

## ТЕКСТОВЫЕ КОРПУСЫ

УДК: 811.512.212, 81'342.2

**С. В. Андросова, О. Н. Морозова, Т. В. Кравец**

*Амурский государственный университет*

### **Корпус звучащей эвенкийской речи\***

В эпоху глобализации при господстве информационных технологий проблема экологии языков стоит особенно остро. Эвенкийский язык – один из языков, находящихся под угрозой исчезновения. Эвенкийско-русский билингвизм для подавляющего большинства эвенков Амурской области уже стал русско-эвенкийским, а количество эвенков, свободно говорящих на родном языке, мизерно. Информационные технологии как часть благ цивилизации, запустившие этот деструктивный процесс, имеют прекрасный потенциал его замедлить и, возможно, даже обернуть вспять через создание электронных аннотированных звуковых корпусов свободного доступа. Особый статус эвенкийского языка и недоступность или отсутствие звуковых подробно аннотированных материалов по говорам его восточного наречия послужили импульсом для начала работы над соответствующим корпусом.

Нами создана информационная система (веб-приложение написано на языке Ruby с использованием инструментария Rails) свободного доступа к базе данных по селемджинскому, джелтулакскому и зейскому говорам. Интерфейс базы данных представлен стандартными страницами: главная, новости, корпус, о проекте, ресурсы и контакты. На главную страницу систематически выкладываются новости о пополнении базы новыми речевыми образцами и о мероприятиях, связанных с эвенкийским и близкородственным ороchonским языками, в частности об олимпиаде по указанным языкам, проводимую на базе Амурского государственного университета. На странице «Корпус», по настоянию самих дикторов, дается необходимая персональная информация (ФИО, профессия, род, говор). Корпус предназначен для широкого круга пользователей – от носителей языка до методистов, ведущих обучение данному языку, и исследователей-лингвистов, занимающихся изучением данного языка. Нами предложены технические решения ряда неоднозначных вопросов, связанных с особенностями сегментации и транскрибирования. Одна из таких проблем – наличие в ряде случаев поствокального

---

\* Исследование поддержано грантом РФФИ 17-04-12004в.

*Андросова Светлана Викторовна* – доктор филологических наук, доцент, профессор кафедры иностранных языков Амурского государственного университета.

Контактная информация: ул. Игнатьевское шоссе, д. 21, г. Благовещенск, 675028, Российская Федерация, e-mail: androsova\_s@mail.ru, тел.: (4162)394585.

*Морозова Ольга Николаевна* – кандидат филологических наук, доцент кафедры иностранных языков Амурского государственного университета, заведующая кафедрой.

Контактная информация: ул. Игнатьевское шоссе, д. 21, г. Благовещенск, 675028, Российская Федерация, e-mail: morozova\_olga06@mail.ru, тел.: (4162)394585.

*Кравец Татьяна Владимировна* – кандидат филологических наук, доцент кафедры иностранных языков ФГБОУ ВО Амурский государственный университет.

Контактная информация: ул. Игнатьевское шоссе, д. 21, г. Благовещенск, 675028, Российская Федерация, e-mail: t\_kravets@mail.ru, тел: (4162)394585.

турбулентного и импульсного шума на переходном участке от гласного к согласному. В настоящее время ведется активная работа над созданием аналогичного корпуса звучащей ороchonской речи.

*Ключевые слова:* исчезающий язык, тип пользователя, аннотирование, транскрипция, сегментный уровень, супрасегментный уровень

## 1. Введение

Эпоха изоляции, которая в течение многих веков поддерживала языковое многообразие, канула в Лету. Ей на смену пришла эпоха глобализации, до зубов вооруженная информационными технологиями. В этой реальности проблема экологии языков стоит особенно остро. Ни для кого не секрет, что примерно каждые две недели на нашей планете исчезает один язык. Еще совсем недавно лингвистическое многообразие было представлено 6800 языками мира, теперь их уже на сотню меньше. Эвенкийский язык – один из тех, которые грозят исчезнуть через два поколения. Поголовный билингвизм, многообразие говоров при своеобразии письменной формы и сложностях стандартизации на лексическом и фонетическом уровнях языка – факторы, неуклонно ускоряющие данный процесс. Численность эвенкийского этноса катастрофически сокращается. Эвенкийско-русский билингвизм для подавляющего большинства эвенков уже стал русско-эвенкийским, а количество эвенков, свободно говорящих на родном языке, мизерно. Такая ситуация типична для Амурской области российского Дальнего востока (см. подробнее об этом в работе О. Н. Морозовой [Морозова, 2014]). Количество представителей данного этноса, которые учились говорить на эвенкийском языке с младенчества в кругу семьи и далее общались преимущественно на нём, ничтожно мало по сравнению с числом тех, которые изучали свой национальный язык лишь со школьной скамьи как иностранный с довольно ограниченным количеством часов, отведённых на освоение данной учебной дисциплины. Языковая ситуация в местах компактного проживания амурских эвенков носит диспропорциональный, несбалансированный характер, выражающийся в абсолютном функциональном доминировании русского языка [Процукович, 2015, с. 88-89]. Возможно, повышение социального статуса и благосостояния, важное для любого человека в принципе, многим эвенкам видится в том, чтобы стать частью русскоязычной цивилизации. Вместе с тем, эвенки ощущают, что блага её сомнительны, если цена этому – забвение национального языка.

Информационные технологии как часть благ цивилизации, запустившие этот деструктивный процесс, имеют прекрасный потенциал его замедлить и, возможно, даже обернуть вспять. Создание звуковых корпусов для исчезающих языков можно по праву назвать первостепенной задачей современной прикладной лингвистики. Эта задача активно выполняется как ведущими отечественными, так и зарубежными специалистами. Достаточно обратиться к проектам по звуковому корпусу эвенкийского языка, разрабатываемым российскими специалистами [Афанасьева, Раднаева, 2011; Казакевич, 2013, 2016]. Между тем, общедоступный подробно аннотированный корпус говоров восточного наречия эвенкийского языка отсутствует. Таким образом, особый статус эвенкийского языка и недоступность или отсутствие звуковых подробно аннотированных материалов по говорам его восточного наречия побудили нас начать работу над соответствующим корпусом.

## 2. Корпус звучащей эвенкийской речи (аннотированный)

### 2.1. Технические решения и интерфейс

Нами создана информационная система (интерфейс написан на языке Ruby) свободного доступа к базе данных по говорам восточного наречия эвенков, проживающих на территории Амурской области [Речевой корпус ..., 2016-2017] (более подробно с техническими решениями можно ознакомиться в нашей предыдущей работе [Морозова и др., 2017]).

К настоящему моменту интерфейс базы данных представлен стандартными страницами: главная, новости, корпус, о проекте, ресурсы и контакты (см. рис. 1).

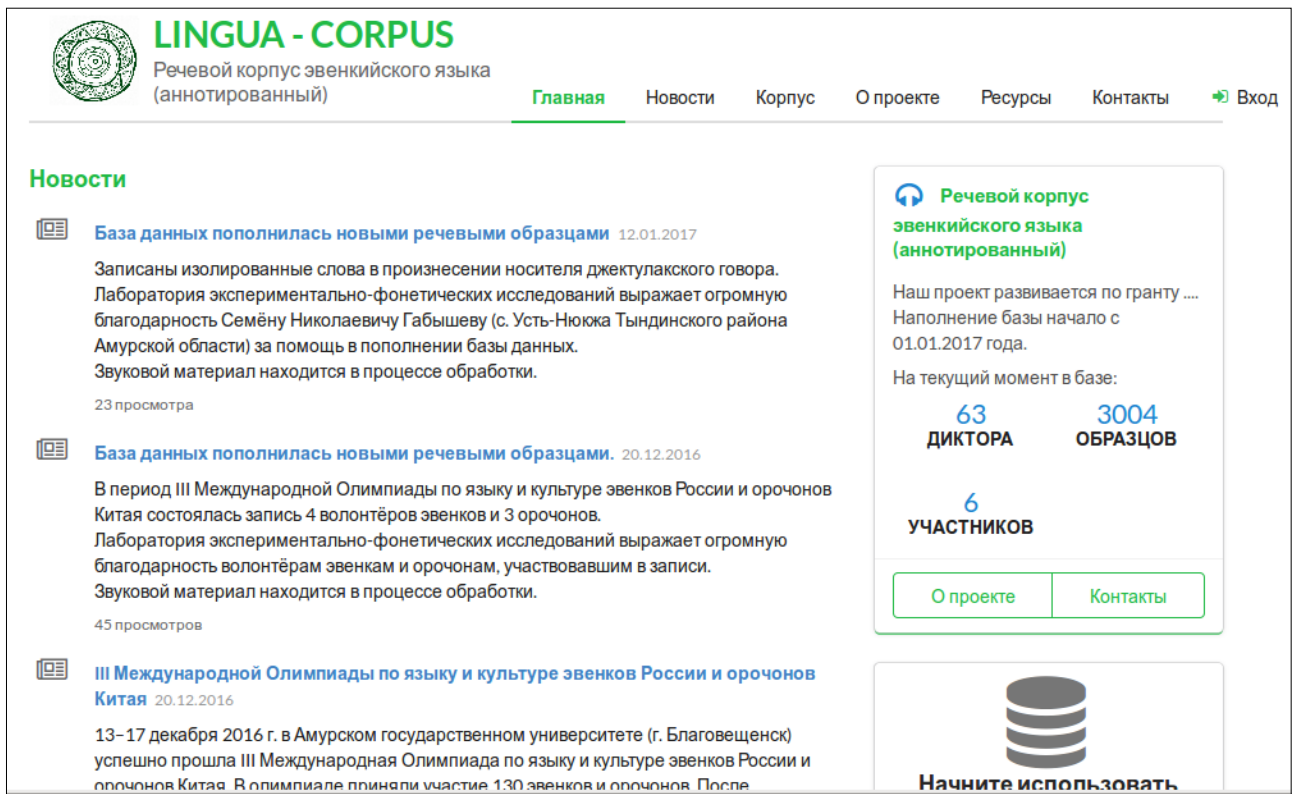


Рис. 1. Главная страница Корпуса  
Fig. 1. Corpus main page

На главную страницу систематически выкладываются новости о пополнении базы новыми речевыми образцами и о мероприятиях, связанных с эвенкийским и близкородственным ороchonским языками. В частности, отражены данные по ежегодной олимпиаде по языку эвенков России и близкородственному языку ороchonов КНР, которая сравнительно недавно приобрела международный статус. Здесь же представлена информация о количестве задействованных дикторов и полученных от них речевых образцов. Следует отметить, что активное пополнение базы данных осуществляется именно во время проведения данной олимпиады, а не только в ходе полевых экспедиций.

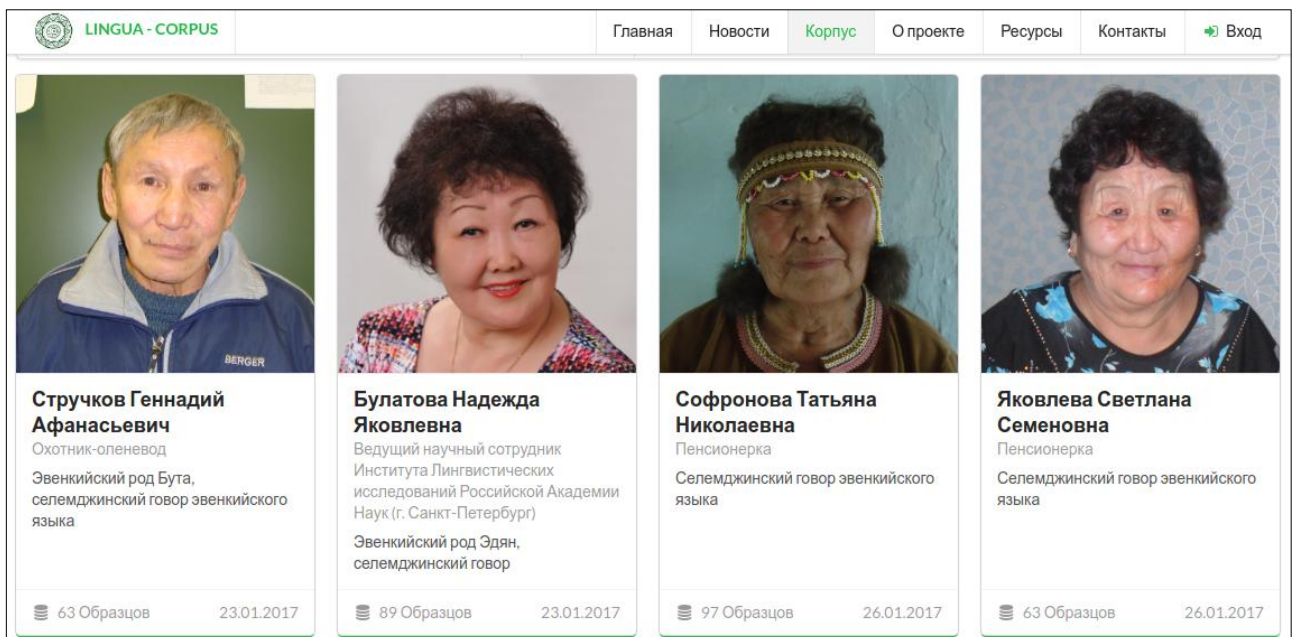


Рис. 2. Карточки дикторов  
Fig. 2. Subjects' pages

На странице «Корпус» представлены карточки дикторов. Обычно участие дикторов в речевых корпусах анонимно. Однако наши дикторы настаивали на внесение персональных данных в корпус и на том, чтобы эти данные были общедоступны. Часть дикторов также не возражали против размещения их фотографий (см. рис. 2). В каждой карточке приведены ФИО, профессия диктора, название эвенкийского рода и говор, количество аннотированных образцов на определенную дату.

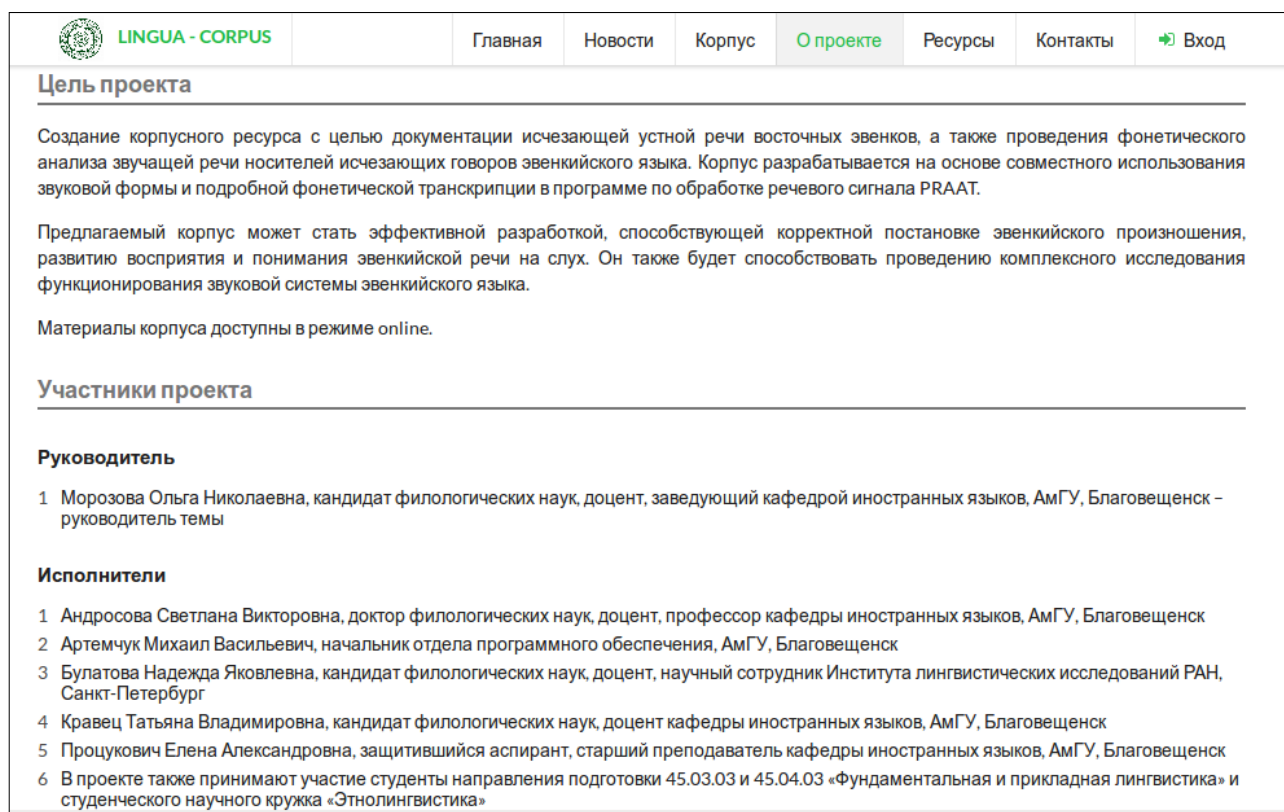
Имеется специальная страница «О проекте», на которой кратко представлены его цели и задачи, указаны используемые говоры, типы речевого и текстового материала и типы речевого сигнала, дана информация об участниках проекта (см. рис. 3-4).



The screenshot shows the website interface for 'LINGUA - CORPUS'. The header includes the logo, the title 'LINGUA - CORPUS', and the subtitle 'Речевой корпус эвенкийского языка (аннотированный)'. A navigation menu contains links for 'Главная', 'Новости', 'Корпус', 'О проекте' (highlighted), 'Ресурсы', 'Контакты', and 'Вход'. The main content area is titled 'Речевой корпус эвенкийского языка (аннотированный)' and 'О проекте'. The text describes the project's goals, mentioning the revival of languages and the development of a corpus for the disappearing dialects of the eastern Evenki language. It lists the types of speech material, text material, and speech signals included in the corpus.

Рис. 3. Общая информация о проекте

Fig. 3. General project info



The screenshot shows the 'Цель проекта' (Project goal) and 'Участники проекта' (Project participants) sections. The 'Цель проекта' section describes the goal of creating a corpus resource for documenting disappearing oral speech of eastern Evenki languages and conducting phonetic analysis. The 'Участники проекта' section lists the project leader, Olga Nikolaevna Morozova, and six executors, including Svetlana Viktorovna Androsova, Mikhail Vasilevich Artemchuk, and others.

Рис. 4. Цель и участники проекта

Fig. 4. Aim and participants

Важно подчеркнуть, что в корпусе представлен весьма разноплановый материал. Во-первых, это начитанные изолированные слова в трехкратном повторении, что в определенной степени имитирует просодию фразы с конечной деклинацией. Имеются и более мелкие единицы – начитанные слоги. Во-вторых, это начитанные одно- и многосинтагменные фразы, часть из которых организована по типу довольно популярных в экспериментальной лингвистике рамочных конструкций. В-третьих, это связанные тексты, как начитанные так и спонтанные, как диалогического, так и монологического характера. Наконец, имеются и образцы эвенкийского устного народного творчества (сказания, героический эпос), произведенные теми немногими эвенками-старожилами, которые все еще прекрасно владеют языком своих предков.

Сегментация записанного материала производится студентами направления подготовки «Фундаментальная и прикладная лингвистика», а также аспирантами направления «Языкознание и литературоведение» профиля «Теория языка» в ЛЭФИ АмГУ под руководством высококвалифицированных специалистов по акустическому анализу речевого сигнала. Сегментация производилась с опорой на стандартные принципы, изложенные, в частности, в работе П. А. Скредина [Скредин, 1999].

Корпус предназначен для широкого круга пользователей – от носителей языка до методистов, ведущих обучение данному языку, и исследователей-лингвистов, занимающихся изучением данного языка. В помощь пользователям имеется страница с ресурсами свободного доступа (см. рис. 5), куда помещены две мультиплатформенные (для операционных систем Windows, Linux и Mac) программы для обработки и анализа звукового сигнала. Здесь же в ресурсах для начинающих специалистов, интересующихся акустическим анализом, предлагается соответствующее пособие [Андросова, 2014], в котором кратко изложены основы работы в PRAAT и базовые акустические сведения с анализом примеров осциллографической и спектрографической картины конкретных звуков – согласных и гласных – и некоторых их модификаций. Многие эти сведения носят общезвонетический характер и могут быть применены и для анализа реализаций эвенкийских фонетических единиц.



**Рис. 5.** Ресурсы  
**Fig. 5.** Resources

## 2.2. Проблемы аннотирования и их решения

### 2.2.1. Общие принципы

Один из важных, на наш взгляд, общих принципов – оптимальное количество уровней в аннотации. Оно естественным образом определяется характером материала – меньшее количество уровней для более мелких единиц и большее – для более крупных (ср. рис. 6 с пятью интервальными уровнями и рис. 7 – с шестью). На уровнях с фонемного по аллофонный использовалась транскрипция разной степени детализации. Были использованы знаки международного фонетического алфавита (далее – МФА).

### 2.2.2. Просодия

При аннотировании разметка просодических особенностей заключалась в обозначении участков пауз (на данный момент – без указания типа паузы) и упрощенной фиксации мелодического компонента. Словесное ударение не отмечалось ввиду того, что авторы настоящей статьи, осознавая всю дискуссионность вопроса (см., напр., возможность музыкального характера эвенкийского ударения, выявленная на материале дисиллабов и трисиллабов, начитанных с интонацией перечисления [Андреева, 2001]), придерживаются мнения об отсутствии такового в эвенкийском языке. Основанием этому служат результаты недавно проведенного нами эксперимента на материале рамочных конструкций и списка изолированных слов с целью выявить отличия в реализации фонетических характеристик ударности в изолированно произнесенном слове и слове в потоке речи в зависимости от структуры предложения [Морозова и др., 2014, с. 89-91]. Анализ мелодических характеристик показал наличие разных типов движения основного тона на одном и том же слове, но реализованном на начальном / срединном / конечном участках фразы и, таким образом, не подтвердил наличие в эвенкийском языке музыкального ударения. Возможно, такая разница с ранее полученными результатами Т. Е. Андреевой обусловлена, во-первых, разницей говоров, во-вторых, особенностью организации экспериментального материала. Все же, представляется необходимым говорить с большой осторожностью о тоновой природе словесного ударения в эвенкийском языке и сходстве типа акцентуации со слогоморфемными языками Юго-Восточной Азии и языками Африки (в которых морфема и слог не совпадают).

В связи с вышеизложенным, производилось только транскрибирование мелодического компонента. Поскольку набор транскрипционных знаков МФА в PRAAT для обозначения особенностей движения основного тона (далее – ОТ) весьма ограничен, пришлось прибегнуть к тире и разнонаправленным слэшам для обозначения соответственно ровного и восходящего и нисходящего тонов. На данном этапе не ясно, возможна ли организация разметки ОТ, единая для всех речевых образцов. Так, для троекратно повторенных изолированных слов оказалось достаточным указать с помощью разнонаправленных знаков «\» и «/» нисходящее или восходящее движение ОТ на участке, соответствующем слову (см. разграничительные метки и знаки для ОТ на уровне «intonation» на рис. 6 и 7). Однако такая стратегия оказалась неподходящей для материала героического эпоса, прежде всего, в силу специфики певческой речи: разные движения ОТ на разных слогах в пределах одного слова и нередкие случаи разнонаправленного движения ОТ в пределах одного слога. Поэтому для таких речевых образцов принято решение производить мелодическую разметку не на участках слов, а на участках слогов (см. разграничительные метки и знаки для ОТ на уровне «intonation» на рис. 7).

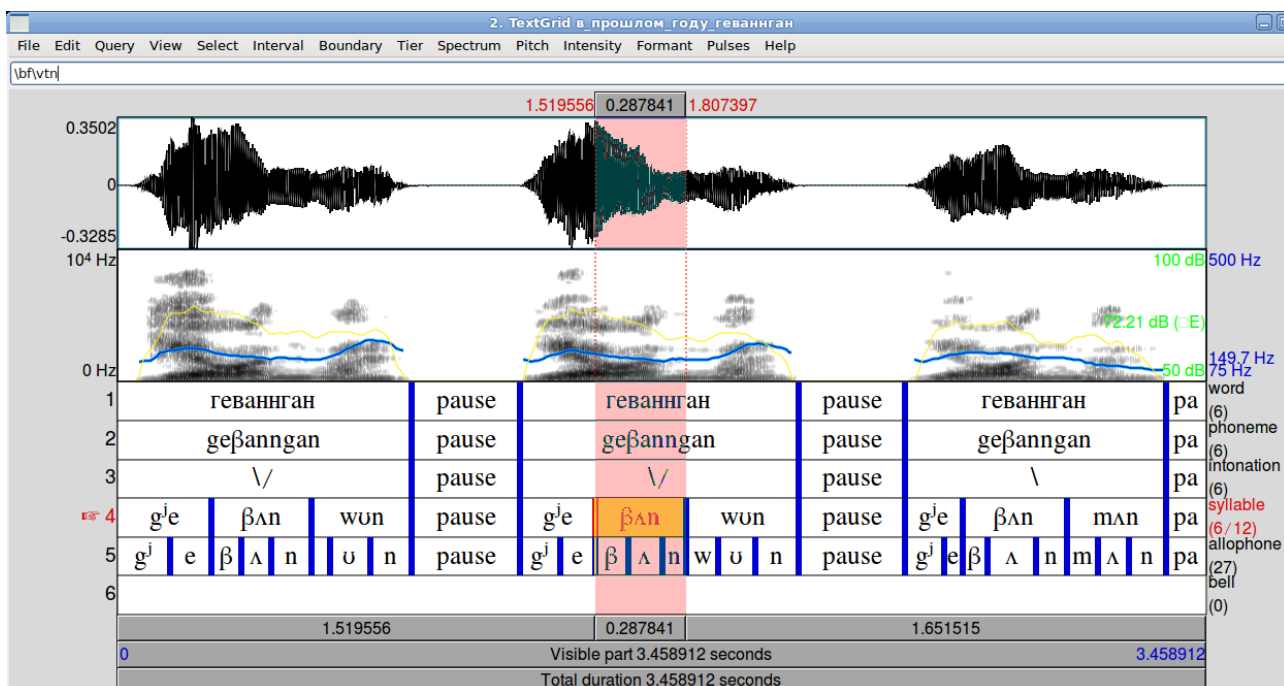


Рис. 6. Аннотация слова *геваннган* (= в прошлом году)

Fig. 6. Annotation of the word *gevanngan* (= last year)

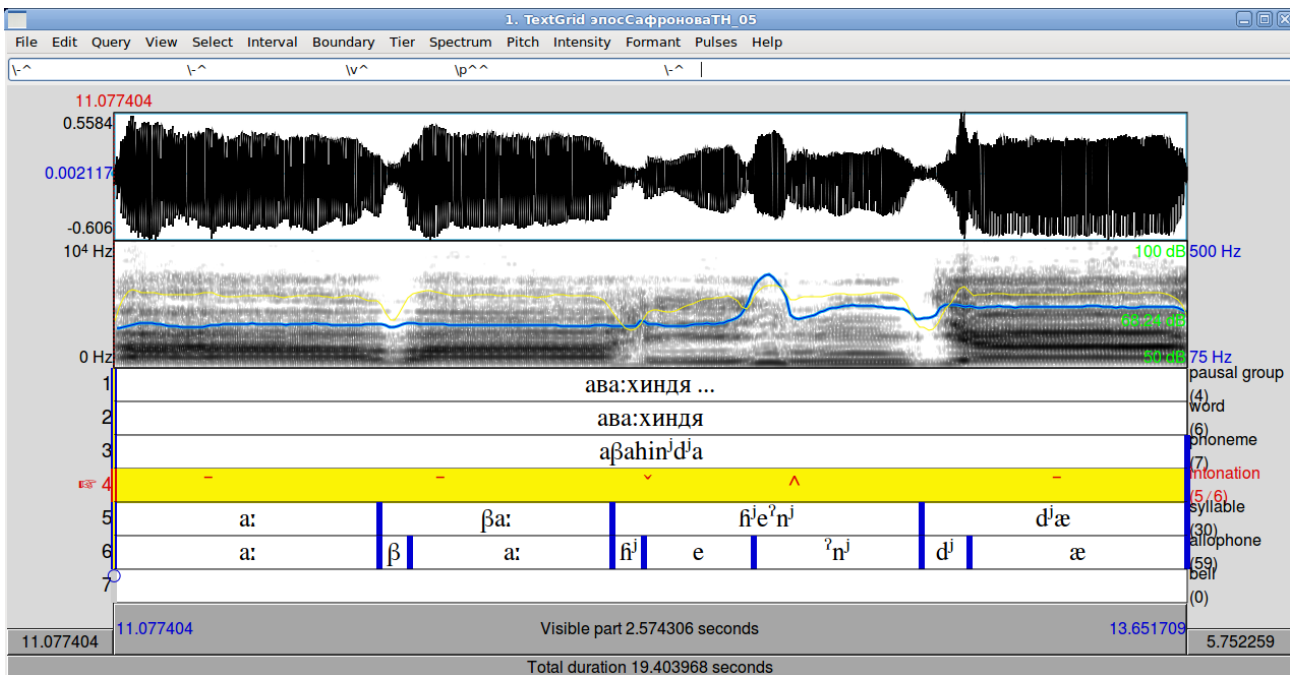


Рис. 7. Фрагмент аннотации героического эпоса

Fig. 7. A piece of heroic epic poem annotation

### 2.2.3. Сегментные единицы

Проблемы аннотирования на сегментных уровнях (фонемный и аллофонный) связаны с особенностями формирования письменной формы языка и перипетий становления эвенкийской орфоэпии. Одна из таких проблем – орфография и каноническая фонемная транскрипция на основе экающего произношения с предыдущим твердым согласным в примерах типа *сиксэ*, вступающие в противоречие с акающим произношением с предыдущим мягким согласным: /s¹kxε/ в западных говорах vs /s¹kx¹æ/. Сложно надеяться на вариантную орфографию – *сиксэ* vs *сикс¹я* – однако остается вопрос о вариантности фонемного или аллофонного состава этого и подобных слов. Помимо этого, при трехкратном повторении изолированных слов диктор иногда мог реализовывать их с несколько отличающимся фонемным составом (ср. три реализации последнего слога в слове *геванган* на рис. 6). Возможное объяснение этому заключается в том, что дикторы уже недостаточно хорошо помнят некоторые слова своего национального языка.

Другая проблема, с которой пришлось столкнуться – наличие турбулентного и импульсного шума, возможно, гортанного, в целом ряде случаев при переходе от гласного к следующему согласному (см. участки, отграниченные метками слева и справа на рис. 8–10). Данная особенность была отмечена в речи всех записанных дикторов. В связи с обнаруженным шумом в перспективе следует поднять вопрос об особенностях фонетического слога в изучаемых говорах восточного наречия эвенкийского языка. На данном этапе решено считать указанный шум окончанием гласного и, возможно, признаком плотного примыкания. Необходимо выяснить, насколько систематично он реализуется и какими факторами определяется его наличие или отсутствие.

Кроме того, пришлось столкнуться с совершенно естественным для речи на любом языке аллофонным варьированием, которое для эвенкийского языка описано очень неполно. Среди часто встречающихся феноменов были отмечены импловивность смычно-взрывных согласных, спيرانтизация заднеязычного звонкого взрывного, многочисленные случаи варьирования сонорных по локусу, систематические вокализации поствокальных сонорных и многое другое.

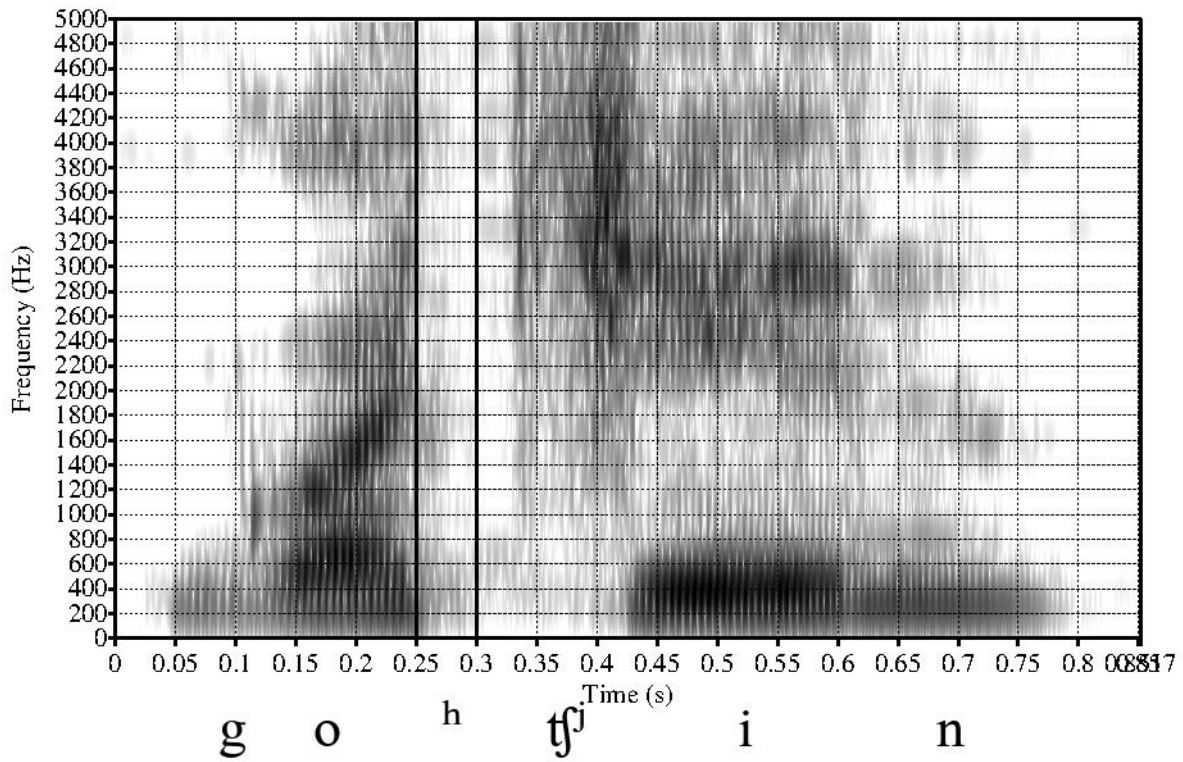


Рис. 8. Поствокальный шум в слове *gochin* (= в следующем году)  
Fig. 8. Post-vocal noise in the word *gochin* (= next year)

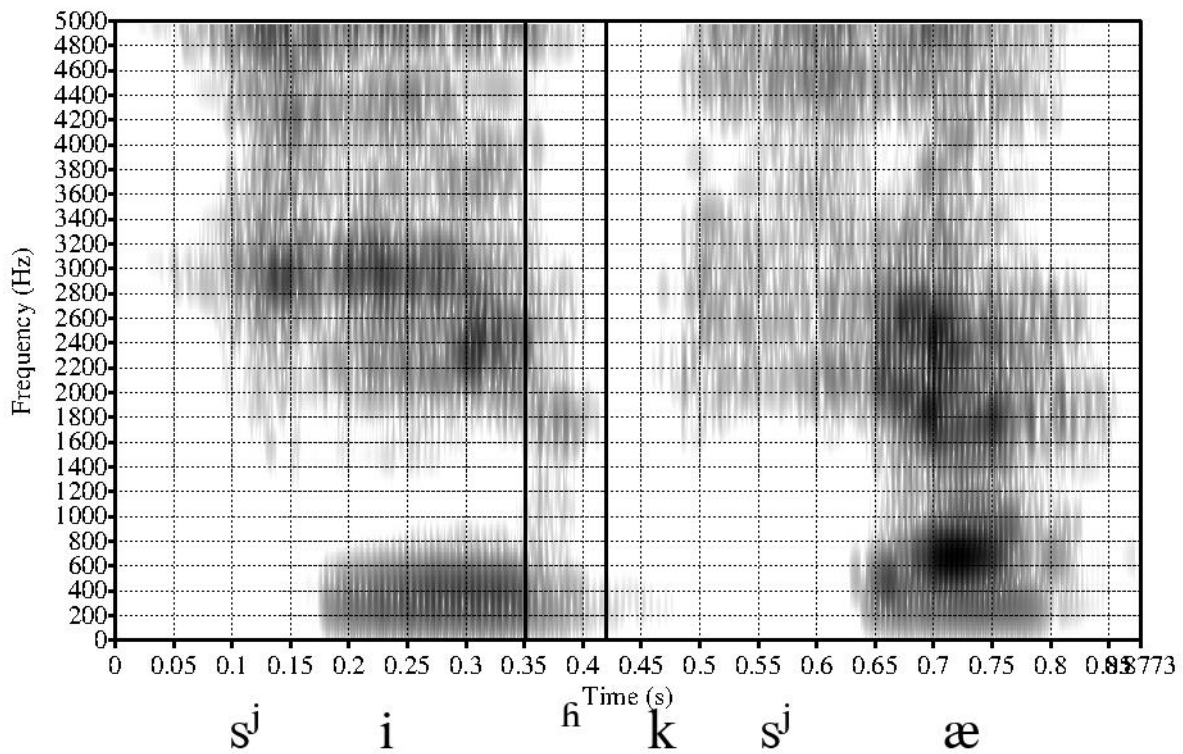
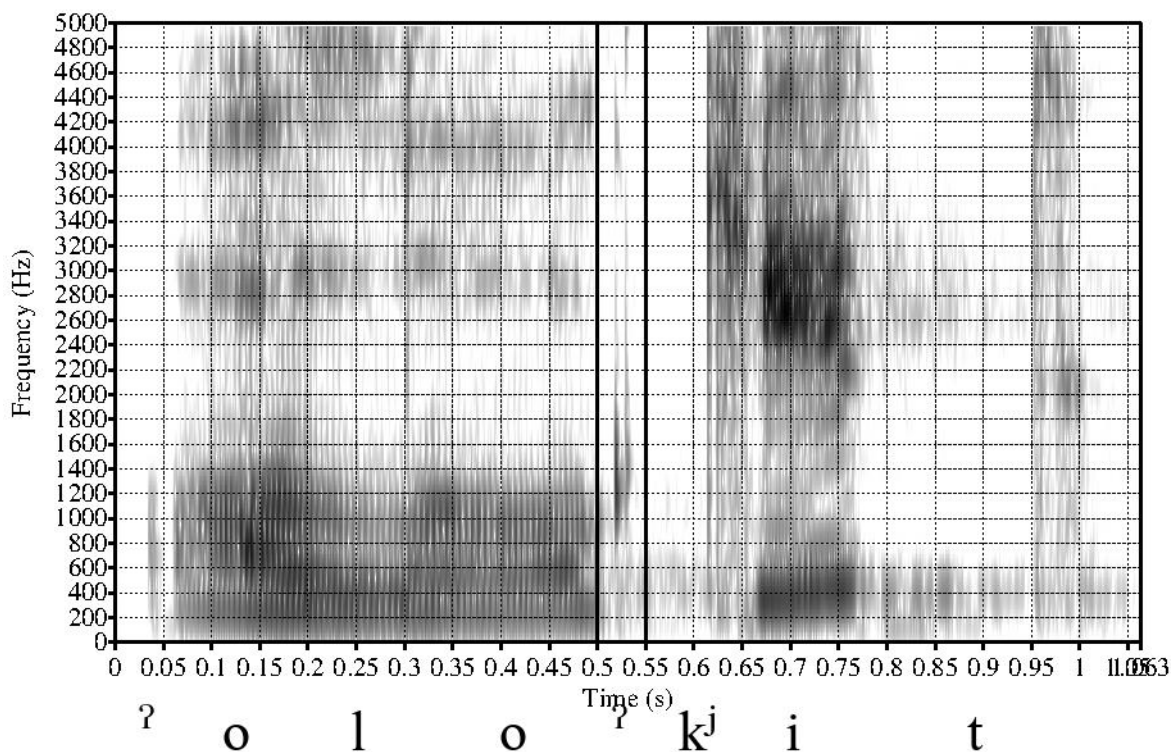


Рис. 9. Поствокальный шум в слове *sikse* (= вечером)  
Fig. 9. Post-vocal noise in the word *sikse* (= in the evening)





**Рис. 10.** Поствокальный шум в слове *олокит* (= брод)  
**Fig. 10.** Post vocal noise in the word *olokit* (= ford)

### 3. Заключение и перспективы

Предложенные в проекте корпуса по трем говорам восточных эвенков технические и лингвистические решения не претендуют на то, чтобы считаться единственно возможными, а ответы на поставленные непростые вопросы – полными. Однако смеем надеяться, что проделанная работа внесет вклад в ревитализацию эвенкийского языка и стимулирует желание представителей среднего и молодого поколений эвенков активно его использовать.

В настоящее время также ведется активная работа над созданием аналогичного корпуса звучащей орончонской речи. Поставленная задача чрезвычайно актуальна, поскольку данный язык, близкородственный эвенкийскому, также находится на грани исчезновения, будучи, несмотря на все усилия китайского правительства, практически вытесненным доминирующим китайским языком [Мэн Шусянь, 2017; Ян Лихуа, 2016]. На данный момент осуществляются накопление и акустический анализ речевых образцов по аналогии с эвенкийским материалом.

### Список литературы

- Андреева Т. Е. Словесное ударение в эвенкийском языке (экспериментально-фонетическое исследование на материале говоров эвенков Якутии). Новосибирск: Наука, 2001. 151 с.
- Андросова С. В. Акустический анализ речевого сигнала : учеб.-метод. пособие. Благовещенск : Изд-во Амур. гос. ун-та, 2014.
- Афанасьева Е. Ф., Раднаева Л. Д. Звуковой корпус данных современного эвенкийского языка // Проблемы изучения и сохранения языков и культур народов России. Материалы секции XXXIX Международной филологической конференции 15–20 марта 2010 г., Санкт-Петербург [отв. ред. Л. Д. Раднаева]. Санкт-Петербург, 2011. С. 3–8.
- Казакевич О. А. Мультимедийный размеченный корпус текстов на говорах западных эвенков [Электронный ресурс]. Режим доступа : <http://languedoc.philol.msu.ru> .
- Казакевич О. А., Клячко Е. Л. Создание мультимедийного аннотированного корпуса текстов как исследовательская процедура [Электронный ресурс] // 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](http://corpora.phil.spbu.ru/Works2013/Казакевич.pdf)

Кравец Т. В. Создание мультимедийного корпуса звучащей речи амурских эвенков: цели, задачи, методы и перспективы // Теоретическая и прикладная лингвистика. 2016. Вып. 2. № 1. С. 41–49.

Морозова О. Н., Андросова С. В., Артемчук М. В. Разработка корпуса звучащей эвенкийской речи // Анализ разговорной русской речи (АР<sup>3</sup>-2017) : тр. Седьмого междисциплинарного семинара / науч. Ред. Д. А. Кочаров, П. А. Скрелин. СПб, 2017. С. 72–77.

Морозова О. Н., Лаврилье А., Болелая А. Н. Некоторые особенности реализации гласных и согласных эвенкийского языка и обозначение на письме // Актуальные проблемы фонетики и методики преподавания иностранных языков. Благовещенск. Амурский гос. ун-т, 2014. С. 84–97.

Морозова О. Н. Уровень владения эвенкийским языком в Зейском, Селемджинском и Мазановском районах Амурской области // Обучение иностранному языку на современном этапе студентов высших и средних общеобразовательных учреждений на современном этапе. Материалы Всероссийской 90 научно-методической видеоконференции. Благовещенск: Амурский гос. ун-т, 2014. С. 207–214.

Мэн Шусянь. Общее описание ороchonского языка в Китае // Теоретическая и прикладная лингвистика. 2017. Вып. 3. № 1. С. 67–86.

Процукович Е. А. Уровень владения эвенкийским языком в местах компактного проживания эвенков Амурской области // Теоретическая и прикладная лингвистика. 2015. Вып. 1. № 2. С. 85–93.

Речевой корпус эвенкийского языка (аннотированный). 2016–2017. URL : <https://linguacorpora.amursu.ru/>

Скрелин П. А. Сегментация и транскрипция. СПб. : Изд-во С.-Петерб. ун-та, 1999.

Ян Лихуа. Исследование современного применения ороchonского языка в провинции Хейлуцзян (КНР) // Филологические науки. Вопросы теории и практики. 2016. № 8. Ч. 2. С. 189–194.

**S. V. Androsova, O. N. Morozova, T. V. Kravets**

*Amur State University, Russian Federation*

*androsova.sv@amursu.ru , morozova\_olga06@mail.ru, t\_kravetc@mail.ru*

### **The corpus of oral Evenki speech**

In the era of globalization with the dominant role of information technologies we are urged to address the issue of language ecology. The Evenki language is the one which is listed among endangered languages. Evenki-Russian bilingualism for the vast majority of the Evenki people in the Amur region has already become Russian-Evenki, and the Evenks who can fluently speak their native language have become scarce. Although IT as a crucial part of civilization benefits have triggered the destructive process, they have a great potential to slow it down or even turn it around through developing on-line free-access oral speech corpora. Endangered language status and the lack of annotated oral speech materials on Eastern Evenki dialects gave us an incentive to start developing such a corpus.

We have developed an information system with the web application written in Ruby on Rails. The system provides free access to the database on Selemdzha, Dzheltulak, and Zea Evenki dialects. The interface is comprised by a number of standard pages: the main page, news, corpus, project information, resources and contacts. In the main page a user can find the information concerning new speech samples added to the database and about the events where the Evenki and Orochon languages are involved. These include the Olympiad in Evenki and Orochon hosted at the Amur State University. The Corpus page provides necessary personal information about the recorded subjects: name, occupation, clan affiliation, spoken dialect. This information is published upon the subjects desire. Some of them even provided photos to be placed on their pages. The corpus is meant for a wide range of users – from native Evenki speakers without any linguistic training to specialists in methods of teaching Evenki and researchers who explore it. We offered a number of technical solutions of certain challenging issues concerning segmentation and transcription peculiarities. One of those was post-vocal turbulent and impulse-like noise in the transition vowel-consonant phase. At present we are also working on developing the Orochon oral speech corpus on the analogy with the Evenki one.

*Keywords:* endangered Evenki language, user type, annotation, transcription, segmental level, suprasegmental level

## References

- Andreeva T.E. *Slovesnoe udarenie v evenkiyskom yazyke (eksperimental'no-foneticheskoe issledovanie na materiale govorov evenkov Yakutii)* [Lexical stress in the Evenki language (Experimental phonetic study of the Evenki dialects of Yakutia)]. Novosibirsk: Nauka, 2001.
- Androsova S.V. *Akusticheskiy analiz rechevogo signala* [Speech acoustics]: A handbook. Blagoveshchensk: Amur State University Press, 2014.
- Afanas'eva E.F., Radnaeva L.D. *Zvukovoy korpus dannykh sovremennogo evenkiyskogo yazyka* [Sound corpus of modern Evenki]. In: L. D. Radnaeva (ed.) Proceedings of XXXIX International Philological Conference, Session "Problemy izucheniya i sokhraneniya yazykov i kul'tur narodov Rossii" [Challenges of exploring and preserving languages and culture of the peoples of Russia]. March 15-20, 2010. St-Petersburg, 2011, pp. 3–8.
- Kazakevich O. A. *Mul'timediynnyy razmechennyy korpus tekstov na govorakh zapadnykh evenkov* [Multimedia annotated corpus of texts in Western Evenki dialects]. URL : <http://languedoc.philol.msu.ru> .
- Kazakevich O.A., Klyachko E.L. *Sozdanie mul'timediynogo annotirovannogo korpusa tekstov kak issledovatel'skaya protsedura* [Developing multimedia annotated corpus of texts as a research procedure]. In: Proceedings of the International Conference «Korpusnaya lingvistika-2013» [Corpus Linguistics-2013]. St-Petersburg State University, Philological Faculty, 2013. Pp. 292–300. URL: [corpora.phil.spbu.ru/Works2013/Kazakevich.pdf](http://corpora.phil.spbu.ru/Works2013/Kazakevich.pdf)
- Kravets T.V. *Sozdanie mul'timediynogo korpusa zvuchashchey rechi amurskikh evenkov: tseli, zadachi, metody i perspektivy* [Creating multimedia corpus of oral speech of the Amur Evenki: Objectives, tasks, methods and perspectives]. In: Teoreticheskaya i prikladnaya lingvistika [Theoretical and Applied Linguistics]. 2016, Vol. 2 (1), pp. 41–49.
- Morozova O.N., Androsova S.V., Artemchuk M.V. *Razrabotka korpusa zvuchashchey evenkiyskoy rechi*. In: D. A. Kocharov, P. A. Skrelin (eds.) Proceedings of the 7th Interdisciplinary Seminar "Analiz razgovornoy russkoy rechi (AR3-2017)" [Russian oral speech analysis (AR3-)]. St-Petersburg State University, 2017, pp. 72–77.
- Morozova O.N., Lavril'e A., Bolelaya A.N. *Nekotorye osobennosti realizatsii glasnykh i soglasnykh evenkiyskogo yazyka i oboznachenie na pis'me* [Some peculiarities of Evenki vowels and consonants realization and the way to write them]. In: Aktual'nye problemy fonetiki i metodiki prepodavaniya inostrannykh yazykov [Challenging issues of phonetics and foreign languages teaching methods]. Blagoveshchensk, Amur State University, 2014, pp. 84–97.
- Morozova O.N. *Uroven' vladeniya evenkiyskim yazykom v Zeyskom, Selemdzhinskom i Mazanovskom rayonakh Amurskoy oblasti*. In: Obuchenie inostrannomu yazyku na sovremennom etape studentov vysshikh i srednikh obshcheobrazovatel'nykh uchrezhdeniy na sovremennom etape. Materialy Vserossiyskoy 90 nauchno-metodicheskoy videokonferentsii. Blagoveshchensk: Amurskiy gos. un-t, 2014, pp. 207–214.
- Meng Shuxian. *Obshchee opisanie orochohnskogo yazyka v Kitae* [General description of the Orochon language in China]. In: Teoreticheskaya i prikladnaya lingvistika [Theoretical and Applied Linguistics]. 2017, Vol. 3 (1), pp. 67–86.
- Protsukovich E.A. *Uroven' vladeniya evenkiyskim yazykom v mestakh kompaktnogo prozhivaniya evenkov Amurskoy oblasti* [The level of the Evenki language skills in the Amur Evenks residences]. In: Teoreticheskaya i prikladnaya lingvistika [Theoretical and Applied Linguistics]. 2015, Vol. 1 (2), pp. 85–93.
- Rechevoy korpus evenkiyskogo yazyka (annotirovanny)*. 2016–2017. URL : <https://linguacorporus.amursu.ru/>
- Skrelin P.A. *Segmentatsia i transkripsia* [Segmentation and transcription]. – St-Petersburg : St-Petersburg University Press, 1999.
- Yan Likhua (2016). *Issledovanie sovremennogo primeneniya orochohnskogo yazyka v provintsii Heilunshian (KNR)* [Study of the contemporary use of the Oroqen language in the province Heilongjiang (People's Republic of China)]. *Filologicheskie Nauki. Voprosy teorii i praktiki* [Philological Sciences. Issues of Theory and Practice], 8 (62). Part 2, pp. 189–194.