

Полевые материалы в электронных изданиях

А.А. Галлямов, Б.В. Орехов

Об электронном научном издании «Фольклорный архив Башкирского государственного университета»

Статья посвящена описанию электронного ресурса, представляющего тексты на башкирском языке, записанные в фольклорных экспедициях, организованных Башкирским государственным университетом в течение последних десятилетий. Издание сопровождается подробной метаинформацией, указателем словоформ, географической картой и визуализацией количественных параметров. Сам ресурс и статья о нем завершены несколько лет назад, за это время появилось много аналогичных электронных изданий, выстроенных на тех же принципах.

Ключевые слова: фольклор, электронное издание, геоинформационные системы, digital humanities.

За многие годы своего существования Башкирский государственный университет (далее – БашГУ) накопил значительное количество самых разнообразных архивных фондов. Значительную часть этого научного наследия составляют фольклорные архивы, хранящиеся на нескольких кафедрах факультета башкирской филологии и журналистики и филологического факультета БашГУ.

Состояние фольклорного архива, видимо, можно назвать вполне удовлетворительным, хотя предстоит еще значительная работа по паспортизации каждого записанного произведения, выявлению, насколько это возможно, всех участников фольклорных экспедиций, составлению детальной описи каждого из имеющихся архивов, а также сводной описи всех архивных материалов БашГУ. Все это необходимо сделать для того, чтобы в конечном счете предоставить каждому желающему, интересующемуся историей, культурой, фольклором региона, возможность обращаться непосредственно

к текстам устного народного творчества в том виде, в каком они были записаны на территории Башкортостана и соседних регионов: Пермского края, Свердловской, Челябинской, Курганской, Оренбургской областей, Татарстана.

Публикация фольклорных архивов должна стать несомненным вкладом в развитие краеведения; она позволит проводить историко-филологические, историко-культурные, антропологические, этнографические и, разумеется, собственно лингвистические и фольклористические исследования в более полном объеме, с вовлечением в работу новых специалистов. Тем более что, судя по имеющимся данным, значительный пласт фольклорных архивов создан и в других вузах республики, где хранятся материалы по фольклору русских, татар и других народов Башкортостана: удмуртов, чувашей, марийцев.

В современных условиях в первую очередь, конечно, нужно думать об общедоступном электронном издании (находящемся в свободном доступе в сети Интернет), у которого множество преимуществ: легкость редактирования текстов в случае обнаружения ошибок, создание специальных интерактивных надстроек, что невозможно в бумажном издании, и др. В то же время пока еще невелико доверие к текстам, представленным в таком формате. Действительно, зачастую те, кто готовит произведения мировой литературы к размещению во Всемирной сети, делают это небрежно. Так что необходимо соблюсти и все требования текстологии и филологической культуры.

Именно такое издание было создано нами в лаборатории компьютерной филологии БашГУ.

Электронное издание фольклорного архива БашГУ в первую очередь преследует научно-исследовательские цели: предоставить исследователям из любого научного центра возможность непосредственного обращения к текстам, сопровождаемым всей необходимой информацией, представляемой в паспорте произведения. Однако думается, что электронное научное издание будет интересно не только специалистам в области фольклора и диалектологии, но и самому широкому кругу лиц. В частности, преподавателям общеобразовательных и средних специальных учебных заведений при проведении занятий по родному языку и литературе, краеведению и т. д., поскольку на сайте издания предоставляется необходимый инструментарий для работы с текстами: поиск по базе данных по заданным параметрам, наглядное представление информации в виде указанных на интерактивной карте районов и населенных пунктов и многое другое.

На сегодняшний день основное внимание уделено представлению в сети Интернет фольклорных материалов, история собирания, публикаций и изучения которого в Республике Башкортостан составляет без малого две сотни лет. Большой вклад в деятельность по сбору фольклорных произведений внесли такие ученые и деятели культуры, как М. Бурангулов, С. Галин, Л. Бараг, Ф. Надршина, Б. Ахметшин, Г. Хусаинов и др.

Важную роль в работе по сбору и публикации фольклорного материала в республике и сопредельных территориях сыграл Институт истории языка и литературы, вошедший в 1951 г. в состав Башкирского филиала АН СССР. В конце 1950-х гг. к этой деятельности в форме фольклорных экспедиций подключаются преподаватели и студенты БашГУ и других вузов республики. Работа получает новый виток развития, издаются обобщающие сборники фольклорных произведений: трехтомное собрание «Башкирское народное творчество» («Башкорт халык ижады», 1954–1959 гг.)¹, сборники «Башкирские народные песни» («Башкорт халык йырлары», 1954 г.)², «Башкирские легенды» («Башкорт легендалары», 1969 г.)³, «Народные сказки, легенды, предания и были, записанные в Башкирии на русском языке в 1960–1966 гг.» (1969 г.)⁴, «Материалы и исследования по фольклору Башкирии и Урала» (1974 г.)⁵, «Сказки, легенды, предания Башкирии в новых записях на русском языке» (1975 г.)⁶. С 1974 по 2006 г. выходит межвузовский научный сборник «Фольклор народов РСФСР» (с 1991 – «Фольклор народов России»).

С конца 1960-х гг. началась работа по изданию фундаментального многотомного научного свода «Башкирское народное творчество» («Башкорт халык ижады»), вариант на башкирском языке вышел в 18 томах, на русском – в 13 томах. Институтом истории языка и литературы начата подготовка издания свода памятников башкирского фольклора в 55 томах. В 2002 г. кабинет министров Республики Башкортостан утвердил разработанную институтом программу по изучению, возрождению и развитию фольклора народов Республики Башкортостан, согласно которой предусмотрено издание в виде научных сводов лучших образцов как башкирского фольклора, так и фольклора других народов республики. Программа предполагает создание республиканского фонда фольклорных материалов, их учет и систематизацию на основе применения компьютерной техники, подготовку и издание памятников устного и музыкального народного творчества в соответствии с требованиями международных стандартов.

Современное развитие компьютерных технологий и сети Интернет в определенной степени корректирует данную программу

в сторону ее обогащения и упрощения доступа к фольклорным фондам. До сего времени к богатейшим фольклорным архивам БашГУ имели доступ только специалисты. К тому же в перечисленных выше изданиях нашли отражение далеко не все памятники народного творчества. Материалы научных экспедиций большей частью оседают в архивах и не всегда доступны для всестороннего научного изучения. В связи с чем возникает необходимость вовлечения их в научно-исследовательский оборот путем оцифровки, предварительной обработки, классификации и предоставления общего доступа в сети Интернет. Этот факт и послужил отправной точкой в создании электронного научного издания (далее – ЭНИ) фольклорных фондов и архивов БашГУ.

ЭНИ является web-ориентированным электронным ресурсом. В отличие от начального периода массовой компьютеризации, когда программные продукты распространялись преимущественно на портативных носителях (флоппи-дисках, CD, DVD и магнитных накопителях), современная идеология и практика распространения информации предполагает наличие общедоступного ресурса в сети Интернет. Воспользоваться возможностями ЭНИ может каждый, введя в программе для навигации по Интернету адрес <http://lcpb.bashedu.ru/index.php?go=editions>.

Электронные научные издания в последние годы стали активно развиваться и приобрели свои образцы и антипримеры. На сегодняшний день главным ориентиром для составителей ЭНИ является Фундаментальная электронная библиотека «Русская литература и фольклор» (далее – ФЭБ). В ЭНИ ФЭБ публикуются только материалы, уже изданные на бумаге: в книгах и авторитетных периодических изданиях. ЭНИ БашГУ предполагает широкий доступ к материалам, которые до сих пор были доступны только специалистам.

Электронное научное издание «Фольклорный архив Башкирского государственного университета» – это постоянно пополняемая полнотекстовая сетевая многофункциональная коллекция фольклорных записей, собранных в Республике Башкортостан и соседних областях студентами и сотрудниками университета на протяжении последних шестидесяти с лишним лет. От ФЭБ это издание отличается прежде всего категориальным распределением содержимого (контента). Если ФЭБ представляет как критические издания собственно художественных и фольклорных текстов, так и авторитетные аналитические работы, в ЭНИ «Фольклорный архив Башкирского государственного университета» на сегодняшний день в текстовом виде представлены только сами фольклорные

записи. Размещение в сети работ башкирских фольклористов – дело отдаленного будущего.

Цели проекта:

- 1) аккумуляция оцифрованных версий фольклорных произведений, собранных студентами и преподавателями БашГУ в фольклорных экспедициях с 1940-х гг. по настоящее время;
- 2) предоставление всем заинтересованным лицам свободного доступа к фольклорным архивам БашГУ;
- 3) повышение эффективности научных исследований фольклорных произведений с помощью многофункциональной программно-информационной среды;
- 4) популяризация народного творчества.

ЭНИ предоставляет следующие пользовательские функциональные возможности:

- 1) доступ пользователей к информации;
- 2) поиск отдельных произведений по базе ЭНИ;
- 3) лексический поиск с помощью указателя словоформ;
- 4) работа со словарем частотности, связанным со страницами, на которых отображаются контексты употребления словоформ;
- 5) экспорт информации;
- 6) оформление полноценных библиографических ссылок согласно ГОСТу;
- 7) визуализация информации о распределении записей по жанрам и районам.

Тексты в ЭНИ публикуются в том виде, в котором они зафиксированы в архиве, то есть никакой редакторской (и тем более литературной) правки, кроме исправления явных ошибок, не проводится. Во всех случаях сохраняются орфография и пунктуация источников. Важнейший принцип, положенный в основу издания – точность и аутентичность передачи исходной текстовой информации. Собранные в ЭНИ тексты публикуются впервые.

Основной единицей ЭНИ является фольклорная запись – текст, записанный участниками экспедиции в конкретном населенном пункте от конкретного информанта. Эта метатекстовая информация прикладывается к каждой записи и может служить параметром пользовательского поиска.

В ЭНИ используется двуязычная система (в перспективе планируется использование мультязычной системы): фольклорные тексты представлены только на языке оригинала, описания и элементы интерфейса – на русском языке.

ЭНИ фольклорных фондов и архивов является динамической системой, так как задача представления в электронном виде всего

фольклорного фонда БашГУ недостижима моментально. Отправным пунктом оформления ЭНИ фольклорных фондов и архивов БашГУ послужил фольклорный фонд кафедры башкирской литературы и фольклора факультета башкирской филологии и журналистики БашГУ.

Начало создания данного архива было положено в конце 1950-х гг., после того как Башкирский государственный педагогический институт им. К.А. Тимирязева (БГПИ) был преобразован в Башкирский государственный университет и новая университетская кафедра начала организацию фольклорных экспедиций студентов. В разное время фольклорной практикой студентов-филологов руководили такие преподаватели кафедры, как доц. Г.С. Амантаев, проф. Р.Н. Баимов, проф. С.А. Галин, проф. М.Х. Идельбаев, проф. Т.А. Кильмухаметов, проф. А.Н. Киреев, ст. преп. Г.М. Кутуева, доц. М.Х. Мингазетдинов, ст. преп. А.Н. Сафиуллин, доц. Ф.Ш. Сибгатов, проф. А.М. Сулейманов, доц. А.И. Чанышев. На сегодняшний день фонд кафедры состоит из 145 томов машинописных и 53 томов рукописных текстов, которые и предполагается перевести в форму ЭНИ в первую очередь.

Записи представлены в текстовом формате, снабжены метаинформацией и ссылкой на источник (номер дела и листа в архиве). Метаинформация включает название текста, его жанровую принадлежность, место и год записи, язык текста, источник публикации, имя собирателя и информанта.

Помимо самих текстов и необходимого для публикации фольклорных записей метаинформации, ЭНИ оснащено специфическими для электронного издания сервисами. *Интерактивное оглавление* позволяет выстраивать опубликованные тексты в удобном пользователю порядке, группируя их по заглавию, жанру, году или месту записи. Приложенные к каждой публикации *интерактивные карты* с отмеченным на них местом записи текста созданы на основе сервиса Google Maps.

Автоматически составляемый *указатель словоформ* издания: каждую представленную в указателе словоформу сопровождают данные о ее частотности и ссылка, по которой можно посмотреть все контексты употребления формы в ЭНИ. Со страницы, отражающей контекст употребления словоформы, можно перейти непосредственно к тексту в издании, где искомая словоформа будет подсвечена. Имеется также и специальный *словарь частотности* словоформ издания: указатель обновляется автоматически и не зависит от графика добавления новых текстов.

Автоматически генерируемая библиографическая информация. Ссылка для цитирования текстов издания в научных работах размещается внизу каждого документа под заголовком «Ссылка на эту публикацию». Ссылка соответствует ГОСТ 7.0.5-2008 «Система стандартов по информации, библиотечному и издательскому делу. Библиографическая ссылка. Общие требования и правила составления» и включает в себя информацию о заглавии текста, названии и годе издания, а также о дате добавления текста и дате обращения пользователя.

Экспорт в формат XML. (Extensible Markup Language – рекомендуемый научным сообществом формат филологических электронных публикаций. Все помещенные в ЭНИ тексты доступны для пользователя в этом формате в соответствии со стандартами международного консорциума TEI (Text Encoding Initiative).

Кроме прочего, сервисы ЭНИ позволяют строить автоматически генерируемые *таблицы и графики распределения текстов по районам записи и жанрам*; автоматически создаваемые *интерактивные карты-оглавления* с отмеченными на них местами записи текстов. Такое оглавление включает название текста (оно одновременно является ссылкой, по которой можно перейти к самой записи), жанровое определение, указание на место записи, сокращенное до района (полную информацию, включающую название населенного пункта, можно посмотреть на странице записи в блоке метаинформации), год записи, язык текста и источник публикации. В ряде случаев название дается публикатором, тогда в оглавлении это специально не отмечается, но на странице записи такие названия помещены в угловые скобки. Жанровые определения даются в сокращенном виде, полностью они представлены на странице записи.

Список сокращений в названиях жанров:

- песня б – бытовая песня;
- песня л – лирическая песня;
- сказка б – бытовая сказка;
- сказка в – волшебная сказка;
- сказка ж – сказка о животных.

По щелчку на заголовке столбца с названием текста, жанровым определением, местом записи, годом, языком или источником можно изменить порядок расположения записей в оглавлении, выстроив его в прямом (один щелчок) или обратном (два щелчка) алфавитном или числовом порядке этого столбца.

В ЭНИ на данный момент представлены тексты, записанные в Салаватском, Кугарчинском, Зианчуринском, Баймакском, Бурзянском, Мечетлинском, Абзелиловском, Давлекановском, Ку-

юргазинском, Альшеевском и других районах республики, а также в Оренбургской области. Районы работы фольклористов обозначены на карте на одной из интернет-страниц ресурса.

Пользователями ЭНИ могут быть как специалисты-филологи, так и школьные учителя, а также любители народного творчества. Созданная электронная платформа для публикации фольклорных текстов позволит в будущем как расширять и совершенствовать функциональность издания, так и проводить систематические исследования на основе оцифрованной коллекции. Причем это смогут делать не только специалисты в области фольклора, но и все интересующиеся народным творчеством.

Главное направление, в котором будет развиваться сопровождающий издание инструментарий исследователя, – это разнообразные поисковые механизмы. Уже сейчас в коллекции можно искать словоформы, находить контексты их употребления и знакомиться с данными о частотности.

В частности, не только университетские, но и школьные преподаватели могут уже в настоящее время проводить занятия по устному народному творчеству или краеведению с привлечением данных, представленных в ЭНИ. Например, поиск дает возможность выделить только те произведения, которые были собраны в том или ином районе или населенном пункте, что позволяет сопоставлять их с текстами, собранными в других районах, выявлять их жанровую специфику. Может оказаться, что для какого-то одного района более характерны байты, в то время как для другого – сказки. Значительную помощь электронный ресурс может оказать и учителям родного языка, поскольку тексты записаны без литературной обработки и сохраняют все диалектные лексические, орфоэпические, грамматические особенности, что дает возможность организовать работу школьных лингвистических и литературоведческих кружков и проводить различного рода внеклассные занятия, посвященные истории и культуре родного края, а также углубленно подходить к изучению различных грамматических закономерностей башкирского языка.

Однако с точки зрения фольклориста – это «микроуровень». Гораздо более важными являются параметры более высокого, чем лингвистический, уровня, относящиеся к организации текста, его места в традиции. Проще всего в этом смысле с повествовательными текстами: для них давно разработаны и каталогизированы классифицирующие признаки. Для сказок это прежде всего персонажи с социальным статусом и/или определенными функциями (мужик, поп, солдат, мачеха, младший сын и т. д.), объекты, определяющие

пространственные параметры повествования (река, гора, озеро и т. д.), зооморфные персонажи (лиса, медведь, конь и т. п.), а также специфические предметы, формирующие сказочные мотивы (веретено, меч, сундук и т. п.). Эти признаки будут задействованы и в организации материалов ЭНИ.

В лирических жанрах такие параметры теряют свою релевантность. Однако во всех без исключения текстах так или иначе проявляют себя ключевые понятия языковой и фольклорной картины мира. Следующим исследовательским этапом после собственно подготовки и размещения текстов на сайте должно стать выявление этих понятий и системных связей между ними. Именно ЭНИ позволит в будущем на основе лексикостатистики, прикладных аспектов лингвистики текста, data mining и information extraction проводить исследования, которые будут не только классифицировать имеющиеся записи, но и давать ответы на насущные вопросы культурологии и фольклористики.

Корпусная лингвистика также предлагает свой инструментарий для исследования фольклорной коллекции. Так, фонд ЭНИ, дополненный материалами академического издания башкирского фольклора, может послужить базой для словаря постоянных эпитетов и прочих текстовых клише, которые легко вычлениаются благодаря методике нахождения коллокаций, т. е. типичного и постоянного лексического окружения какого-либо слова, образующего вместе с ним устойчивое словосочетание: например, *в тридевятом царстве в тридесятом государстве, сдать хвосты; борон-борон заманда, тублыгы ла агас зур булмай араһынан үтәр юл булмай*.

ЭНИ также позволит в будущем создавать разного рода справочные ресурсы: указатели, словари и атласы (что возможно благодаря географической составляющей издания). Своего рода новаторской разработкой станут, как мы надеемся, сервисы, основанные на географической части проекта. Как уже отмечалось выше, каждый текст в ЭНИ имеет картографическую привязку. Для пользователя это означает отметку на карте на странице записи. Однако за визуальным планом стоят данные о географических координатах. Например: село Малояз имеет в этой системе координаты 55.177789, 58.157158, деревня Акчура – 51.696991, 57.514141 и т. д. Соответственно на координатную сетку могут быть наложены и мотивы, и грамматические формы, и прочие филологически релевантные параметры текста. Например, вычисляя разницу между координатами и группируя их, будет возможно составить карту распространенности мотивов и словоформ (в том числе диалектных, а также архаичных и прочих маркированных)

по всему ареалу башкирского языка. Таким образом, с увеличением объема ЭНИ появится автоматически составляемый фольклорный и диалектологический атлас.

Коллектив лаборатории компьютерной филологии БашГУ планирует испытать на коллекции фольклорных текстов активно разрабатываемые в современных зарубежных информационно-гуманитарных исследованиях методики автоматического извлечения мотивов, которые в будущем будут систематизированы в форме указателей по жанрам и местам бытования, что позволит создать базу данных по башкирскому фольклору и обеспечит исследователей необходимым инструментарием для проведения научно-исследовательских изысканий.

Таким образом, думаем, что материалы, представленные на сайте лаборатории по адресу <http://lcph.bashedu.ru>, будут интересны широкому кругу пользователей.

Примечания

- ¹ Башкорт халык ижады. Өфө: Башкортостан китап нәшриәте, 1954–59.
- ² Башкорт халык йырлары. Өфө: Башкортостан китап нәшриәте, 1954.
- ³ Башкорт легендалары. Өфө: БДУ нәшриәте, 1969.
- ⁴ Народные сказки, легенды, предания и были, записанные в Башкирии на русском языке в 1960–1966 гг. Уфа: Башкнигоиздат, 1969.
- ⁵ Материалы и исследования по фольклору Башкирии и Урала. Уфа: Изд-во Башкирского ун-та, 1974.
- ⁶ Сказки, легенды и предания Башкирии в новых записях. Уфа: Башкнигоиздат, 1975.