

УДК 811.2.512
UDC 811.2.512

Кравец Татьяна Владимировна
Амурский государственный университет
г. Благовещенск, Российская Федерация
Tatyana V. Kravets
Amur State University
Blagoveshchensk, Russian Federation
e-mail: t_kravetc@mail.ru

СОЗДАНИЕ МУЛЬТИМЕДИЙНОГО КОРПУСА ЗВУЧАЩЕЙ РЕЧИ
АМУРСКИХ ЭВЕНКОВ: ЦЕЛИ, ЗАДАЧИ, МЕТОДЫ
И ПЕРСПЕКТИВЫ
CREATING MULTIMEDIA CORPUS OF ORAL SPEECH
OF THE AMUR EVENKI: OBJECTIVES, TASKS, METHODS
AND PERSPECTIVES

Аннотация

Статья посвящена разработке мультимедийного корпуса звучащей речи эвенков Приамурья. Подчеркивается, что исчезающие диалекты амурских эвенков впервые представлены в формате базы данных. На сегодняшний день фоноархив образцов речи Амурских эвенков отсутствует. Проект мультимедийного архива эвенкийского языка создается на базе лаборатории экспериментально-фонетических исследований сотрудниками кафедры иностранных языков Амурского государственного университета с использованием программ PRAAT и «Wave Assistant». Мультимедийный фонд основан на результатах звукозаписи диалектов восточных эвенков – селемджинского, желтулакского и зейского говоров. Документация звуковой формы языка эвенков выполняется в цифровом формате. Автор раскрывает цель, задачи, методы и перспективы создания мультимедийного корпуса звучащей речи эвенков. В статье рассматриваются основные этапы и принципы разработки обсуждаемой мультимедийной базы данных.

Abstract

The article deals with creating multimedia corpus of oral speech of the Evenki of the Amur region. It is emphasized that endangered dialects of the Amur Evenki will be presented in the database format for the first time. At present there is no phonetic archive of oral speech samples of the Amur Evenki. The project of the multimedia Evenki archive has started on the basis of the Laboratory of experimental phonetic research and is performed by the staff of the Department of Foreign Languages of the Amur State University. PRAAT and Wave Assistant computer programs are used. The multimedia corpus is based on the results of sound records of the East-Evenki Dialects: Dzheltulak, Selemdzha, and Zeya Evenki dialects. Documenting the sound form of the language of Evenki speakers is carried out in a digital format. Objectives, methods, and perspectives for creating multimedia corpus of oral Evenki speech are revealed. The article considers the main stages and principles of working out the multimedia database.

Ключевые слова: джелтулакский, селемджинский и зейский говоры эвенкийского языка, мультимедийный фонд, звучащая речь, исчезающие диалекты Амурских эвенков.

Keywords: Dzheltulak, Selemdzha and Zeya Evenki dialects, multimedia fund, oral speech, endangered dialects of the Amur Evenki.

Введение

Последние десятилетия, в связи с проблемой возрождения языков и сохранения информации по языковому разнообразию и культурно-историческим ценностям человечества, разработка звуковых фондов на материале исчезающих языков приобрела мировое значение. Документированием языков, находящихся под угрозой исчезновения, занимаются многие мировые организации: Alaska Native Language Archive (ANLA), Asociación Cultural Parola, Association for Cultural Equity, Center for American Indian Languages, Endangered Language Alliance, First Peoples' Cultural Council, Grassroots Indigenous Multimedia, Indigenous Language Institute, Living Tongues Institute for Endangered Languages, Volkswagen Foundation (The DoBES Project), PanLex (The Long Now Foundation), The Endangered Languages Catalogue team at the University of Hawaii at Manoa, The John Carter Brown Library at Brown University, The Language Archive at the Max Planck Institute for Psycholinguistics, The LINGUIST List, The Miami Project, The Rosetta Project (The Long Now Foundation), Transparent Language, Inc., UCLA Phonetics Laboratory, World Oral Literature Project, Summer Institute of Linguistics (SIL) и многие другие. Перечисленные организации являются участниками проекта «Endangered Languages» под руководством компании Google и объединены в «Союз за лингвистическое разнообразие» (Alliance for Linguistic Diversity) [Endangered Languages Project...].

В России большое внимание уделяется языкам коренных малочисленных народов, проживающих в Российской Федерации. Одним из таких языков является эвенкийский, ситуация с которым все чаще вызывает тревогу лингвистов в связи с тенденцией к его утрате и потенциальной угрозой исчезновения звуковой формы этого языка. По данным мирового атласа языков [Atlas..., 2010], 72% эвенков бегло говорят на русском, а не на своём родном языке.

Для решения проблемы сохранения, развития и исследования языков малочисленных народов России предпринимаются определённые шаги. Так, в Бурятском государственном университете создаётся мультимедийная база по баргузинскому, джелтулакскому и каларскому говорам эвенкийского языка [Раднаева, Афанасьева, 2012, с. 91]. В институте филологии сибирского отделения РАН [Институт филологии...] изучаются томмотский, баунтовский, полигусовский говоры эвенкийского языка. Отделение лингвистики Института эволюционной антропологии имени Макса Планка участвует в трёх проектах по изучению тюркских языков Сибири, к сожалению, не связанных напрямую с эвенкийским языком. В поле зрения одного из проектов находится якутский язык, которому не грозит

исчезновение, но который сам угрожает языку более малочисленных народов, живущих на территории Якутии – эвенов и эвенков [Endangered Languages of Indigenous Peoples of Siberia]. Достаточно большой опыт успешной ревитализации языков коренных народностей Сахалина – нивхского, уильтинского, эвенкийского и нанайского языков – накоплен Ассоциацией коренных малочисленных народов Севера, Сибири и Дальнего Востока Российской Федерации (АКМНСС и ДВ РФ). К методам ревитализации, хорошо зарекомендовавшим себя в различных частях мира, относятся: программа «мастер-ученик», образование «языковых гнёзд» и изучение родного (этнического) языка как второго или унаследованного языка [Ассоциация].

Богатый опыт по составлению мультимедийной базы кетского и селькупского языков имеется у сотрудников Лаборатории автоматизированных лексикографических систем Научно-исследовательского вычислительного центра Московского государственного университета имени М. В. Ломоносова под руководством О. А. Казакевич. [Лаборатория...]. В 2012 году на базе мультимедийного эвенкийского архива коллективом лаборатории автоматизированных лексикографических систем НИВЦ МГУ при поддержке РФФИ был создан проект «Мультимедийный размеченный корпус текстов на говорах западных эвенков», в результате которого собрано около 35 тысяч словоупотреблений [Казакевич, Клячко, 2013, с. 292].

Работа по документации языков сегодня во многом опирается на технические и программные средства, которые упрощают транскрипцию и анализ данных, выпуск грамматик и словарей, а также архивацию исследуемых текстов. Труд лингвистов играет все большую роль в проектах оздоровления языков, которые запрашивают языковые сообщества.

Создание электронных информационных ресурсов способствует сохранению и популяризации исчезающих языков и позволяет обеспечить доступ пользователей к информации на родном языке и общению с соотечественниками. По мнению Колетт Гринвальд, работы по созданию подобных переводчиков и словарей для языков народов России ведутся, в основном, усилиями отдельных энтузиастов или в рамках проектов по документированию языков. Крупные компании не участвуют в такой работе, так как из-за использования специфических символов в алфавитах языков народов России работа по созданию электронных переводчиков и словарей является очень сложной с технической точки зрения и нерентабельной [Гринвальд, 2014].

Регистрация и сохранение звуковой формы как никогда актуальны и для эвенкийского языка Амурской области. Изучением эвенкийского языка в области диалектологии занимается Н. Я. Булатова, научный труд которой «Говоры эвенков Амурской области» [Булатова, 1987] является единственной работой по языковой системе эвенкийских говоров Приамурья. В Амурском регионе осуществляются попытки сохранить уникальные говоры Амурских эвенков в печатном виде силами местных учёных-филологов. В 2006 году Научно-методический центр лингвистики и межкультурной

коммуникации БГПУ подготовил «Словарь джелтулакского говора эвенков Амурской области» [Болдырев, 2009, 2010 а] и «Словарь зейского говора эвенков Амурской области» [Болдырев, 2010 б]. В 2008 году подготовлен уникальный двухтомный «Словарь селемджинского говора эвенков Амурской области» В 2013 году вышла из печати первая часть уникального диалектного словаря селемджинского говора эвенков Амурской области [Болдырев et al., 2013].

Тем не менее, системного описания звуковой формы амурских говоров эвенкийского языка – сегментной фонетики и просодики – до сих пор нет. Актуальность задачи по сбору образцов эвенкийской звучащей речи и, в перспективе, формирование звукового архива данного говора, определяется тем, что в настоящее время для говора амурских эвенков, как и для большинства языков народов России, отсутствует корпус представительной звучащей речи, на основе которого возможно осуществление многоуровневого лингвистического исследования.

На сегодняшний день язык амурских эвенков является малоизученным. Фоноархив образцов речи эвенков Амурской области, доступный для широкого круга заинтересованных лиц, не собран. Для тщательного изучения говора Амурских эвенков необходимо осуществить сбор речевого материала путём звукозаписи, документирования звуковой формы исследуемого языка с помощью компьютерных технологий.

Цели проекта

В данном направлении работает лаборатория экспериментально-фонетических исследований на базе кафедры иностранных языков Амурского государственного университета, сотрудники которой занимаются разработкой представительного корпуса звучащей речи для лингвистического исследования джелтулакского, селемджинского и зейского говоров эвенкийского языка в Амурском регионе. Основной целью проекта является документирование звуковой формы эвенкийского языка в цифровом виде для сохранения звуковой стороны говоров эвенков Приамурья, проживающих в пяти национальных селах, расположенных на территории Тындинского, Джелтулакского, Нюкжинского, Зейско-Учурского, Верхне-Селемджинского, Верхне-Буреинского и Мазановского районов.

Конечным результатом работы над проектом «Мультимедийный корпус звучащей речи эвенков Приамурья» будет программный продукт на цифровом носителе, содержащий 30-часовой фоно и видеоархив звуковых образцов речи эвенков Приамурья, систематизированный и аннотированный в компьютерном пакете программ по обработке звукового сигнала PRAAT [Морозова, 2014, с. 91].

Описание корпуса

Мультимедийный фонд представляет собой фонотеку изолированно произнесенных звуков, фонотеки слогов и слов, а также образцы эвенкийской речи на материале различных говоров в виде отдельных файлов, со-

проводящихся словарным и текстовым описанием. В корпусе будут представлены: 1) звуковой словарь амурских говоров эвенкийского языка (лексический блок: изолированные слова, в произнесении носителей селенгинского говора (ивановская и норская группы), зейского и желтулакского говоров эвенкийского языка; 2) фольклорные тексты и бытовые рассказы в звуковом и текстовом форматах: нарративный (монологический и диалогический) блок, включающий эпосы, охотничьи песни, бытовые рассказы и воспоминания по следующим темам: о себе (автобиографии), о семье, о родственниках, о жизненном укладе, об эвенкийском образе жизни (о праздниках, обрядах, традициях), об эвенкийской культуре, о традиционных эвенкийских видах деятельности, о флоре и фауне, об эвенкийской кухне и так далее.

На материале корпуса проводится экспериментально-фонетический анализ с описанием акустических свойств звуковых единиц. Звукозапись осуществляется в условиях лаборатории экспериментально-фонетических исследований сотрудниками кафедры иностранных языков Амурского государственного университета с использованием программ PRAAT и Wave Assistant. Записи содержат образцы спонтанной речи носителей эвенкийского языка.

Целевой группой пользователей, для которой планируется разработка мультимедийного корпуса, является эвенкийское сообщество как Амурской области, так и других регионов проживания эвенков (в том числе эвенки Китая (провинции Хэйлунцзян (уезд Нэхэ) и Эвенкийского автономного хошуна района Внутренняя Монголия) и эвенки Монголии (Селенгинского аймака)) [Морозова, 2014, с. 91–92].

Задачи

Для достижения заявленной темы планируется следующее.

1. Составление представительного списка изолированных слов, включающего реализацию в сильной позиции всего инвентаря гласных и согласных фонем эвенкийского языка.

2. Составление списка вопросов на бытовые темы для получения образцов монологической спонтанной эвенкийской речи.

3. Составление анкеты, включающей перечень «лингвистически значимых» признаков: пол, возраст, уровень образования, знание языков, род занятий, место рождения и место наиболее длительного проживания и другие.

4. Обработка полученного звукового материала при помощи компьютерных программ «PRAAT», Wave Assistant для составления звуковой базы данных речи носителей эвенкийского языка.

Методы

В процессе работы над созданием мультимедийной базы звучащей речи эвенков будут применяться такие методы исследования как 1) лингвистические методы определения представительности языкового материала; 2) методы и подходы полевой лингвистики (интервьюирование, анкетиро-

вание и др.); 3) методы речевых технологий (комплекс методов звукозаписи, обработки и анализа речевого сигнала); 4) методы прикладной и корпусной лингвистики.

Перспективы

Мультимедийный корпус образцов эвенкийской речи послужит гарантией сохранения для последующих поколений находящейся на грани исчезновения живой речи носителей селемджинского, джелтулакского и зейского говоров эвенкийского языка.

Материалы мультимедийного фонда могут быть использованы как для лингвистического анализа звуковых единиц эвенкийского языка, так и для образовательных целей, в том числе – разработки словарей и учебников по эвенкийскому языку. Аутентичные аудио- и видеозаписи речи эвенков Приамурья станут неотъемлемой частью учебного процесса обучения эвенкийскому языку и культуре в общеобразовательных школах в районах компактного проживания эвенкийского этноса.

Представительный звуковой материал может быть использован для проведения многоуровневых лингвистических исследований, для подготовки справочной литературы и обучающих мультимедийных продуктов, а также для решения задач современных речевых технологий.

Электронный формат репрезентации звучащей речи эвенков окажет положительное влияние на развитие мотивации и поддержание интереса у эвенкийской молодежи к изучению родной речи.

Данные мультимедийной базы найдут своё применение в исследовательской работе ведущих научных центров нашей страны (Научно-исследовательский вычислительный центр МГУ им. Ломоносова, филологический факультет СПбГУ, РГПУ им. Герцена, Институт лингвистических исследований) и за рубежом: (институт Макса Планка (Германия), Версальский университет (Франция), Сеульский и Чонбукский национальные университеты (Южная Корея) [Морозова, 2014, с. 92].

Таким образом, разработка корпуса звучащей речи Амурских эвенков на материале находящихся под угрозой исчезновения селемджинского, джелтулакского и зейского говоров эвенкийского языка направлена на решение глобальной проблемы по сохранению, развитию и исследованию языков малочисленных народов России.

Список литературы

1. Болдырев, Б. В. Словарь селемджинского говора эвенков Амурской области [Текст] / Б. В. Болдырев, Г. В. Быкова, Л. А. Соловьева. – Благовещенск: Изд-во БГПУ, 2013. – Ч. 1. – 481 с.
2. Болдырев, Б. В. Словарь джелтулакского говора эвенков Амурской области. Ч. 1 [Текст] / Б. В. Болдырев. – Благовещенск : Изд-во БГПУ, 2009. – 608 с.
3. Болдырев, Б. В. Словарь джелтулакского говора эвенков Амурской области. Ч. 2 [Текст] / Б. В. Болдырев. – Благовещенск : Изд-во БГПУ, 2010 а. – 236 с.
4. Болдырев, Б. В. Словарь зейского говора эвенков Амурской области. [Текст] / Б. В. Болдырев. – Благовещенск : Изд-во БГПУ, 2010 б. – 426 с.

5. Булатова, Н. Я. Говоры эвенков Амурской области [Текст] / Н. Я. Булатова. – Ленинград : Наука, 1987. – 168 с.
6. Гринвальд, К. Исчезающие языки // ИНОСМИ.РУ [Электронный ресурс] / К. Гринвальд. – Режим доступа : <http://inosmi.ru/world/20140125/216849905.html> (дата обращения: 29.02.2016 г.)
7. Груздева, Е. Сохранение и возрождение языков коренных народов Сахалина [Электронный ресурс] Е. Груздева, Ю. Янхунен // Ассоциация коренных малочисленных народов Севера, Сибири и Дальнего Востока Российской Федерации (АКМНСС и ДВ РФ) Russian (Association of Indigenous Peoples of the North, Siberia and Far East (RAIPON)). – Режим доступа : <http://www.raipon.info/info/actual/1294> (дата обращения: 05.03.2016 г.)
8. Институт филологии сибирского отделения РАН [Электронный ресурс]. – Режим доступа : <http://www.philology.nsc.ru/> (дата обращения: 29.02.2016 г.)
9. Казакевич, О. А. Создание мультимедийного аннотированного корпуса текстов как исследовательская процедура [Электронный ресурс] / О. А. Казакевич, Е. Л. Клячко // Труды Международной конференции «Корпусная лингвистика-2013». СПб. : С.-Петербургский гос. ун-т, Филологический фак., 2013. – С. 292–300. – Режим доступа : corpora.phil.spbu.ru/Works2013/Казакевич.pdf (дата обращения: 29.02.2016 г.)
10. Лаборатория автоматизированных лексикографических систем Научно-исследовательского вычислительного центра Московского государственного университета имени М. В. Ломоносова. [Электронный ресурс]. – Режим доступа : <http://www.lcl.srcc.msu.ru/staff.htm> (дата обращения: 01.03.2016 г.)
11. Морозова, О. Н. «Мультимедийный корпус звучащей речи эвенков Приамурья» в процессе обучения эвенкийскому языку [Текст] / О. Н. Морозова // Лингвистические и дидактические аспекты в преподавании иностранных языков : сб. науч. статей науч.-практ. конф. с междунар. участием. – Петрозаводск : Изд-во ПетрГУ, 2014. – С. 91–93.
12. Отделение лингвистики Института эволюционной антропологии им. Макса Планка (Endangered Languages of Indigenous Peoples of Siberia) [Electronic Resource]. – URL: http://lingsib.iea.ras.ru/ru/round_table/papers/pakendorf.shtml. (дата обращения: 06.02.2016 г.)
13. Раднаева, Л. Д., Афанасьева, Е.Ф. Разработка звукового фонда современного эвенкийского языка в Бурятском государственном университете [Электронный ресурс] / Л. Д. Раднаева, Е. Ф. Афанасьева // Вестник Бурятского государственного университета. – 2012. – № 10 – С. 90–94. – Режим доступа : <http://cyberleninka.ru/article/n/razrabotka-zvukovogo-fonda-sovremennogo-evenkiyskogo-yazyka-v-buryatskom-gosudarstvennom-universitete> (дата обращения: 29.02.2016 г.)
14. Atlas of the World's Languages in Danger [Electronic Resource] / Christopher Moseley (ed.). – UNESCO, 2010. – URL : <http://www.unesco.org/culture/en/endangeredlanguages/atlas> (дата обращения: 29.02.2016 г.)
15. Endangered Languages Project [Electronic Resource]. – URL : <http://www.endangeredlanguages.com/lang/1186> (дата обращения: 02.03.2016 г.)

References

1. Boldyrev, B. V., Bykova, G. V., Solov'eva, L. A. (2013). *Slovar' selemdzhinskogo govora evenkov Amurskoy oblasti* [Dictionary of the Selemdzha Evenki dialect of the Amur region]. Blagoveshchensk: Blagoveshchensk State Pedagogical University Press.
2. Boldyrev, B. V. (2009). *Slovar' dzheltulakskogo govora evenkov Amurskoy oblasti. Chast' 1.* [Dictionary of the Dzheltulak Evenki dialect of the Amur region. Part 1]. Blagoveshchensk: Blagoveshchensk State Pedagogical University Press.
3. Boldyrev, B. V. (2010 a). *Slovar' dzheltulakskogo govora evenkov Amurskoy oblasti. Chast' 2.* [Dictionary of the Dzheltulak Evenki dialect of the Amur region. Part 2]. Blagoveshchensk: Blagoveshchensk State Pedagogical University Press.
4. Boldyrev, B. V. (2010 b). *Slovar' zeyskogo govora evenkov Amurskoy oblasti.* [Dictionary of the Zeya Evenki dialect of the Amur region]. Blagoveshchensk: Blagoveshchensk State Pedagogical University Press.
5. Bulatova, N. Ya. (1987). *Govory evenkov Amurskoy oblasti* [Evenki Dialects of the Amur Region]. Leningrad: Nauka.
6. Grinevald, C. Les langues menaces. *INOSMI.RU*. Retrieved 29.02.2016 from <http://inosmi.ru/world/20140125/216849905.html>.
7. Gruzdeva, E., Yankhunen Yu. Sokhranenie i vrozozhdenie yazykov korennykh narodov Sakhalina [Preservation and revival of the Languages of native peoples of Sakhalin]. *Assotsiatsiya korennykh malochislennykh narodov Severa, Sibiri i Dal'nego Vostoka Rossiyskoy Federatsii (AKMNSS i DV RF)* [Association of Native Peoples of the North, Siberia and Far East of Russian Federation] (RAIPON)), Retrieved 05.03.2016 from <http://www.raipon.info/info/actual/1294/>.
8. Institut filologii sibirskogo otdeleniya RAN [Institute of Philology of the Siberian Branch of the Russian Academy of Sciences]. Retrieved 29.02.2016 from <http://www.philology.nsc.ru/>.
9. Kazakevich, O. A., Klyachko, E. L. (2013). Sozdanie mul'timediynogo annotirovannogo korpusa tekstov kak issledovatel'skaya protsedura [Creating multimedia annotated corpus of texts as a research procedure]. *Proc. International Conference «Corpus Linguistics»*. St-Petersburg, 292–300. Retrieved 29.02.2016 from corpora.phil.spbu.ru/Works2013/Kazakevich.pdf.
10. Laboratoriya avtomatizirovannykh leksikograficheskikh sistem Nauchno-issledovatel'skogo vychislitel'nogo tsentra Moskovskogo gosudarstvennogo universiteta imeni M. V. Lomonosova. [Laboratory of automatic Lexicography systems of the Research Computing Center of Lomonosov Moscow State University]. Retrieved 01.03.2016 from <http://www.lcl.srcc.msu.ru/staff.htm>.
11. Morozova, O. N. (2014). Mul'timediynyy korpus zvuchashchey rechi evenkov Priamur'ya v protsesse obucheniya evenkiyskomu yazyku [Multimedia corpus of oral Evenki speech of the Amur region in the Evenki language learning]. *Proc. Scientific-Practical Conference with International Participation*. Petrozavodsk, 91–93.
12. Department of Linguistics of Max Planck Institute for Evolutionary Anthropology. (Endangered Languages of Indigenous Peoples of Siberia), Retrieved 06.02.2016 from http://lingsib.iea.ras.ru/ru/round_table/papers/pakendorf.shtml.

13. Radnaeva, L. D., Afanas'eva, E. F. (2012). Razrabotka zvukovogo fonda sovremennogo evenkiyskoggo yazyka v Buryatskom gosudarstvennom universitete [The development of a oral speech corpus of the modern Evenki language in the Buryat State University]. *Vestnik Buryatskogo gosudarstvennogo universiteta* [Bulletin of the Buryat State University]. Retrieved 29.02.2016 from <http://cyberleninka.ru/article/n/razrabotka-zvukovogo-fonda-sovremennogo-evenkiyskogo-yazyka-v-buryatskom-gosudarstvennom-universitete>.
14. Atlas of the World's Languages in Danger / Christopher Moseley (2010). UNESCO. Retrieved 29.02.2016 from <http://www.unesco.org/culture/en/endangeredlanguages/atlas>.
15. Endangered Languages Project. Retrieved 02.03.2016 from <http://www.endangeredlanguages.com/lang/1186>.