском языке.

Антропоцентризм диалектной речи проявляется в богатстве приемов и тактик, которыми владеет сельский житель и которые явственно проявляют себя в процессе их восприятия носителем литературного языка; оппозиция диалектная норма и литературная небезынтересна не только для разрешения вопросов коммуникативного и стилистического планов; это все осложняется определенными культурными ценностями, объемами соответствующих языковых знаний. В результате этого архив электронной библиотеки позволит уточнить, конкретизировать аспекты изучения диалектной языковой личности в преломлении к новым исследовательским парадигмам современного языкознания. Например, это можно увидеть при изучении статусно-ролевых действий диалектных носителей, когда в тексте они могут передавать в разнообразной тембровой окраске фрагменты речи своих родных, подруг, односельчан, особым образом моделируя их присутствие в диапазоне собственной речи.

Так называемый познавательный аспект языковой личности в ресурсах электронной библиотеки можно почувствовать не только через порождающийся диалектный текст, но и путем сравнения текстов разных носителей, принадлежащих к иным диалектным зонам. Обращаясь по ходу повествования к неносителям говора, наши информанты в качестве эффективного способа успешно демонстрировали технику ассоциативного наращивания смысла за счет повтора, интонационного выделения и специфики концептуализации смыслов (особенно, когда активизировалась модель диалогизации в общении).

Личностно-ориентированное общение в диалектном коммуникативном пространстве указывает на своеобразие поведенческого аспекта языковой личности; они называют самые важные смысловые каналы с ориентацией на собственное видение мира, но способны динамично коррелировать их с установками участников акта общения. Отсюда актуальность изучения природы прецедентных текстов в

диалектном пространстве. Созданная нами электронная библиотека также содержит материал, используя который, мы можем уточнить ракурсы онтологии диалектного текста через специфику просодии диалектного слова.

Диалектные тексты в составе электронной библиотеки русских говоров уникальны еще и потому, что в связи с рядом причин экстралингвистического плана они содержат базу данных, которая не может быть адекватна современному положению дел. Изменились времена и система ценностей, которыми они руководствовались, условия окружающего их речевого контекста, и мы имеем дело с моделированием уже иного порядка.

Идеи Казанской лингвистической школы подтверждаются природой диалектного материала, который оцифрован и может быть использован специалистами разных областей знания.

## 5 Заключение

Электронная библиотека говоров является частью мультимедийного диалектологического корпуса русского языка. Библиотека рассчитана на широкий круг пользователей. Говоры представлены в ней в четырех формах: звучащей речи, облегченно транскрибированного текста, отражающего специфику говора орфографического текста, и в форме словаря словоформ. Такое представление материала позволяет проводить исследования самодостаточности говора как коммуникативной системы жителей данного региона, а также социальных особенностей жизни села и фонетических различий говоров.

Работа по созданию библиотеки не завершена, на настоящий момент обработано чуть более половины имеющихся записей. Как и разрабатываемый диалектологический корпус, библиотека говоров многоаспектна, она может быть использована для проведения исследований речевой культуры и речевого поведения носителей русского языка. Материалы библиотеки могут найти применение в разработке систем автоматического распознавания и синтеза речи, при построении устройств идентификации диктора по голосу.

---

<sup>1</sup> Работа выполнена при финансовой поддержке РФФИ (проект 08-06-00160)

## Литература

- [1] Кульшарипова Р.Э., Ибрагимов Т.И. Аспекты описания электронной библиотеки говоров русского языка; по материалам диалектологических экспедиций Казанского государственного университета в 1969-1990 гг.// Русская и сопоставительная филология. – Казань: Изд-во Казан. ун-та, 2008. – С. 66-71.
- [2] Кульшарипова Р.Э., Ибрагимов Т.И., Салимов Ф.И., Сайхунов М.Р., Хусаинов Р.Р. К созданию электронной библиотеки русских говоров по материалам диалектологических экспедиций КГУ// Web Journal of Formal, Computational & Cognitive Linguistics. – 2008. – No 10. – [http://fccl.ksu.ru/issue10/Salimov\\_el\\_libr.rtf](http://fccl.ksu.ru/issue10/Salimov_el_libr.rtf).
- [3] Кульшарипова Р.Э., Ибрагимов Т.И., Салимов Ф.И. Библиотека русских говоров: речевое поведение говорящего в диалектной речи// Русская и сопоставительная филология. – Казань: Изд-во Казан. ун-та, 2009. – С. 78-84.
- [4] Кульшарипова Р.Э. Электронная библиотека русских говоров: аспекты изучения// Актуальные проблемы русской диалектологии и исследования старообрядчества: Тез. докл. – М.: Институт русского языка им. В.В. Виноградова РАН, 2009. – С. 135-137.

### **Elektronic library of the russian dialects: ideas, problems, decisions**

R.F. Kulsharipova, T.I. Ibragimov, F.I. Salimov

The paper is devoted to the studying of the possibilities of the electronic library of the Russian dialects in the theoretical and applied aspects. The archive of the electronic library of the Russian dialects is very interesting for working out the basis of the dialectal speech culture by means of the computer technologies and for the enrichment of the dialectal sub-corpora of the Russian.