Омский государственный технический университет

СОЗДАНИЕ ЗВУЧАЩЕГО СЛОВАРЯ ДЛЯ ТЕВРИЗСКОГО ГОВОРА ЯЗЫКА СИБИРСКИХ ТАТАР

IX Международная конференция по компьютерной обработке тюркских языков "TurkLang 2021"

Убалехт Иван Павлович

Кызыл, 2021

1. Диалекты языка сибирских татар
2. Тевризский говор тоболо-иртышского диалекта
3. Данные для звучащего словаря (полевая работа в 2020 и 2021 гг.)
4. Создание звучащего словаря для тевризского говора языка сибирских татар

1. Диалекты языка сибирских татар

1. Тоболо-иртышский диалект
   - Тюменский говор
   - Тобольский говор
   - Заболотный говор
   - Тевризский говор
   - Тарский говор

2. Барабинский диалект
3. Томский диалект

1. Диалекты языка сибирских татар

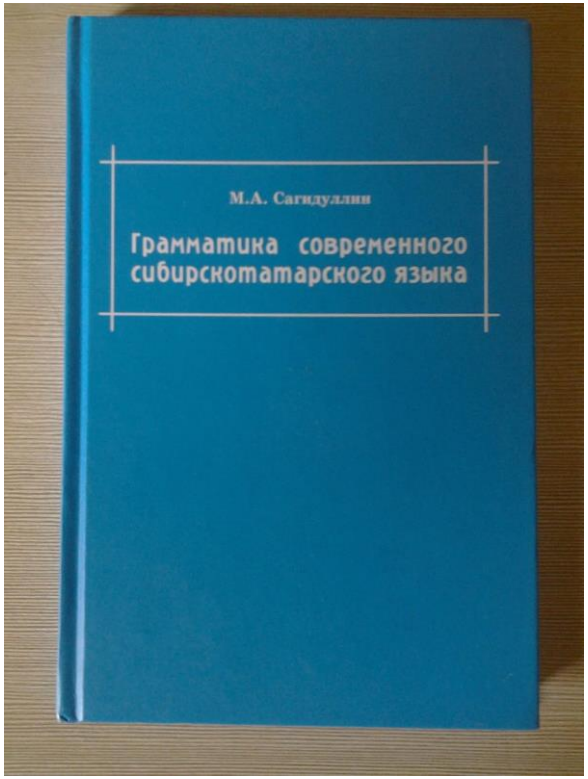
В настоящее время складывается литературная норма на основе тюменского и тобольского говоров тоболо-иртышского диалекта.

Исследуемый нами тевризский говор отличается от этой нормы и является довольно плохо исследованным.

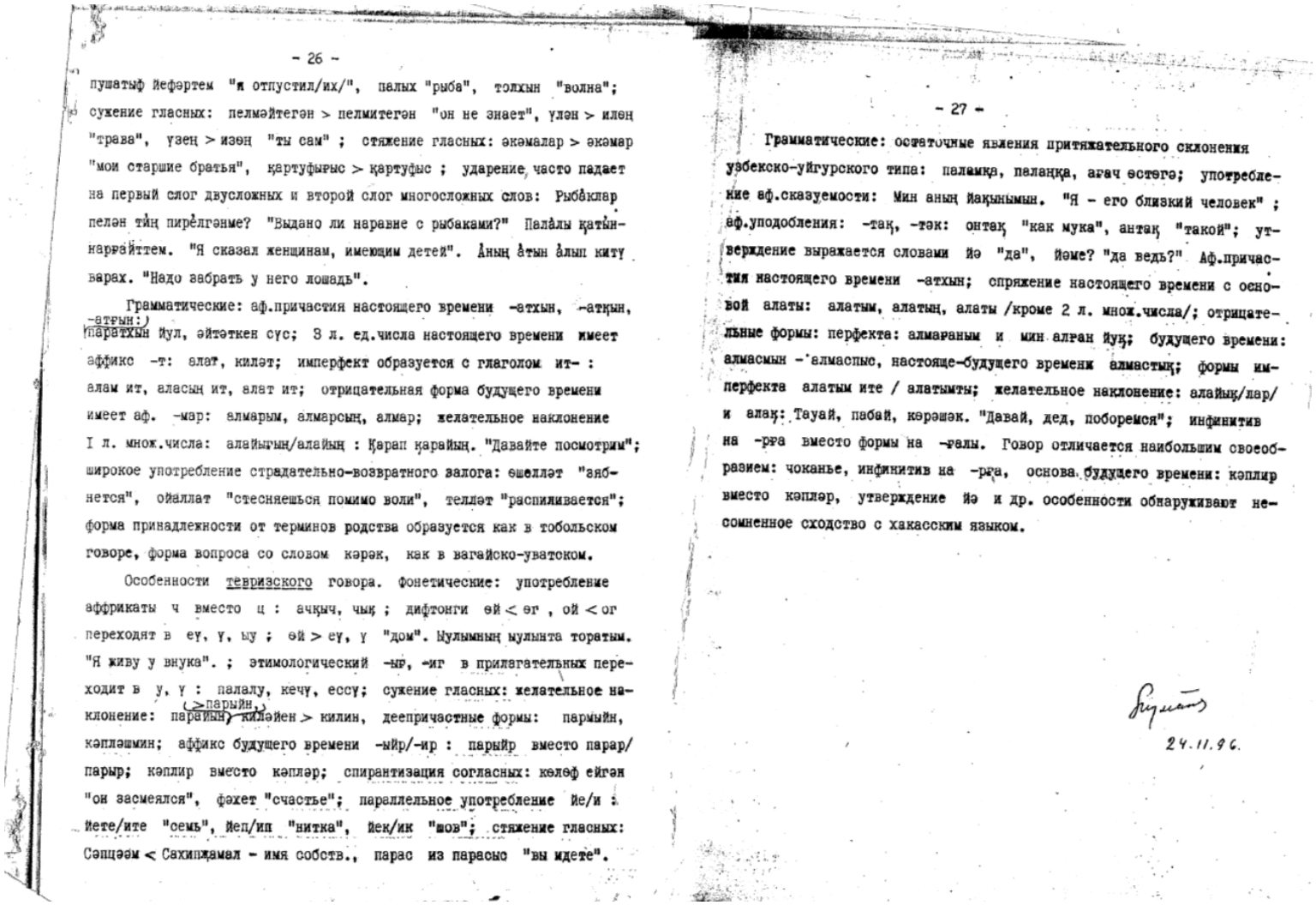
Описание тевризского говора существует, существует «Словарь диалектов сибирских татар» Тумашевой, в этом словаре присутствует некоторое количество лексики из Тевризского говора.

Аудио записей тевризского говора в открытом доступе не было обнаружено.

Существуют ли вообще аудио данные тевризского говора?



2. Тевризский говор тоболо-иртышского диалекта



- 26 -

пушатыф йефәртем "я отпустил/их/", палых "рыба", толкын "волна"; сужение гласных: пелмәйтегән > пелмитегән "он не знает", үлән > илән "трава", үзең > изең "ты сам" ; стяжение гласных: әкәмалар > әкәмар "мои старшие братья", қартуфыыс > қартуфыс ; ударение часто падает на первый слог двусложных и второй слог многосложных слов: Рыбаклар пелән тин пирелгәнме? "Выдано ли наравне с рыбаками?" Палалы қатын-нарәйттем. "Я сказал женщинам, имеющим детей". Аның атын алып китү варах. "Надо забрать у него лошадь".

Грамматические: аф.причастия настоящего времени -атхын, -атқын, -атрын; паратхын йул, әйтәткен сүс; 3 л. ед.числа настоящего времени имеет аффикс -т: алат, киләт; имперфект образуется с глаголом ит- : алам ит, аласың ит, алат ит; отрицательная форма будущего времени имеет аф. -мәр: алмарым, алмарсың, алмар; желательное наклонение I л. множ.числа: алайыгың/алайың : Қарап қарайың. "Давайте посмотрим"; широкое употребление страдательно-возвратного залога: өшеллят "зябнется", ойаллат "стесняешься помимо воли", теллят "распиливается"; форма принадлежности от терминов родства образуется как в тобольском говоре, форма вопроса со словом кәрәк, как в вагайско-уватском.

Особенности тевризского говора. Фонетические: употребление аффрикаты ч вместо ц : ачқыч, чық ; дифтонги өй < өг , ой < ог переходят в еү, ү, ыу ; өй > еү, ү "дом". Ыулымның ыулында торатым. "Я живу у внука". ; этимологический -ыр, -иг в прилагательных переходит в у, ү : палалу, кечү, ессү; сужение гласных: желательное наклонение: парайың (>парыйн) киләйен > килин, деепричастные формы: пармыйн, кәпләшмин; аффикс будущего времени -ыйр/-ир : парыйр вместо парар/парыр; көплир вместо көпләр; спирантизация согласных: көләф ейган "он засмеялся", фәхет "счастье"; параллельное употребление йе/и : йете/ите "семь", йеп/ип "нитка", йек/ик "нов"; стяжение гласных: Сәпцәәм < Сахипжамал - имя собств., парас из парасың "вы идете".

- 27 -

Грамматические: остаточные явления притяжательного склонения узбекско-уйгурского типа: паламқа, паланқа, агач өстәгә; употребление аф.сказуемости: Мин аның йақынымын. "Я - его близкий человек" ; аф.уподобления: -тақ, -тәк: онтақ "как мука", антақ "такой"; утверждение выражается словами йә "да", йәме? "да ведь?" Аф.причастия настоящего времени -атхын; спряжение настоящего времени с основой алаты: алатым, алатың, алаты /кроме 2 л. множ.числа/; отрицательные формы: перфекта: алмаранын и мин алған йук; будущего времени: алмасмын - алмаспыс, настояще-будущего времени алмастық; формы имперфекта алатым ите / алатымты; желательное наклонение: алайық/лар/ и алақ: Тауай, пабай, көрәшәк. "Давай, дед, поборемся"; инфинитив на -рға вместо формы на -ралы. Говор отличается наибольшим своеобразием: чоканье, инфинитив на -рға, основа будущего времени: көплир вместо көпләр, утверждение йә и др. особенности обнаруживают несомненное сходство с хакасским языком.

24.11.96.

Тумашева Д.Г. д.филол.н, академик АН РТ
«Язык сибирских (тоболо-иртышских) татар», Тюмень 1997

2. Тевризский говор тоболо-иртышского диалекта

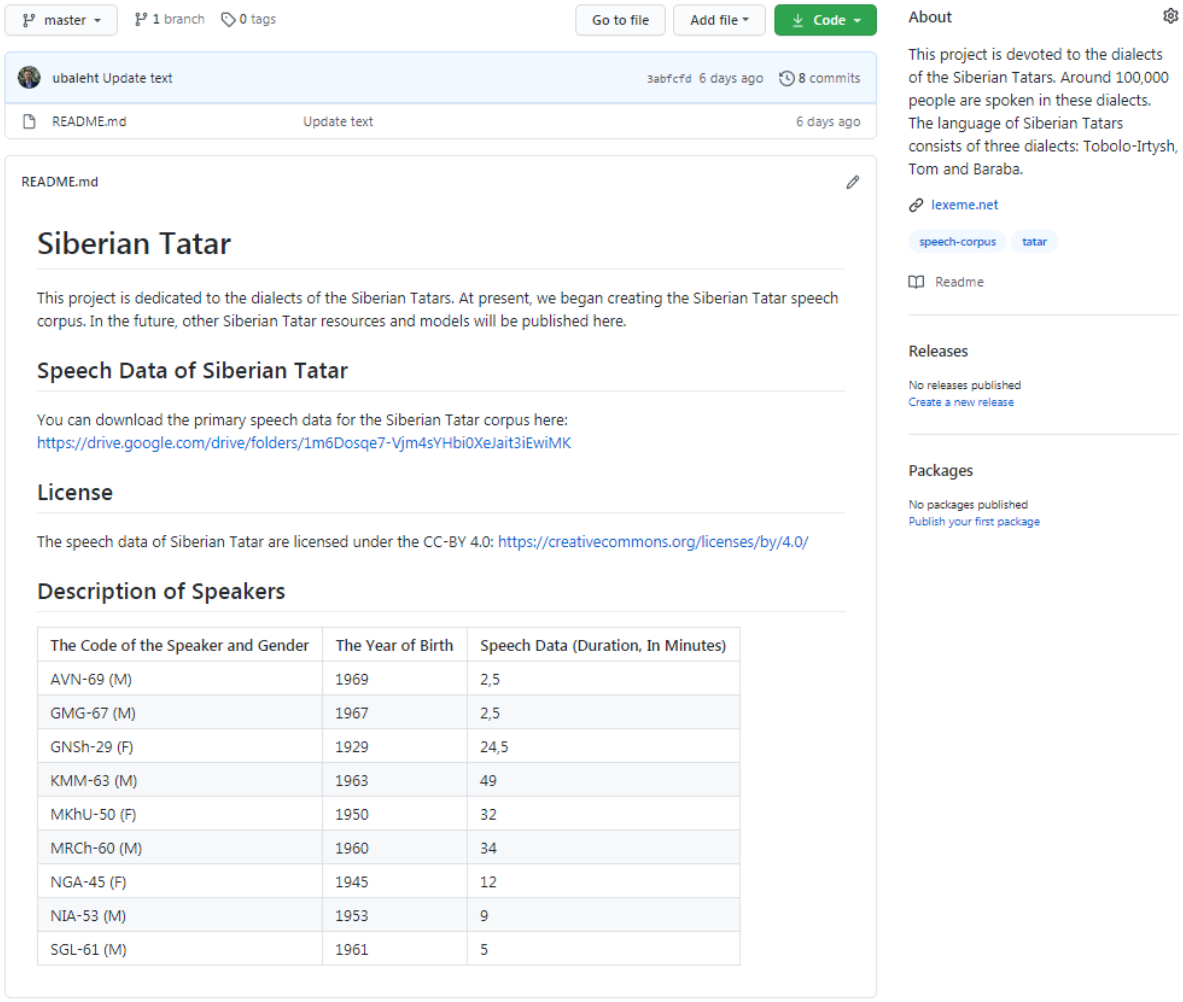
на -рға вместо формы на -ғалы. Говор отличается наибольшим своеобразием: чоканье, инфинитив на -рға, основа будущего времени: кәплир вместо кәпләр, утверждение йә и др. особенности обнаруживают несомненное сходство с хакасским языком.

“Говор отличается наибольшим своеобразием: чокание, инфинитив на –рға, основа будущего времени: кәплир, вместо кәпләр, утверждение йә, и др. особенности обнаруживают несомненное сходство с хакасским языком”

Тумашева Д.Г. д.филол.н, академик АН РТ
«Язык сибирских (тоболо-иртышских) татар», Тюмень 1997

3. Сбор данных, полевая работа в 2020 и 2021 гг.

Рабочий репозиторий проекта на GitHub:
https://github.com/ubaleht/SiberianTatar



master · 1 branch · 0 tags · Go to file · Add file · Code

ubaleht Update text — 3abfcfd 6 days ago — 8 commits

README.md — Update text — 6 days ago

README.md

Siberian Tatar

This project is dedicated to the dialects of the Siberian Tatars. At present, we began creating the Siberian Tatar speech corpus. In the future, other Siberian Tatar resources and models will be published here.

Speech Data of Siberian Tatar

You can download the primary speech data for the Siberian Tatar corpus here:
https://drive.google.com/drive/folders/1m6Dosqe7-Vjm4sYHbi0XeJait3iEwiMK

License

The speech data of Siberian Tatar are licensed under the CC-BY 4.0: https://creativecommons.org/licenses/by/4.0/

Description of Speakers

The Code of the Speaker and Gender	The Year of Birth	Speech Data (Duration, In Minutes)
AVN-69 (M)	1969	2,5
GMG-67 (M)	1967	2,5
GNSh-29 (F)	1929	24,5
KMM-63 (M)	1963	49
MKhU-50 (F)	1950	32
MRCh-60 (M)	1960	34
NGA-45 (F)	1945	12
NIA-53 (M)	1953	9
SGL-61 (M)	1961	5

About

This project is devoted to the dialects of the Siberian Tatars. Around 100,000 people are spoken in these dialects. The language of Siberian Tatars consists of three dialects: Tobolo-Irtysh, Tom and Baraba.

lexeme.net

speech-corpus · tatar

Readme

Releases

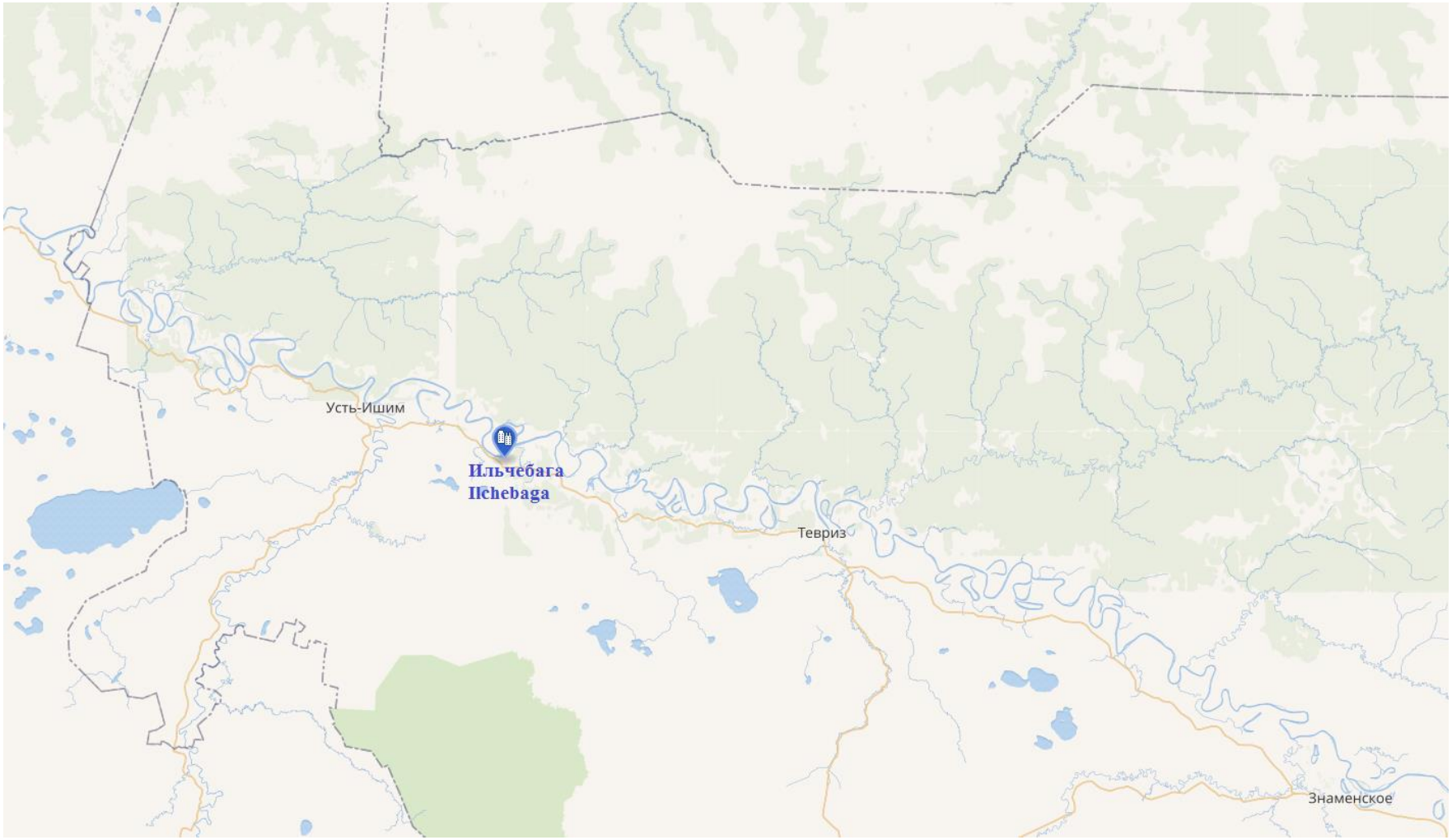
No releases published
Create a new release

Packages

No packages published
Publish your first package

3. Полевая работа в 2020 и 2021 гг.

Тевризский говор тоболо-иртышского диалекта распространён в Усть-Ишимском, Тевризском, Знаменском районах Омской области, а также вероятно на прилегающей территории Тюменской области



3. Полевая работа в 2020 г.

2020 году экспедиция в деревню Ильчебага Усть-Ишимского района Омской области.

Таблица 10

Этнический состав татар Усть-Ишимского района Омской области (по данным родословных 1970-х гг.), чел.

Населенный пункт	Общее число родословных	Считают себя по происхождению						Фактический состав опрошенных татар					
		сибирскими татарами		поволжско-приуральскими татарами		просто татарами		Сибирские татары		Поволжско-приуральские татары		Выходцы из смешанных семей	
		Количество	%	Количество	%	Количество	%	Количество	%	Количество	%	Количество	%
Ашеваны	194	190	98,0	4	2,0	—	—	178	91,8	4	2,0	12	6,2
Большая Тебендя	124	117	94,4	7	5,6	—	—	111	89,5	7	5,6	6	4,9
Ильчебага	153	120	78,4	24	15,7	9	5,9	120	78,4	24	15,7	9	5,9
Кайнаул	80	75	93,8	5	6,2	—	—	75	93,8	5	6,2	—	—
Лешаково	42	42	100,0	—	—	—	—	40	95,2	—	—	2	4,8
Малая Тебендя	4	4	100,0	—	—	—	—	4	100,0	—	—	—	—
Саургачи	117	104	88,9	10	8,5	3	2,6	104	88,9	10	8,5	3	2,6
Тюрметяки	51	45	88,2	6	11,8	—	—	45	88,2	6	11,8	—	—
Хутор	16	16	100,0	—	—	—	—	16	100,0	—	—	—	—
Итого...	781	713	91,4	56	7,2	12	1,4	693	88,6	56	7,2	32	4,2

С.Н. Корусенко, Н.В. Кулешова «Генеология и этническая история Барабинских и курдакско-саргатских татар», 1999

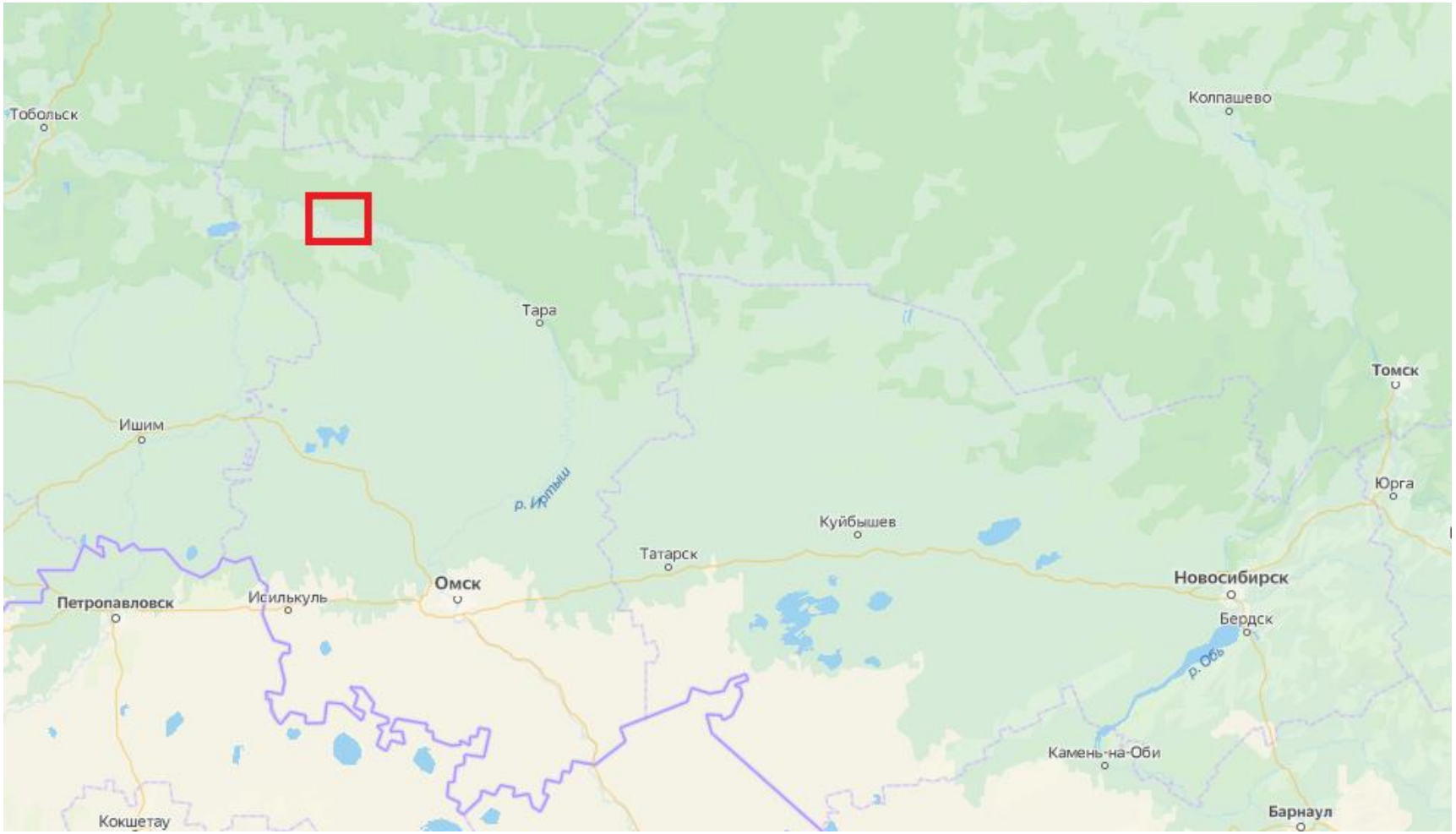
3. Полевая работа в 2020 г.

В 2020 году речевые данные были собраны у 10 информантов в деревне Ильчебага. Аудио Данные были опубликованы под Свободной лицензией CC BY 4.0 в рабочем репозитории проекта на GitHub.

The code of the speaker and gender	The Year of Birth	The current place of residence	The place of the birth	Birthplaces of parents or/and ethnic of parents	Speech data (duration, in minutes)
AVN-69 (M)	1969	Ilchebaga	Ilchebaga	Both parents: Siberian Tatars	2,5
GMG-67 (M)	1967	Ilchebaga	Ilchebaga	Both parents: Volga Tatars	2,5
GNSh-29 (F)	1929	Ilchebaga	Ilchebaga	Three grandparents: Volga Tatars, one grandparent: Siberian Tatars	24,5
KMM-63 (M)	1963	Ilchebaga	Tavinsk	Both parents: Tavinsk (Siberian Tatars)	49
MKhU-50 (F)	1950	Ilchebaga	No data	Father: Tavinsk, mother: Tebendya, both Siberian Tatars	32
MRCh-60 (M)	1960	Ilchebaga	No data	Three grandparents: Erbagul, one grandparent: Ilchebaga	34
NGA-45 (F)	1945	Ilchebaga	Yarkovo	Father: Volga Tatars, mother: Siberian Tatars	12
NIA-53 (M)	1953	Ust-Ishim	Kuchum	Father: Kuchm (Siberian Tatars), mother: Volga Tatars	9
SGL-61 (M)	1961	Ilchebaga	No data	No data	5
Anonym Speaker (M)	No data	Ilchebaga	No data	Mother: Siberian Tatars, father: Bukharian Tatars	4

3. Полевая работа в 2021 гг.

Летом 2021 была совершена экспедиция в деревни Ильчебага и Тавинск Усть-Ишимского района Омской области.



3. Полевая работа в 2021 г.

Результаты экспедиции 2021 года

Работа велась по анкетам предоставленным командой Анны Владимировны Дыбо:
- 100 и 200 список Сводеша с контекстами
- Экспериментальная анкета по татарским диалектам
- Записывалась спонтанная речь, включая диалоги
- Записывались отдельные слова не из списка Сводеша
- Записывалось видео

Данные собраны в двух деревнях, от 13 носителей, собрана также социолингвистическая информация

Результаты экспедиций 2020 и 2021 гг.

Данные собраны от 22 носителей
Общий объём аудио данных – около 7 часов

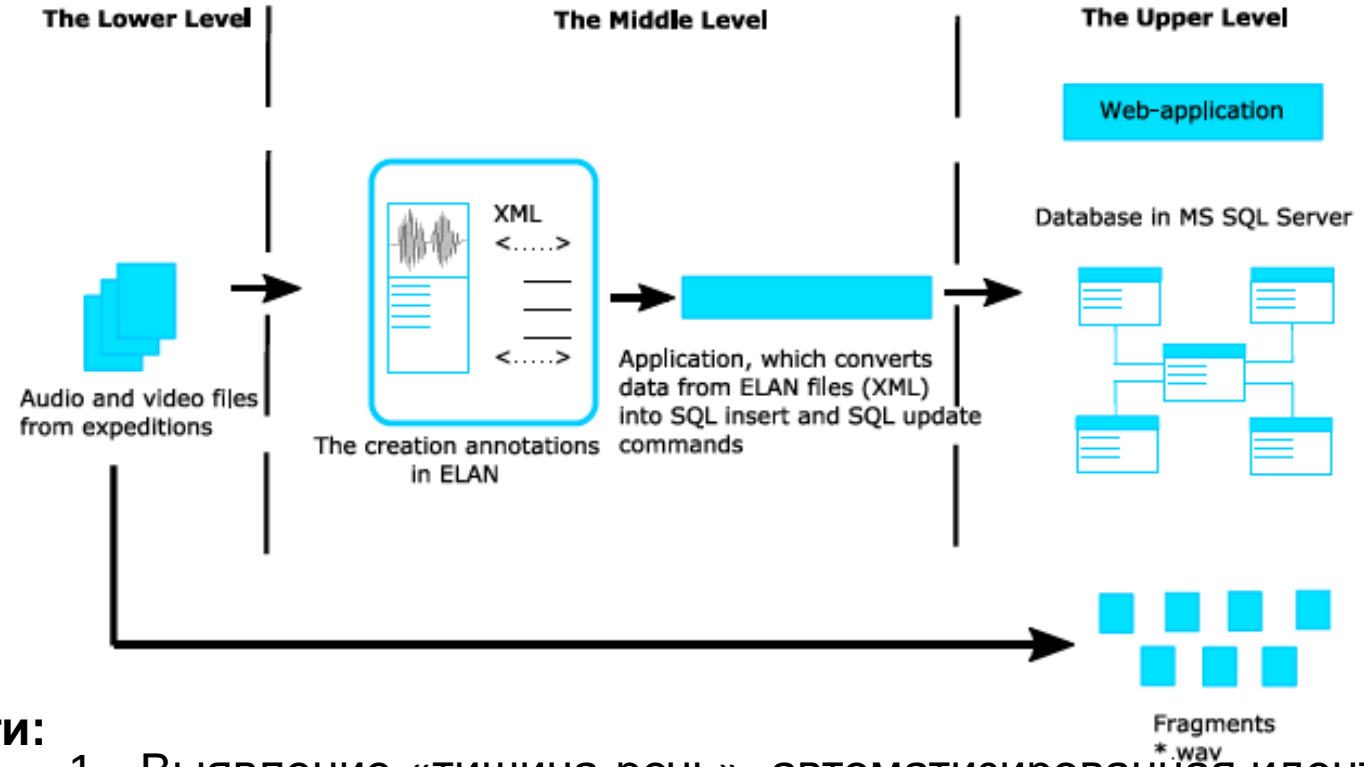
4. Создание звучащего словаря тевризского говора языка сибирских татар

Цель: Создание сервиса для работы с аудиоданными диалектов сибирских татар, имеющего удобный интерфейс для предоставления данных другим группам учёных и пользователей
- интерфейс на уровне «сырых» аудиоданных
- на уровне аннотированных аудиоданных
- На уровне запросов к данным с использованием web-интерфейсов.

Текущая цель: Предоставить удобный доступ к множеству отдельных слов тевризского говора языка сибирских татар. То есть обеспечить функционал озвученного словаря.

Так как тевризский говор не имеет стабильной системы письменности, то озвученный словарь важным первым шагом для ввода в научный оборот лексического материала тевризского говора языка сибирских татар.

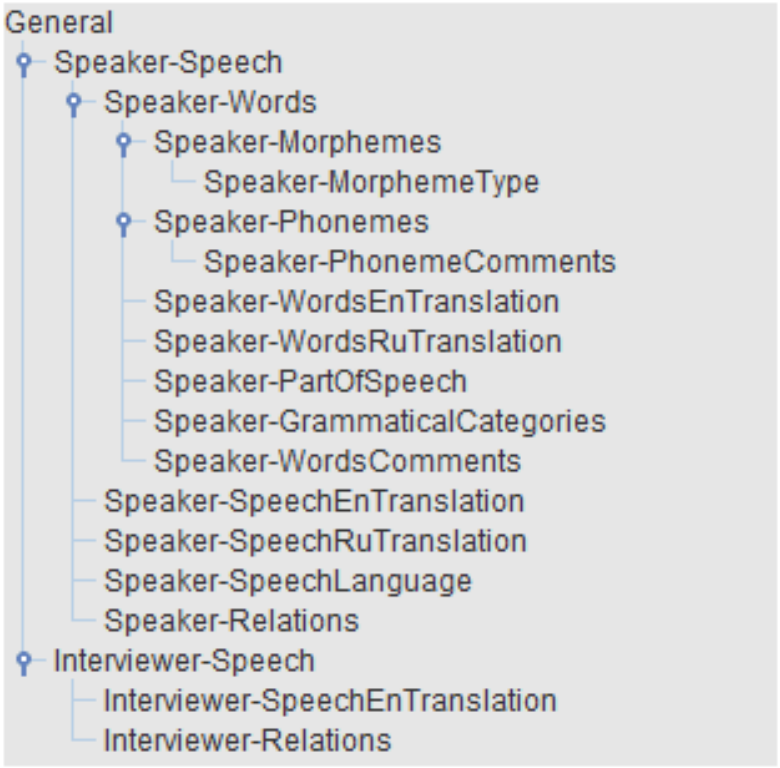
4. Создание звучащего словаря тевризского говора языка сибирских татар



Шаги:

1. Выявление «тишина-речь», автоматизированная идентификация звучащей русской и татарской речи. Возможно решение задачи распознание речи на сибирскотатарском языке с использованием разработок для волжского варианта татарского языка.
2. Автоматическое или полуавтоматическое аннотирование, конвертация аннотаций в записи в базе данных
3. Разработка веб-приложения звучащего словаря

4. Создание звучащего словаря тевризского говора языка сибирских татар



Список уровней в ELAN файлах аннотаций,
ELAN файлы это обычные XML файлы

4. Создание звучащего словаря тевризского говора языка сибирских татар

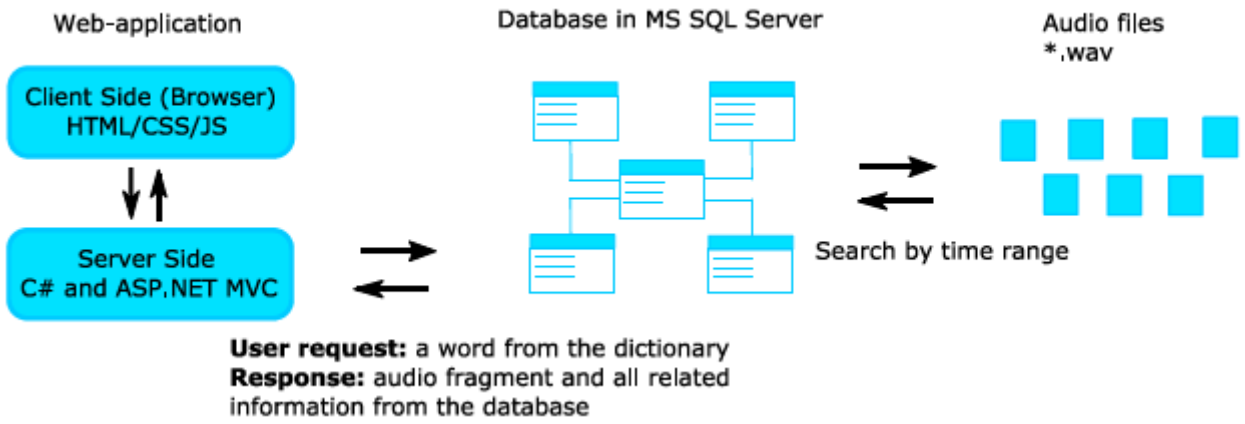


Схема работы звучащего словаря

Используемые технологии: .NET Framework, ASP.NET MVC, C#, MS SQL Server

Контакты

Омский государственный технический университет
Кафедра «Автоматизированные системы обработки информации и управления»

Убалехт Иван Павлович

ubaleht@gmail.com

8 905 923 81 34