

УДК 81-25

DOI: 10.31040/2222-8349-2022-0-2-71-77

**ОТДЕЛ ПРИКЛАДНОЙ ЛИНГВИСТИКИ И ДИАЛЕКТОЛОГИИ:  
ВЕХИ СТАНОВЛЕНИЯ**© **З.А. Сиразитдинов**

Рассматриваются этапы становления отдела прикладной лингвистики и диалектологии. Данный отдел создан на базе лаборатории лингвистики и информационных технологий ИИЯЛ УФИЦ РАН. За годы существования лаборатория внесла существенный вклад в поддержку башкирского языка в компьютерных технологиях. Были разработаны стандарты кодовой таблицы и раскладки клавиатуры для башкирского языка, локализованы ряд программных продуктов, создавших «инфраструктуру» для национального языка в IT-технологиях. Сотрудниками подразделения созданы лингвистические информационные системы: машинный фонд башкирского языка, корпусные проекты, терминологический банк данных, – которые функционируют в сети Интернет в свободном доступе.

Башкирскими диалектологами в начале и середине XX в. была проделана значительная работа: собран полевой материал по опорным пунктам говоров, изданы диалектологические словари, описана диалектная система, подготовлен атлас говоров национального языка. Но в связи с уходом поколения диалектологов и ошибками, допущенными предыдущими руководителями ИИЯЛ в кадровом вопросе, с 90-х гг. по настоящее время систематическое научное исследование башкирских говоров и диалектов в ИИЯЛ не велось. Количество экспедиций с единой программой сбора самих материалов резко уменьшилось. Материалы прежних экспедиций не были оцифрованы, большая часть их была утрачена.

За более чем 30 лет в говорах и диалектах произошли сдвиги и изменения, которые не фиксировались и не изучались. Современная диалектология выдвигает новые подходы к исследованию говоров, появилась новая методология обработки полевых материалов.

Учитывая необходимость восстановления исследований по башкирской диалектологии, в особенности незамедлительного начала сбора полевых материалов, поскольку идет процесс исчезновения самих носителей говоров, создания на их основе информационных ресурсов по диалектам, дирекцией института и руководством УФИЦ было решено усилить кадровый состав подразделения специалистами по диалектологии и преобразовать лабораторию в отдел. В статье рассматриваются новые научные задачи, стоящие перед отделом.

Ключевые слова: башкирский язык, диалектология, тюркские языки, лингвистический корпус, корпусная лингвистика, лингвистические базы данных.

Отдел прикладной лингвистики и диалектологии учрежден в Институте истории, языка и литературы в январе 2022 г. по инициативе директора института Зуғуры Ягануровны Рахматуллиной на базе ранее существовавшей лаборатории лингвистики и информационных технологий при отделе языкознания приказом руководителя УФИЦ РАН В.П. Захарова от 14 января 2022 г. (№17-1252). Сама лаборатория лингвистики и информационных технологий была выделена в январе 2002 г. из отдела языка согласно пункту 5.2 (Организация научно-исследовательской лаборатории «Башкирский язык в информационных технологиях») Государственной программы сохранения, изучения

и развития языков народов Республики Башкортостан на 2000–2005 гг.

До организации лаборатории сотрудники З.А. Сиразитдинов, Л.А. Бускунбаева, А.В. Гильманова входили в состав группы «Автоматизация лексикографии», образованной в 1991 г. при отделе языка. В становлении этой группы большую роль сыграли казахские ученые, где аналогичное направление лингвистики было создано гораздо раньше. Директор Института языкознания АН КазССР академик А.Т. Кайдаров и руководитель казахской группы «Автоматизация лексикографии» д.ф.н. А.Х. Джубанов приняли в 1983 г. на стажировку З.А. Сиразитдинова. В стенах этого института была начата

работа по вводу в ЭВМ текста романа «Кан» классика башкирской литературы Д. Юлтыя. В дальнейшем они стали руководителем и научным консультантом его диссертационной работы «Лексика романа «Кан» Д. Юлтыя (квантитативно-стилистический анализ).

В дальнейшем З.А. Сиразитдиновым были введены в эвм тексты всех произведений классика башкирской литературы по трехтомному изданию и подготовлен частотный словарь произведений Д. Юлтыя.

Основной задачей группы автоматизации лексикографии в ИИЯЛ была разработка частотных словарей по функциональным стилям башкирского языка. В группе работали к.ф.н. З.А. Сиразитдинов, младшие научные сотрудники И.У. Надергулов и А.Г. Рахимова. Усилиями этих сотрудников были введены в эвм тексты выборки из публицистической и научной стилей башкирского языка объемом в 400 000 словоупотреблений. В дальнейшем в работе участвовали м.н.с А.В. Гильманова и лаборанты-исследователи З.Б. Валеева, Г. Шарафутдинова. Они участвовали во вводе и обработке текстов выборки из прозаических стилей. На базе компьютерной обработки этих и предыдущих материалов были составлены частотные словари функциональных стилей башкирского языка [1–3], которые стали третьими по счету после словарей русского и казахского языков среди аналогичных лексикографических разработок по языкам народов СССР. Данные словари позволили выделить и изучить лексику по частотным зонам функциональных стилей национального языка, активно использовались при разработке толкового словаря современного башкирского литературного языка. В эти же годы З.А. Сиразитдиновым и И.У. Надергуловым была начата работа по созданию архитектуры Машинного фонда башкирского языка (МФБЯ), концепция которой была изложена на совместном заседании президиумов УНЦ РАН и АН РБ в 1999 г.

Лаборатория лингвистики и информационных технологий начала свою работу в следующем составе: старший научный сотрудник З.А. Сиразитдинов, младший научный сотрудник Л.А. Бускунбаева, старшие лаборанты-исследователи А.Ш. Ишмухаметова, Г.Г. Шамсутдинова, Т.Р. Тузбекова, Р. Ханнанова, Л.Р. Хамитова. В период 2002–2004 гг. лаборатория участвовала в локализации программных продуктов FineReader, Лексикон, Диско-командир, Хамелеон, Mozilla по заданию Правительства Республики Башкортостан.

Эти программные средства создали предпосылки для широкого внедрения цифровых технологий в башкирский язык. Так текстовый анализатор FineReader компании ABBY-HOUSE позволил сканировать и распознавать тексты на башкирском языке. В локализованной версии FineReader-5.0 имелся встроенный башкирский морфологический корректор, позволяющий исправлять орфографические ошибки. К сожалению, в последующих версиях этот сервис отсутствует, требуется дополнительное соглашение с компанией на уровне Правительства республики. Текстовый редактор Лексикон компании Арсеналь открыл возможность вводить башкирские тексты с клавиатуры, производить орфографическую корректировку, распечатывать и сохранять в виде файла, пригодного для работы в редакторах Word, OpenOffice. Данный редактор был разработан для поддержки документооборота на башкирском языке в общеобразовательных учреждениях, офисах, органах управления. Все локализованные программные продукты имели как башкирский, так и русский интерфейс.

В ходе локализации этих программных продуктов были определены стандарты кодовой таблицы и раскладки башкирской клавиатуры, разработаны формальная грамматическая модель словоизменительной системы и модель слогоделения башкирского языка.

В 2002–2003 гг. З.А. Сиразитдиновым совместно с представителем БГУ к.ф.н. В.И. Луценко был создан функционирующий на отдельных компьютерах сотрудников лаборатории прототип МФБЯ. Сотрудниками лаборатории были осуществлены записи чтения от 40 информаторов отрывка художественного произведения, выделены амплитудно-частотные характеристики и сонограммы фонем башкирского языка, которые в дальнейшем вошли в подфонд экспериментально-фонетических данных МФБЯ.

В 2004 г. лаборатория совместно с Интернет-центром БГУ реализовала Лингвистический портал в сети Интернет, освещающий работу лаборатории и представляющий пользователям информацию по башкирскому языку. Портал включал в себя такие возможности как перевод русских слов на башкирский с выбором до 8 словарей с общей базой данных в 100 тыс. лексем, башкирский ЧАТ, сведения о деятельности лаборатории, информацию о локализованных программных продуктах.

В 2006 г. лаборатория тесно сотрудничала с Майкрософт: компании были переданы разра-

ботанные стандарты башкирской раскладки клавиатуры и кодировки для башкирского языка; осуществлены переводы на башкирский язык некоторых системных сообщений, проведена совместная работа для определения правильности сортировки башкирских текстов. В этом же году З.А. Сиразитдинов и программист-математик БГУ А.И. Полянин начали работу над программной оболочкой МФБЯ.

В дальнейшем сотрудники лаборатории по заданию Правительства республики работали с компанией ParaType над разработкой 18 башкирских шрифтов форматов Tuetype и Opentype. Шрифты были бесплатно розданы госучреждениям с лицензионными соглашениями.

В последующие годы Л.А. Бускунбаева, А.Ш. Ишмухаметова, Г.Г. Шамсутдинова, А.Д. Ибрагимова работали над созданием баз данных подфондов генеральной картотеки и лексикографии. В подфонд лексикографии были введены 58 изданных в разные годы лексикографических трудов по башкирскому языку общим объемом в 382 701 словарная статья.

Частая смена состава лаборатории объяснялась нестабильностью финансирования и недопониманием дирекцией института актуальности самих работ. Отсутствие должной материально-технической базы самой лаборатории привело к тому, что Машинный фонд был запущен не в институте, а на сервере БГУ. Сотрудники лаборатории благодарны директору Интернет-центра к.т.н. Амиру Даутовичу Максютову за содержание доменного имени и Интернет-трафика, организацию технического обслуживания МФБЯ в период 2006–2012 гг.

В рассматриваемый период в лаборатории активно велась и научная работа. З.А. Сиразитдиновым была разработана формальная грамматика словоизменительной системы башкирского языка, в которой проведено теоретическое обобщение работ лаборатории в процессе создания системы автоматической проверки орфографии башкирского языка для программ Лексикон и FineReader [4]. Л.А. Бускунбаевой были исследованы компрессионные единицы башкирской разговорной речи, где языковая экономия выражена наиболее полно [5]. В дальнейшем ею была успешно защищена кандидатская диссертация по данной проблеме.

К 2012 г. МФБЯ уже состоял из 7 подфондов: подфонд генеральной картотеки; лексикографический подфонд; грамматический подфонд; подфонд каталога рукописных книг; под-

фонд каталога старопечатных книг; экспериментально-фонетический подфонд; диалектологический подфонд (URL:<http://www.mfb12.ru>).

В 2012 г. сотрудники лаборатории получили грант РГНФ на разработку терминологического банка башкирского языка. З.А. Сиразитдинов, Л.Г. Мигранова, А.Д. Ибрагимова составили базы данного банка (<http://www.bashterm.ru>). Для систематизации национальной терминологии был использован рубрикатор ГРНТИ (Государственный рубрикатор научно-технической информации, прежнее наименование – Рубрикатор ГАСНТИ), представляющий собой универсальную иерархическую классификацию научно-технической информации. Выделенные по рубрике тематические группы образуют отдельные базы общего терминологического банка данных башкирского языка. Банк данных построен по общепринятым нормам в отечественной практике создания таких систем. Терминологический банк данных башкирского языка используется в работе терминологических комиссий при Правительстве и администрациях городов РБ.

С 2013 г. лаборатория в составе З.А. Сиразитдинов, Л.А. Бускунбаева, А.Д. Ибрагимова, А.Ш. Ишмухаметова, Г.Г. Шамсутдинова вела работу над проектами корпусов башкирского языка. Ими были разработаны пилотные проекты корпусов прозы, публицистики и фольклор. В целях экономии средств на покупку доменного имени и на интернет-трафик корпусные проекты функционируют в составе МФБЯ. Сегодня общий объем корпусных проектов составляет более 30 млн словоупотреблений.

Корпус прозы включает 1300 текстов башкирских писателей. Каждый текст имеет следующие метаразметки: автор, название, объем в словоформах, время создания, тип текста (повесть, притча, рассказ, роман, сказка, триллер, эпопея, эссе и т.п.).

Корпус публицистики включает тексты республиканских газет «Башкортостан», «Киске Өфө», «Йэшлек», журналов «Ағизел», «Башкортостан кызы», «Шоңкар» за определенные годы. Публицистические тексты представлены в базе данных с указанием следующих метаразметок: дата, названия издания, фио автора текста, жанр, тематика.

Корпус фольклора объемом в 855 876 словоупотреблений включает тексты 135 эпосов (с вариантами), 383 сказок, 10907 пословиц, 1290 поговорок, 985 примет и многое другое.

Все корпуса позволяют создавать подкорпусы по метаразметкам, получать частотные словари словоформ и лексем по этим подкорпусам.

Опыт работы лаборатории по созданию лингвистических информационных систем и баз данных за предыдущие годы был обобщен в коллективной монографии [6].

В 2017–2021 гг. сотрудники лаборатории были заняты исполнением государственной темы «Создание корпуса диалектных текстов башкирского языка». Над темой работали к.ф.н. З.А. Сиразитдинов, к.ф.н. Л.А. Бускунбаева, м.н.с. А.Ш. Ишмухаметова, аспирант Г.Г. Шамсутдинова. По данной теме планировался сбор и обработка обширного материала говоров восточного диалекта башкирского языка с охватом широкого круга информаторов по половозрастным группам с учетом уровня образования. К сожалению, пандемия и отсутствие финансирования экспедиций со стороны дирекции не позволили охватить все говоры диалекта: были собраны 765 полевых аудиозаписей от 75 информаторов за счет личных поездок исполнителей темы в районы проживания носителей кызылского, учалинского и айского говоров.

Для записи информаторов определены 15 тем: свадьба, свадебные обычаи; повседневные и праздничные блюда; домашние животные; дети и близкие родственники; приусадебное хозяйство; дом, постройки; топонимия в окрестностях поселения; история села, школы, рода; повседневная жизнь; времена года, погода; малые формы фольклора; поездка в райцентр; игры детей; друзья; животный мир около поселения.

Каждая аудиозапись сопровождается экстралингвистической информацией, включающей данные об информаторе: имя, отчество, фамилия, гендерные характеристики, образование, возраст, язык обучения, язык общения в семье, национальность информатора, место проживания, место последнего проживания до переезда в данное место (в случае переезда), время проживания до переезда, время записи.

Тексты транскрипций представлены в текстологической базе диалектологического подфонда МФБЯ (<http://mfbl2.ru/mfbl/bashdial>). Сегментированный и размеченный аудиоматериал полевых записей из-за большого объема (общий объем 765 аудиофайлов составляет 18.5 Гб) не может быть выставлен в сети Интернет. Данные аудиоматериалы хранятся в

ИИЯЛ УФИЦ РАН, могут быть предоставлены всем заинтересованным научным учреждениям по договоренности.

За годы своего существования лаборатория лингвистики и информационных технологий активно сотрудничала с профильными лабораториями, отделами и институтами не только России, но и зарубежными. Так сотрудничество с отделом прикладной лингвистики Института языкознания им. А. Байтурсынова и кафедрой кыргызского языка Кыргызского государственного университета им. К. Карасаева по проблемам автоматического анализа тюркской словоформы завершилось изданием совместной монографии [7], получившей положительный отзыв в научном сообществе.

Международным признанием научной деятельности лаборатории явились II Международная научная конференция «Электронная письменность народов Российской Федерации» (Уфа, 11–12 декабря 2019) и VIII Международная конференция по компьютерной обработке тюркских языков «Turklang-2020» (Уфа, 18–20 октября, 2020), проведенные сотрудниками в ИИЯЛ по резолюции предыдущих конференций.

Сегодня в отделе трудятся кандидаты филологических наук З.А. Сиразитдинов, Л.А. Бускунбаева, Л.К. Ишкильдина, М.Р. Валиева, научный сотрудник Р.Н. Каримова, младшие научные сотрудники А.Ш. Ишмухаметова и Г.Г. Шамсутдинова.

Сотрудники отдела прикладной лингвистики и диалектологии работают над двумя зарегистрированными в ЕГИСУ НИОКТР научными темами: «Развитие и лексико-грамматическое исследование корпуса письменных памятников башкирского языка», «Создание лингвистических баз данных и изучение состояния караидельского и таныпского говоров северо-западного диалекта башкирского языка».

По первой теме планируется создание корпуса текстов газет южных районов Республики Башкортостан, лексика и синтаксис которых очень близки к живому разговорному языку, отражают особенности местных говоров. Создание такого корпуса, внесет вклад в развитие и обогащение литературного языка.

По теме «Создание лингвистических баз данных и изучение состояния караидельского и таныпского говоров северо-западного диалекта башкирского языка» планируется сбор новых и обработка старых архивных полевых материалов по говорам башкирского языка.

Башкирская диалектология всегда занимала центральное место в научной деятельности ИИЯЛ. Исходным моментом в изучении башкирской диалектологии принято считать исследование говоров в 20–30-х гг. XX в. Экспедиции организованные с участием Г.С. Амантаева, Т.Г. Байшева, Г.Д. Давлетшина, З.Ш. Шакирова, Н.Г. Ишбулатова и др. позволили выявить основные границы и черты говоров и диалектов языка.

С начала 1950-х гг. башкирская диалектология приступила к системному изучению фонетики, морфологии, лексики говоров. Появление монографических трудов, в особенности Т.Г. Байшева [8], вывело башкирскую диалектологическую науку на всесоюзную тюркологическую арену.

С 1960 г. началось систематическое изучение диалектов с помощью программы «Анкета для собирания материалов по говорам башкирского языка», предусматривающей единообразный сбор диалектного материала. Полевые материалы, собранные во время экспедиций 60–70-х гг., заложили начало диалектной лексикографии. Выход в свет трехтомного словаря говоров явился значительным событием в башкирском языкознании. Материалы опросников экспедиций 60–70-х гг. были изучены в сравнительно-сопоставительном аспекте и легли в основу диссертационных и монографических работ [9–12].

В 1973 г. начался сбор материалов для Атласа башкирского языка, который был издан в 2005 г. В 1991 г. вышла фундаментальная монография С.Ф. Миржановой по северо-западному диалекту башкирского языка [13].

В последние годы кандидатами филологических наук Л.К. Ишкильдиной и М.Р. Валиевой в научных работах были затронут отдельные проблемы диалектной фонетики и лексики в говорах башкирского языка. Л.К. Ишкильдина на основе экспериментального изучения раскрыла некоторые аспекты артикуляторных особенностей согласных и гласных звуков диалектной системы башкирского языка [14]. М.Р. Валиевой выявлены древнейшие признаки болгарского языка в говорах, проведено сравнительно-историческое исследование башкирского и болгарского языков [15].

Но в целом с 90-х гг. по настоящее время фронтальные научные исследования башкирских говоров и диалектов в ИИЯЛ не велись. Количество экспедиций с единой программой сбора самих материалов резко уменьшилось.

Полевые материалы с охватом широкого круга информаторов по поло-возрастным и другим социальным группам отсутствуют. 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 среди которых главными являются максимально полный охват носителей говоров (населенных пунктов), рассмотрение частных диалектных систем как элементы общей системы национального языка, учет в одном говоре разных вариантов одного и того же звена языковой системы, корпусный подход, статистический анализ материалов и др.

Научная деятельность отдела в области башкирской диалектологии направлена на возобновление исследований говоров в русле вышеуказанных подходов.

Основой для таких исследований должны стать ГИС система и корпус аудиоматериалов носителей говоров. Для этого планируются экспедиции в районы для широкого охвата всех населенных пунктов с привлечением волонтеров. Создание диалектологических информационных систем не должно основываться на устаревшую концепцию «опорных пунктов». Работа в этом направлении сотрудниками отдела уже начата.

## Литература

1. Сиразитдинов З.А. Частотный словарь башкирского языка. Т. 1 (наука). Уфа: Гилем, 1997. 330 с. (на башк. яз.)
2. Сиразитдинов З.А. Частотный словарь башкирского языка. Т. 2 (проза). Уфа: Гилем, 2002. 413 с. (на башк. яз.)
3. Сиразитдинов З.А. Шамсутдинова Г.Г. Частотный словарь башкирского языка. Т. 3 (публицистика). Уфа: Гилем, 2006. 384 с. (на башк. яз.)
4. Сиразитдинов З.А. Моделирование грамматики башкирского языка. Словоизменительная система. Уфа: Гилем, 2006. 160 с.
5. Бускунбаева Л.А. Закономерности речевой экономии и их отражение в башкирском языке. Уфа: Гилем, 2008. 139 с.
6. Сиразитдинов З.А., Бускунбаева Л.А., Ишмухаметова А.Ш., Ибрагимова А.Д. Информационные системы и базы данных башкирского языка. Уфа: Книжная палата РБ, 2013. 116 с.
7. Жубанов А.Х., Садыков Т.С., Сиразитдинов З.А. Синтез тюрской словоформы. Алматы:

Лингвострановедческий инновационный центр КИЕ., 2019. 248 с.

8. Баишев Т.Г. Башкирские диалекты в их отношении к литературному языку. М.: МГУ, 1955. 121 с.

9. Максютлова Н.Х. Восточный диалект башкирского языка в сравнительно-историческом освещении. М.: Наука, 1976. 293 с.

10. Надергулов У.Ф. Говор башкир Саратовской и Куйбышевской области: автореф. дис. ... канд. филол. наук. Алма-Ата, 1979. 28 с.

11. Миржанова С.Ф. Южный диалект башкирского языка. М.: Наука, 1979. 272 с.

12. Дильмухаметов М.И. Говор среднеуральских башкир: автореф. дис. ... канд. филол. наук. Уфа, 1980. 27 с.

13. Миржанова С.Ф. Северо-западный диалект башкирского языка. Уфа, 1991. 295 с.

14. Ишкильдина Л.К., Уртегешев Н.С., Хисамитдинова Ф.Г. Атлас артикуляторных настроек гласных башкирского языка. Уфа: ИИЯЛ УФИЦ РАН, 2021. 298.

15. Валиева М.Р. Проблема болгарского влияния на развитие башкирского языка. Уфа, 2016. 23 с.

#### References

1. Sirazitdinov Z.A. Chastotnyy slovar' bashkirskogo yazyka, 1997, vol. 1, 330 p. (na bash. yaz.).

2. Sirazitdinov Z.A. Chastotnyy slovar' bashkirskogo yazyka, 2002, vol. 2 (proza), 413 p. (na bash. yaz.).

3. Sirazitdinov Z.A. Shamsutdinova G.G. Chastotnyy slovar' bashkirskogo yazyka, 2006, vol. 3 (publitsistika), 384 p. (na bash. yaz.).

4. Sirazitdinov Z.A. Modelirovaniye grammatiki bashkirskogo yazyka. Slovoizmenitel'naya sistema, 2006, 160 p.

5. Buskunbayeva L.A. Zakonomernosti rechevoy ekonomii i ikh otrazheniye v bashkirskom yazyke, 2008, 139 p.

6. Sirazitdinov Z.A., Buskunbayeva L.A., Ishmukhametova A.SH., Ibragimova A.D. Informatsionnyye sistemy i bazy dannykh bashkirskogo yazyka, 2013, 116 p.

7. Zhubanov A.KH., Sadykov T.S., Sirazitdinov Z.A. Sintez tyurskoy slovoformy. Almaty: Lingvostranovedcheskiy innovatsionnyy tsentr KIYe., 2019, 248 p.

8. Baishiev T.G. Bashkirskiye dialekty v ikh otnoшенii k literaturnomu yazyku, 1955, 121 p.

9. Maksyutova N.KH. Vostochnyy dialekt bashkirskogo yazyka v sravnitel'no-istoricheskom osveshchenii, 1976, 293 p.

10. Nadergulov U.F. Govor bashkir Saratovskoy i Kuybyshevskoy oblasti. Avtoref. dis. na soisk. uchen. step. kand. filol. nauk, 1979, 28 p.

11. Mirzhanova S.F. Yuzhnyy dialekt bashkirskogo yazyka, 1979, 272 p.

12. Dil'mukhametov M.I. Govor sredneural'skikh bashkir. Avtoref. dis. na soisk. uchen. step. kand. filol., 1980, 27 p.

13. Mirzhanova S.F. Severo-zapadnyy dialekt bashkirskogo yazyka, 1991, 295 p.

14. Ishkil'dina L.K., Urtegeshev N.S., Khisamitdinova F.G. Atlas artikulyatornykh nastroyek glasnykh bashkirskogo yazyka, 2021, 298 p.

15. Valiyeva M.R. Problema bulgarskogo vliyaniya na razvitiye bashkirskogo yazyka, 2016, 23 p.



#### DEPARTMENT OF APPLIED LINGUISTICS AND DIALECTOLOGY: MILESTONES

© Z.A. Sirazitdinov

Order of the Badge of Honour Institute of History, Language and Literature –  
Subdivision of the Ufa Federal Research Centre of the Russian Academy of Sciences  
71, prospekt Oktyabrya, 450054, Ufa, Russian Federation

The article discusses the stages of formation of the Department of Applied Linguistics and Dialectology. This department was created on the basis of the Laboratory of Linguistics and Information Technologies of the Institute of Nuclear Physics of the Ural Federal Research Center of the Russian Academy of Sciences. Over the years of its existence, the laboratory has made a significant contribution to the support of the Bashkir language in computer technology. Standards for the code table and keyboard layout for the Bashkir language were developed, a number of software products were localized that created an «infrastructure» for the national language in IT technologies. The employees of the division created linguistic information systems: a machine fund of the Bashkir language, corpus projects, a terminological data bank, which operate on the Internet in the public domain.

Significant work was done by Bashkir dialectologists in the early and middle of the 20th century: field material was collected on the reference points of dialects, dialectological dictionaries were published, the dialect system was described, and an atlas of dialects of the national language was prepared. But due to the departure of the generation of dialectologists and the mistakes made by the previous leaders of the IIL in the personnel issue, from the 90s to the present, a systematic scientific study of the Bashkir dialects and dialects has not been conducted at the IIL. The number of expeditions with a unified program for collecting the materials themselves has sharply decreased. The materials of previous expeditions were not digitized, most of them were lost.

For more than 30 years, there have been shifts and changes in dialects and dialects that have not been recorded or studied. Modern dialectology puts forward new approaches to the study of dialects, a new methodology for processing field materials has appeared.

Taking into account the need to restore research on Bashkir dialectology, in particular, the immediate start of collecting field materials, since the process of the disappearance of the speakers themselves is underway, creating information resources on dialects on their basis, the directorate of the institute and the leadership of the UFIC, it was decided to strengthen the staff of the unit with specialists in dialectology and transform the laboratory to the department. The article deals with new scientific tasks facing the department.

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