

# КЕТСКИЙ ПРОЕКТ: ГОДОВОЙ ЦИКЛ РАБОТ И РЕЗУЛЬТАТЫ ПЕРВОГО КРУГА<sup>1</sup>

*О. А. Казакевич*

[kazak@orc.ru](mailto:kazak@orc.ru)

*И. В. Самарина*

[ira\\_samarina@hotmail.com](mailto:ira_samarina@hotmail.com)

*И. Л. Иткин*

[joseph@e-tools.com](mailto:joseph@e-tools.com)

*Т. Б. Базаряцкая*

[reutt@rambler.ru](mailto:reutt@rambler.ru)

*Т. Е. Реутт*

[reutt@rambler.ru](mailto:reutt@rambler.ru)

В докладе рассказывается о работе над проектом «Мультимедийная база данных кетского языка», реализуемом в НИВЦ МГУ при поддержке РГНФ (грант № 04-04-12028в), о некоторых полученных на первом этапе результатах и о возникших проблемах. Затрагиваются как вопросы программного обеспечения, так и методика сбора и обработки материалов для информационного наполнения базы.

## **1. Введение**

Данный проект продолжает на новом языковом материале начатую нашей группой в 2001 г. работу по комплексной документации исчезающих языков с использованием современных технологий аудио- и видеозаписи с параллельным созданием компьютерного мультимедийного архива (базы данных) для хранения полученных полевых материалов и обеспечения удобного к ним доступа при последующей обработке<sup>2</sup>. Выбор кетского языка для очередного проекта не случаен. Языковая ситуация у кетов сегодня такова, что если не создать качественный аудио- и видеоархив фиксаций его функционирования в ближайшие несколько лет, этого не удастся сделать уже никогда: только в старшем поколении кетов еще остались носители этнического языка, ни молодежь, ни представители средних возрастов своим этническим языком уже не владеют, а говорят исключительно по-русски. Кетский язык представлен несколькими локальными говорами, объединяемыми в три диалекта: северный,

центральный и южный. Наша задача – по возможности представить в базе материалы всех кетских говоров.

Два основных направления работы группы - это программная реализация мультимедийной базы и обеспечение ее информационного наполнения.

## **2. Программное обеспечение<sup>3</sup>**

Как уже говорилось, конечной целью нашей работы является не только сбор и упорядочивание данных полевых исследований по исчезающим языкам, но и внедрение программного окружения, которое позволит упростить этот процесс. Разработка такого окружения требует использования не только стандартного обще-функционального программного обеспечения, но и создания собственной предметно-ориентированной информационной системы.

Информационная система должна решить несколько задач:

- 1) хранение и систематизация имеющихся данных, как текстового, так и мультимедийного формата;
- 2) удобство добавления новых данных, обеспечение возможностей коллективной работы;
- 3) широкие возможности по созданию отчетов и трансляции их в разные форматы представления данных.

---

<sup>1</sup> Доклад подготовлен в рамках проекта «Мультимедийная база данных кетского языка», реализуемого при финансовой поддержке РГНФ, грант № 04-04-12028в.

<sup>2</sup> Информацию о первом в этой серии селькупском проекте можно найти в [Казакевич 2001; Казакевич и др. 2002; 2004].

---

<sup>3</sup> Программное обеспечение создано при техническом содействии OrkHard Group (<http://www.orkhard.com>).

Наша информационная система - мультимедийная база данных для хранения и редактирования результатов полевых исследований. Это специализированный программный комплекс для решения указанных задач. Система основана на использовании технологии клиент-сервер и имеет двухуровневую архитектуру - уровень базы данных и уровень графического интерфейса.

### 2.1. Уровень базы данных

Реляционная база данных, используемая в системе содержит всю текстовую информацию - словари, сведения об информантах, сведения о языках и диалектах. База данных также содержит таблицы, связывающие текстовую информацию с мультимедиа данными, которые физически хранятся отдельно в файловой системе. В качестве СУБД мы используем MSDE - бесплатную редакцию Microsoft SQL сервера. Эта редакция СУБД специально предназначена для включения в виде merge модуля в состав других программных продуктов. Для составления отчетов используется функциональность хранимой процедуры

sp\_makewebtask для генерации web страниц по результатам выполнения одного или нескольких SQL запросов. Одним из достоинств системы основанной на реляционной базе данных является возможность расширения списка запросов к нашей базе данных. Для этого нужно закодировать запрос на языке T-SQL и при необходимости добавить html шаблон. На нашем сайте представлены примеры html отчетов полученных в результате выполнения некоторых самых простых запросов (например получить фонетическую транскрипцию и аудио данные для каждого информанта на выделенном наборе слов). Примеры отчетов находятся в разделах Словарь и Дикторы. Нормализованность структуры БД позволяет обеспечить целостность данных, что очень важно для дальнейшего расширения базы за счет новых данных. Для доступа к данным используется технология ADO.NET. Диаграмма (Рис. 1) иллюстрирует схему базы данных. Данная схема позволяет не только учесть особенности специфичные для каждого конкретного диалекта, но и проводить локализацию самой информационной системы.

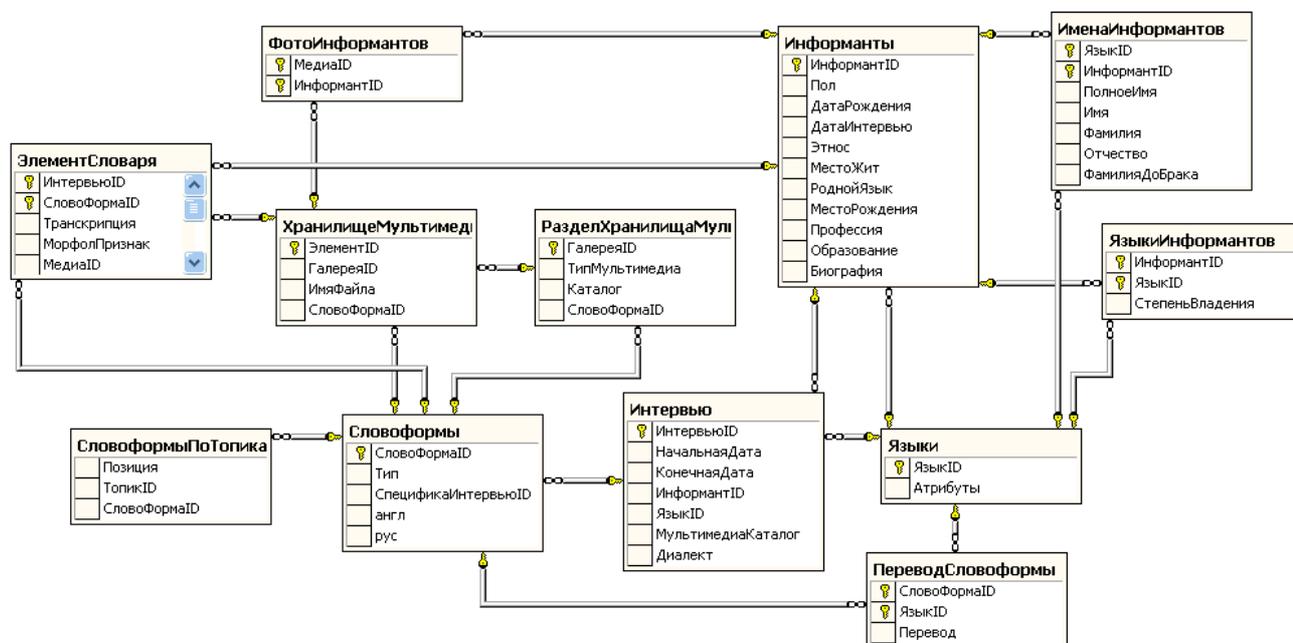


Рис. 1. Схема базы данных

### 2.2. Уровень графического интерфейса

В настоящий момент уровень находится в разработке. Текущая версия может быть использована для получения отчетов и простого редактирования данных. Уровень графического интерфейса создан с использованием языка C# и Microsoft Visual Studio.NET 2003 и основан на использовании Windows Forms. Помимо простоты выбор среды разработки, обусловлен также, желанием использовать также технологию ASP.NET расширив возможности онлайнного доступа к системе.

### 3. Информационное обеспечение

При всей важности программного обеспечения создаваемой базы наиболее существенным нам представляется качество ее информационного наполнения. В рамках проекта летом 2004 г. была проведена экспедиция (в составе двух лингвистов) в поселок Суломай Байкитского района Эвенкийского АО. Во время экспедиции велось комплексное документирование говора суломайских кетов.

Комплексная документация языка включала сбор информации следующих типов:

- демографическая и социологическая информация;
- социолингвистическая информация;
- лингвистическая информация;
- фольклор;
- информация о традиционной и современной культуре.

Источником социолингвистической информации послужили поселковые похозяйственные книги, анкетирование жителей и интервью. Лингвистическая информация собиралась посредством озвучивания (аудиозаписи) заранее подготовленного словаря-тезауруса, аудио- и видеозаписи текстов (с последующей их расшифровкой с помощью информантов) и аудиозаписи грамматических анкет.

Изучение социолингвистической ситуации является необходимым компонентом работ по сбору материалов и описанию исчезающих языков. С этой целью в пос. Суломае было проведено анкетирование 36,3% всех жителей (44,8% жителей старше 13 лет). Кроме того, для получения более полных данных по поло-возрастному и этническому составу, образовательному цензу, а также для прослеживания динамики изменения этих параметров мы использовали похозяйственные книги за разные периоды.

По данным похозяйственной книги 2002-2004 гг., в Суломае проживает 226 человек, в том числе 128 мужчин и 98 женщин. Следует отметить, что тенденция преобладания мужского населения сохраняется во всех возрастных категориях, за исключением самых старшей группы (1917–1939 гг. рождения), где женщин больше, чем мужчин. Кеты составляют около 70% жителей поселка, остальное население – преимущественно русские (при этом следует учитывать, что в смешанных браках дети обычно записывались кетами).

Всего было проанкетировано 82 жителя Суломае; среди них кеты составили 80,5% (66 чел.), русские 11,0% (9 чел.), кроме того, в выборку

попали двое татар и трое эвенков; 2 человека свою этническую принадлежность указать затруднились.

В качестве родного языка кетский указали 52,4% опрошенных (43 чел.), при этом ответы часто сопровождались следующими ремарками: "Так вообще кетский, но говорю по-русски"; "Родным не владею, по-русски говорю"; "Я здесь вырос среди кетов и считаю себя кетом, поэтому мой родной язык – кетский". Русский язык считают родным 39,0% респондентов (32 чел.) ("Не могу сказать, что кетский, так как абсолютно его не знаю"; "Больше по-русски разговариваю, хороший русский язык" и т.д.). Представители более молодых поколений в качестве родного языка чаще указывают русский. Четверо (4,9%) указали в качестве родного два языка (кетский и русский, кетский и эвенкийский, эвенкийский и русский). Двое респондентов (эвенки) назвали в качестве родного языка эвенкийский, а один затруднился указать свой родной язык ("Кетский или русский, не знаю как сказать, с детства разговаривала по-русски"). Из приведенных ответов видно, что понятие "родной язык" респонденты толкуют по-разному.

По степени владения кетским языком всех респондентов можно разделить на три группы (см. Табл. 1): группа А – владеют кетским языком, группа Б – владеют кетским языком в малой степени (в эту группу были включены респонденты, которые хотя бы понимают и/или могут произнести несколько слов), группа В – не владеют кетским языком. Русским языком владеет все население Суломае. Ограниченно владеют русским языком 5 человек, все они относятся к старшей возрастной группе (1926-1940 гг. рождения). Знание других языков в Суломае распространено мало: одна женщина-кетка 1946 г.р. понимает по-селькупски, четверо в разной степени владеют эвенкийским языком (двое эвенков, кетов), один мужчина-кет 1969 г.р. на бытовом уровне знает якутский язык.

Социолингвистическое обследование Суломаешний раз показало, насколько неотложна наша работа.

Год рождения	Всего (чел.)	Группа А. Владеют языком	Группа Б. Владеют языком в малой степени	Группа Б.1. Немного говорят	Группа Б.2. Понимают отдельные слова	Группа В. Не владеют языком
1917	1	1	0	0	0	0
1926-1939	6	4	1	0	1	1
1940	5	4	1	1	0	0
1941-1950	8	2	2	2	0	4
1951-1960	14	0	14	9	5	0
1961-1970	24	0	21	9	12	3
1971-1980	15	0	7	3	4	8
1981-1990	9	0	2	0	2	7
Всего:	82	11	48	21	27	23

Таблица 1. Распределение жителей Суломае по возрасту и степени владения кетским языком (выборка 82 чел.)

Лингвистический материал, привезенный из Суломай, состоит из материалов для озвученного словаря кетских говоров, включающих некоторую грамматическую информацию, и 10 текстов. Нам удалось сделать довольно полные записи озвученного словаря от четырех носителей суломайского говора (все они женщины). Еще от двух человек, слабо владеющих кетским языком, были записаны отдельные фрагменты словаря. Запись велась на цифровые мини-дисконные диктофоны, дающие высокое качество звучания. Кроме того, мы записали 10 текстов (сказки, легенды, истории жизни, фрагменты шаманских камланий). Запись текстов велась параллельно на мини-дисконные диктофоны и цифровую видеокамеру. Объем аудио-записей, сделанных во время экспедиции, составляет 58 часов, объем видео-записей – 15 часов. Самый полный вариант словаря и самые ценные записи фольклора были сделаны от Ольги Васильевны Латиковой (1917 г. рождения), информантки высочайшей квалификации, работавшей в 1960-1970-е гг. с Е.А. Крейновичем, А.П. Дульзоном, Е.А. Алексеенко. Собственно, кроме нее практически никто в поселке кетского фольклора на кетском языке сегодня не знает. От нее же был записан и ценный этнографический материал. По-сути она одна в Суломае в полной мере является носителем кетской традиционной культуры и нередуцированного кетского языка.

Продолжением экспедиции явилась поездка в г. Кострому для работы с живущим там суломайским кетом (1936 г. рождения), владеющим своим этническим языком. От него была сделана запись единственного в нашей базе достаточно полного мужского варианта озвученного словаря суломайского говора.

Запись материалов для озвученного словаря велась по заранее составленному эталонному словнику, организованному по тезаурусному принципу и содержащему в основном базовую лексику кетского языка. Словник составлен по материалам существующих кетских словарей, прежде всего «Сравнительного словаря енисейских языков» Г.К. Вернера [Werner 2002]. Объем словника – около 1500 единиц. При составлении словника мы ориентировались прежде всего на специфику кетской, а не русской лексики: в состав словника включались названия предметов материальной культуры, представителей фауны и флоры среды обитания кетов, реалий, играющих важную роль в их повседневной жизни, несмотря на то, что в лексической системе русского языка данные единицы нередко относятся к периферии. В состав словника включен 100-словный список Сводеша. При записи материалов для озвученного словаря кетскому информанту-диктору в качестве стимула предъявлялось русское слово. В случае если информант не мог вспомнить кетский

эквивалент слова, ему давался второй стимул – кетское слово.

В настоящее время все записанные во время экспедиции в Суломай и поездки в Кострому аудиоматериалы перенесены с записывающих устройств на компьютер и начата их обработка. Подготовлена первая версия озвученного словаря кетских говоров, содержащая материалы суломайского говора. Словарь содержит 300 словарных статей. Каждое слово произносится пятью дикторами. В словаре дается транскрипция каждого произнесения и в случае, если произносится форма слова, отличная от обычной словарной, дается грамматический комментарий.

При подготовке словаря использовался редактор аудиофайлов SoundForge 7, позволяющий изменять параметры уже записанных звуковых файлов, если они отличаются от требуемых, а при необходимости уменьшать или увеличивать громкость аудиофайлов, выравнивая уровень громкости. С помощью этого редактора было проведено вычленение требуемых звуковых фрагментов из исходной записи сеансов работы с информантами и организация этих фрагментов в отдельные файлы. Каждый звуковой файл словарной базы данных содержит отдельную словоформу или пример употребления словоформы. Как словоформа, так и пример произносятся информантом три раза. Каждый файл имеет уникальное имя.

Звуковые файлы в базе данных хранятся в формате \*wav с параметрами записи 22050 гц, 16 бит моно. Выбор параметров записи обусловлен тем, что используемый для последующего фонетико-акустического анализа звуковой редактор Speech Analyzer (Winccil) работает только с файлами указанных параметров.

На сегодняшний день в базе данных кроме озвученного словаря кетских говоров представлены экспедиционные фотоматериалы, в том числе фотографии всех дикторов и их краткая лингвистическая биография, сведения о языковой ситуации у кетов, а также сведения о кетской фонетической системе. Демонстрационная версия базы данных представлена на Интернет-сайте проекта <http://sdld.narod.ru> и на сайте НИВЦ МГУ по адресу <http://minlang.src.msou.ru>. На следующем этапе в базу будут введены тексты в трех представлениях (видеоряд, звук и транскрипция, а также частичное глоссирование).

#### ***4. Нужна ли наша работа кетам?***

Потенциальными пользователями базы данных являются в первую очередь лингвисты, специализирующиеся в самых разных областях (от лингвистической типологии до общей и экспериментальной фонетики, социо- и этнолингвистики), фольклористы, этнологи, культурологи, создатели аудио-визуальных учебных пособий для вузов и школ; кроме того, база может

быть использована для получения информации о состоянии и функционировании кетского языка и как источник аудио- и видеоматериалов по кетскому языку и фольклору представителями общественных организаций, способствующих сохранению языкового и культурного многообразия и оказывающих поддержку языкам и культурам, находящимся под угрозой исчезновения как внутри России, так и за рубежом, преподавателями и студентами вузов, средствами массовой информации (прежде всего региональными телерадиокомпаниями на территории Сибири и Дальнего Востока).

Но нужна ли наша мультимедийная база кетам? Самое интересное, что, оказывается, нужна. Во всяком случае, такое впечатление у нас сложилось от общения с жителями Суломая. Нам уже неоднократно приходилось убеждаться в том, что работа лингвистов в языковой общности повышает престиж языка у большинства членов этой общности. В еще большей степени повышению интереса к этническому языку (и тем самым повышению престижа этого языка), в том числе и у «безъязыкой» молодежи, способствует соединение языка с новыми, «модными» технологиями - компьютером и тем более Интернетом. Даже если в поселке всего один компьютер, а ни о каком доступе в Интернет речи пока не идет, тот факт, что этнический язык представлен в Интернете, становится для молодых людей источником

положительных эмоций, а таких источников в суровой жизни северных поселков не так уж много.

#### **Список литературы:**

- 1) Казакевич О.А. Мультимедийная база данных исчезающего языка // Труды Международного семинара Диалог'2001 по компьютерной лингвистике и ее приложениям. Том 1. Аксаково, 2001. С. 108-110.
- 2) Казакевич О.А., Самарина И.В, Трушков Д.Л. Озвученный словарь говоров исчезающего языка // Компьютерная лингвистика и интеллектуальные технологии. Труды Международного семинара Диалог'2002 (Протвино, 6-11 июня 2002 г.). Т. 2. Прикладные проблемы. М.: Наука, 2002. С. 245-249.
- 3) Казакевич О.А., Захаров Л.М., Самарина И.В, Трушков Д.Л. Корпусная лингвистика, компьютерная лексикография, мультимедийные технологии и исчезающие языки (Проект завершен – да здравствует новый проект!) // Компьютерная лингвистика и интеллектуальные технологии. Труды международной конференции Диалог'2004 («Верхневолжский», 2-7 июня 2004 г.). М.: Наука, 2004. С. 252-257.
- 4) Werner H. Vergleichendes Wörterbuch der Jenissej-Sprachen. Bd. 1-3. Wiesbaden: Harrassowitz Verlag, 2002.