

ЭЛЕКТРОННАЯ ВЕРСИЯ АТЛАСА ТАТАРСКИХ НАРОДНЫХ ГОВОРОВ¹

Ф.И. Салимов, Д.Б. Рамазанова, А.Г. Пилюгин, Р.Ф. Салимов

Описан опыт первого этапа работы по созданию электронного атласа татарских народных говоров. Обсуждаются основные принципы построения и содержание электронной версии атласа.

Ключевые слова. Электронный атлас, диалектология, лингвистическая география, ГИС, диалекты татарского языка

История создания атласа

Печатная версия “Атласа татарских народных говоров” в двух томах вышла в 1989 году [1; 2]. Издание атласа явилось завершением этапа многолетнего кропотливого труда коллектива диалектологов Института языка, литературы и истории им. Г.Ибрагимова АН РТ (ИЯЛИ АН РТ) и представляет собой итог исследования говоров татарского языка методом лингвистической географии². Изучение говоров методом лингвистической географии [3; 4; 5; 6; 7] позволяет объективно и на научной основе уточнить классификацию говоров, установить границы и пределы их распространения, уточнить ареал распространения многих важных с точки зрения языка явлений, в совокупности с историческими материалами раскрыть историю формирования большинства говоров и их носителей, показать на конкретном материале результаты многовековых связей народа с носителями родственных и неродственных языков.

Атлас татарских народных говоров включает в себя все основные районы расселения татар и отражает сведения по фонетике, морфологии, лексике и синтаксису татарского языка, собранные на основе специально разработанной

¹ Работа выполнена при финансовой поддержке РГНФ в рамках проекта Электронный атлас татарских народных говоров (проект №11-04-12020в)

² Лингвистическая география – раздел диалектологии, изучающий закономерности распространения диалектных явлений и их литературных соответствий путем сопоставления лингвистических карт.

программы по собиранию материалов для диалектологического атласа (1959 г.).
Материалы для атласа собирались в течение 1958 – 1980 гг. (первые два тома) и 1997 - 2008 гг. для 3-го тома во время многочисленных экспедиций, организованных ИЯЛИ АН РТ. Первые два тома охватывают татарские народные говоры, распространенные на территории областей и республик Среднего Поволжья и Приуралья; третий том - говоры татарского населения Западной Сибири. При сборе информации происходила тщательная выверка собираемого материала, его интерпретация, паспортизация. Расстояния между охватываемыми для исследования населенными пунктами не превышало 10-15км.

В структурном отношении атлас состоит из 215 карт языковых явлений, вспомогательных карт и сводных карт изоглосс. На картах отражены основные признаки диалектного различия, закономерности и характер распространения диалектных явлений. Тем самым эти карты дают яркие и дифференцированные факты для лингвистической географии, теоретическое обобщение которых – дело будущего. Сводные карты изоглосс представлены в Атласе 17 картами. В них при помощи линий (изоглосс) выделены и обобщены некоторые языковые явления, имеющие сходные очертания.

В атласе каждая карта снабжена комментариями, без которых невозможно было бы чтение карт и понимание их содержания. В картах использована система условных знаков (в виде геометрических фигур³ белого и черного цветов), отражающих территорию распространения отдельных или противопоставленных диалектных явлений. В легендах карт каждый знак привязывается к определенному лингвистическому факту. Кроме того комментарий содержит дополнительную информацию, относящуюся к теме данной карты, но не получившую в ней отражение.

В ходе осуществления лингвогеографического этапа была проведена работа по определению состава диалектной системы татарского языка. При этом были выявлены не только новые говоры и подтверждены ранее известные

³ Далее, ярлыков

классификационные признаки татарских диалектов, но и раскрыты дополнительные специфические черты. В результате появилась возможность по-новому взглянуть на накопленные знания и, используя созданный базис, приступить к обобщающим исследованиям в различных аспектах, к вычленению лингвистических пространств, противопоставленных по определенным признакам остальному лингвистическому континууму, что имеет первостепенное значение, прежде всего, для исследований по исторической диалектологии татарского языка.

Недостатки печатной версии атласа

Издание атласа в 1989 году было ярким достижением татарской диалектологии, на базе атласа были проведены многочисленные исследования и получено достаточно много результатов [8; 9; 10; 11; 12]. Однако в силу специфики книжной формы представления информации можно указать ряд связанных с такой формой изложения недостатков:

Существующая книжная версия атласа представляет собой больших размеров альбом, весящий более 16 кг. Это обстоятельство создает значительные трудности при работе над картами, особенно в случае проведения сравнительных исследований языковых явлений говором, расположенных на определенной территории. Кроме того, в виду малочисленности тиража первых двух томов, исследователи имеют ограниченный доступ к картам атласа: работу над атласом можно проводить, в основном, в библиотеках. Это затрудняет работу лингвистов при необходимости работы с несколькими источниками одновременно. Качество печати карт, с учетом временного старения бумажного носителя, оставляет желать лучшего: некоторые ярлыки и надписи, особенно на местах сгибов карт, износились и плохо различимы. Плотность размещения ярлыков и соответствующих номеров населенных пунктов на картах достаточно высока. Это создает определенные трудности в идентификации соответствующего населенного пункта (в определенных случаях даже создатели атласа затруднялись определить, какой именно идентификационный номер соответствует нанесенному на карту ярлыку). Работать с номерами в

качестве идентификаторов ярлыков языковых явлений не очень удобно, поскольку эти номера необходимо соотносить со справочником населенных пунктов, который включает в общей сложности 1031 наименований. При этом приходится решать следующие задачи:

- по идентифицирующему номеру на карте восстановить наименование населенного пункта, который помечен этим номером;
- по наименованию населенного пункта найти на карте соответствующее языковое явление для этого населенного пункта, а для этого необходим идентифицирующий номер.

В книжной версии для решения этой задачи в комментариях к атласу предлагается два варианта представления списка населенных пунктов: в порядке возрастания нумерации и в алфавитном порядке. Тем не менее, учитывая достаточно большой объем справочника, поиск соответствующей единицы в словаре не очень удобен, отвлекает усилия исследователя от решения основной задачи на выполнения вспомогательных операций по поиску.

Печатная версия атласа представляет собой временной срез состояния административного деления РФ на момент проведения экспедиций и на момент выхода в свет первых двух томов атласа. Понятно, что административное деление слабо связано с лингвистическими особенностями говоров (это отмечается и авторами атласа: в первых двух томах для группировки языковых явлений была предложена определенная километровая сетка). С другой стороны, наложение границ административного деления на языковые карты позволяет понять исторические и миграционные явления размещения носителей говоров в пределах административных образований. С момента выхода первых томов атласа административные границы районов и субъектов изменились, некоторые населенные пункты слились под единым именем или перестали существовать. Это еще раз подчеркивает, что книжная версия атласа не может отражать динамическую временную составляющую и быстро теряет свою актуальность.

Печатная форма атласа представляет собой карту-схему, в которой местоположение населенных пунктов наносилось приблизительно, без точного определения географических координат. Причиной такого положения вещей являлось наличие жесткой цензуры на момент издания атласа. В то время географические карты с указанием координат населенных пунктов были практически недоступны, поэтому населенные пункты наносились на карту-схему с учетом направлений и расстояний от базовых опорных пунктов (в качестве которых выступали крупные населенные пункты) вручную. Понятно, что такой подход может дать только приблизительное расположение населенных пунктов друг относительно друга. Перепроверка местоположений населенных пунктов в ходе построения электронного варианта показало наличие ошибок в картах атласа.

Лингвистическая география и новые информационные технологии

Развитие новых информационных технологий значительно расширило возможности применения новых подходов для анализа и представления лингвистической информации. В частности развитие и внедрение географических информационных систем (ГИС), предназначенных для сбора, хранения, анализа и графической визуализации пространственных данных и связанной с ними информации, позволило создавать электронные аналоги карт. ГИС включают в себя возможности систем управления базами данных (СУБД), редакторов графики и аналитических средств, и широко применяются в картографии, экологии, экономике, муниципальном управлении и во многих других областях. Применение этих систем в лингвогеографии позволило снять многие отмеченные выше проблемы традиционного подхода, сделать лингвистические карты динамичными, доступными широкому кругу пользователей и проводить совместные исследования диалектов на единой базе. К сожалению, отметим, что количество работ в этом направлении в нашей стране незначительно [13; 14]

4. Электронная версия атласа

В 2011 году при финансовой поддержке гранта РГНФ инициативной группой сотрудников института прикладной семиотики АН РТ, Казанского федерального университета совместно с лингвистами ИЯЛИ АН РТ была начата работа по созданию электронного атласа татарских говоров. Проект был задуман как Интернет-ресурс, результаты которого будут открыты для всех пользователей, имеющих доступ к Интернету. Кроме материалов первых двух томов атласа в создаваемый ресурс с согласия руководства ИЯЛИ предполагалось включить неопубликованный материал по татарским говорам Западной Сибири. Электронный атлас состоит из картографической и атрибутивной баз данных. Картографическая база содержит пространственную информацию, включающую следующие слои: административное деление РФ по состоянию на 1995 год, основные водные объекты, опорные пункты, в которых проводились исследования. Разнесение информации по слоям позволяет проводить фильтрацию информации и показывать карты в виде удобном конечному пользователю. Включение в карты слоя с водными объектами продиктовано тем, что очень часто изоглоссы по языковым явлениям проходят по границам водоемов. Атрибутивная база включает в себя информацию по распределению языковых явлений по опорным пунктам, а также содержит дополнительную информацию по фонетическим, морфологическим и лексическим особенностям диалектов татарского языка. В основу создания электронного атласа были положены следующие требования:

1. Электронная версия атласа должна как можно более полно и точно отражать весь материал, содержащийся в печатной версии атласа. Это касается как способов представления информации на картах, так и дополнительной информации, содержащейся в комментариях к атласу. В частности, в книжной версии атласа была принята локальная нумерация населенных пунктов по томам. При этом номера различных населенных пунктов в различных регионах совпадали. При создании электронного атласа разработчики не стали вводить сквозную нумерацию по той

причине, что в изданных работах по диалектологии, созданных на базе атласа имеются ссылки на нумерацию, принятую в книжной версии. Для удобства работы с подобной литературой номера книжной версии были дополнены префиксом номера тома.

2. Должны быть использованы удобные средства представления информации, позволяющие переходы по гиперссылкам от одних составляющих объектов в составе атласа – к другим. Кроме того пользователь должен иметь возможность параллельно работать с картами и комментариями к ним, в реальном режиме времени получать информацию по населенным пунктам, отображенным на карте.
3. Атлас должен содержать библиографический указатель по публикациям, посвященным татарским говорам, который можно просматривать в алфавитном порядке по желанию пользователя, как по авторам, так и по названию работы, по году издания или по виду издания (книга, статья, тезисы). В библиографию дополнительно по сравнению с книжной версией предполагается включить ссылки на статьи по данной тематике, имеющиеся в сети ИНТЕРНЕТ. Кроме того, предполагается включить электронные версии некоторых публикаций по татарской диалектологии, опубликованных в ИЯЛИ АН РТ.
4. Интерфейс языка запросов системы должен быть понятен конечному пользователю; библиотека запросов должна включать основные типы запросов, которые используются лингвистами при решении задач с использованием книжной версии атласа.
5. Должны быть предоставлены средства для корректировки информации привилегированным пользователям, позволяющие поддерживать информацию в актуальном состоянии.
6. Должны быть предоставлены определенные инструменты для работы лингвиста, которые позволяют осуществлять фильтрацию (и/или) сортировку необходимой для работы информации, предоставлению

пользовательских пометок на карте, распечатку отдельной карты (или ее части).

При создании электронной версии атласа необходимо было решить ряд задач:

1. Для каждой из 215 карт восстановить информацию о распределении языковых явлений по опорным пунктам, в которых проводились исследования. К сожалению, эта информация после выхода книжной версии атласа 1989 года была утеряна и сохранилась только в виде ярлыков на картах атласа. Для восстановления этой информации пришлось по каждой карте атласа, восстановить и внести в базу данных информацию о привязке изучаемых языковых явлений к населенным пунктам. При этом ввиду плотности размещения ярлыков, временами возникала неоднозначная трактовка, которая разрешалась экспертным советом лингвистов, участвовавших в создании печатной версии атласа. Эта работа отняла достаточно много времени. На текущий момент создана база данных, содержащая информацию по распределению языковых явлений для первого тома атласа.
2. Картографическая база атласа предполагает включение в базу данных географических координат населенных пунктов. Ввиду того, что со времени проведения экспедиций некоторые населенные пункты перестали существовать или поменяли свои наименования, была проведена ревизия по всем населенным пунктам, включенным в атлас. Поскольку самой доступной и актуальной является база КЛАДР населенных пунктов, публикуемая налоговой инспекцией РФ, каждый населенный пункт был проверен на его присутствие в КЛАДР. При этом поиск осуществлялся по русскоязычным наименованиям населенных пунктов с использованием их административной принадлежности. Информацию по отсутствующим в КЛАДР населенным пунктам дополнительно собиралась поиском через ИНТЕРНЕТ. В процессе такой работы было выявлены населенные пункты, которые на сегодняшний

день прекратили свое существование. Поскольку географические координаты таких поселений найти не удалось, они были нанесены на карту приблизительно (источником информации об их расположении служили карты опубликованной версии атласа). Такие населенные пункты выводятся на картах электронной версии другим цветом. Кроме того, пользователь имеет возможность по выбору посмотреть как русскоязычное, так и татаро-язычное наименование населенного пункта. Поскольку разработчики не удалось найти единой электронной карты, с указанием координат всех населенных пунктов, включенных в состав атласа, информацию по координатам населенных пунктов и границам административных образований пришлось собирать из разных (в основном, региональных) источников. При этом возникли проблемы, связанные с различием представлений границ на разных источниках, которые разрешались с использованием ручной корректировки.

3. При нанесении ярлыков, обозначающих различные языковые явления, авторы электронного атласа старались сохранить преемственность и поддерживать систему обозначений, предложенную авторами книжной версии. Однако выяснилось, что современные шрифты не содержат более половины соответствующих знаков. С помощью графических редакторов была создана база ярлыков, каждый из которых был представлен несколькими цветами для визуализации некоторых особенностей, относящихся к населенному пункту или языковому явлению.
4. Электронная версия атласа поддерживает транскрипцию, принятую для книжной версии. Учитывая, что транскрипция содержит многоэтажные записи, возникла проблема визуализации таких символов в браузерах. Был отвергнут графический вариант представления таких символов, (например, представления текстов в pdf формате), поскольку в этом случае появляются накладные расходы связанные с необходимостью дополнительного времени для считывания графических файлов на компьютер клиента. Был также отвергнут выбор специальных форматов

(в частности, формат редактора WORD), для просмотра которых требуется специализированные инструменты, которые могут отсутствовать на компьютере клиента. Предлагаемое решение предполагает перевод всех текстов в html формат с соответствующим оформлением транскрипционных знаков.

5. Для авторизованных пользователей зарезервирован слой, в который можно вносить изменения. В частности, пользователь может прорисовывать границы изоглосс языковых явлений с дальнейшим выводом получившейся карты на печать.

Для реализации электронной версии атласа был использован следующий инструментарий: WEB-server Apache, СУБД PostGreSQL 9 и MAPServer 6.0.1 с библиотекой OpenLayers 2.10. Текст комментариев к атласу был разбит на части и представлен в нескольких таблицах атрибутивной базы, хранящих информацию об административной принадлежности населенных пунктов, о распределении языковых явлений по населенным пунктам, о привязках соответствующих текстов (комментариев и вопросов программы по собиранию материалов для диалектологического атласа татарских говоров) к картам, библиографическую информацию. Электронная версия атласа предусматривает возможность корректировки информации в атрибутивной базе данных для привилегированных пользователей (в данной версии – это администратор системы). Для осуществления обратной связи с остальными пользователями предусмотрена гостевая книга, в которой каждый пользователь системы может помещать информацию об обнаруженных ошибках, о своих пожеланиях по доработке системы и новых результатах, которые, по его мнению, должны присутствовать в электронном атласе татарских говоров. После анализа этой информации специалистами она может быть внесена в базы атласа администратором системы.

Хранение данных в виде электронной базы данных является особенно полезным для задач статистического анализа. Электронный атлас можно рассматривать как объект, заданный в трехмерном пространстве, где каждая

карта, являясь плоскостью, несет в себе информацию по одному языковому явлению. Набор карт создает объемный образ, и запросы по выбранному набору позволяют сопоставить распределения соответствующих языковых явлений, что может оказаться полезным при решении задач кластеризации говоров. В рамках проекта предполагается создать мастер языка запросов, позволяющий пользователю генерировать запросы как по одному языковому явлению, так и по выбранному набору.

На данный момент проект не завершен. Проведена работа по сбору информации, необходимой для создания электронного атласа, создана и частично заполнена атрибутивная база данных, создан макет электронного атласа.

* * * * *

1. Атлас татарских народных говоров Среднего Поволжья и Приуралья. / Под ред. Н. Б. Бургановой, Л. Т. Махмутовой, Ф. С. Баязитовой, Д. Б. Рамазановой, З. Р. Садыковой, Т. Х. Хайрутдиновой: в 2-х т.. - Казань, Татпрокаттехприбор, 1989 - 240 с.
2. Комментарии к Атласу татарских народных говоров среднего Поволжья и Приуралья. Казань, Татпрокаттехприбор, 1989. - 300 с.
3. Диалектологический атлас русского языка. Центр Европейской части СССР. Под ред. Р. И. Аванесова и С. В. Бромлей. Фонетика. - М., Наука, 1986. - Вып. 1. - 318 с.
4. Диалектологический атлас русского языка. Центр Европейской части СССР. Под ред. Р. И. Аванесова и С. В. Бромлей, Морфология. М., Наука, 1986; - Вып. 2. - 165 с.
5. *Насибуллин Р.Ш., Максимов С.А., Семенов В.Г. Отставнова Г.В.* Диалектологический атлас удмуртского языка, Карты и комментарии, Ижевск: НИЦ «Регулярная и хаотическая динамика», 2009. - Вып. 1 - 260с.
6. Дыялекталагічны Атлас беларускай мовы. - Минск: Выдавецтва Акадэміі навук БССР, 1963. -338 карт

7. Диалектологический атлас башкирского языка. - Уфа: Изд-во "Гилем", 2005. - 234 с., 169 карт.
8. *Арсланов Л. Ш.* Татары Нижнего Поволжья и Ставрополья (язык, фольклор, словарь).- Набережные Челны: Камаз, 1995.- 195 с.
9. *Алишина Х. Ч.* Тоболо-иртышский диалект языка сибирских татар.- Казань: Изд-во Казан. пед. ин-та, 1994. - 119 с.
10. *Рамазанова Д. Б., Хайрутдинова Т. Х.* К вопросу о формировании байкибашевского говора мишарского диалекта татарского языка // Исследования по исторической диалектологии татарского языка. - Казань, 1985. - Вып. 3.- С. 62 - 92.
11. *Рамазанова Д. Б.* Татарская диалектная лексикография и диалектная лексикология на рубеже III тысячелетия // Актуальные вопросы татарского языкознания. - Казань, 2002.- С. 101 - 106;
12. *Рамазанова Д. Б.* К вопросу изучения сибирско-татарских диалектов методом лингвистической географии / X Сулеймановские чтения / Всероссийская научно-практическая конференция (Тюмень, 18 - 19 мая 2007 г.) : Материалы и доклады. - Тюмень: Ситипресс, 2007. - С. 71 - 73.
13. *Соломенников И. С.* Лексический атлас русских говоров Удмуртии в Интернет // Современные информационные технологии и письменное наследие: от древних рукописей к электронным текстам. // Материалы международной научной конференции Ижевск, 13 - 17 июля 2006 г. - Ижевск.: Изд-во ИЖГТУ С. 161-163
14. *Земляков Д.В., Кузнецова Е.В.* Сетевые технологии при обработке диалектного материала: «Лексический атлас Волгоградской области» // Электронный научно-образовательный журнал ВГПУ «Грани познания» 2011, №2. URL: <http://www.grani.vspu.ru/files/publics/1311757938.pdf> (Дата обращения: 19.11.2011)

**THE ELECTRONIC VERSION
OF THE TATAR NATIONAL DIALECTS ATLAS**

F.I.Salimov, D.B.Ramazanova, A.G. Pilugin, R.F. Salimov

The first stage of work on creation of the electronic tatar national dialects atlas is described. The main principles of construction and maintains of the electronic version of the atlas are discussed.

Key words: Digital atlas, dialectology, linguistic geography, DB, GIS, tatar language dialects

* * * * *

Салимов Фарид Ибрагимович - кандидат физико - математических наук, доцент кафедры теоретической кибернетики Института вычислительной математики и информационных технологий Казанского (Приволжского) федерального университета, главный научный сотрудник НИИ Института прикладной семиотики АН РТ.

E-mail:Farid.Salimov@ksu.ru

Рамазанова Дария Байрамовна – доктор филологических наук, главный научный сотрудник отдела лексикологии и диалектологии Института языка, литературы и истории им. Г.Ибрагимова АН РТ.

E-mail:rdb@mail.ru

Пилюгин Александр Георгиевич – старший преподаватель кафедры моделирования экосистем Института экологии и географии Казанского (Приволжского) федерального университета.

E-mail: pag@kzn.ru

Салимов Рустем Фаридович – аспирант кафедры математической статистики Института вычислительной математики и информационных технологий Казанского (Приволжского) федерального университета.

E-mail: Rust1k@mail.ru

Поступила в редакцию 23.11.2011