



Published in the Russian Federation
Bulletin of the Kalmyk Institute for Humanities
of the Russian Academy of Sciences
Has been issued since 2008
ISSN: 2075-7794; E-ISSN: 2410-7670
Vol. 23, Is. 1, pp. 212–219, 2016
DOI 10.22162/2075-7794-2016-23-1-212-219
Journal homepage: <http://kigiran.com/pubs/vestnik>

UDC 811.512.141'28

Dialectological Base of the Machine Fund of the Bashkir Language as a Tool for Research Bashkir Dialect Continuum

Zinnur A. Sirazitdinov¹, Lily A. Buskunbaeva², Ramilya N. Karimova³

¹ Ph. D. of Philology, Leading Scientist, Head of the Laboratory of Linguistics and Informational Technologies at the Department of Linguistics, Institute of History, Language and Literature, the Ufa Scientific Center, Russian Academy of Sciences (Ufa, Russian Federation). E-mail: sazin11@mail.ru

² Ph. D. of Philology, Scientist of the Laboratory of Linguistics and Informational Technologies at the Department of Linguistics, Institute of History, Language and Literature, the Ufa Scientific Center, Russian Academy of Sciences (Ufa, Russian Federation). E-mail: buskl@yandex.ru

³ Scientist of the Department of Linguistics, Institute of History, Language and Literature, the Ufa Scientific Center, Russian Academy of Sciences (Ufa, Russian Federation). E-mail: 1967-ram@mail.ru

Abstract

This article deals with the problem of development of the dialectological base of the Machine fund of the Bashkir language. Along with the written monuments, folklore material, dialect is one of the most important sources for the study of the historical development and formation of the literary language. The main task of dialectologists is not only the collection but also the storage of dialect materials of the Bashkir language.

Taking into account the importance of translation of dialect materials in electronic format and opening of free access to a wide audience to the information provided, the staff of the Laboratory of Linguistics and Information Technologies of the Institute of History, Language and Literature of the Ufa Scientific Center established a dialectological base as a part of the Machine fund of the Bashkir language consisting of three separate databases — lexical database, the database of dialectological atlas and textological database.

The lexical database includes information about the dialectal lexicon of the Bashkir language. The database was developed on the basis of dialectal dictionaries compiled and published by the staff of the Institute of History, Language and Literature. The volume of the database is more than 52 000 dialectal units.

The database of dialectological atlas is developed on the basis of materials of Dialectological atlas of the Bashkir language. The database allows to select the types of linguistic phenomena (phonetic, morphological, syntactic, lexical), for each type specific isoglosses were identified. Isoglosses allocated 250 strong points of the republic and neighboring regions.

The textological database represents illustrative materials collected during numerous expeditions by the staff of the Institute of History, Language and Literature. To date, the database contains more than 500 texts in all dialects of the Bashkir language.

Introduced in dialectological base material forms the basis of synchronic and diachronic study of the language features of dialects and sub-dialects of the Bashkir language.

Keywords: the Bashkir language, Turkic languages, Machine Fund of the Bashkir language, information systems, dialectology, dialectological base.

К настоящему времени башкирская диалектология располагает достаточными сведениями и материалами, освещающими современное состояние народного разговорного языка. Изданы обобщающие монографические работы по всем трем диалектам башкирского языка [Баишев 2006; Максютова 1976; Максютова 1996; Миржанова 2006; Миржанова 1979]. Опубликованы диалектологические словари, охватившие лексику всех диалектных зон башкирского языка [Башкорт... 1967; Башкорт... 1970; Башкорт... 1987; Башкорт... 2002]. Созданный на основе использования метода лингвистической географии, «Диалектологический атлас башкирского языка» [Диалектологический... 2005] представляет собой 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.

Однако многие проблемы башкирской диалектологии до сих пор остаются вне поля зрения диалектологов, отдельные говоры нуждаются в специальном исследовании. Относительно слабо изучен диалектный синтаксис, процесс взаимовлияния и взаимообогащения диалектов, требуют более детального изучения говоры северо-западного диалекта, в сопоставительном аспекте диалектные данные башкирского языка с тюркскими и финно-угорскими языками.

Изучение диалектного материала является одним из факторов для сохранения не только родного языка, но и этнической самобытности, национальной культуры в условиях глобализации и урбанизации современного общества. Диалектный материал выступает в качестве одного из важнейших источников для изучения исторического развития и становления литературного языка, сравнительно-сопоставительных исследований языков. Интерес к диалектным материалам определяется и тем, что они позволяют проследить историю самого народа.

Однако в эпоху высоких технологий и информатизации общества, которая вовлекает в свою орбиту и филологическую науку, наличие диалектных материалов только в печатной форме имеет ряд недостатков. Во-первых, в силу небольшого тиража бумажных изданий исследователи имеют к ним ограниченный доступ, во-вторых, поиск нужной информации занимает намного

больше информации, чем при работе с электронной версией.

На сегодняшний день многие языки России переводят свои ресурсы по диалектологиям в электронный формат: проводятся работы по оцифровке диалектологических словарей и атласов, создаются базы филологических данных и целые информационные ресурсы с поисковыми системами в сети Интернет [Куканова, Бембеев, Убушаев, Манджиева 2013; Салимов, Рамазанова, Пилюгин, Салимов 2011; Соломенников 2006]. Вовлечение национального языка в информационную сферу открывает широкие перспективы для сравнительных исследований лингвистических реалий родственных языков и быстрый доступ к материалам исследования большого объема.

Учитывая актуальность перевода диалектных материалов в электронный формат и открытия свободного доступа широкой аудитории к представленной информации, сотрудниками Лаборатории лингвистики и информационных технологий был создан в рамках Машинного фонда башкирского языка диалектологический подфонд, состоящий из трех самостоятельных баз — лексической базы данных, базы данных диалектологического атласа и текстологической базы данных.

Перед началом создания данного подфонда была проделана огромная работа по разработке основных положений проекта, сканирование и перевод в нужный формат диалектных материалов [Каримова 2011; Каримова, Сиразитдинов, Бускунбаева 2010; Сиразитдинов, Бускунбаева, Ишмухаметова, Ибрагимов 2013].

Представленный в базе данных диалектный материал дает реальную картину всего многообразия лексического состава говоров башкирского языка, фонетических, морфологических и синтаксических модификаций.

Лексическая база данных. В данной базе представлена информация о диалектной лексике башкирского языка. База данных разработана по составленным и изданным сотрудниками института диалектным словарям [Башкорт... 1967; Башкорт... 1970; Башкорт... 1987; Башкорт... 2002], охватывающим все многообразие лексической системы башкирских диалектов. Созданные по дифференциальному принципу, словари включают в свой состав специфические диалектные слова, семантические и фонетические диалектизмы.

На сегодняшний день объем базы данных составляет свыше 52 тыс. диалектных единиц с шестью информационными полями: само диалектное слово, часть речи, диалект, говор, литературная норма, русский перевод.

Пример представления диалектной лексики по данной структуре:

Тупар — *исем, көнъяк диалект, урта һөйләш, ер еләге* 'клубника'.

Имән бармак — *исем, төньяк-көнбайыш диалекты, танып һөйләше, һук бармак* 'указательный палец'.

В базе реализованы поиск, сортировка и выдача результатов по шести полям: по диалектному слову, части речи, указателю диалекту и говору, литературной норме и русскому переводу.

Система позволяет по лексическим записям базы данных (диалектное слово, литературная норма, русский перевод) осуществлять поиск информации в следующих режимах: по точному совпадению слова, по совпадению фрагмента с началом слова, по совпадению фрагмента с концом слова, по совпадению фрагмента с любой частью слова. Сочетание различных записей с разными режимами поиска позволяет получить многообразные выборки из базы данных, отсортированных в соответствии с целями лингвистического эксперимента.

Например, в лексической базе при поиске литературного слова *камка* 'божья коровка' база выдает 148 диалектизмов. Только в восточном диалекте встречается 17 вариантов данного слова: *алла быуай һыйыры* (среднеуральский говор), *көнйылткыс* (кизильский), *камәгортто* (кизильский), *койашапай* (кизильский), *койашкорт* (среднеуральский), *койашымбак* (среднеуральский), *кызыл һыйыр* (аргаяшский, миасский), *кызылгандала* (кизильский), *осакай* (аргаяшский), *сыбар һыйыр* (аргаяшский), *эбей сыуак* (кизильский, миасский) и т. д.

База данных диалектологического атласа. Представленная база разработана на основе материалов «Диалектологического атласа башкирского языка» [Диалектологический... 2005], составленного в результате лингвогеографического изучения башкирского языка диалектологами Института истории, языка и литературы УНЦ РАН в 1976–1987 гг. Атлас включает в себя богатый диалектный материал по фонетике, морфологии, лексике и синтаксису. В ходе экспедиционных исследований были обследованы 400 опорных пунктов не

только в Республике Башкортостан, но и в соседних регионах, где компактно проживает башкирское население (в Курганской, Оренбургской, Пермской, Самарской, Саратовской, Свердловской, Челябинской областях), 250 из которых нанесены на карту-бланковку.

Основным преимуществом атласа является наличие лингвистических карт, наглядно демонстрирующих ареал распространения тех или иных языковых (фонетических, лексических, морфологических и синтаксических) явлений диалектов башкирского языка.

Электронная версия данного издания размещена в диалектологической базе Машинного фонда башкирского языка. Пользователи благодаря понятному и удобному интерфейсу базы значительно сократят время поиска нужной информации.

На сегодняшний день наблюдается тенденция к постепенному размыванию исторически сложившихся границ не только отдельных диалектных явлений, но и самих говоров башкирского языка. Встроенные интерактивные карты позволяют решать не только лингвистические задачи, но и исторические, географические, этнографические и т. д.

База данных диалектологического атласа включает фонетические, морфологические, лексические и синтаксические явления.

Для удобства пользования и наглядности разработан интерфейс, состоящий из трехуровневого меню:

1) меню выбора типа языкового соответствия (*фонетический, морфологический, лексический и синтаксический уровни*);

2) меню выбора подтипа языкового соответствия (например, для *синтаксических соответствий: по формам употребления направительного падежа, форме изафета и т. д.*);

3) меню выбора конкретного языкового соответствия (*в виде конкретного противопоставления лингвистических явлений в разных позициях в опорных словах и конструкциях*).

При выборе конкретного лингвистического соответствия выдается изоглосса, включающая список из 250 опорных пунктов, по которым проходят графические обозначения изоглоссных явлений. На рис. 2 продемонстрирована выдача списка опорных пунктов, характеризующих изоглоссы по фонетическому соответствию *ә-а* в начале слова в опорном слове **атай/этей**.

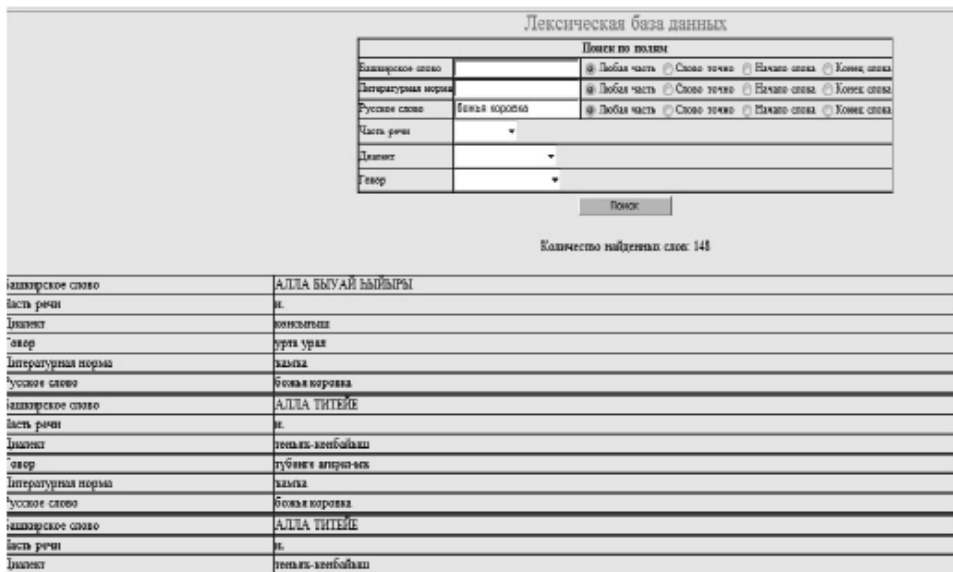


Рис. 1. Интерфейс лексической базы данных диалектологического подфонда. Поиск по русскому переводу ‘божья коровка’

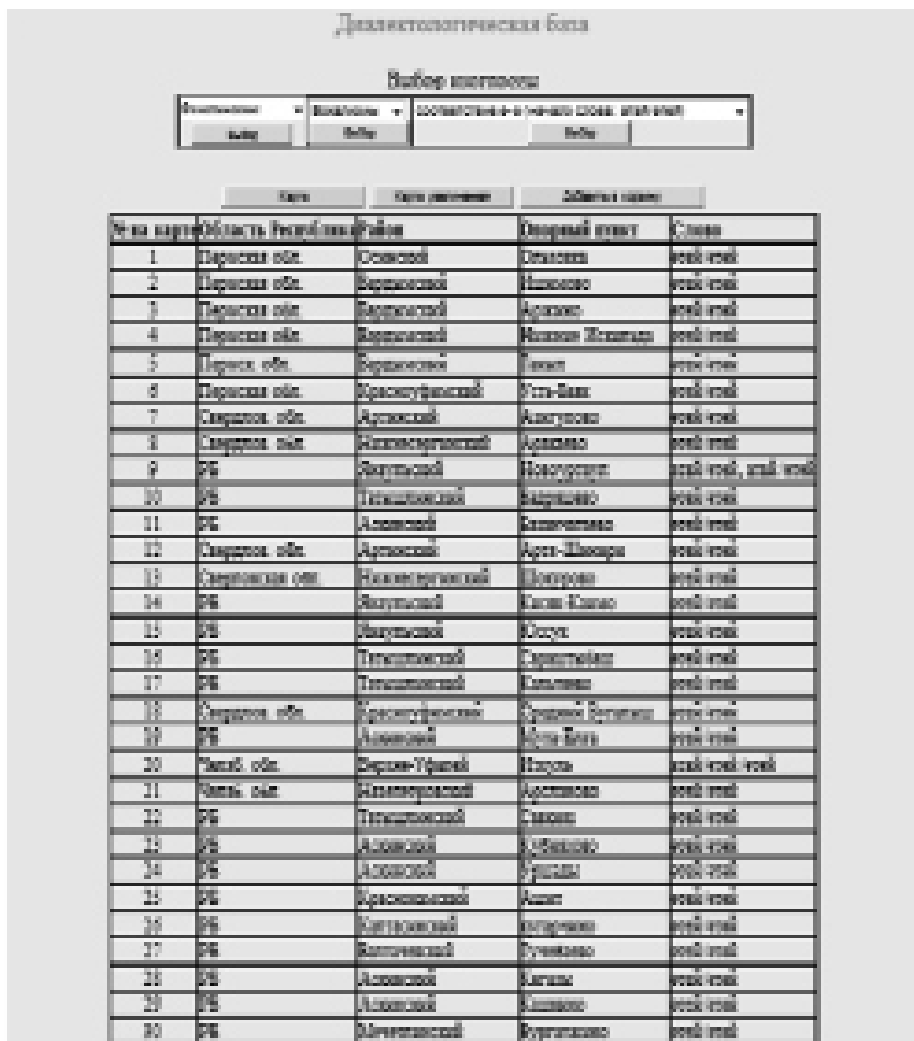


Рис. 2. Выдача списка населенных пунктов, характеризующих данную изоглоссу

Пользователю предоставлена возможность просмотра карты с изоглоссами, что позволяет наиболее наглядно показать территорию распространения того или иного языкового явления (см. рис. 3).

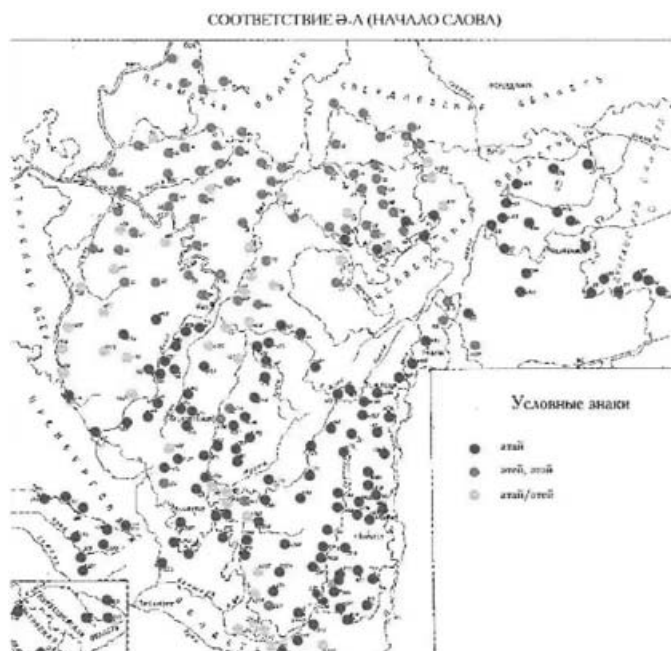


Рис. 3. Пример выдачи карты с помеченными населенными пунктами распространения фонетического соответствия а-е (населенные пункты, относящиеся к одной изоглоссе, выделены цветными обозначениями)

Текстологическая база. В данной базе представлен иллюстративный материал, собранный во время многочисленных экспедиций сотрудниками Института истории, языка и литературы Уфимского научного центра РАН, отражает живую речь носителей говоров башкирского языка.

На сегодняшний день база содержит более 500 текстов по всем говорам башкирского языка. Объем представленных текстов колеблется в пределах от нескольких

предложений до нескольких страниц формата А4.

В текстологической базе есть возможность поиска текста по лингвистическим и экстралингвистическим параметрам: диалекту, говору, году записи, образованию информанта, полу информанта, возрасту информанта, национальности. Программное обеспечение позволяет производить поиск текстов по любому из вышеуказанных параметров либо по их комбинации.



Рис. 4. Пример выдачи текста, где в качестве параметра указан ирригиз-камеликский говор

Текстологическая база данных дает возможность выбора иллюстрационного материала для лингвистических исследований за короткий промежуток времени.

Таким образом, в данных базах представлен богатый материал по башкирской диалектологии, который открывает широкие возможности для изучения диалектов и говоров башкирского языка в фонетическом, лексическом, грамматическом и стилистическом аспектах.

Литература

- Башиев Т. Г. Башкирские диалекты в их отношении к литературному языку. Уфа: Гилем, 2006. 132 с.
- Башкорт һөйләштәренен һүзлеге. Өфө: Башкортостан китап нәшриәте, 1967. Т. 1. 352 б.
- Башкорт һөйләштәренен һүзлеге. Өфө: Башкортостан китап нәшриәте, 1970. Т. 2. 326 б.
- Башкорт һөйләштәренен һүзлеге. Өфө: Башкортостан китап нәшриәте, 1987. Т. 3. 232 б.
- Башкорт теленен диалекттары һүзлеге. Өфө: Китап, 2002. 432 б.
- Диалектологический атлас башкирского языка. Уфа: Гилем, 2005. 234 с.
- Каримова Р. Н. О проблемах восстановления, сохранения диалектологических аудиоматериалов 60-х годов XX века // Актуальные проблемы диалектологии языков народов России: материалы XI Межрегиональной конференции. Уфа: ИИЯЛ УНЦ РАН, 2011. С. 114–117.
- Каримова Р. Н., Сиразитдинов З. А., Бускунбаева Л. А. О текстологической базе данных диалектологического подфонда «Машинного фонда башкирского языка» // Актуальные проблемы диалектологии языков народов России: материалы X региональной конференции. Уфа: ИИЯЛ УНЦ РАН, 2010. С. 186–188.
- Куканова В. В., Бембеев Е. В., Убушаев Н. Н., Манджиева Б. Б. Устные тексты на калмыцком языке: запись и расшифровка // Вестник калмыцкого университета. 2013. № 3 (19). С. 56–64.
- Максютова Н. Х. Восточный диалект башкирского языка: в сравнительно-историческом освещении. М.: Наука, 1976. 292 с.
- Максютова Н. Х. Башкирские говоры, находящиеся в иноязычном окружении. Уфа: Китап, 1996. 287 с.
- Миржанова С. Ф. Северо-западный диалект башкирского языка. Уфа: Китап, 2006. 290 с.
- Миржанова С. Ф. Южный диалект башкирского языка. М.: Наука, 1979. 274 с.
- Салимов Ф. И., Рамазанова Д. Б., Пилюгин А. Г., Салимов Р. Ф. Электронная версия атласа татарских народных говоров // Филология и культура. 2011. Вып. 26. С. 205–210.
- Соломенников И. С. Лексический атлас русских говоров Удмуртии в Интернет // Современные информационные технологии и письменное наследие: от древних рукописей к электронным текстам: материалы междунар. науч. конф. (Ижевск, 13–17 июля 2006 г.). Ижевск: Изд-во ИжГТУ, 2006. С. 161–163.
- 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.

References

- Baishev T. G. *Bashkirskie dialekty v ih otnoshenii k literaturnomu jazyku* [Bashkir dialects in their relation to the literary language]. Ufa, Gilem Publ., 2006, 132 p. (In Russ.).
- Bashkort *hөjlәштәrenen һүzlege*. Өфө, Башкортостан китап нәшриәте, 1967, Т. 1, 352 b. (In Bashkir).
- Bashkort *hөjlәштәrenen һүzlege*. Өфө, Башкортостан китап нәшриәте, 1970, Т. 2, 326 b. (In Bashkir).
- Bashkort *hөjlәштәrenen һүzlege*. Өфө, Башкортостан китап нәшриәте, 1987, Т. 3, 232 b. (In Bashkir).
- Bashkort *telenen dialektтары һүzlege*. Өфө, Китап, 2002, 432 b. (In Bashkir).
- Dialektologicheskij atlas bashkirskogo jazyka* [Dialectological Atlas of the Bashkir language]. Ufa, Gilem Publ., 2005, 234 p. (In Russ.).
- Karimova R. N. *O problemah vosstanovlenija, sohraneniya dialektologicheskikh audiomaterialov 60-h godov XX veka* [On the problems of restoration, conservation of dialectological audio records of the 60-ies of the 20th century]. *Aktual'nye problemy dialektologii jazykov narodov Rossii: materialy XI Mezhregional'noj konferencii* (Proceed. of the XI Intern. Conf. "Actual problems of dialectology of languages of the peoples of Russia"). Ufa, IJaL UNC RAN Publ., 2011, pp. 114–117 (In Russ.).
- Karimova R. N., Sirazitdinov Z. A., Buskunbaeva L. A. *O tekstologicheskoy baze dannyh dialektologicheskogo podfonda «Mashinnogo fonda bashkirskogo jazyka»* [On textual

- database of dialectological sub fund “Machine Fund of the Bashkir language”]. *Aktual'nye problemy dialektologii jazykov narodov Rossii: materialy X regional'noj konferencii* [Proceed. of the Xth Regional Conf. “Actual problems of dialectology of languages of the peoples of Russia”]. Ufa, IJaL UNC RAN Publ., 2010, pp. 186–188 (In Russ.).
- Kukanova V. V., Bembeev E. V., Ubushaev N. N., Mandzhieva B. B. *Ustnye teksty na kalmyckom jazyke: zapis' i rasshifrovka* [Oral texts in the Kalmyk language: recording and transcript]. *Vestnik kalmyckogo universiteta* (Bulletin of Kalmyk State University), 2013, no. 3 (19), pp. 56–64 (In Russ.).
- Maksjutova N. H. *Vostochnyj dialekt bashkirskogo jazyka: v sravnitel'no-istoricheskom osveshhenii* [Eastern dialect of the Bashkir language in comparative-historical aspect]. Moscow, Nauka Publ., 1976, 292 p. (In Russ.).
- Maksjutova N. H. *Bashkirskie govory, nahodjashhiesja v inojazychnom okruzenii* [Bashkir dialects, being in a foreign environment]. Ufa, Kitap Publ., 1996, 287 p. (In Russ.).
- Mirzhanova S. F. *Severo-zapadnyj dialekt bashkirskogo jazyka* [The North-Western dialect of the Bashkir language]. Ufa, Kitap Publ., 2006, 290 p. (In Russ.).
- Mirzhanova S. F. *Juzhnyj dialekt bashkirskogo jazyka* [The Southern dialect of the Bashkir language]. Moscow, Nauka Publ., 1979, 274 p. (In Russ.).
- Salimov F. I., Ramazanova D. B., Piljugin A. G., Salimov R. F. *Jelektronnaja versija atlasa tatarskih narodnyh govorov* [The electronic version of the Tatar national dialects atlas] *Filologija i kul'tura* (Philology and culture), 2011, Issue 26, pp. 205–210 (In Russ.).
- Solomennikov I. S. *Leksicheskij atlas russkih govorov Udmurtii v Internet* [Lexical Atlas of Russian dialects of Udmurtia in the Internet]. *Sovremennye informacionnye tehnologii i pis'mennoe nasledie: ot drevnih rukopisej k jelektronnym tekstam: materialy mezhdunar. nauch. konf. (Izhevsk, 13–17 ijulja 2006 g.)* [Proceed. of the International. Scient. Conf. “Modern information technologies and written heritage: from ancient manuscripts to electronic texts” (Izhevsk, 13–17 July 2006)]. Izhevsk, IzhGTU Publ., 2006, pp. 161–163 (In Russ.).
- Sirazitdinov Z. A., Buskunbaeva L. A., Ishmuhametova A. Sh., Ibragimova A. D. *Informacionnye sistemy i bazy dannyh bashkirskogo jazyka* [Information systems and databases of the Bashkir language]. Ufa, Knizhnaja palata RB, 2013, 116 p. (In Russ.).

УДК 811.512.141'28

ДИАЛЕКТОЛОГИЧЕСКАЯ БАЗА МАШИННОГО ФОНДА БАШКИРСКОГО ЯЗЫКА КАК ИНСТРУМЕНТ ИССЛЕДОВАНИЯ БАШКИРСКОГО ДИАЛЕКТНОГО КОНТИНУУМА

Зиннур Амирович Сиразитдинов¹, Лилия Айсовна Бускунбаева²,
Рамиля Нигматьяновна Каримова³

¹ кандидат филологических наук, ведущий научный сотрудник, заведующий лабораторией лингвистики и информационных технологий отдела языкознания Института истории, языка и литературы Уфимского научного центра РАН (Уфа, Российская Федерация). E-mail: sazin11@mail.ru

² кандидат филологических наук, научный сотрудник лаборатории лингвистики и информационных технологий отдела языкознания Института истории, языка и литературы Уфимского научного центра РАН (Уфа, Российская Федерация). E-mail: buskl@yandex.ru

³ научный сотрудник отдела языкознания Института истории, языка и литературы Уфимского научного центра РАН (Уфа, Российская Федерация). E-mail: 1967-ram@mail.ru

Аннотация. В настоящей статье рассматривается проблема разработки диалектологической базы Машинного фонда башкирского языка. Наряду с письменными памятниками, фольклорным материалом, диалектизмы являются одним из важнейших источников для исследования исторического развития и становления литературного языка. Сегодня влияние

литературного языка на диалекты очень сильное, что проявляется, прежде всего, в активизации процесса исчезновения диалектной лексики. Основной задачей языковедов является не только сбор диалектных материалов, но и сохранение данного пласта языка. Диалектная система башкирского языка достаточно хорошо исследована, в то же время многие проблемы остаются не исследованными.

Для решения данных задач сотрудниками Лаборатории лингвистики и информационных технологий Института истории, языка и литературы Уфимского научного центра РАН была разработана интегрированная в Машинный фонд башкирского языка Диалектологическая база, состоящая из трех самостоятельных баз — лексической базы данных, базы данных диалектологического атласа и текстологической базы данных. Эти базы построены на основе изданных печатных продукций с указанием источников. Все существующие базы имеют свои поисковые аппараты, меню пользователя представлено на трех языках: башкирском, русском и английском.

Лексическая база данных включает информацию о диалектной лексике башкирского языка на базе изданных в ИИЯЛ диалектных словарей. База данных диалектологического атласа разработана на основе материалов диалектологического атласа башкирского языка. В текстологической базе представлен иллюстративный материал, собранный во время многочисленных экспедиций сотрудниками ИИЯЛ. На сегодняшний день база содержит более 500 текстов по всем говорам башкирского языка. В данных базах представлен богатый материал по башкирской диалектологии, который может быть использован как в научных целях, так и в образовательном процессе.

Ключевые слова: башкирский язык, тюркские языки, Машинный фонд башкирского языка, информационные системы, диалектология, диалектологическая база.