

З.А. Сиразитдинов¹, Л.А. Бускунбаева¹, А.Ш. Ишмухаметова²

¹к.ф.н., Уфимский федеральный исследовательский центр Российской академии наук, Уфа, Башкортостан/Россия (e-mail: sazin11@mail.ru)

¹к.ф.н., Уфимский федеральный исследовательский центр Российской академии наук, Уфа, Башкортостан/Россия (e-mail: buskl@yandex.ru)

²МНС, Уфимский федеральный исследовательский центр Российской академии наук, Уфа, Башкортостан/Россия (e-mail: ishmuhametova_anita@mail.ru)

Об обработке звуковых материалов для диалектологического аудиокорпуса башкирского языка

Аннотация

В статье рассматриваются принятые в лаборатории лингвистики и информационных технологий ИИЯЛ УФИЦ РАН принципы транскрибирования аудиоматериалов полевых экспедиций. Данная работа начата по проекту составления аудиокорпуса по говорам восточного диалекта башкирского языка. Транскрипция выполняется в «полуорфографической записи», которая нашла широкое распространение в речевых корпусных разработках.

Авторами разработана единая система графической передачи диалектных особенностей с использованием специальных знаков и буквосочетаний для обозначения тех звуков, которых невозможно передать при помощи башкирской орфографии.

Единицей описания в корпусе являются не только слова, но и знаки сегментирования, символы, обозначающие паралингвистические элементы речи, такие как смех, кашель, вздохи, стоны, причмокивание, плач и др., хезитационные явления, сопровождающие живую речь.

При разработке аннотирования на уровне высказываний в орфографической записи был применен опыт создателей диалектологических и мультимедийных корпусов русского и калмыцкого языков.

Просодические характеристики речи, как ударение, интонация, тон, на данном этапе транскрибирования диалектного материала не учитывались. Явления элизии и ассимиляции, то есть изменения, происходящие в потоке речи на стыке слов, учитываются.

В диалектном корпусе такая многоуровневая разметка, включающая и паралингвистические элементы, хезитационные явления, окажется бесценным источником при исследовании языка с точки зрения когнитивных, социологических и психологических подходов и повысит информативность корпуса и расширит круг задач, решаемых с помощью корпуса.

Ключевые слова: лингвистика, башкирский язык, диалектология, корпусная лингвистика, транскрибирование, аудиоматериалы

Z.A. Syrazitdinov¹, L.A. Buskunbaeva¹, A.Sh. Ishmukhametov²

¹Candidate of Philology, Ufa Federal Research Center of the Russian Academy of Sciences, Ufa, Bashkortostan / Russia (e-mail: sazin11@mail.ru)

¹Candidate of Philology, Ufa Federal Research Center of the Russian Academy of Sciences, Ufa, Bashkortostan / Russia (e-mail: buskl@yandex.ru)

²Junior Researcher, Ufa Federal Research Center of the Russian Academy of Sciences, Ufa, Bashkortostan / Russia (e-mail: ishmukhametova_anita@mail.ru)

On the Processing of Sound Materials for Dialectological Audio Enclosures of the Bashkir Language

Abstract

The article discusses the principles of transcribing audiomaterials of field expeditions adopted in the laboratory of linguistics and information technologies of the Institute of Physical Ural Physics Institute of the Russian Academy of Sciences. This work has been started on the project of compiling an audio corpus according to the dialects of the eastern dialect of the Bashkir language. Transcription is performed in a “semi-morphological record”, which is widely used in speech hulls.

The authors have developed a unified system of graphic transmission of dialect features using special characters and letter combinations to designate those sounds that cannot be transmitted using Bashkir spelling.

The unit of description in the corpus is not only words, but also segmentation marks, symbols denoting paralinguistic elements of speech, such as laughter, coughing, sighing, moaning, smacking, crying, etc., khizitatsionnye phenomena accompanying live speech.

When developing annotations at the level of statements in spelling, the experience of the creators of dialectological and multimedia corps of Russian and Kalmyk languages were applied.

Prosodic characteristics of speech, such as stress, intonation, tone, were not taken into account at this stage of transcribing dialect material. The phenomena of elision and assimilation, that is, changes occurring in the flow of speech at the junction of words, are taken into account.

In the dialect case, such multilevel markup, including paralinguistic elements, citation phenomena, will prove to be an invaluable source for studying language from the point of view of cognitive, sociological and psychological approaches and will increase the information content of the case and will expand the range of tasks solved with the help of the case.

Keywords: linguistics, Bashkir language, dialectology, corpus linguistics, transcribing, audio materials

В Лаборатории лингвистики и информационных технологий ИИЯЛ УФИЦ РАН начата работа над проектом диалектного фонетического корпуса башкирского языка. Проект предусматривает сбор и обработку обширного материала с охватом широкого круга информантов по поло-возрастным и другим социальным группам с перспективой создания представительного корпуса, включающего в

себя аудиоматериалы и их транскрипции по говорам восточного, южного и северо-западного диалектов башкирского языка.

Проект реализуется в три этапа. На первом этапе объектом исследования стал восточный диалект башкирского языка, ареалом распространения которого являются Абзелиловский, Баймакский, Белокатайский, Белорецкий, Бурзянский, Кигинский, Мечетлинский, Салаватский, Учалинский районы Республики Башкортостан, а также прилегающие к ним отдельные районы Челябинской, Курганской и Свердловской областей.

Отметим, что работы по созданию речевых диалектных корпусов стали популярными и востребованными в современной лингвистике. В России известны Саратовский диалектологический корпус, Мультимедийный корпус диалектных текстов Устьянского района Архангельской области, диалектные подкорпусы в составе Национального корпуса русского языка, ижемского говора коми языка, Национального корпуса калмыцкого языка и др. [Корпусы].

Следует отдельно отметить уникальный проект, разрабатываемый учеными Института языкознания РАН в сотрудничестве с Институтом системного программирования РАН, который представляет собой информационную систему, содержащую обширную информацию по уральским и алтайским языкам на базе экспедиционных аудиозаписей: корпусы, мультимедийные словари и пр. [Lingvodoc].

На данном этапе осуществляется сбор диалектного материала (запись информанта на диктофон на определенные темы), обработка материала для базы данных, которая включает членение аудиоматериала на коммуникативные эпизоды (фразы, синтагмы), транскрибирование и аннотирование. Вся работа по транскрибированию, синхронизации транскрипта с аудиоматериалом, представлению его литературного варианта и перевода на русский язык осуществляется в программе аннотирования ELAN. На рис.1 представлен пример аудиоматериала с транскрибированием, литературным вариантом транскрипции и его переводом на русский язык.

Транскрипция выполняется в «полуорфографической записи», широко распространенной транскрипции в речевых корпусных разработках [1, с. 4]. Эта транскрипция близка к фонематической, максимально приближена к орфографии современного башкирского языка. Была разработана единая система графической передачи

диалектных особенностей с использованием специальных знаков и буквосочетаний для обозначения тех звуков, которых невозможно передать при помощи башкирской орфографии.

Гласные фонемы:

Наряду с фонемами э, о, ө, у, ү, и, ы, э(е) выделены следующие гласные:

а* – гортанный звук (кара – ка*ра),

а^ – огубленный звук (алма – а^лма).

Согласные фонемы:

Наряду с фонемами б, й, ж, з, ҙ, д, к, ҡ, ғ, г, л, м, н, ң, п, р, с, ҫ, т, ф, х, ш, һ выделены следующие согласные:

v – билабиальная шумная согласная в (вакыт – вакыт, вәғәзә – вәғәзә),

б* – билабиальный вариант согласной [б] (арба – арб*а),

к* – гортанный вариант согласной [к] (калын – к*алын),

ф* – билабиальная согласная, близкая к [ф] (афын – аф*ын),

ҫ* – твердая верхнезубная согласная,

у^ – огубленный вариант гласной [у] (юк – йу^к),

ү^ – огубленный вариант гласной [ү] (күп – кү^п),

и* – гласная близкая к [е] (кил – ки*л),

йа – вместо буквы [я] (як – йак),

йу – вместо буквы [ю] (юл – йул),

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

Для различения аллофонов использованы знаки (*, ^) с учетом их ввода в ELAN.

Надо отметить, что вышеуказанная транскрипция применяется только при транскрибировании башкирских слов. Вкрапления из русского языка представлены в оригинальной орфографической записи, если они под влиянием башкирского языка не подвергались видоизменениям: *ана шул (э-э) нейзәрҙә беззең (э-э) буш вакытында / теге # насиф # вакыты бәттә нисә / китереб беззең капка төб*әнә өйөб куйа торғандар ийе инде // 'после посева / когда не было в них необходимости / приносили к нам и складывали возле ворот //'*. Октоторпом (#) в транскрибированных текстах отмечается переключение языковых кодов (с башкирского на русский).

Единицей описания в корпусе являются не только слова, но и знаки сегментирования, символы, обозначающие паралингвистические элементы речи, такие как смех, кашель, вздохи, стоны, причмокивание, плач и др., хезитационные явления, сопровождающие живую речь.

При разработке аннотирования на уровне высказываний в орфографической записи был применен опыт создателей диалектологических и мультимедийных корпусов русского и калмыцкого языков [2; 3; 4].

Особое внимание уделяется и непреднамеренным остановкам информационного потока в процессе коммуникации, которые обусловлены целым рядом факторов как индивидуальных, психологических, так и физиологических. Как заметил американский паузолог О'Коннор, «паузы могут сказать о человеке не меньше, чем слова, что в разговоре на них уходит 40—50% времени.» [Цит. по: Введенская и др. 2005]. Поэтому разработчиками было решено зафиксировать паузы хезитации при транскрибировании диалектных текстов.

В большинстве случаев такие паузы возникают в случае наличия у говорящего определенных сложностей «в планировании текущего высказывания – чем больше этих сложностей, тем более вероятно появление паузы хезитации и тем больше ее продолжительность» [5].

В потоке речи информанта различают короткую паузу () и длительную (..).

Паузы хезитации могут быть и заполненными некоторыми звуками, в таком случае соответствующие буквенные символы ставятся внутри скобок: напр., (ы-ы).

В потоке речи в связи со спонтанностью и неподготовленностью устной речи нередко наблюдается употребление неправильных конструкций в грамматическом плане, повторов, вставных конструкций, самокоррекций, обрывов [6, с. 24]. Иначе говоря, неподготовленная устная речь самоорганизуется в процессе порождения, в отличие от речи письменной, где наблюдается только фиксированный результат [7, с. 27].

В процессе транскрибирования аудиофайлов вышеуказанные хезитационные явления учитываются и передаются специальными знаками.

Самокоррекция возникает в тех случаях, когда информант решает, что по той или иной причине определенный фрагмент порожденного им высказывания не соответствует тому, что он намеревался произнести. Информант немедленно реагирует на ошибку, например, на неправильное слово в соответствующем контексте

(лексико-семантические ошибки) или некорректную грамматическую форму (морфо-синтаксические ошибки), и в последующем исправляет неверный элемент корректным с помощью замены: *Этийемде (э-э) сугышка [алып кителгэн] / алып киткэндэр март айында // 'Отца [был отправлен] отправили на фронт в марте'* [информант 1].

Особое место занимают в потоке речи обрывы (..), когда информант начинает произносить то или иное слово, по какой-то причине, не закончив ее, начинает вновь: *анан азак (ы-ы) / уралдан бер # участок # алыб / ул уралдагы учасканы / (//) өй.. (//) / беззең өйзән ул бер + умбиш километр йер / урал аша үтэргә кирәк / урал аша / шул // 'потом взяли на урале покосный участок/ до которого было около пятнадцати километров от нашего дома / и надо было перейти через урал / вот'* [информант 5]

Просодические характеристики речи, как ударение, интонация, тон, на данном этапе транскрибирования диалектного материала не учитывались.

Языковые явления, такие как элизия и ассимиляция, то есть изменения, происходящие в потоке речи на стыке слов, тоже выделяются специальным знаком (+). Например: + *оланайымдар беззең менән йәшәй (ммм) — оло апайымдар беззең менән йәшәй 'старшая сестра со своей семьей живет с нами'* [информант 10].

Если в письменной речи структурно-смысловое членение высказывания осуществляется с помощью пунктуационных средств, то в устной речи звуковой поток делится на синтагмы (/) (смысловое целое, отделенное небольшой паузой) и фразы (//) (законченное целое, которое может состоять из группы синтагм, но может состоять и из одной синтагмы, и которое нормально характеризуется конечным понижением тона» [8, с. 87].

Специальные знаки транскрибирования:

/ – знак сегментирования синтагмы;

// – знак сегментирования повествовательного высказывания (членение на фразы и синтагмы осуществляется с учетом интонационно-синтаксических характеристик отрезков звуковой цепи);

! ? – знак сегментирования вопросительных и восклицательных фраз;

(..) – знак длительной паузы в потоке речи информатора;

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

- (//) – обрыв высказывания перед последующим словом;
 [] – самокоррекция (неправильное слово или грамматическая форма, которые в последующем исправляются информантом);
 .. – обрыв слова перед последующим словом;
 \$ Н – не поддающееся расшифровке или неуверенно расшифрованное слово или словосочетание;
 \$ С – паралингвистический элемент речи – смех;
 \$ К – паралингвистический элемент речи – кашель;
 \$ В – паралингвистический элемент речи – вздох, стон, причмокивание и др.
 # ... # – переключение языковых кодов (с башкирского на русский);
 + – для указания языковых явлений, как элизия и ассимиляция (+баралманы – бара алманы, +кайтыб*ара – кайтып бара).

В диалектном корпусе такая многоуровневая разметка, включающая и паралингвистические элементы, хезитационные явления окажется бесценным источником при исследовании языка с точки зрения когнитивных, социологических и психологических подходов и повысит информативность корпуса и расширит круг задач, решаемых с помощью корпуса.

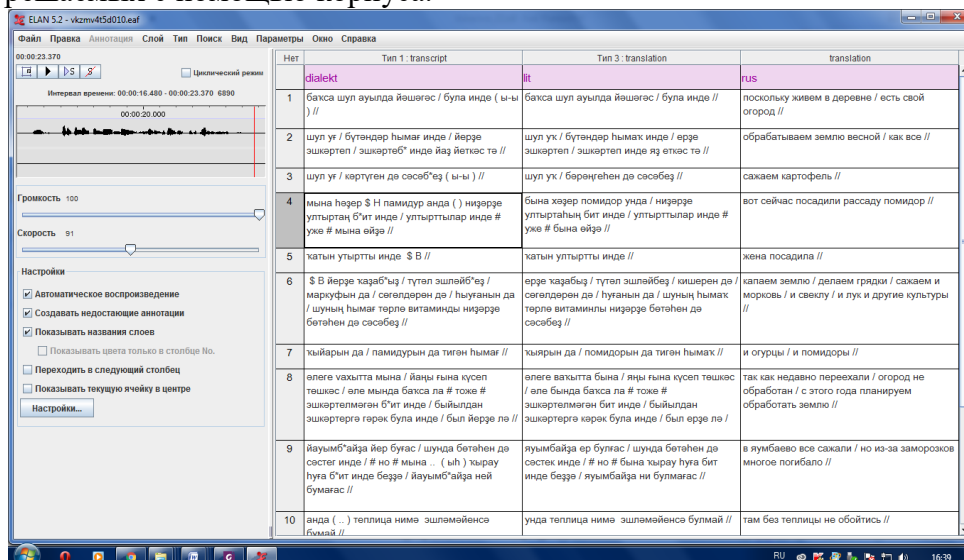


Рис. 1. Пример представления аудиоматериала с транскрибированием, литературным вариантом транскрипции и его переводом на русский язык в системе ELAN.

На сегодняшний день обработаны аудиофайлы 30 информантов по Бурзянскому, Учалинскому и Дуванскому районам Республики Башкортостан. Предстоит сбор полевых материалов по населенным пунктам остальных районов, которые были определены как опорные для построения Диалектологического атласа башкирского языка.

В дальнейшем планируется разработка корпуса-менеджера для осуществления поисковых запросов по транслитерации и литературному эквиваленту.

Материалы диалектного корпуса внесут вклад в пополнение лексикографического фонда по диалектам башкирского языка за счет выявления новых лексических единиц, расширения семантических значений уже зафиксированных диалектизмов.

Сегодня повсеместно наблюдается нарушение орфоэпических норм национального языка носителями всех слоев городского населения и молодым поколением на селе. Создание аудиокорпуса разговорной речи носителей диалектов позволит выделить общие орфоэпические нормы, откроет возможность для разработки орфоэпического словаря, который, к сожалению, до сих пор отсутствует. Корпус разговорной речи носителей диалектов сыграет роль в сохранении своеобразия звучащей речи.

Корпус восточного диалекта башкирского языка также предоставит возможность свободного доступа лингвистов различной специализации к первичному диалектному материалу, выступит в качестве бесценного источника для изучения отдельного говора, установления территории распространения того или иного языкового явления, станет базой для изучения исторического развития и становления литературного языка, социолингвистического анализа, сравнительно-сопоставительных исследований языков.

ЛИТЕРАТУРА

1. Кибрик А. А., Подлеская В. И. К созданию корпусов устной русской речи: принципы транскрибирования // Научно-техническая информация. Серия 2: Информационные процессы и системы. 2003, № 10. С. 5–13.
2. Степанова С. Б., Асиновский А. С., Богданова Н. В., Русакова М. В., Шерстинова Т. Ю. Звуковой корпус русского языка повседневного общения «один речевой день»: концепция и состояние формирования // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 4–8 июня 2008 г.). Вып. 7 (14). – М.: 2008. С. 488–495.

3. Куканова В.В., Бембеев Е.В., Убушаев Н.Н., Манджиева Б.Б. Устные тексты на калмыцком языке: запись и расшифровка // Вестник калмыцкого университета. 2013 г. №3 (19). С. 56–64.
4. Блинова О.В. Будешь задумываться / что говорить / а что не говорить: парадокс наблюдателя и речевой контроль в ОРД [Электронный ресурс]. URL: <https://www.slm.uni-hamburg.de/slavistik/forschung/veranstaltungen/symposium-sprachvariation/downloads-symposium-april-2016/blinova.pdf>. (дата обращения: 20.02.2019).
5. Белицкая А.А. О роли хезитационных пауз в спонтанной речи // Филология и литературоведение. 2014. № 2 [Электронный ресурс]. URL: <http://philology.snauka.ru/2014/02/697> (дата обращения: 20.02.2019).
6. Бускунбаева Л. А. Закономерности речевой экономии и их отражение в башкирском языке. Уфа: Гилем, 2008. 140 с.
7. Звуковой корпус как материал для анализа русской речи. Ч. 1. Чтение. Пересказ. Описание /отв. ред. Н. В. Богданова-Бегларян. СПб.: Филол. ф-т СПбГУ, 2013. 532 с.
8. Щерба Л. В. Фонетика французского языка. Очерк французского произношения в сравнении с русским: пособие для студентов факультетов иностранных языков. М.: Изд-во лит. на иностр. языках, 1953 (1937). 257 с.
9. Введенская Л.А., Павлова Л.Г., Катаева Е.Ю. Русский язык и культура речи: Учебное пособие для вузов 12-е изд. Ростов Н/Д: Феникс, 2005. – 544 с.
10. Саратовский мультимедийный диалектный корпус. [электронный ресурс] // URL: <http://ruscorpora.ru/sbornik2005/13letuchy.pdf>. (дата обращения: 25.03.2019); Мультимедийный корпус диалектных текстов Устьянского района Архангельской области [электронный ресурс] URL: <http://www.slavist.de/Pushkino/> (дата обращения: 25.02.2019); Национальный корпус русского языка. [электронный ресурс] URL:<http://www.ruscorpora.ru> (дата обращения: 25.10.2018); Корпус ижемского диалекта коми языка. [электронный ресурс] URL:<http://videocorpora.ru/ru/allmedia> (дата обращения: 25.03.2019); Национальный корпус калмыцкого языка. [электронный ресурс] URL:kalmcorpora.ru (дата обращения: 25.03.2019).
11. Информационная система для описания малочисленных языков народов мира. Создание описаний алтайских и уральских языков России, находящихся на грани исчезновения. [электронный ресурс] // URL: <http://lingvodoc.ispras.ru/info> (дата обращения: 20.03.2019).

REFERENCES

1. Kibrík A. A., Podlesskaia V. I. K sozdaniú korpýsov ýstnoi rýsskoj rechi: printsipy transkribirovaniia // Naúchno-tehnicheskaja informatsiia. Serija 2: Informatsionnye protsessy i sistemy. 2003, № 10. S. 5–13.
2. Stepanova S. B., Asinovskii A. S., Bogdanova N. V., Rýsakova M. V., Sherstinova T. Iú. Zvýkovoi korpýs rýsskogo iazyka povsednevnogo obeniia «odin rechevoi den»: kontseptsija i sostoianie formirovaniia // Kompýuternaia lingvistika i intellektýalnye tehnologii: Po materialam ejegodnoi Mejdýnarodnoi konferentsii «Dialog» (Bekasovo, 4–8 iúnia 2008 g.). Вып. 7 (14). – М.: 2008. S. 488–495.

3. Kýkanova V.V., Bembeev E.V., Ýbýshaev N.N., Mandjueva B.B. Ýstnye teksty na kalmytskom iazyke: zapís ı rasshifrovka // Vestnik kalmytskogo ýnıversıteta. 2013 g. №3 (19). S. 56–64.
4. Blinova O. V. Býdesh zadýmyvatsia / chto govorit / a chto ne govorit: paradoks nablıýdatelia ı rechevoi kontrol v ORD [Elektronnyı resýrs]. URL: <https://www.slm.uni-hamburg.de/slavistik/forschung/veranstaltungen/symposium-sprachvariation/downloads-symposium-april-2016/blinova.pdf>. (data obraeniia: 20.02.2019).
5. Belitskaia A.A. O roli hezitatsionnyh paýz v spontannoı rechi // Filologıa ı literatýrovedenie. 2014. № 2 [Elektronnyı resýrs]. URL: <http://philology.snauka.ru/2014/02/697> (data obraeniia: 20.02.2019).
6. Býskýnbaeva L. A. Zakonomernosti rechevoi ekonomıı ı ih otrajenie v bashkırskom iazyke. Ýfa: Gilem, 2008. 140 s.
7. Zvýkovoı korpýs kak material dlia analiza rýsskoı rechi. Ch. 1. Chtenie. Pereskaz. Opisanie /otv. red. N. V. Bogdanova-Beglarian. SPb.: Filol. f-t SPbGÝ, 2013. 532 s.
8. Sherba L. V. Fonetika frantsýzskogo iazyka. Ocherk frantsýzskogo proiznosheniia v sravnenıı s rýsskim: posobie dlia stýdentov fakýltetov mostrannyh iazykov. M.: Izd-vo lit. na mostr. iazykah, 1953 (1937). 257 s.
9. Vvedenskaia L.A., Pavlova L.G., Kataeva E.Íý. Rýsskii iazyk ı kúltúra rechi: Ýchebnoe posobie dlia výzov 12-e izd. Rostov N/D: Feniks, 2005. — 544 s.
10. Saratovskii múltimedıunnyı dialektnyı korpýs. [elektronnyı resýrs] // URL: <http://ruscorpora.ru/sbornik2005/13letuchy.pdf>. (data obraeniia: 25.03.2019); Mýltimedıunnyı korpýs dialektnyh tekstov Ýstianskogo raiona Arhangel'skoi oblasti [elektronnyı resýrs] URL: <http://www.slavist.de/Pushkino/> (data obraeniia: 25.02.2019); Natsionalnyı korpýs rýsskogo iazyka. [elektronnyı resýrs] URL:<http://www.ruscorpora.ru> (data obraeniia: 25.10.2018); Korpýs ijemskogo dialekta komı iazyka. [elektronnyı resýrs] URL:<http://videocorpora.ru/ru/allmedia> (data obraeniia: 25.03.2019); Natsionalnyı korpýs kalmytskogo iazyka. [elektronnyı resýrs] URL:kalmcorpora.ru (data obraeniia: 25.03.2019).
11. Informatsionnaia sistema dlia opisaniia malochislennyh iazykov narodov mira. Sozdanie opisaniı altaiskih ı ýralskih iazykov Rossii, nahodıaihsia na granı ischeznoveniia. [elektronnyı resýrs] // URL: <http://lingvodoc.ispras.ru/info> (data obraeniia: 20.03.2019).

Аңдатпа

Мақалада Уфа федералды ғылыми зерттеулер орталығы Тіл білімі және Әдебиет ҒЗИ (ИИЯЛ УФИЦ РАН) лабораториясында алынған далалық экспедициялар аудиоматериалдарын транскрипциялау принциптері туралы айтылады. Бұл жұмыс башқұрт тілі шығыс диалектісі сөйленістерінің аудиокорпусын құру жобасы бойынша басталды. Жұмыстың транскрипциясы сөйлеу корпусы талдауларында кең қолданысқа ие болған «жартылай орфографиялық жазба» түрінде орындалды.

Авторлар башқұрт тілі орфографиясына сай келмейтін диалектілік дыбыс ерекшеліктерін жазба тәсілімен берудің бірыңғай жүйесін жасаған. Олар ол дыбыстарды білдіру үшін арнайы таңбалар мен әріп тіркестерін қолданған.

Корпустағы сипаттама бірлігі сөз ғана емес, сонымен бірге күлкі, жөтел, күрсіну, күңірену, таңдайын тақылдату, жылау, т.б. тәрізді сөйлеудің паралингвистикалық

элементтерін білдіретін және басқа да тірі сөзде қабаттаса қолданылатын хезитациялық құбылыстар – сегменттеу белгілері, символдар болып табылады.

Орфографиялық жазбадағы айтылым деңгейіндегі аңдатпаны құру кезінде орыс және қалмақ тілдерінің диалектологиялық және мультимедиялық корпусын құрушылардың тәжірибесі негізге алынды. Белгілі бір уақыттағы диалектілік материалды транскрипциялауда екпін, интонация, тон сияқты сөйлеудің просодиялық сипаттамаларына мән берілген жоқ. Элизия мен ассимиляция құбылыстары, яғни, сөйлеу ағынында сөз аралық түйісу кезінде пайда болатын өзгерістер зерттеу нысанына енгізілді. Құрамында паралингвистикалық элементтер, хезитациялық құбылыстар тәрізді көп деңгейлі өлшемдері бар диалектілік корпус когнитивтік, социологиялық және психологиялық тұрғысынан алғанда зерттеу кезінде баға жетпес қайнар көз болады. Сонымен бірге корпусстың ақпараттылығын ұлғайтады және корпус көмегі арқылы шешілетін міндеттер ауқымын кеңейтеді.

Кілт сөздер: лингвистика, башқұрт тілі, диалектология, корпуслық лингвистика, транскрипциялау, аудиоматериалдар

(З.А.Сиразитдинов, Л.А.Бускунбаева, А.Ш.Ишмухаметова. Башқұрт тілінің диалектологиялық аудиокорпусына қажетті дыбыстық материалдардың талдамасы)

Özet

Makalede Başkurt Özerk Cumhuriyeti Ufa İlmî Araştırmalar Merkezi Dil Bilimi ve Edebiyat Enstitüsü laboratuvarında saha araştırmalarından elde edilen ses malzemelerinin transkripsiyon ilkeleri üzerinde durulmuştur. Bu çalışmaya Başkurt Türkçesinin doğu diyalektine ait ses laboratuvarını kurma planı doğrultusunda başlanmıştır. Çalışmanın transkripsiyonu diyalekt araştırmalarında oldukça geniş kullanıma sahip “yarı orfografik yazma” şeklinde gerçekleştirilmiştir.

Yazarlar Başkurt Türkçesi imlasına uymayan diyalektik ses özelliklerini yazma yöntemiyle vermenin metodlarını belirleyerek bu sesleri göstermek için özel damgalar ve harfler için ek işaretler kullanmışlardır.

Ses laboratuvarında sadece kelimeler değil, gülme, aksırma, tıksırma, hapşırma gibi konuşmanın paralinguistik elementlerini gösteren semboller de yer almaktadır.

Çalışmada Rus ve Kalmuk dillerinin diyalektolojik ve multimedya laboratuvarlarını kuran ekollerin tecrübeleri esas alınmıştır. Belli bir diyalektolojik malzemeyi transkripsiyonunda aksan, tonlama, vurgu gibi faktörler dikkate alınmamıştır. Sesin bir söylenişte çıkarılması ile asimilasyona bağlı değişiklikler, yani konuşma sırasında söz arasında meydana gelen değişiklikler araştırmaya dahil edilmiştir. Diyalektolojik ses laboratuvarı sosyolojik ve psikolojik açıdan bakıldığında araştırma açısından paha biçilemez bir kaynak olacaktır. Bununla birlikte ses laboratuvarının bilgi içeriğini artırmayı ve laboratuvar yardımıyla çözülen problemlerin kapsamını genişletecektir.

Anahtar Kelimeler: linguistik, Başkurt Türkçesi, diyalektoloji, transkripsiyon, ses malzemeleri

(Z. A. Sirazitdinov, L. A. Buskunbayeva, A. Ş. İşmuhametova, Başkurt Türkçesinin Diyalektolojik Ses Laboratuvarı İçin Gerekli Malzemelerin Araştırılması)