

## К задаче создания этнолингвистического словаря татарского языка<sup>1</sup>

Салимов Ф.И., Баязитова Ф.С., Салимов Р.Ф

Казанский федеральный университет,  
Институт прикладной семиотики АН Республики Татарстан,  
Институт языка, литературы и искусства АН Республики Татарстан, г.Казань  
farid.salimov@kpfu.ru

**Аннотация.** Рассматривается задача создания этнолингвистического электронного терминологического словаря. Электронный ресурс включает в себя коллекцию образцов живой диалектной речи, собранную в полевых экспедициях. Ресурс привязан к электронному атласу татарских народных говоров. **Ключевые слова** Этнолингвистика, термин, словарь, татарские диалекты, татарский язык

Проблема сохранения и изучения дошедшего из глубины веков культурного наследия является одной из важнейших задач современной науки. При этом, большое значение имеет задача изучения языковых особенностей народов, населявших ту или иную территорию, реконструкции и восстановления древнейших обрядов, обычаев и верований [1]. В последние десятилетия татарскими диалектологами разрабатывается культурно-языковая и обрядовая терминология, сохранившаяся в диалектах татарского языка. Решение этой задачи требует сбора и систематизации очень объемного полевого материала, сохранившегося в коллективной народной памяти.

К настоящему времени усилиями группы ученых института языка, литературы и искусства АН Республики Татарстан (ИЯЛИ АН РТ) собран богатый этнолингвистический материал, относящийся к традиционным обрядам и обычаям татарского населения. Материал собирался на обширной географической территории, включающей в себя более 25 регионов Российской Федерации в течение более чем полувекового интервала времени. Результаты исследований были систематизированы и опубликованы в трудах татарских диалектологов в виде монографий и многочисленных журнальных статей. При этом особая роль отводилась обрядовой терминологии, которая рассматривалась в контексте собранных в полевых экспедициях материалах, воспроизводивших своеобразие и богатую колоритность живой народной речи. В этом направлении можно отметить многочисленные труды Ф.С. Баязитовой, которые характеризуются анализом и систематизацией обширного этнолингвистического материала, связанного с различными событиями жизни человека, описанием обрядовой лексики, мифологии, календарных обычаев. Большинство книг Баязитовой построены по тематически-гнездовому принципу, в основе которого лежит разбиение собранного фольклорного и диалектного материала по тематическим группам с выделением терминов, относящихся к определенному разделу, и описания их семантики. Особенно ценным в публикациях являются многочисленные иллюстративные примеры,

---

<sup>1</sup> Работа выполнена при финансовой поддержке РГНФ (проект № 14-04-12024)

показывающие употребление терминов в различных диалектах татарского языка. Однако книжная форма представления информации затрудняет цельное восприятие приведенного материала, прежде всего, в виду его значительного объема и специфики бумажного изложения. Кроме того, все книги Баязитовой написаны исключительно на татарском языке, что значительно сужает круг возможных исследователей и затрудняет изучение взаимовлияния различных языковых культур, сосуществующих в рамках одной географической территории.

При поддержке фонда РГНФ решалась задача создания электронного этнолингвистического словаря татарского языка, основанного на материале книг, связанных с различными этапами жизни человека (рождением, свадебными церемониями, смертью) [2,3,4]. Создаваемый ресурс представляет собой базу данных, включающую в себя терминологический словарь, описание семантики включенных в словарь терминов, а также коллекцию образцов практического использования включенных в словарь терминов в диалектной речи татар. Кроме того, предполагалось связать термины словаря с базой данных электронного атласа татарских народных говоров с целью визуализации на картах атласа географии распространения терминов в рамках диалектов, в которых они употребляются.

Поскольку основным источником для создания базы данных выступали книги, то прежде всего необходимо было решить задачу сегментации изложенного в них материала выделением самого термина, его семантического описания, диалектов, в которых термин употребляется, а также примеров его употребления в живой диалектной речи. При решении этой задачи возникли определенные трудности. Книги Баязитовой, в большей степени, ориентированы на их чтение человеком. При написании книг изначально не предполагалась их обработка в автоматическом режиме. В текстах книг перемешаны как сами термины, так и примеры их употребления в различных диалектах, их семантическое толкование и другая информация. В книгах не поддерживается единый язык разметки одинаковых по структурному назначению частей текста, многие одинаковые в семантическом отношении фрагменты оформлены различным образом, в выделенных терминах не везде проставлены ссылки на диалекты и говоры, где они встречаются.

Был проведен тщательный анализ оформления фрагментов текста книг. Основная задача состояла в описании системы признаков, характеризующих различные фрагменты текста: термины, ссылки на диалекты и говоры, в которых они употребляются, их семантическое описание, а также примеров, которые служат для иллюстрации употребления соответствующих терминов в различных диалектах и говорах. Далее на базе проведенного анализа была создана специальная процедура сегментации, которая сканировала текст книги, и далее, с учетом правил оформления абзацев, разбивала анализируемый фрагмент на

составные части. К сожалению, подобный подход не гарантировал точности разбиения. По этой причине результаты работы процедуры сегментации сбрасывались в специальные таблицы, и далее обрабатывались вручную. Точность определения границ выделяемых сегментов определялась лингвистом, при обнаружении ошибок производилась их ручная корректировка. Эта работа была достаточно трудоемкой и потребовала просмотра большого по объему материала.

В создаваемый словарь была включена информация по фонетической транскрипции терминов. В книгах Баязитовой используется предложенная автором облегченная транскрипция. Для возможности использования терминов и примеров в диалектических корпусах была предложена транслитерация терминов на письменные формы татарского языка с использованием символов Международного фонетического алфавита (МФА)[5].

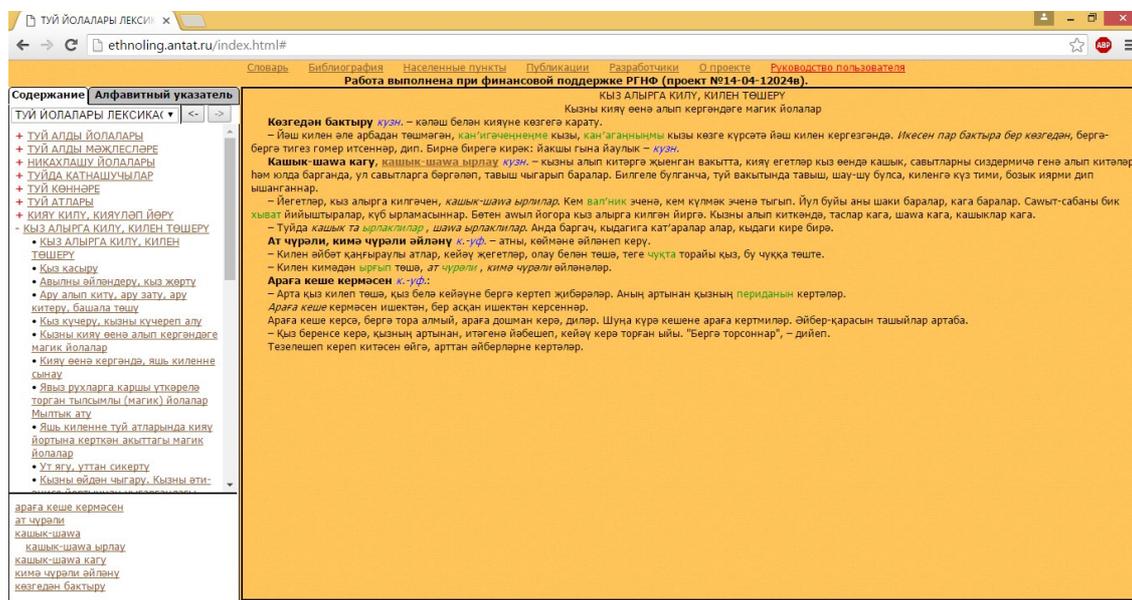
Кроме того, был осуществлен перевод семантического описания части терминов на русский язык. Работа по переводу была вызвана стремлением создать доступ к базе данных более широкому кругу пользователей (в частности и тем, кто плохо владеет татарским языком) для различных сопоставительных исследований обрядовой терминологии. Эта задача оказалась достаточно сложной даже для опытных лингвистов. Описание каждого термина с одной стороны должно быть предельно точным (отражать характерные особенности связанных с термином обрядов, их специфику, национальный колорит, историю, этнографию), с другой стороны – быть достаточно кратким. Эти условия, являясь достаточно противоречивыми, серьезно осложнили работу в этом направлении. За два года работы над проектом удалось обработать только словарь свадебных терминов (около 6000 наименований), отложив подобную работу для описания родильных терминов на третий год.

В создаваемую базу данных была включена обширная коллекция диалектных текстов, собранных Ф.С.Баязитовой в этнолингвистических экспедициях. Включенные в словарь термины и примеры из этой коллекции связаны между собой определенными отношениями, которые позволяют для различных диалектов татарского языка демонстрировать примеры употребления термина в живой речи татар. Была разработана специальная программа согласованной загрузки данных из обработанных текстовых фрагментов. В состав базы данных были включены дополнительные словари по диалектизмам татарского языка, которые играют вспомогательную роль, но встречаются в текстах книг и расшифровываются в виде сносок. Дополнительно в базу данных включена информация по информантам, по населенным пунктам, а также библиография по научным источникам, которые были использованы при создании книг. При анализе списка населенных пунктов добавлена информация по диалектам и говорам, характеризующим данный населенный пункт, эта информация сравнивалась с подобной информацией для населенных пунктов, описанных в

составе атласа татарских говоров. Следует отметить, что из более чем 200 населенных пунктов, где собиралась информация по свадебной лексике, только лишь половина описана в составе атласа. Это объясняется тем, что информация по этнолингвистическим материалам в основном собиралась в отдаленных, малонаселенных селах, в которых в силу их удаленности от "цивилизации", сохранились сведения о первобытном языке и культуре. Однако это обстоятельство создало определенные сложности при определении географических координат соответствующих населенных пунктов. Со времени проведения экспедиций, многие населенные пункты прекратили свое существование, и в ряде случаев было довольно трудно найти по ним необходимую информацию.

Одним из основных отличий электронной реализации словаря в отличие от книжной является возможность выбора различных форм представления информации в зависимости от вида решаемых задач. Как уже упоминалось, словарь построен на базе трех книг, отличающихся тематикой. Пользователь имеет возможность выбирать источник информации и тем самым расширять или сужать объем набора терминов базового словаря. Доступ к словарю осуществляется в двух вариантах: тематическом и алфавитном.

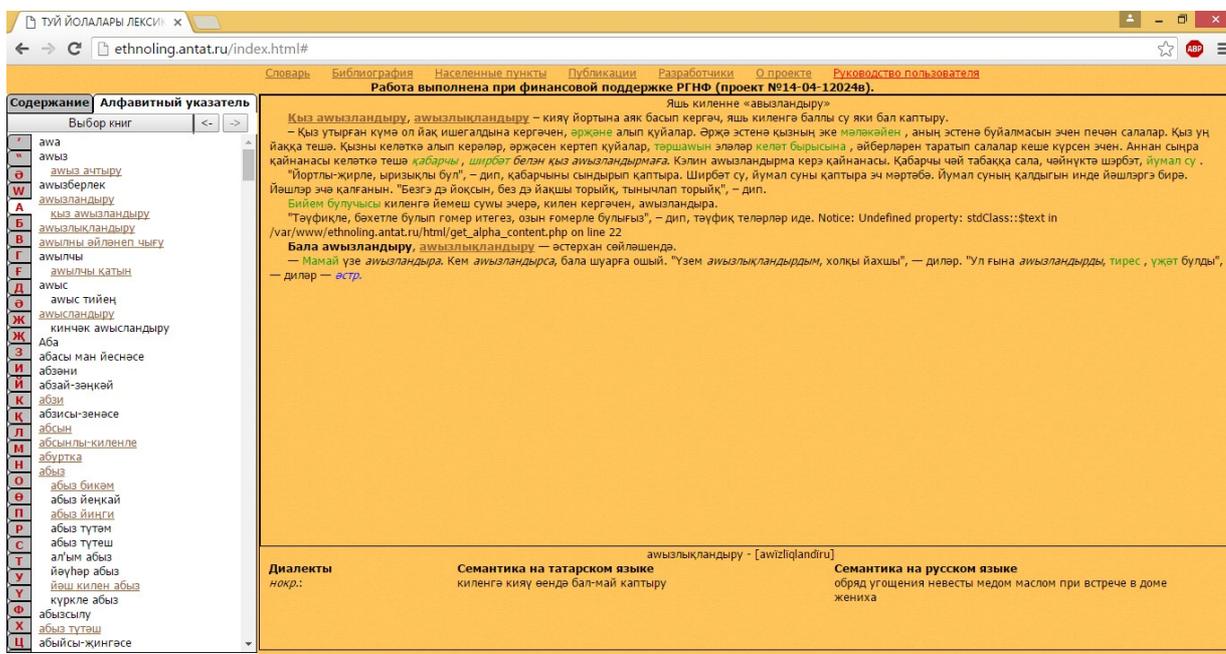
Первый способ предоставляет иерархический доступ к терминам, разбитым по предметным областям, который в целом соответствует способу представления информации изложенной в книжной версии.



В правой части экрана для каждого выбранного термина появляется набор примеров употребления этого термина в диалектной речи с указанием диалектов, которые представляю гиперссылку и выделены синим цветом. При необходимости, пользователь может перейти по ссылке на карты электронного атласа татарских народных говоров[6,7], на которых можно отследить населенные пункты, которые связаны с выбранным диалектом. Более точная локализация (зависимость термина от географии его записи) населенных пунктов

невозможна, поскольку эта информация отсутствует в книжной версии. Зеленым цветом на экране выделены диалектизмы, которые встречаются в сносках.

Второй способ реализует доступ к терминам в алфавитном порядке. В отличие от предыдущего варианта, в этом случае в нижней части экрана появляется дополнительная информация (фонетическая транскрипция термина, семантическое описание на татарском и русском языках).



Описанные выше варианты доступа реализуются через специально построенную систему запросов. Система программно поддерживает стек запросов: пользователь имеет возможность возвращаться на предыдущие шаги, перемещаясь по клавишам-стрелкам в верхнем углу активного окна.

Программа создается как интернет-ресурс. Доступ к словарю можно получить по адресу ethnoling.antat.ru. Материалы словаря могут быть использованы в этнолингвистических исследованиях, при создании диалектологических корпусов для татарского языка, а также при обучении татарскому языку

### Литература

1. Толстой Н.И. Язык и народная культура (очерки по славянской мифологии и этнолингвистике). –М.,1995, 509с.
2. Баязитова Ф.С. (2011) Туй йолалары лексикасы (Пирле сЙйлШ М фольклор текстлары яссылыгында). – Казан: Паравитта, 2011. – 976 б.
- 3.Баязитова Ф.С. (2012)Халык традициялре лексикасы: бишек туге (йола М фольклор текстлары яссылыгында). – Казан: Алма-Лит, – 331 б.
4. Баязитова Ф.С. (2015) Халык традициялре лексикасы: согы туй (дини фольклор М Пирле сЙйлШ текстлары яссылыгында). – Казан: 711 б.
5. У.Ш.Байчур "Звуковой строй татарского языка", изд-во КГУ, 1959г.
- 6 Салимов Ф.И., Рамазанова Д.Б., Пилюгин А.Г., Салимов Р.Ф. (2012) Электронная версия атласа татарских народных говоров// Вестник татарского государственного педагогического университета, Казань, с.205-210
7. Салимов Ф.И. Пилюгин Г.А. Ершов С.А. (2012) Электронный атлас татарских народных говоров как инструмент исследования - Труды татарской школы по компьютерной и когнитивной лингвистике, TEL-2012, ФЭН, АН РТ, Казань, с.48-54