

УДК 004.658:801.3=512.141

¹З. А. Сиразитдинов, ²Л. А. Бускунбаева,
³А. Ш. Ишмухаметова, ⁴Г. Г. Шамсутдинова

*Ордена «Знак Почета» Институт истории, языка и литературы
УФИЦ РАН, проспект Октября, 71, Уфа, 450054, Россия
¹sazin11@mail.ru, ²buskl@yandex.ru
³ishmuhametova_anita@mail.ru, ⁴moon07@yandex.ru*

Проблемы разработки диалектного корпуса башкирского языка

В данной статье рассматривается проблема разработки диалектного корпуса башкирского языка, принципы сбора и обработки экспедиционных материалов по восточному диалекту.

В корпусе полевой материал, записанный в виде аудиозаписи на диктофон, транскрибируется, представляется в литературной форме и переводится на русский язык. Оснащенный экстралингвистической разметкой корпус предоставит сведения о диалекте (диалект, говор), информанте (ФИО, пол, возраст, национальность, образование и др.), тематике беседы, записи (время записи). Разнообразие диалектных текстов с точки зрения половозрастного, социального, профессионального состава информантов позволит проводить наблюдения над современным состоянием речи того или иного говора или диалекта башкирского языка.

Ключевые слова: башкирский язык, тюркские языки, диалектология, диалект, лингвистический корпус, корпусная лингвистика.

Каждый говор, каждый диалект является важнейшей частью башкирского языка. Однако сегодня влияние литературного языка на диалекты очень сильное, что проявляется, прежде всего, в активизации процесса исчезновения диалектных особенностей языка. Следовательно, одной из основных задач языковедов является сбор диалектных материалов с широким охватом носителей для дальнейшего всестороннего анализа последующим поколением лингвистов.

Изучение языковых особенностей диалектов и говоров на материале полевых записей, при этом устанавливая характер фонетических, лексических и грамматических черт каждого говора, представляет большую ценность не только для диалектологических исследований, но и для смежных дисциплин, изучающих историю развития диалектов, историю взаимоотношений с соседними народами, этнографию башкирского народа, географию и т. д.

Вопрос более детального изучения башкирских диалектов не раз ставился башкирскими учеными. Например, известный диалектолог Н. Х. Максютлова призывает обратить взор диалектолога на «историю формирования диалектов, выяснение внутренней истории диалектных явлений в связи с историей этноса. В башкирской диалектологической науке еще не изучены языковые явления, связанные с наличием различных этнических компонентов как тюркского, так и нетюркского происхождения; не раскрыты древнетюркский, древнебашкирский пласты явлений и их трансформации, не объяснены причины происхождения явлений и признаков, составляющих специфику в грамматическом, лексическом и синтаксическом строе башкирского языка» [Максютлова, 1996, 10–11].

Башкирский языковед-тюрколог А. А. Юлдашев отмечает неопределимую роль диалектного материала «для воссоздания истории диалектов и башкирского языка в целом» [Юлдашев, 1959, 213–214].

Известный ученый Р. З. Шакуров поднимает вопрос о необходимости сравнительного изучения фонетики, грамматики и лексики говоров башкирского

языка. «Такие исследования приобрели бы большую значимость для определения связанных друг с другом глубокими корнями диалектов и говоров как единой системы, поистине уникального и неповторимого места, которое занимает башкирский язык среди языков других народов Евразии» [Шакуров, 2012, 43].

Для осуществления данных целей нужен богатый фактический материал по диалектам и говорам башкирского языка.

Масштабное исследование башкирских диалектов началось в 20—30-е годы XX века. Большой вклад в изучение диалектов внесли видные башкирские ученые Г. С. Амантаев, Т. Г. Баишев, Г. Я. Давлетшин, З. Ш. Шакиров и др., которые в ходе многочисленных экспедиций выявили основные границы и черты говоров и диалектов языка. Большое значение имели научные отчеты Т. Г. Баишева и З. Ш. Шакирова [Архив1; Архив2], 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.

В 50-е годы башкирская диалектология приступила к системному изучению фонетики, лексики и морфологии говоров. Появление монографических трудов Т. Г. Баишева [Баишев, 1953], Дж. Г. Киекбаева [Киекбаев, 1958] 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» [Ишбулатов и др., 1961], 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 языкознании [Словарь башкирских говоров, 1967; Словарь башкирских говоров, 1970; Словарь башкирских говоров, 1987]. Материалы этих словарей в значительной части вошли в реестры двухтомного толкового и многотомного академического словарей башкирского языка, обогатив литературный язык.

Материалы опросников экспедиций 60–70-х годов были изучены в сравнительно-сопоставительном аспекте и легли в основу диссертационных и монографических исследований [Ишбулатов, 1974; Максютлова, 1976; Миржанова, 1979; Миржанова, 1991]. Эти работы послужили основой для перехода к лингвогеографическому изучению говоров и диалектов башкирского языка.

В 1973 г. башкирскими диалектологами была разработана единая 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.

Материалы Атласа и вышеуказанных словарей башкирских говоров введены исполнителями данной научной темы в диалектологическую базу данных Машинного фонда башкирского языка [Сиразитдинов, 2008] и размещены в открытом доступе по адресу: mfbl2.ru.

Полевые материалы прежних лет до сих пор привлекают внимание исследователей в исторических исследованиях башкирского языка, являются источником этнолингвистического анализа определения диалектной языковой идентичности и мифологического воззрения башкирского народа [Батыршина, 2007; Ишегулова, 2011; Ишкильдина, 2013; Каримова, 2012; Муратова, 2008; Сулейманова, 2006; Хисамитдинова, 2010].

К сожалению, с 80-х годов XX в. 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 информантов по

поло-возрастным и другим социальным группам отсутствует. Большинство материалов прежних экспедиций не было оцифровано, большая часть утрачена.

За более чем 30 лет в говорах и диалектах, несомненно, произошли сдвиги и изменения, которые требуют фиксации и научного анализа. Современная лингвистика выдвигает новые подходы к исследованию говоров языка, способствующие более быстрой и объективной обработке языковых данных. Все это актуализирует задачу сбора и анализа полевых материалов сегодняшнего дня на новых методологиях и подходах по говорам башкирского языка. Например, использование электронных лингвистических баз данных не только значительно ускоряет и оптимизирует трудоёмкий процесс сбора и обработки диалектного материала, но и ведёт к смене научной парадигмы в лингвистике [Кузнецова, 2011, 96].

Одним из ведущих направлений современной лингвистики является корпусное исследование, объектом которого являются речевые материалы, реализованные в виде как письменных текстов, так и устных (фонетических) массивов данных. На сегодняшний день в мире насчитывается более тысячи корпусов, их количество растёт с каждым годом экспоненциально.

Количество и объёмы речевых корпусов значительно отстают от текстовых, но темп составления таких корпусов с каждым годом растёт. Так, по английскому [LDC Top Ten Corpora], немецкому [Фонетический немецкого разговорного языка], китайскому [Chinese Broadcast Conversation Speech], японскому [The Corpus of Spontaneous Japanese], турецкому [The Spoken Turkish Corpus], эстонскому [Фонетический корпус спонтанной эстонской речи], русскому [Фонетический корпус “Один день русского языка”] языкам реализованы корпусы, содержащие как мультимедийные данные, так и транскрипции речи.

Сформированные идеи и подходы корпусной лингвистики активно используются применительно и к диалектологическим материалам, при составлении 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, посвящённых разновидности немецкого языка (Sprachvarietäten) на территории Германии и за ее пределами. В России известны корпус ижимского говора коми языка [Коми медиатека], корпусы говоров русского языка [Саратовский мультимедийный диалектный корпус; Мультимедийный корпус диалектных текстов Устьянского района Архангельской области; Диалектный подкорпус Национального корпуса русского языка] и др.

Сбалансированный и репрезентативный корпус диалектных текстов призван решать многие исследовательские задачи на новом уровне, выступая в свою очередь не только как справочный материал.

Создание электронных диалектных корпусов позволяет зафиксировать полевые материалы в единой базе данных и пополнять по мере фиксации.

Но в то же время корпусы диалектных текстов существенно отличаются от корпусов, созданных на письменном материале литературного языка, и составляются по определенным критериям. Создание такого типа корпусов сопряжено с целым рядом сложностей, куда относятся системные языковые отличия от литературного языка; исключительно устный характер диалектной коммуникации, вариативность на всех уровнях, затрудняющая идентификацию единиц в корпусе; отсутствие единообразия при фиксации диалектной речи и различные способы организации информации [Москвина, 2014]. Ввиду специфичности диалектного материала, процесс подготовки данных текстов и включение их в корпус более трудоёмкий и занимает больше времени и усилий, поскольку включает в себя запись информанта, расшифровку аудиофайла, транскрибирование, ввод в базу данных и разработку экстралингвистических и лингвистических разметок. Как отмечают разработчики Саратовского диалектологического корпуса “это не механическая, а исследовательская

работа, серьезный и очень ответственный авторский труд» [Крючкова, 2011, 363].

Задача создания корпуса башкирских диалектов впервые поднимается в башкирской лингвистической науке. В то же время у коллектива имеется большой научный задел по разработке лингвистических корпусов башкирского языка. На сегодняшний день в сети Интернет функционируют корпуса прозы, публицистики и фольклора, снабженные экстралингвистической и лингвистической (морфологической и семантической) системой разметок [Сиразитдинов, 2012; Сиразитдинов 2012А; Сиразитдинов, 2014].

Разработка диалектного электронного корпуса является одной из актуальных задач современной филологии. С одной стороны, диалектный материал, собранный усилиями сотрудников Лаборатории лингвистики и информационных технологий в ходе экспедиций по опорным пунктам не только Республики Башкортостан, но и соседних регионов, где компактно проживает башкирское население, будет размещен в одной базе данных, позволяющей открыть доступ широкому кругу пользователей, интересующимся данной проблематикой. С другой стороны, исследование диалектных текстов, опираясь на богатый фактический полевой материал, является одним из факторов для сохранения не только родного языка, но и этнической самобытности, национальной культуры в условиях глобализации и урбанизации современного общества.

Для получения объективных результатов по диалектам и говорам башкирского языка необходимо создание репрезентативного корпуса диалектных текстов, максимально охватывающий все говоры и содержащий богатую металингвистическую разметку представленного текста, как возраст, гендерная принадлежность, уровень образования, возраст, язык общения, национальность информантов, разнообразие тем для беседы и т.д. В корпусе будет представлена аудиозапись, расшифровка звуковых файлов в виде транскрибированных текстов, их литературный вариант и русский перевод диалектного текста.

Такой корпус будет иметь не только ознакомительный характер, но и выступать в качестве бесценного источника для изучения отдельного говора, установления территории распространения того или иного языкового явления. Богатый диалектный материал, содержащийся в корпусе, станет базой для изучения в том числе и исторического развития и становления литературного языка, сравнительно-сопоставительных исследований языков.

На данном этапе работы сотрудники Лаборатории лингвистики и информационных технологий начали сбор и обработку речевого материала по восточному диалекту башкирского языка.

По каждому населенному пункту производится запись с охватом половозрастных групп. Выделяются следующие возрастные группы: дошкольный и начальный класс (до 11 лет), средний школьный (от 11–15 лет), старший и студенческий возраст (16–25 лет), средний возраст (25–45 лет), старший возраст (45–65 лет), пожилой возраст (от 65 лет).

При записи полевого материала учитывается образование информанта: начальное, среднее, высшее.

Аудиозапись сопровождается информацией, которая составляет экстралингвистическую разметку аудиофайла и включает следующее:

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

– время проживания до последнего места проживания (в случае переезда);

– время записи.

Выделяется тип общения: монолог, диалог, полилог.

Для записи информантов определены следующие 15 тем:

– свадьба, свадебные обычаи;

– питание, блюда (что любят готовить на каждый день, что на праздниках);

– домашний скот, его содержание;

– дети и близкие родственники (система родства);

– приусадебное хозяйство (огород, сад);

– дом (когда построен, кто строил, крыша, рамы);

– топонимия в окрестностях поселения;

– история села, школы, рода;

– работа, школа, друзья;

– времена года, погода;

– фольклор (какие частушки, пословицы, поговорки, сказки знают);

– поездка в райцентр (по каким делам, каким транспортом пользуются);

– 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).

Запись осуществляется на цифровой диктофон в несжатом формате (wav, 16бит/22kHz – 16бит/48kHz), при отсутствии посторонних звуков, хотя отдельными исследователями и отмечается, что записи в потенциально шумных местах также представляют интерес, поскольку «они отражают разные коммуникативные ситуации» [Куканова, 2013, 57]. Первичная обработка аудиозаписей (очистка от посторонних шумов и длительных пауз), паспортизация производится в программе SoundForge.

Паспортизация файлов и экстралингвистическая разметка частично включается в имена аудиофайлов:

1. v – восточный диалект, aj – айский говор, ar – аргаяшский говор, sl – сальютский говор, mi – миасский говор, kz – кизильский говор, uc – учалинский говор;

2. m – мужчина, w – женщина;

3. образование: n – начальное образование, s – среднее образование, v – высшее образование;

4. 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);

5. t1 – t15 – темы;

Например, файл varwvt84d003 – означает принадлежность звукового файла восточному диалекту, аргаяшскому говору, в котором осуществлена запись на тему “история села, школы, рода” женщиной среднего возраста с высшим образованием.

Дополнительные коды информанта: от d001 до d999. Этот код является именем текстового файла, в котором указывается имя, фамилия, отчество, язык обучения, язык общения в семье, национальность информанта, место последнего долгого проживания до переезда в данное место (в случае переезда), время проживания до последнего места проживания (в случае переезда), время проживания в данном месте, дата записи. Все данные разделяются знаком @. Например: ЗилдаАхмадиевнаМавлетова@б@б@б@Мишкино Учалинский район РБ@10@25@12.08.2017 – имя этого файла d005.txt.

Следует отметить, что многие проблемы башкирской диалектологии до сих пор остаются вне поля зрения лингвистов, например, относительно слабо изучен диалектный синтаксис, процесс взаимовлияния и взаимообогащения диалектов, требуют более детального изучения говоры северо-западного диалекта. Ждут своей очереди проблемы исследования в сопоставительном

аспекте диалектных данных башкирского языка с тюркскими, финно-угорскими и другими языками.

Таким образом, создаваемый корпус будет включать в себя богатый диалектный материал, что станет бесценным источником для последующих изысканий в области фонетики, лексики и грамматики башкирских говоров и диалектов башкирского языка.

Список литературы

- Архив 1 — Архив Уфимского научного центра РАН (УНЦ РАН). Фонд 3, оп. 9, ед. хр. 17.
 Архив 2 — Архив Уфимского научного центра РАН (УНЦ РАН). Фонд 3, оп. 20, ед. хр. 12.
 Баишев Т. Г. Башкирские диалекты в их отношении к литературному языку. Уфа, 1953. 132 с.
 Батыршина Г. Р. Лексика послеродовых обрядов башкир (этнолингвистический анализ) // Вопросы филологии. 2007. № 4. С. 167–169.
 Диалектный подкорпус национального корпуса русского языка URL: <http://www.ruscorpora.ru/search-dialect.html> (дата обращения: 25.04.2018).
 Ишбулатов Н. Х., Максютлова Н. Х., Миржанова С. Ф. Анкета для собирания материалов по говорам башкирского языка. Уфа, 1961. 56 с.
 Ишбулатов Н. Х. Диалектная система башкирского языка в сравнительно-историческом освещении: Автореф. дисс. ... д-ра филол. наук. Уфа, 1974.
 Ишегулова А. М. Диалектная языковая личность // Труды Института истории, языка и литературы Уфимского научного центра РАН. Уфа, 2011. С. 184–186.
 Ишкильдина Л. К. Историческое развитие консонантизма башкирского языка (на материале диалектов). Автореферат. дис. ... канд. филол. наук. Казань, 2013. 28 с.
 Каримова Р. Н. Проблемы и перспективы изучения диалектов языков народов России (сообщение о XI межрегиональной конференции «Актуальные проблемы диалектологии языков народов России») // Известия Уфимского научного центра Российской академии наук. 2012. № 1. С. 69–72.
 Киембаев Дж. Г. Башкирские диалекты и краткое введение в их историю. Уфа, 1958.
 Коми медиатека. URL: <http://videocorpora.ru/ru/allmedia> (дата обращения: 25.12.2017).
 Корпус китайского языка. (LIVACSynchronousCorpus). URL: <http://www.rcl.cityu.edu.hk/livac> (дата обращения: 25.04.2018).
 Крючкова О. Ю., Гольдин В. Е. Корпус русской диалектной речи: концепция и параметры оценки // Компьютерная лингвистика и интеллектуальные технологии: Материалы Международной конференции «Диалог-2011». Бекасово, 2011. С. 359–367.
 Кузнецова Е. В. Проблемы и перспективы изучения лингвогеографического ландшафта региона (на примере Волгоградской области) // Гуманитарный вектор. 2011. № 4 (28). С. 96–101.
 Куканова В. В., Бембеев Е. В., Убушаев Н. Н., Манджиев Б. Б. Устные тексты на калмыцком языке: запись и расшифровка // Вестник Калмыцкого университета. 2013. №3(19). С. 56–64.
 Максютлова Н. Х. Восточный диалект башкирского языка в сравнительно-историческом освещении. М.: Наука, 1976.
 Максютлова Н. Х. Башкирские говоры, находящиеся в иноязычном окружении. Уфа: Китап, 1996.
 Миржанова С. Ф. Южный диалект башкирского языка. М.: Наука, 1979.
 Миржанова С. Ф. Северо-западный диалект башкирского языка. Уфа. 1991.
 Москвина Т. Н. Методы и подходы корпусной лингвистики в исследованиях семантики диалектной лексики // Современные проблемы науки и образования. 2014. № 6.; URL: <http://science-education.ru/ru/article/view?id=15784> (дата обращения: 25.04.2018)
 Мультимедийный корпус диалектных текстов Устьянского района Архангельской области URL: <http://www.slavist.de/Pushkino/> (дата обращения: 25.01.2018).
 Мурагова Р. Т. Символика чисел в мифологических представлениях башкир о мире // Актуальные проблемы истории, языка и культуры Башкортостана. Уфа, 2008. С. 178–183.
 Corpus diriferimento della lingua italiana scritta con temporanea (“CoLFIS”) (корпус итальянского языка) URL: <http://www.ge.ilc.cnr.it/dizionari.php> (дата обращения: 25.04.2018).
 Саратовский мультимедийный диалектный корпус. ruscorpora.ru/sbornik2005-13letuchy.pdf. (дата обращения: 25.04.2018).

Сиразитдинов З.А. О диалектологической подсистеме машинного фонда башкирского языка // Сборник материалов Всероссийской научно-практической конференции Республика Башкортостан «Теория и практика башкирского языка и литературы в свете современных достижений филологических наук» Уфа, 2008. С. 178–182.

Сиразитдинов З. А., Бускунбаева Л. А., Бардыбаева А. Д., Ишмухаметова А. Ш., Рафикова Л. Н. О разработке национального корпуса башкирского языка // Роль и место национальной библиотеки в социокультурном пространстве: материалы Международной научно-практической конференции. Уфа, 2012. С. 118–126.

Сиразитдинов З. А., Бускунбаева Л. А., Ишмухаметова А. Ш., Ибрагимова А. Д., Мигранова Л. Г. Корпус текстов периодической печати на башкирском языке // Актуальные проблемы диалектологии языков народов России: Материалы XII региональной конференции. Уфа, 2012. С. 139–141.

Сиразитдинов З.А., Бускунбаева Л.А., Ишмухаметова А.Ш., Ибрагимова А.Д. О создании корпуса башкирского фольклора // Урал-Алтай: через века в будущее: Материалы VI Всероссийской тюркологической конференции (с международным участием). Уфа, 2014. С. 86–89.

Словарь башкирских говоров. Т. 1. Восточный диалект / Н. Х. Максютова и др. Уфа, 1967. 300 с.

Словарь башкирских говоров. Т. 2. Южный диалект / Н.Х.Максютова и др. Уфа, 1970. 327 с.

Словарь башкирских говоров. Т. 3. Западный диалект / Н.Х. Максютова и др. Уфа, 1987. 231 с.

Сулейманова Р.А. Антропонимы современного башкирского языка и тенденции их развития. Дисс. канд. филол. наук. Уфа, 2006. 189 с.

Фонетический корпус “Один день русского языка” URL:<http://www.model.org.-spbu.ru> (дата обращения: 25.04.2018, доступ ограничен).

Фонетический корпус спонтанной эстонской речи. URL:<http://www.murre.ut.ee-/phonetic-corpus> (дата обращения: 25.04.2018).

Фонетический корпус немецкого разговорного языка URL:http://dsav-wiss.ids-mannheim.de/korpora/pf/pf_doku.htm (дата обращения: 25.04.2018).

Шакуров Р. З. Диалектная система башкирского языка // Ватандаш. 2012. № 8. С. 40–61.

Хисамитдинова Ф. Г. Мифологический словарь башкирского языка. – М., 2010. 452 с.

Юлдашев А. А. К методике сбора диалектологического материала по башкирскому языку на современном этапе // Башкирский диалектологический сборник. Уфа, 1959. С. 213–221.

British National Corpus (BNC). URL:<http://www.natcorp.ox.ac.uk> (дата обращения: 25.12.2017).

Český národní korpus (ČNK) (чешский национальный корпус). URL:<http://ucnk.ff.cuni.cz> (дата обращения: 25.04.2018).

Chinese Broadcast Conversation Speech (мультимедийный корпус китайского языка). URL: <http://www ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId= LDC2013S04> (дата обращения: 25.04.2018).

Corpus del español (корпус испанского языка). URL:<http://www.corpusdelespanol.org> (дата обращения: 25.04.2018).

Das Deutsche Referenzkorpus (DeReKo) URL:<http://www.ids-mannheim.de/kl/projekte/korpora> (дата обращения: 25.04.2018).

LDC TopTenCorpora (мультимедийные корпуса английского языка). URL: <http://www ldc.upenn.edu/Catalog/topTen.jsp> (дата обращения: 17.12.2017).

Magyar Nemzeti Szövegtár (корпус венгерского языка). URL:<http://corpus.nytud.-hu/mnsz> (дата обращения: 25.04.2018).

The corpus of contemporary american english (COCA). URL:<http://corpus.byu.edu/coca> (дата обращения: 25.04.2018).

The Spoken Turkish Corpus (мультимедийный корпус разговорного турецкого языка). URL:<http://stc.org.tr> (дата обращения: 25.04.2018).

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: 25.04.2018).

¹Z. A. Sirazitdinov, ²L. A. Buskunbaeva,
³A. SH. Ishmuhametova, ⁴G. G. Shamsutdinova

Order of the Badge of Honour
 Institute of History, Language and Literature UFRC
 Russian Academy of Sciences, October Avenue, 71,
 Ufa, 450054, Russian Federation
¹sazin11@mail.ru, ²buskl@yandex.ru
³ishmuhametova_anita@mail.ru, ⁴moon07@yandex.ru

Problems of development of the dialect corpus of the Bashkir language

This article deals with the problem of developing the dialect corpus of the Bashkir language, the principles of collecting and processing expedition materials in the eastern dialect.

In the case, the field material recorded as an audio recording on the recorder will be transcribed, presented in a literary form and translated into Russian. Equipped with extralinguistic marking, the corpus will provide information on dialect (dialect, sub-dialect), informant (name, sex, age, nationality, education, etc.), topics of conversation, recording (recording time). The variety of dialect texts from the point of view of the age, social, professional composition of informants will allow us to observe the modern state of speech of a dialect or dialect of the Bashkir language.

Keywords: the Bashkir language, Turkic languages, dialectology, dialect, linguistic corpus, corpus linguistics.

References

Arhiv 1 — Arhiv Ufimskogo federal'nogo issledovatel'skogo centra RAN (UFIC RAN) [The Archives of the Ufa Federal Research Centre of the RAS (UFRC RAS)]. Fund 3, bordereau 9, individual file 17.

Arhiv 2 — Arhiv Ufimskogo federal'nogo issledovatel'skogo centra RAN (UFIC RAN) [The Archives of the Ufa Federal Research Centre of the RAS (UFRC RAS)]. Fund 3, bordereau 20, individual file 12.

Baishev T. G. *Bashkirskie dialekty v ihotnoshenii k literaturnomu yazyku* [Bashkir dialects in their relation to the literary language]. Ufa, 1953, 132 p.

Batyrshina G. R. *Leksika poslerodovyh obryadov bashkir (ehtnolingvisticheskij analiz)* [Vocabulary of post-natal rites of Bashkirs (ethnolinguistic analysis)]. In: *Voprosy filologii* [Questions of philology]. 2007, № 4, pp. 167–169.

Dialektnyj podkorpus nacional'nogo korpusa russkogo yazyka [Dialectic subcorpus of the national corpus of the Russian language]. URL: <http://www.ruscorpora.ru/search-dialect.html> (date of circulation: 25.04.2018).

Ishbulatov N. Kh., Maksyutova N. Kh., Mirzhanova S. F. *Anketa dlya sobiraniya materialov po govoram bashkirskogo yazyka* [Questionnaire for collecting materials on the dialects of the Bashkir language]. Ufa, 1961, 56 p.

Ishbulatov N. Kh. *Dialektnaya sistema bashkirskogo yazyka v sravnitel'no-istoricheskom osveshchenii* [Dialect system of the Bashkir language in comparative-historical illumination]. Author's abstract. diss. ... Dr. Philol. sciences]. Ufa, 1974.

Ishegulova A. M. *Dialektnaya yazykovaya lichnost'* [Dialectal language personality]. In: *Trudy Instituta istorii, yazyka i literatury Ufimskogo nauchnogo centra RAN* [Proceedings of the Institute of History, Language and Literature of the Ufa Science Center of the Russian Academy of Sciences]. Ufa, 2011, pp. 184–186.

Ishkildina L. K. *Istoricheskoe razvitiye konsonantizma bashkirskogo yazyka (na materiale dialektov)* [Historical development of the consonantism of the Bashkir language (on the basis of dialects)]. Abstract. diss. ... cand. philol. sciences. Kazan, 2013. 28 p.

Karimova R. N. *Problemy i perspektivy izucheniya dialektov yazykov narodov Rossii (soobshchenie o XI mezhhregional'noj konferencii «Aktual'nye problem dialektologii yazykov narodov Rossii»)* [Problems and perspectives of the study of dialects of the languages of the peoples of Russia (report on the XI interregional conference "Actual problems of the dialectology of the languages of the peoples of Russia")]. In: *Izvestiya Ufimskogo nauchnogo centra Rossijskoj akademii nauk* [Izvestiya Ufa Scientific Center of the Russian Academy of Sciences]. 2012, № 1, P. 69–72.

Kiekbayev Dzh. G. *Bashkirskie dialekty i kratkoe vvedenie v iistoriyu* [Bashkir dialects and a brief introduction to their history]. Ufa, 1958.

Komi mediateka [*Komi Media Library*]. URL:<http://videocorpora.ru/ru/allmedia> (date of circulation: 25.04.2018).

Korpus kitajskogo yazyka [The Corpus of the Chinese language] (LIVAC Synchronous Corpus). URL:<http://www.rcl.cityu.edu.hk/livac> (date of circulation: 25/12/2017).

Kryuchkova O. Yu., Goldin V. E. *Korpus russkoj dialektnoj rechi: koncepciya i parametry ocenki* [Corpus of Russian dialect speech: the concept and parameters of evaluation]. In: *Komp'yuternaya lingvistika i intellektual'nye tekhnologii: Materialy Mezhdunarodnoj konferencii «Dialog-2011»* [Computer linguistics and intellectual technologies: Materials of the International Conference "Dialogue-2011"]. Bekasovo, 2011, P. 359–367.

Kuznecova E. V. *Problemy i perspektivy izucheniya lingvogeograficheskogo landshafta regiona (na primere Volgogradskoj oblasti)* [Problems and prospects of studying the linguistic landscape of the region (on the example of the Volgograd region)]. In: *Gumani-tarnyj vector* [Humanitarian vector]. 2011, № 4 (28), pp. 96–101.

Kukanova V. V., Bembeev E. V., Ubushaev N. N., Mandzhiev B. B. *Ustnye teksty na kalmyckom yazyke: zapis' i rasshifrovka* [Oral texts in the Kalmyk language: record and decoding]. In: *Vestnik Kalmyckogo universiteta* [Bulletin of the Kalmyk University]. 2013, №3 (19), pp. 56–64.

Maksyutova N. Kh. *Vostochnyj dialect bashkirskogo yazyka v sravnitel'no-istoricheskom osveshchenii* [Oriental dialect of the Bashkir language in comparative-historical coverage]. Moscow, Nauka, 1976.

Maksyutova N. Kh. *Bashkirskie govory, nahodyashchiesya v inoyazychnom okruzhenii* [Bashkir dialects, located in a foreign environment]. Ufa, Kitap, 1996.

Mirzhanova S. F. *Yuzhnyj dialect bashkirskogo yazyka* [Southern dialect of the Bashkir language]. Moscow, Nauka, 1979.

Mirzhanova S.F. *Severo-zapadnyj dialect bashkirskogo yazyka* [North-western dialect of the Bashkir language]. Ufa, 1991.

Moskvina T.N. *Metody i podhody korpusnoj lingvistiki v issledovaniyah semantiki dialektnojleksiki* [Methods and approaches of corpus linguistics in studies of the semantics of dialectal vocabulary]. In: *Sovremennye problem nauki i obrazovaniya* [Modern problems of science and education]. 2014, № 6.; URL: <http://science-education.ru/ru/article/view?id=15784> (date of circulation: 25.04.2018)

Mul'timedijnyj korpus dialektnyh tekstov Ust'yanskogo rajona Arhangel'skojo blasti [Multimedia corpus of dialect texts of the Ustyanskiy district of the Arkhangelsk region]. URL: <http://www.slavist.de/Pushkino/> (date of circulation: 25.04.2018).

Muratova R.T. *Simvolika chisel v mifologicheskikh predstavleniyah bashkir o mire* [Symbolics of numbers in mythological representations of Bashkirs about the world]. In: *Aktual'nye problem istorii, yazyka i kul'tury Bashkortostana* [Actual problems of history, language and culture of Bashkortostan]. Ufa, 2008. pp. 178–183.

Corpus diriferimento dellaling ua italiana scritta contemporanea ("CoLFIS") (korpustal'yanskogoyazyka) [Reference Corpus of Italian written contemporary ("CoLFIS") (corpus of the Italian language)]. URL: <http://www.ge.ilc.cnr.it/dizionari.php> (date of circulation: 25.04.2018).

Saratovskij mul'timedijnyj dialektnyj korpus [Saratov multimedia dialect corpus]. ruscorporu.ru/sbornik2005/13letuchy.pdf. (date of circulation: 25.01.2018).

Sirazitdinov Z. A. *O dialektologicheskoy podsisteme mashinnogo fonda bashkirskogo yazyka* [On the Dialectological Subsystem of the Machine Fund of the Bashkir Language]. In: *Sbornik materialov Vserossijskoj nauchno-prakticheskoy konferencii Respublika Bashkortostan «Teoriya i praktika bashkirskogo yazyka i literatury v svete sovremennyh dostizhenij filologicheskikh nauk»* [Collected Materials of the All-Russian Scientific and Practical Conference Republic of Bashkortostan "Theory and Practice of the Bashkir Language and Literature in the Light of Modern Achievements in Philology"]. Ufa, 2008, pp. 178–182.

Sirazitdinov Z.A., Buskunbaeva L.A., Bardybaeva A.D., Ishmuhametova A.SH., Rafikova L.N. *O razrabotke nacional'nogo korpusa bashkirskogo yazyka* [About the development of the national building of the Bashkir language]. In: *Rol' i mestonacional'noj biblioteki v sociokul'turnom prostranstve: materialy Mezhdunarodnoj nauchno-prakticheskoy konferencii* [Role and place of the national library in the sociocultural space: materials of the International scientific and practical conference]. Ufa, 2012, pp. 118–126.

Sirazitdinov Z.A., Buskunbaeva L.A., Ishmuhametova A.SH., Ibragimova A.D., Miganova L.G. *Korpus tekstov periodicheskoy pechati nabashkirskom yazyke* [Corpus of texts of the periodical press in the Bashkir language]. In: *Aktual'nye problem dialektologii yazykov narodovRossii: Materialy XII regional'noj konferencii* [Actual problems of dialectology of languages of the peoples of Russia: Materials of the XII regional conference]. Ufa, 2012, pp. 139–141.

Sirazitdinov Z.A., Buskunbaeva L.A., Ishmuhametova A.SH., Ibragimova A.D. *O sozdanii korpusa bashkirskogo fol'klora* [On the creation of the Bashkir folklore case]. In: *Ural-Altaj: cherez veka v budushchee: materialy VI Vserossijskoj tyurkologicheskoy konferencii (s mezhdunarodnym uchastiem)* [Ural-Altai: through the centuries into the future: materials of the VI All-Russian Turkic Conference (with international participation)]. Ufa, 2014, pp. 86–89.

Slovar' bashkirskih govorov. T. 1. Vostochnyj dialekt [Dictionary of Bashkir dialects. T. 1. Eastern dialect] / N. Kh. Maksyutova etc. Ufa, 1967, 300 p.

Slovar' bashkirskih govorov. T. 2. Yuzhnyjdialekt [Dictionary of Bashkir dialects. T. 2. Southern dialect] / N. Kh. Maksyutova etc. Ufa, 1970, 327 p.

Slovar' bashkirskih govorov. T. 3. Zapadnyjdialekt [Dictionary of Bashkir dialects. T. 3. Western dialect] / N. Kh. Maksyutova etc. Ufa, 1987, 231 p.

Sulejmanova R.A. *Antroponimy sovremennogo bashkirskogo yazyka i tendencii ih razvitiya* [Anthroponyms of the modern Bashkir language and tendencies of their development]. Diss. for the competition uch. degree of Cand. philol. sciences. Ufa, 2006, 189 p.

Foneticheskij korpus "Odin den' russkogo yazyka" [Phonetic corpus "One day of the Russian language"]. URL: <http://www.model.org.spbu.ru> (date of circulation: 25.04.2018).

Foneticheskij korpus spontannoje ehstonskoje rechi [The phonetic corpus of spontaneous Estonian speech]. URL: <http://www.murre.ut.ee/phonetic-corpus> (date of circulation: 25.04.2018).

Foneticheskij korpus nemeckogo razgovornogo yazyka [Phonetic corpus of German speaking language]. URL: http://dsav-wiss.ids-mannheim.de/korpora/pf/pf_doku.htm (date of circulation: 25.04.2018).

Shakurov R. Z. *Dialektnaya sistema bashkirskogo yazyka* [Dialect system of the Bashkir language]. In: *Vatandash*, 2012, № 8, pp. 40–61.

Hisamitdinova F.G. *Mifologicheskij slovar' bashkirskogo yazyka* [Mythological dictionary of the Bashkir language]. Moscow, 2010, 452 p.

Yuldashev A. A. *K metodike sbora dialektologicheskogo materiala po bashkirskomu yazyku na sovremennom ehstape* [To the method of collecting dialectological material on the Bashkir language at the present stage]. In: *Bashkirskij dialektologicheskij sbornik* [Bashkir dialectological collection]. Ufa, 1959, pp. 213–221.

British National Corpus (BNC). URL: <http://www.natcorp.ox.ac.uk> (date of circulation: 25.04.2018).

Český národní korpus (ČNK) (cheshskij nacional'nyj korpus) [Czech national corpus]. URL: <http://ucnk.ff.cuni.cz> (date of circulation: 25.04.2018).

Chinese Broadcast Conversation Speech (mul'timedijnyj korpus kitajskogo yazyka) [multimedia of Chinese language]. URL: <http://www ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2013S04> (date of circulation: 25.04.2018).

Corpus del español (korpus ispanского yazyka) [corpus of the Spanish language]. URL: <http://www.corpusdelespanol.org> (date of circulation: 25.04.2018).

Das Deutsche Referenzkorpus (DeReKo) URL: <http://www.ids-mannheim.de/kl/projekte/korpora> (date of circulation: 25.04.2018).

LDC Top Ten Corpora (mul'timedijnye korpusy anglijskogo yazyka) [multimedia corpora of the English language]. URL: <http://www ldc.upenn.edu/Catalog/topTen.jsp> (corpus: 25.04.2018).

Magyar Nemzeti Szövegtár (korpus vengerskogo yazyka) [corpus of the Hungarian language]. URL: <http://corpus.nytud.hu/mnsz> (date of circulation: 25.04.2018).

The corpus of contemporary american english (COCA). URL: <http://corpus.byu.edu/coca> (date of circulation: 25.04.2018).

The Spoken Turkish Corpus (mul'timedijnyj korpus razgovornogo tureckogo yazyka) [multimedia corpus of spoken Turkish language]. URL: <http://stc.org.tr> (date of circulation: 25.04.2018).

The Corpus of Spontaneous Japanese (mul'timedijnyj korpus yaponskogo yazyka) [multimedia corpus of Japanese language]. URL: <http://www.ninjal.ac.jp/products-katsudo/seika/corpus/public/index.html> (date of circulation: 25.04.2018).