

Концепция и структура Web-сайта "Каталог звукозаписей Санкт-Петербургских коллекций"

Бондарко Л.В., Вольская Н.Б., Скрелин П.А., Шерстинова Т.Ю., Кастров А.Ю.

Кафедра фонетики Санкт-Петербургского государственного университета

Фонограммархив института Русской Литературы РАН

E-mail: lvbon@lb1082.spb.edu, paul@phonet.lang.pu.ru, tanya@ts4306.spb.edu

1 Введение

"Каталог звукозаписей Санкт-Петербургских коллекций" разрабатывается в настоящее время на кафедре фонетики филологического факультета СПбГУ совместно с Фонограммархивом института Русской Литературы РАН (Пушкинского Дома) в рамках реализации международного научного проекта "Звуковые архивы Санкт-Петербургских коллекций в World Wide Web", проддерживаемого фондом INTAS (Nr. 1705). На сайте планируется представить каталог уникальных архивных аудиозаписей из коллекции Фонограммархива и других Санкт-Петербургских фонотек.

2 Коллекции звукозаписей Фонограммархива

Единственный в России фольклорный Фонограммархив Института Русской литературы Российской Академии наук был учрежден в начале XX столетия как Фонографический архив Первого Славянского Отделения Библиотеки Императорской Академии Наук по инициативе ее директора, академика А.А.Шахматова. Основу фонда Фонографического архива составили записи образцов фольклора славянских народов России и Восточной Европы, собранные в фольклорно-лингвистических экспедициях А.А.Шахматовым, Э.А.Вольгером, Н.С.Державиным и другими учеными.

После длительной консервации (с 1917 года) фонды Фонографического архива были объединены с обширной фонотекой Музея Антропологии и Этнографии, включавшей фольклорные и диалектологические звукозаписи народностей Сибири, Дальнего Востока и Средней Азии; материалами Фоноархива Государственного Института Истории Искусств, а также звуковыми коллекциями Ленинградской Государственной Консерватории. Таким образом, к концу 30-х годов Фонограммархив стал обладателем обширного и разнообразного фонда фольклорных звукозаписей от более чем ста народностей. В эти годы широко развернулась научная и собирательская деятельность его сотрудников, продолжающаяся до настоящего времени. В 1939 году Фонограммархив передан Институ-

ту русской литературы АН СССР, где существует и по сей день.

Наряду с записями русского фольклора, составляющими основную часть коллекций Фонограммархива, в его фондах широко представлены звуковые материалы различных наций и народностей, проживающих на территории России (шире - Российской Империи). Таким образом, специфика Фонограммархива заключается не только в большом количестве материалов (около ста пятидесяти тысяч единиц хранения), но и в разнообразии его фондов, сформированных на протяжении почти ста лет. Основные коллекции Фонограммархива к настоящему времени имеют полное научное описание.

Звуковые материалы Архива могут быть сгруппированы следующим образом:

1) Записи, выполненные на фонографических валиках (конец XIX - 50-е годы XX века, самая ранняя из них датируется 1890 годом). Среди них коллекция Е.Э. Линева (Русский Север и Средняя Россия), записи акад. Н.С. Державина от болгар-колонистов, материалы, записанные В.И. Йохельсоном в Сибири и на Дальнем Востоке, записи О.Э. Озаровской на Пинеге и Кулое, материалы комплексной экспедиции Государственного Института Истории Искусств (Заонежье, Пинега, Мезень, Печора). Особый интерес представляют уникальные записи Линева от молокан и духоборов в Закавказье, коллекции В.М.Жирмунского, И.Р.Зиндера и Э.Иогансена (фольклор немцев Поволжья и Санкт-Петербурга), среднеазиатские записи С.Г.Рыбакова, осетинские коллекции Б.А. Галаева и фольклор грузин-гурийцев в записях Э.Т.Эмстеймера и И.Д.Мегрелидзе. По технике звукозаписи к фоноваликам примыкают записи, выполненные на восковых дисках Эдисона в 1930 - 1950 гг.

2) Записи, выполненные на бытовых и репортерских магнитофонах (с 1950-х годов по настоящее время). Среди них материалы экспедиций ИРЛИ на Мезень, Печору, Кулой, в Костромскую, Пермскую, Горьковскую, Астраханскую, Смоленскую области, в Карелию и в районы русско-белорусско-украинского пограничья.

3) Высококачественные стереофонические звукозаписи на профессиональной аппаратуре (с 70-х годов по настоящее время), выполненные от выдающихся аутентичных исполнителей и этнографических коллективов.

Непреходящая научная и культурная ценность Фонограммархива обусловили его значение как источникового центра, содержащего богатейшие материалы для исследований и публикаций.

Первая Всероссийская научная конференция
ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ:
ПЕРСПЕКТИВНЫЕ МЕТОДЫ И ТЕХНОЛОГИИ,
ЭЛЕКТРОННЫЕ КОЛЛЕКЦИИ
19 - 21 октября 1999 г., Санкт-Петербург

3 Общая концепция сайта

Основной задачей Web-сайта "Каталог звукозаписей Санкт-Петербургских коллекций", как и следует из его названия, является представление в мировой сети текстовой информации об архивных аудиозаписях Санкт-Петербурга. В первую очередь будет описано собрание Фонограммархива - крупнейшее в России, - а также две коллекции Санкт-Петербургского университета: кафедры фонетики и кафедры русской литературы филологического факультета, а в будущем - и другие собрания.

Учитывая уникальные размеры архива, создание его полной электронной версии может занять многие годы. В то же время хотелось бы, чтобы уже первая версия каталога в определенной степени отражала все разнообразие коллекций - разные жанры, многочисленные языки и диалекты России и хронологические периоды. Поэтому показалось целесообразным начать работу с каталогизации всех материалов, относящихся к отдельному российскому региону. Выбор исследователей пал на Архангельскую область (губернию), основное население которой составляют русские, ненцы и коми.

В настоящее время на основе картотеки Фонограммархива создается компьютерная база данных, которая затем конвертируется в HTML-формат для представления в Internet. Сами звукозаписи прослушиваются, рестаурируются и переносятся на современные магнитные носители (CD-ROM). Для представления в Internet отбираются образцы записей разных жанров, диалектов и языков, интересные с точки зрения фонетики, лингвистики или фольклора. Далее они расшифровываются, сегментируются на фразы, синтагмы и слова, транскрибируются и подвергаются экспертному фонетическому анализу. Результаты обработки материала и сами звукозаписи становятся доступными пользователям Internet посредством акустической базы данных, представленной на сайте.

4 Технические требования

Оптимальный просмотр сайта обеспечивается броузерами Netscape Navigator и Internet Explorer последних версий. Так как каталог звукозаписей, выставляемых на сайте в настоящее время, охватывает русские, ненецкие и коми записи и поэтому ориентирован на читателя, в той или иной степени владеющего русским языком, большинство сопровождающей информации (тексты, комментарии, описания) дается на русском языке без английского дублирования. Основная кодировка сайта - Windows-1251. На страничке English-Help приводится информация о настройке кодировки на кириллицу для тех пользователей сети, которые ранее в Internet в таком режиме не работали. Сайт использует стандартные кириллические шрифты за одним исключением: для воспроизведения на экране транскрипционных символов необходима установка специального шрифта Times Trn3, который пользователь может получить по сети и установить на своем компьютере. Информация об этом дается на специальной странице сайта - "Технические требования".

Образцы и фрагменты звукозаписей хранятся в настоящее время в wav-формате. Поэтому для их прослушивания через сеть пользователю необходим компьютер со звуковой картой типа SoundBlaster и соответствующий Plug-in броузера или системный плеер. В ближайшее время планируется перевод материала из wav-

формата в формат RealAudio, который позволяет потоковое прослушивание звукового сигнала через сеть в режиме реального времени, минуя стадию его копирования на винчестер клиента. В этом случае от пользователя не требуется установка проигрывателя RealAudio Player, некоторые версии которого распространяются бесплатно фирмой-производителем RealNetworks Inc.

5 Особенности доступа

Коллекционные звукозаписи Фонограммархива являются национальным достоянием России. Поэтому во всемирную сеть выставляются только отдельные демонстрационные записи и их фрагменты. Доступ к звукозаписям может быть осуществлен с помощью персонального пароля, присваиваемого постоянным посетителям сайта, или после заполнения стандартной регистрационной формы пользователя. Представленные образцы разрешается использовать исключительно в научных и культурологических целях. Несанкционированное копирование, тиражирование и любое их использование в коммерческих целях преследуется по закону. Текстовый каталог собрания Фонограммархива выставляется в Internet со свободным доступом.

6 Описание структуры сайта

Начальная структура сайта содержит четыре основных гиперссылки: 1) технические требования; 2) общая информация о проекте; 3) текстовый каталог звукозаписей; 4) акустическая база данных образцов коллекций и их фрагментов. Информация, содержащаяся в первых двух ссылках, идентична той, что представлена в данной статье в соответствующих разделах.

Текстовый каталог звукозаписей на первой странице содержит список коллекций и их разделов, представленных в Internet в электронном виде. В настоящее время работа ведется над коллекцией 002 (Архангельская область, Россия, записи 1927г.). Посредством гиперссылки пользователь попадает в указанный им раздел заданной коллекции. Собственно каталог представляет собой таблицу, полученную в результате преобразования реляционной базы данных, которая заполняется сотрудниками Фонограммархива по данным картотек (поддерживается средой Microsoft Access 98/2000). Строки таблицы - описание конкретных записей, столбцы - параметры описания. Основными параметрами (полями) описания являются: 1) номер коллекции; 2) порядковый номер; 3) архивный шифр; 4) наименование записи; 5) начальная строка записи; 6) жанр; 7) количество исполнителей; 8) исполнители (в базе данных на каждую запись указывается до 12-ти исполнителей, в WWW-каталоге оставлены только первые три); 9) дата записи; 10) место записи (город, село, деревня); 11) район бытования (район, область); 12) национальность исполнителей; 13) язык; 14) качество записи; 15) моно/стерео; 16) скорость записи; 17) комментарий. На отдельной странице приводятся данные об исполнителях (в первую очередь - возраст и возможные комментарии). Для того чтобы процесс загрузки каталога по сети был экономичней, удобней и быстрее, он разбит на страницы по 10 записей на каждую. Навигационное окно в верхнем фрейме страниц позволяет быстрый переход от одного раздела коллекции к другой. Когда количество представленных на сайте звукозаписей станет достаточно представительным, планируется

разработка динамичной поисковой системы, основанной на запросах пользователей.

Для доступа к акустической базе данных образцов коллекции и их фрагментов пользователю необходимо заполнить в режиме on-line регистрационную форму или ввести персональный пароль. После этого раскрывается страница, схожая по структуре с основным каталогом, в центральной части которой находится таблица, описывающая образцы звукозаписей по стандартным параметрам. В дополнение к этому каждая строка таблицы имеет гиперссылку, открывающую доступ пользователя к целому списку документов, сопровождающих звукозапись, и ее прослушиванию. Внешний вид страниц акустической базы имеет довольно традиционный для WWW вид: они состоят из двух вертикальных фреймов - небольшого левого, выполняющего функции меню, и основного правого, в котором располагается собственно информация. По умолчанию здесь появляется орфографическая запись текста. Меню позволяет перейти к транскрипционной записи, которая может быть представлена символьно (для тех пользователей, которые установили на своей компьютере оригинальный транскрипционный шрифт) или в виде графического файла.

Прослушивание звукозаписей может осуществляться в следующих режимах: весь текст целиком, крупными фрагментами (примерно по 1 минуте звучания), отдельными синтагмами и изолированными словами. Последние два режима обеспечивают наиболее динамичную работу с материалом. (Известно, что размер передаваемых по сети файлов сильно влияет на эффективность работы в Internet, а звуковые файлы сами по себе довольно велики и плохо поддаются компрессии). Для прослушивания пользователь указывает в меню требуемый режим воспроизведения, после чего в основном окне появляется орфографическая запись текста с соответствующей разбивкой гиперссылок. Пользователь отмечает интересующий его фрагмент (слово, синтагму и т.д.), и гиперссылка вызывает соответствующий звуковой файл, который прослушивается с помощью браузера или установленного по умолчанию аудио-проигрывателя. Из общего количества записей, насчитывающего около 5000 часов звучания, для представления в Internet отбираются наиболее интересные и отражающие разнообразие жанров для русского, ненецкого и коми языков.

Отдельный пункт меню позволяет "раскрыть" фонетический комментарий, в котором отмечаются и анализируются особенности произношения, выявляются общие тенденции, характерные для языка, регионального диалекта, жанра или временного периода. На основе этого материала уже началось исследование сосуществования и взаимовлияния языков на примере русского, ненецкого и коми на территории Архангельской области России. Таким образом, акустическая база данных звукозаписей, хронологически охватывающих почти столетие, предоставляет ценнейший материал для исследований в области фонетики, диалектологии, фольклора, культурологии, этнографии и ряда других гуманитарных дисциплин, а возможности Internet-технологий откроют доступ к ней для пользователей World Wide Web.