

Федеральное агентство научных организаций

ИНСТИТУТ ИСТОРИИ, ЯЗЫКА И ЛИТЕРАТУРЫ – ОБОСОБЛЕННОЕ
СТРУКТУРНОЕ ПОДРАЗДЕЛЕНИЕ ФЕДЕРАЛЬНОГО
ГОСУДАРСТВЕННОГО БЮДЖЕТНОГО НАУЧНОГО УЧРЕЖДЕНИЯ
УФИМСКОГО ФЕДЕРАЛЬНОГО ИССЛЕДОВАТЕЛЬСКОГО ЦЕНТРА
РОССИЙСКОЙ АКАДЕМИИ НАУК

УДК: 512.141:004.655

Н госрегистрации: AAAA-A17-117040710023-4

«УТВЕРЖДАЮ»
Врио председателя УФИЦ РАН
доктор биологических наук



З.Х.Шигапов

201_ г.

ОТЧЕТ О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ «СОЗДАНИЕ
КОРПУСА ДИАЛЕКТНЫХ ТЕКСТОВ БАШКИРСКОГО ЯЗЫКА»

Программа фундаментальных научных исследований
государственных академий наук на 2013–2020 годы

(Промежуточный)

Протокол Ученого совета
№ 10 от «12» декабря 2017 г.

Врио директора ИИЯЛ УФИЦ
РАН
д-р геогр. наук, проф.
Научный руководитель
к-т филолог. наук

П.Сяянчин
подпись, дата

А.В. Псяянчин

З.А. Сиразитдинов
подпись, дата

З.А. Сиразитдинов

Уфа - 2017

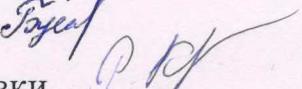
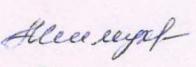
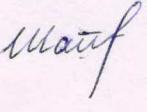
СПИСОК ИСПОЛНИТЕЛЕЙ

Руководитель темы:

кандидат филологических наук З.А.Сиразитдинов



Исполнители темы:

1. кандидат филологических наук З.А.Сиразитдинов 
2. кандидат филологических наук Л.А.Бускунбаева, 
3. младший научный сотрудник Р.Н.Каримова, 0,5 ставки, 
4. младший научный сотрудник А.Ш.Ишмухаметова, 0,25 ставки, 
5. лаборант-исследователь Г.Г.Шамсутдинова, 0,25 ставки. 

СПИСОК ИСПОЛНИТЕЛЕЙ

Руководитель темы:

кандидат филологических наук З.А.Сиразитдинов

Исполнители темы:

1. кандидат филологических наук З.А.Сиразитдинов
2. кандидат филологических наук Л.А.Бускунбаева,
3. младший научный сотрудник Р.Н.Каримова, 0,5 ставки,
4. младший научный сотрудник А.Ш.Ишмухаметова, 0,25 ставки,
5. лаборант-исследователь Г.Г.Шамсутдина, 0,25 ставки.

РЕФЕРАТ

Отчет 18 с., 48 источников.

КЛЮЧЕВЫЕ СЛОВА: филология, тюркология, башкирский язык, корпусная лингвистика, расшифровка речи, фонетическая транскрипция, диалектология, башкирские говоры, лингвистические базы данных.

ОБЪЕКТ ИССЛЕДОВАНИЯ. Объектом данного научного исследования является восточный диалект башкирского языка.

ЦЕЛЬ РАБОТЫ. Целью научной работы является создание речевого корпуса восточного диалекта башкирского языка.

МЕТОДОЛОГИЯ ИССЛЕДОВАНИЯ. При работе над проектом используется методология сбора и компьютерной обработки полевых речевых материалов информаторов.

РЕЗУЛЬТАТЫ РАБОТЫ И ИХ НОВИЗНА. Изучена методология сбора и обработки полевых материалов, составлена инструкция обработки, собраны и обработаны речевые материалы по современному состоянию миасского и кизильского говоров восточного диалекта башкирского языка.

ОБЛАСТЬ ПРИМЕНЕНИЯ. Результаты научно-исследовательской работы найдут применение в диахронных и синхронных исследованиях башкирских языковедов, в историко-сравнительных научных изысканиях тюркологов и сопоставительных исследованиях алтайистов. Обработанные полевые материалы могут использоваться в учебном процессе – при чтении спецкурсов, лекций по фонетике и диалектологии башкирского языка, при написании учебных пособий для ВУЗов. Результаты могут быть полезными при создании и реализации проектов, направленных на сохранение языка и культуры носителей говоров национального языка. Созданный материал может быть использован в решении смежных с филологическим вопросов по культуре, этнографии, фольклору, искусствоведению.

1. ВВЕДЕНИЕ	5
2. ОСНОВНАЯ ЧАСТЬ.....	8
2.1. СБОР ПОЛЕВОГО МАТЕРИАЛА.....	8
2.2. ПРИНЦИПЫ ТРАНСКРИБИРОВАНИЯ.....	9
2.3. СРЕДСТВА РАБОТЫ С КОРПУСОМ АУДИО И ТЕКСТОВЫХ ДАННЫХ..	11
2.4. ПРОВЕДЕННАЯ РАБОТА	12
3. ЗАКЛЮЧЕНИЕ.....	14
4. ИСТОЧНИКИ И ЛИТЕРАТУРА	15

1. ВВЕДЕНИЕ

Башкирский язык является одним из крупных и развитых языков народов России. В его составе различают три диалекта: восточный, южный и северо-западный, каждый из которых подразделяется на говоры, имеющие своеобразные фонетические, лексические и грамматические особенности.

Исходным моментом в изучении башкирской диалектологии принято считать исследования говоров в 20-30х годах 20 века. Экспедиции организованные с участием Г.Амантаева, Т.Баишева, Г.Давлетшина, З.Ш.Шакирова, Г.Ишбулатова и др. позволили выявить основные границы и черты говоров и диалектов языка. Большое значение имели научные отчеты Т.Г.Баишева и З.Ш.Шакирова [1, 2], которые позволили выдвинуть идею создания научной грамматики башкирского языка, основанную на материалах народного языка. Эта идея была реализована Н.К.Дмитриевым в его работе “Грамматика башкирского языка” (М.;Л., 1948), которая в значительной мере опиралась на диалектологический материал башкирских языковедов.

С начала 1950-х годов башкирская диалектология приступила к системному изучению фонетики, морфологии, лексики говоров. Появление монографических трудов Т.Г.Баишева [3], Дж.Г.Киекбаева [4] вывело башкирскую диалектологическую науку на всесоюзную тюркологическую арену.

С 1960 года началось систематическое изучение диалектов с помощью программы «Анкета для собирания материалов по говорам башкирского языка» [5], предусматривающей единообразный сбор диалектного материала. Полевые материалы, собранные во время экспедиций 60-70-х годов, заложили начало диалектной лексикографии. Выход в свет трехтомного словаря говоров явился значительным событием в башкирском языкоznании [6-8]. Материалы словарей полностью вошли в реестр «Русско-башкирского словаря» (1964), обогатив литературный язык.

Материалы опросников экспедиций 60-70-х годов были изучены в сравнительно-сопоставительном аспекте и легли в основу диссертационных и монографических работ

[9-12]. Эти работы послужили основой для перехода к лингвогеографическому изучению говоров и диалектов башкирского языка.

В 1973 г. вышла программа по сбору материала для Атласа башкирского языка. Программа предусматривала сбор материала для изучения фонетического, лексического и грамматического явлений, которые должны были дать линию изоглоссы с точки зрения исторической, ареальной, смысловой нагрузки. Изданный в 2005 г. Атлас отображает данные по 250 опорным пунктам Республики Башкортостан и сопредельных областей. Данные атласа введены исполнителями данной научной темы в базу данных Машинного фонда башкирского языка и доступны в сети Интернет [13].

К сожалению, с 80-х годов и по настоящее время башкирские говоры и диалекты остаются в стороне от фронтальных полевых исследований. Количество экспедиций с единой программой сбора полевого материала резко уменьшилось. Полевых материалов с охватом широкого круга информаторов по полу-возрастным и другим социальным группам отсутствует. Материалы прежних экспедиций не были оцифрованы, большая часть их утрачена.

За более чем 30 лет в говорах и диалектах, несомненно, произошли сдвиги и изменения, которые требуют фиксации и научного анализа. Современная лингвистика выдвигает также и новые подходы к исследованию говоров языка. В ИИЯЛ за эти годы диалектологическая тема вообще не поднималась. Все это актуализирует задачу сбора и анализа полевых материалов сегодняшнего дня на новых методологиях и подходах по говорам башкирского языка.

В современной лингвистике одним из ведущих направлений исследований является корпусное исследование. Объектом корпусной лингвистики являются речевые материалы, реализованные в виде как письменных текстов, так и устных (фонетических) массивов данных. На сегодня в мире насчитывается более тысячи корпусов, их количество растет с каждым годом экспоненциально.

По объему привлеченных текстовых материалов выделяются корпусы немецкого (DeReKo, 5,4 млрд. слов) [14], английского (BNC, 100 млн. слов) [15], американского варианта английского (450 млн. слов) [16], китайского (LIVAC Synchronous Corpus, 1 млрд. слов) [17], венгерского (100 млн. слов) [18], испанского (100 млн. слов) [19],

итальянского (100 млн. слов) [20], чешского (200 млн. слов) [21], русского (НКРЯ, 600 млн. слов) [22] языков.

Количество и объемы речевых корпусов значительно отстают от текстовых, но темп составления таких корпусов с каждым годом увеличивается. Так, по английскому [23], немецкому [24], китайскому [25], японскому [26], турецкому [27], эстонскому [28], русскому [29] языкам реализованы корпусы, содержащие как мультимедийные данные, так и транскрипции речи.

В последнее время возрос интерес к составлению корпусов диалектов. Часть таких корпусов включена в состав национальных корпусов, часть функционирует как самостоятельный корпус. К примеру корпусы диалектов немецкого языка (подкорпусы Мангеймского проекта) охватывают 4 группы корпусов, посвященных разновидностям немецкого языка (*Sprachvarietaten*) на территории Германии и за ее пределами. В России известны корпусы ижимского говора коми языка [30], Саратовский корпус русских говоров [31], Мультимедийный корпус диалектных текстов Устьянского района Архангельской области [32], диалектный подкорпус Национального корпуса русского языка [33] и др.

Интерес к корпусам обусловлен тем, что это направление лингвистики активно влияет на все остальные направления языкознания, изменяет теоретические приоритеты и создает новые идеологии в понимании того, что же представляет собой язык. Корпусы открывают перспективу для новых исследований не только в области лингвистики, но и в смежных областях: в культурологии, философии, в общественных науках (изучение социальных объектов через язык, используя такие параметры текстов, как период, социальная группа, семантический контент текстов) и др.

2. ОСНОВНАЯ ЧАСТЬ

2.1. СБОР ПОЛЕВОГО МАТЕРИАЛА

Восточный диалект башкирского языка включает 5 говоров: айский, сальютский, аргаяшский, миасский, кизильский.

1) Для сбора полевых материалов выбраны 28 населенных пунктов:

Айский говор: 1) д. Яныбаево, д. Нижнеутяшево — Белокатайский район РБ, 2) д. Сагирово, д. Кургашево — Кигинский район РБ, 3) д. Аркаулово, д. Яхино — Салаватский район РБ. Всего 6 населенных пунктов.

Салютский говор: 1) д. Малоказакбаево, д. Курманово, д. Ибрагимово — Челябинская область, Кунашакский район, 2) д. Большекызылбаево — Мечетлинский район РБ, 3) д. Худайбердино — Челябинская область, Аргаяшский район. Всего 5 населенных пунктов.

Аргаяшский говор: 1) д. Сарт-Абдрашево, д. Бикбирдино — Курганская область, Сафакулевский район. 2) д. Тарголово, д. Зайнекеево — Курганская область, Альменевский район, 2) д. Байрамгулово, д. Аязгулово — Челябинская область, Аргаяшский район. Всего 6 населенных пунктов.

Миасский говор: 1) д. Сафарово, д. Расулево, д. Казаккулово, д. Мулдакаево — Учалинский район РБ, 2) д. Халилово — Абзелиловский район РБ. Всего 5 населенных пунктов.

Кизильский говор: 1) д. Давлетово, д. Халилово — Абзелиловский район РБ, 2) д. Кусеево, д. Туркменево — Баймакский район РБ, 3) д. Мурадымово, д. Атиково — Бурзянский район. Всего 6 населенных пунктов.

2) По каждому населенному пункту производится минимальное количество записей в 8 единиц:

2 записи по половому признаку, 4 записи по возрастным группам. Выделяются следующие возрастные группы:

- молодежный возраст 12-25 лет;

- средний 25-45 лет.

- старший возраст 45-65 лет

- пожилой возраст: 65 и старше.

Учитывается образование: начальное, среднее, высшее.

3) Аудиозапись сопровождается информацией, которая составляет экстралингвистическую разметку аудиофайла:

- пол: мужской, женский;
- образование: начальное, среднее (среднее школьное и суз), высшее;
- возраст;
- язык обучения: башкирский, русский, татарский, чувашский;
- язык общения в семье: башкирский, русский, татарский;
- национальность информатора;
- имя, отчество, фамилия;
- место последнего долгого проживания до переезда в данное место;
- время проживания до последнего места проживания;
- время записи.

4) Выделяется тип общения: монолог, диалог, полилог.

5) Для записи информаторов определены следующие темы речевого текста:

- свадьба, свадебные обычаи;
- обычные блюда (что любят готовить на каждый день, что на праздниках);
- какой скот и как содержат;
- дети и близкие родственники (система родства);
- приусадебное хозяйство (огород, сад);
- о доме (когда построен, кто строил, крыша, рамы);
- топонимии в окрестностях поселения;
- история села, школы, рода;

- работа, школа, друзья;
 - времена года, погода;
 - частушки, пословицы, поговорки, сказки (какие знают);
 - поездка в райцентр (по каким делам, каким транспортом ездят);
 - игры детей (какие игры и как играют);
 - друзья детей (какие у детей друзья, где они живут);
 - животный мир около поселения (какие птицы и звери обитают).

2.2. ПРИНЦИПЫ ТРАНСКРИБИРОВАНИЯ ТЕКСТА

Транскрипция выполняется в «полуорфографической записи», широко распространенной транскрипции в речевых корпусных разработках. Эта транскрипция близка в фонематической, максимально приближена к орфографии современного башкирского языка. В нашей транскрипции использованы ряд специальных значков для обозначения звуков, которые невозможно передать при помощи башкирской орфографии.

Гласные фонемы:

Наряду с фонемами ə, o, θ, y, ɪ, ы, э(е) выделены следующие гласные: а̄ – широкий заднеязычный, а° – огубленный звук.

Согласные фонемы:

\hat{y} – билабиальная согласная (вақыт – \hat{y} ақыт),

γ̄ - билабиальная согласная (вэфэзэ - γ̄эфэзэ),

б̄ - билабиальная согласная (арба - арб̄а),

κ - гортанская согласная,

$\hat{\phi}$ - билабиальная согласная, близка к [ɸ],

ζ^{\wedge} - твердая верхнезубная согласная,

y° - огубленная гласная (ю^{\wedge} - $\text{й} y^{\circ} \text{к}$),

Y° -огубленная гласная (куп - $\text{к} \text{Y}^{\circ} \text{п}$),

i^{\wedge} - гласная близка к [e] (кил - $\text{ки}^{\wedge} \text{l}$),

йа – вместо буквы [я],

йу – вместо буквы [ю] (юл – йул),

йе – вместо буквы [e] в начале слова (емеш - йемеш).

При транскрибировании использовались специальные знаки, предложенные Шерстиновой Т.Ю. для аннотирования звукового корпуса русского языка «Один речевой день» [34].

Специальные знаки транскрибирования:

в транскрипции точки и запятые не ставятся;

/ — знак сегментирования синтагмы;

// — знак сегментирования фразы (членение на фразы и синтагмы осуществляется с учетом интонационно-сintаксических характеристик отрезков звуковой цепи);

! ? — вопросительное и восклицательное предложение;

· — знак длительной паузы в потоке речи информатора;

() — нак паузы хезитации, если она заполнена некоторыми звуками, соответствующие буквенные символы ставятся внутри скобок: напр., э (э-э);

*Н – если слово или словосочетание не поддается расшифровке или неуверенно расшифровывается;

... # — переключение языковых кодов (с башкирского на русский);

+ - для указания явления элизии (+ баралманы — бара алманы);

Транскрибирование и синхронизация транскрипта с аудиофайлом осуществляется в программе аннотирования ELAN.

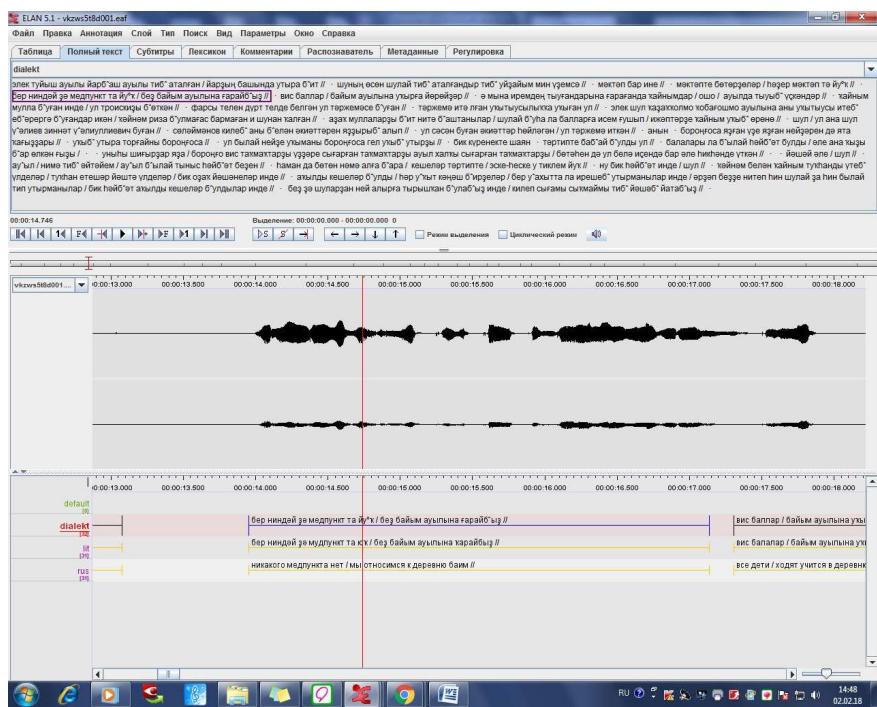


Рис. Скриншот транскрибирования диалектного материала на ELAN.

Речевой материал разбивается на фразовые промежутки, которые в дальнейшем транскрибируются в слое транскрипция и связываются со слоями представления диалектной фразы на современном башкирском литературном языке и переводом на русский язык. В дальнейшем это даст возможность включить материал в самостоятельный диалектный корпус башкирского языка с реализацией поиска по словарной и морфологической базам.

2.3. СРЕДСТВА РАБОТЫ С КОРПУСОМ АУДИО- И ТЕКСТОВЫХ ДАННЫХ

Собранный и аннотированный текстовый материал на первом этапе предполагается внедрить в Машинный фонд башкирского языка. На сегодняшний день разработана система поиска текстовых транскрипций по диалектам, говорам, по полу, образованию, возрасту информатора и дате записи информатора. Пользователь может ознакомиться с текстом, в дальнейшем по идентификатору загрузить с нашего сервера себе аудиофайл и аннотацию для автономной работы в программе ELAN. На Рис. 2. представлен скриншот выдачи запроса с учетом всех полей поиска.

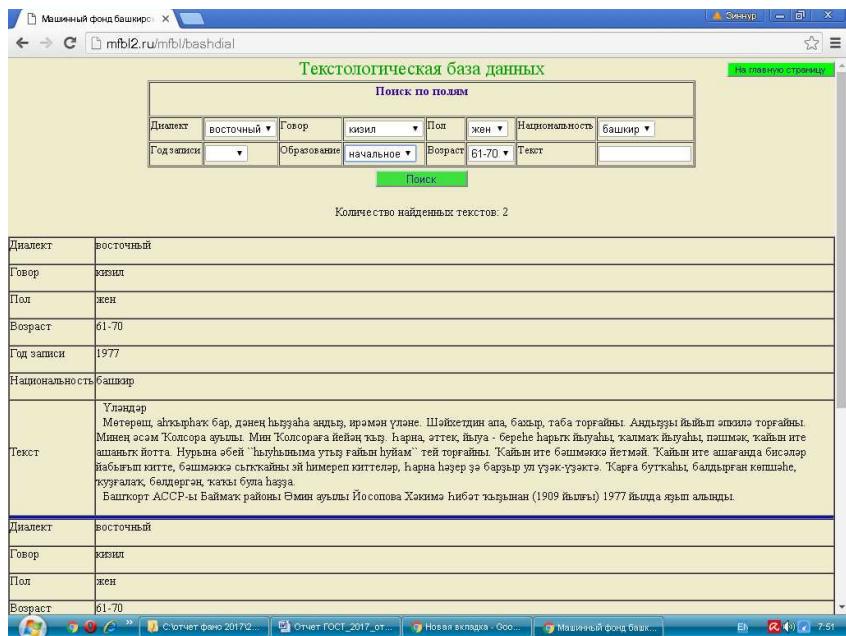


Рис. 2. Пример выдачи текста по сложному запросу.

В дальнейшем планируется разработать специализированный файловый менеджер, позволяющий искать как диалектные, так и литературные их варианты по транскрипции речевого материала. Предполагается выдача аннотированного контекста в размере фразы и аудио сопровождения. Данная работа потребует участия программиста.

2.4. ПРОВЕДЕННАЯ РАБОТА

- 1) Составлен реестр сохранившихся в ИИЯЛ архивных магнитофонных записей диалектологических и фольклорных экспедиций за прошлый век для включения в корпус.**
- 2) Проведена оцифровка и предварительная расшифровка пригодных аудиозаписей по айскому говору восточного диалекта (Мечетлинский и Кигинский районы РБ).**
- 3) Участниками проекта собран новый полевой материал по населенным пунктам Абзелиловского, Учалинского, Белорецкого районов Республики Башкортостан, население которых относится к кизилскому и миасскому говорам**

восточного диалекта башкирского языка. Материал включает 176 записей речи информаторов.

4) Оцифрованный материал прежних экспедиций и вновь собранный материал прошли первичную обработку на удаление шумов и длительных пауз в программе SoundForige.

5) Проведена паспортизация оцифрованного полевого материала.

Паспортизация включается как в само имя аудиофайла, так и в отдельный текстовый файл:

Имя файла включает следующие знаки:

- v- восточный диалект, aj – айский говор , ar – аргаяшский говор, sl – сальютский говор, mi – миасский говор, kz – кизильский говор;

- m – мужчина, w – женщина;

- образование: n – начальное образование, s – среднее образование, v – высшее образование;

- номер возрастной группы: 1- дошкольный и начальный класс (до 11 лет), 2 – средний школьный (от 11- 15 лет), 3 – старший и студенческий возраст (16-25 лет), 4 – средний возраст (25-45 лет), 5 – старший возраст (45-65 лет), 6 – пожилой возраст (от 65 лет).

- t1 - t15 – темы;

- доп коды информатора: от d001 до d999. Этот код является именем текстовых файлов.

В текстовых файлах указывается информация об информаторах: имя, фамилия, отчество, язык обучения, язык общения в семье, национальность информатора, место последнего долгого проживания до переезда в данное место, время проживания до последнего места проживания, время проживания в данном месте. Все данные разделяются знаком @.

Например: Финия Ахметовна Ибрагимова@б@б@б@Москово Учалинский район РБ@10@25 – имя этого файла d001.txt

4) Осуществлена сегментация цифрового речевого материала. Данная процедура проведена в программе ELAN и включает сегментирование речевого потока на фразы.

5) Проведено транскрибирование цифрового материала. Транскрибирование осуществлено в программе ELAN по разработанной нами методике транскрибирования полевого материала.

6) Осуществлено аннотирование цифрового материала. Аннотирование проведено в программе ELAN. В данную процедуру включено аннотирование таких фонетических явлений как элизия, протеза, выделены синтагматические единицы. Добавлены слои аннотирования «литературный» и «русский», в которых даны литературные варианты транскрипции речи и его перевод на русский язык.

7) Составлена база данных всей диалектной лексики башкирского языка с указанием диалектов и говоров объемом в 52000 единиц, получен 1 патент [34].

7) Разработана система ввода материала из формата ELAN в текстовую поисковую систему Машинный фонд башкирского языка (mfb12.ru). По базам данных Машинного фонда башкирского языка в 2017 г. получен 1 патент [35].

3. ЗАКЛЮЧЕНИЕ

Проект «Создание корпуса диалектных текстов башкирского языка» начал в 2017 г. За это время разработаны принципы подготовки и обработки материала. Обработан материал диалектологических и фольклорных экспедиций прошлого века. Обработан речевой материал новых экспедиций по двум говорам (из 5 запланированных на период реализации проекта). Общий объем нового полевого материала 1,5 Гб.

В 2018 г. предстоит:

- оцифровать и обработать оставшиеся материалы прежних экспедиций;

- провести сбор новых полевых материалов по айскому, аргаяшскому и сальютскому говорам;
- транскрибировать и аннотировать цифровые материалы;
- начать ввод материала в базу данных Машинного фонда башкирского языка в раздел “Диалектологическая база данных”.

Запланированная научно-исследовательская работа на 2017 год выполнена. Начат научный анализ полученных материалов. В рамках научного проекта коллективом опубликовано 12 научных статей [36-48] (из них 1 – ваковская, 1- скоповская), получены 2 патента.

4. ЛИТЕРАТУРА

1. Материалы диалектологической экспедиции (язык рабочих-башкир Баймакского, Тубинского заводов и башкир Янаульского и Учалинского районов//Научный архив УНЦ РАН, ф.3, оп.9, д.17.
2. Филми тикшеренеү институтының 1931 йыл июль айында Урал өлкәһенең ялан-катај төбәгө, Башкортостан йөмһүриятенең Арғаяш төбәгенә тел, фольклор тикшеренеүзәре өсөн ебәрелгән отрәттең алдыңғы хисабы//Научный архив УНЦ РАН, ф.3, оп.9, д.12.
3. Баишев Т.Г. Башкирские диалекты в их отношении к литературному языку. Уфа, 1953.
4. Киекбаев Дж.Г. Башкирские диалекты и краткое введение в их историю. Уфа, 1958.
5. Ишбулатов Н.Х., Максютова Н.Х., Миржанова С.Ф. Анкета для собирания материалов по говорам башкирского языка.
6. Башкорт һөйләштәренең һүҙлеге. Т.1. Көнсығыш диалект/Н.Х.Мәксұтова һ.б. Өфө, 1967, 300 б.
7. Башкорт һөйләштәренең һүҙлеге. Т.2. Көньяқ диалект/Н.Х.Мәксұтова һ.б. Өфө, 1970, 327 б.
8. Башкорт һөйләштәренең һүҙлеге. Т.3. Көнбайыш диалекты/Н.Х.Мәксұтова һ.б. Өфө, 1987, 231 б.
9. Ишбулатов Н.Х. Диалектная система башкирского языка в сравнительно-историческом освещении: Автореф. Дисс. ... д-ра филол. Наук. Уфа, 1974.
10. Максютова Н.Х. Восточный диалект башкирского языка в сравнительно-историческом освещении. М.:Наука, 1976, 293 с.
11. Миржанова С.Ф. Южный диалект башкирского языка. М.:Наука, 1979.

12. Миржанова С.Ф. Северо-западный диалект башкирского языка. Уфа, 1991.
13. Машинный фонд башкирского языка. Диалектологический атлас башкирского языка. URL: <http://www.mfbl2.ru> (дата обращения: 24.12.2017).
14. Das Deutsche Referenzkorpus (DeReKo) URL:<http://www.ids-mannheim.de/kl/projekte/korpora> (дата обращения: 25.12.2017).
15. British National Corpus (BNC). URL:<http://www.natcorp.ox.ac.uk> (дата обращения: 25.12.2017).
16. The corpus of contemporary american english (COCA). URL:<http://corpus.byu.edu/coca> (дата обращения: 25.12.2017).
17. Корпус китайского языка. (LIVAC Synchronous Corpus). URL:<http://www.rcl.cityu.edu.hk/livac> (дата обращения: 25.12.2017).
18. Magyar Nemzeti Szövegtár (корпус венгерского языка). URL:<http://corpus.nytud.hu/mnsz> (дата обращения: 25.12.2017).
19. Corpus del español (корпус испанского языка). URL:<http://www.corpusdelespanol.org> (дата обращения: 25.12.2017).
20. Corpus di riferimento della lingua italiana scritta contemporanea (“CoLFIS”) (корпус итальянского языка) URL:<http://www.ge.ilc.cnr.it/dizionari.php> (дата обращения: 25.12.2017).
21. Český národní korpus (ČNK) (чешский национальный корпус). URL:<http://ucnk.ff.cuni.cz> (дата обращения: 25.12.2017).
22. Национальный корпус русского языка. URL:<http://www.ruscorpora.ru> (дата обращения: 25.12.2017).
- 23 LDC Top Ten Corpora (мультимедийные корпусы английского языка). URL: <http://www.ldc.upenn.edu/Catalog/topten.jsp> (дата обращения: 17.12.2017).
24. Фонетический немецкого разговорного языка URL:http://dsav-wiss.ids-mannheim.de/korpora/pf_pf_doku.htm (дата обращения: 25.01.2017).

25. Chinese Broadcast Conversation Speech (мультимедийный корпус китайского языка). URL: <http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2013S04> (дата обращения: 17.12.2017).
26. The Corpus of Spontaneous Japanese (мультимедийный корпус японского языка). URL: <http://www.ninjal.ac.jp/products-k/katsudo/seika/corpus/public/index.html> (дата обращения 25.12.2017).
27. *The Spoken Turkish Corpus* (мультимедийный корпус разговорного турецкого языка). URL:<http://stc.org.tr> (дата обращения: 25.12.2017).
28. Фонетический корпус спонтанной эстонской речи.
URL:<http://www.murre.ut.ee/phnetic-corpus> (дата обращения: 25.12.2017).
29. Фонетический корпус “Один день русского языка”
URL:<http://www.model.org.spbu.ru> (дата обращения: 17.01.2018, доступ ограничен).
30. Корпус ижемского диалекта коми языка URL:<http://videocorpora.ru/ru/allmedia> (дата обращения: 25.12.2017).
31. Саратовский мультимедийный диалектный корпус.
ruscorpora.ru/sbornik2005/13letuchy.pdf. (дата обращения: 25.01.2018).
32. Мультимедийный корпус диалектных текстов Устьянского района Архангельской области URL: <http://www.slavist.de/Pushkino/> (дата обращения: 25.01.2018).
33. Диалектный подкорпус национального корпуса русского языка URL: <http://www.ruscorpora.ru/search-dialect.html> (дата обращения: 25.01.2018).
34. Шерстинова Т. Ю., Степанова С. Б., Рыко А. И. Система аннотирования в звуковом корпусе русского языка «Один речевой день» // Материалы XXXVIII международной филологической конференции. Формальные методы анализа русской речи. 16–20 марта 2009 г. Санкт-Петербург. СПб., 2009.
35. Свидетельство о государственной регистрации базы данных. №2017620490 “Базы данных “Генеральный словарь башкирского языка”. Правообладатель Федеральное государственное бюджетное учреждение науки Институт истории, языка и литературы Уфимского научного центра Российской академии наук. Авторы: Сиразитдинов З.А.,

Бускунбаева Л.А., Ишмухаметова А.Ш., Полянин А.И., Шамсутдинова Г.Г. Дата гос. Регистрации 2 мая 2017 г.

36. Свидетельство о государственной регистрации базы данных. №2017620474 “Лексическая база данных (диалектологическая)”. Правообладатель Федеральное государственное бюджетное учреждение науки Институт истории, языка и литературы Уфимского научного центра Российской академии наук. Авторы: Сиразитдинов З.А., Бускунбаева Л.А., Ишмухаметова А.Ш., Полянин А.И., Шамсутдинова Г.Г. Дата гос. Регистрации 27 апреля 2017 г.

37. Сиразитдинов З.А., Бускунбаева Л.А., Ишмухаметова А.Ш., Шамсутдинова Г.Г. Разработка корпуса текстов афористических жанров башкирского языка // Al-Farabi 1st international congress on social sciences (сборник тезисов). – Газиантеп, 2017. С. 201.

38. Сиразитдинов З.А., Бускунбаева Л.А., Ишмухаметова А.Ш., Шамсутдинова Г.Г. К проблеме создания корпуса текстов афористических жанров башкирского фольклора // Актуальные проблемы диалектологии языков народов России: Материалы XVII Всероссийской научной конференции. Уфа, 2017. – С. 60-63.

39. Сиразитдинов З.А., Бускунбаева Л.А., Ишмухаметова А.Ш., Шамсутдинова Г.Г. **Башҡорт фольклорының афористик жанрҙарына қараган текстар корпусын төҙөү проблемаһына қарата//Сборник материалов научной конференции: «Цифровые технологии в языкознании: создание национального корпуса казахского языка», Алматы, 17 ноября 2017 года. С.160-164.**

40. Сиразитдинов З.А., Бускунбаева Л.А., Ишмухаметова А.Ш., Шамсутдинова Г.Г. **О подкорпусе афористических жанров башкирского фольклора//Сборник материалов V Международной конференции по компьютерной обработке тюркских языков TurkLang’2017. Казань 18-21 октября 2017. С.329-335.**

41. Сиразитдинов З.А., Бускунбаева Л.А., Шамсутдинова Г.Г. Структура словарей в лексикографическом подфонде машинного фонда башкирского языка//Электронная письменность народов Российской Федерации: опыт, проблемы и перспективы. Сыктывкар, 2017. С.190-195.

42. Бускунбаева Л.А., Сиразитдинов З.А., Ишмухаметова А.Ш. Состав и структура корпуса публицистики башкирского языка//Электронная письменность народов Российской Федерации: опыт, проблемы и перспективы. Сыктывкар, 2017. С.39-44.

43. Сиразитдинов З. А. Информационные технологии в реализации Закона “О языках народов Республики Башкортостан: этапы и перспективы//Сборник материалов Всероссийской научно-практической конференции “Актуальные вопросы алтайского языкознания: проблемы развития литературного языка, совершенствования современной орфографии, посвященной 115-летию Т.М.Тоощаковой. Горно-Алтайск, 26-28 апреля, 2017. С.136-152.

44. Buskunbaeva L.A., Sirazitdinov Z.A., Ishmukhametova A.Sh., Shamsutdinova G.G. THE CREATION OF CORPUS OF APHORISTIC TEXTS OF THE BASHKIR LANGUAGE //В сборнике: ICSSCA-V 5th International Congress on Social Sciences, China to Adriatic Congress book of full texts. – Tokyo, 2017. С. 82-88.

45. Шамсутдинова Г.Г., Ишмухаметова А.Ш., Бускунбаева Л.А. Структура и состав базы данных загадок в подкорпусе текстов афористических жанров башкирского фольклора// Вестник Калмыцкого института гуманитарных исследований РАН. Элиста, 2017. №4. С. 118-125.

46. Норманская Ю.В., Экба З.Н., Каримова Р.Н. К вопросу о диалектных особенностях первых башкирских книг на кириллице//Актуальные проблемы диалектологии языков народов России Материалы XVII Всероссийской научной конференции. 2017. С. 195-197.

47. Норманская Ю.В., Экба З.Н., Каримова Р.Н. О диалектной принадлежности «Букваря для башкир» А.Г.Бессонова // Дмитриевские чтения Программа и тезисы XXV Международной научной конференции, посвященной памяти Дмитрия Михайловича Насилова. 2017. С. 5.

48. Норманская Ю.В., Каримова Р.Н., Экба З.Н. В.В.Катаринский — автор первой кириллической книги на башкирском языке? // Урало-алтайские исследования. 2017. № 2 (25). С. 46-52.