

Создание индексов Общеславянского лингвистического атласа с помощью компьютерной программы *Atlas Project**

Перед составителями Общеславянского лингвистического атласа встает задача не только создания разных типов карт, но и обработки собранных диалектных материалов, которые подвергаются картографической интерпретации [Израелян 2006: 48]. Идея создания и использования компьютерной программы под названием *Atlas Project*, которая призвана помочь в обработке диалектных материалов ОЛА, (составление индексов), возникла еще в начале 2000-х гг., но по причинам различного характера осуществилась лишь в настоящее время.

Первые попытки использования ЭВМ в обработке данных ОЛА были предприняты Словацкой комиссией, в частности, Павлом Жиго, который обратил внимание на преимущество компьютерной обработки материалов Атласа [Жиго 1996].

Сейчас, когда работа над базой данных русских диалектных материалов ОЛА близится к завершению, значение программы, способной преобразовывать «старые» – уже введенные – данные, вводить новые, а также генерировать индексы и тем самым значительно повышать возможности составителей ОЛА, трудно переоценить.

Реализация указанной базы данных ОЛА, а также программного продукта *Atlas Project*, осуществлялись на основе разработок, обеспечивающих возможности совместного доступа участников проекта к базе данных ОЛА, объединения различных информационных потоков проекта (согласованные данные, требующие ревизии; удаленно вводимые данные, резервные копии и т. д.) и создания специальных средств управления данными проекта ОЛА.

Программа *Atlas Project*, позволяющая вводить, хранить, а также редактировать диалектные материалы Общеславянского лингвистического атласа была разработана в 2007 году. Она позволила создать методику работы, которая включает:

- а) единые правила интерпретации, ввода и верификации исходных материалов ОЛА;
- б) форму внутреннего представления и способы хранения данных (включая систему служебных индексов);
- в) общую технологию подготовки различного рода индексов, которые в дальнейшем будут опубликованы в томах ОЛА.

* Статья подготовлена при финансовой поддержке Российского гуманитарного научного фонда (проект № 07-04-00129В)

Технические возможности программы *Atlas Project* позволили создать оптимальную базу данных (*MS Access*) для проекта ОЛА, перенести в новую базу уже существующие данные; работать автономно, то есть, через встроенный в операционную систему *Windows* драйвер без установки *MS Access*, входящей в пакет *Microsoft Office*, в будущем – с минимальными изменениями – перенести проект на более мощную базу данных, например, *ORACLE* (планируется соединение всех славянских диалектных материалов в единую SQL-базу данных *Oracle*).

Ниже приведен интерфейс программы к новой базе данных (см. рис. 1).

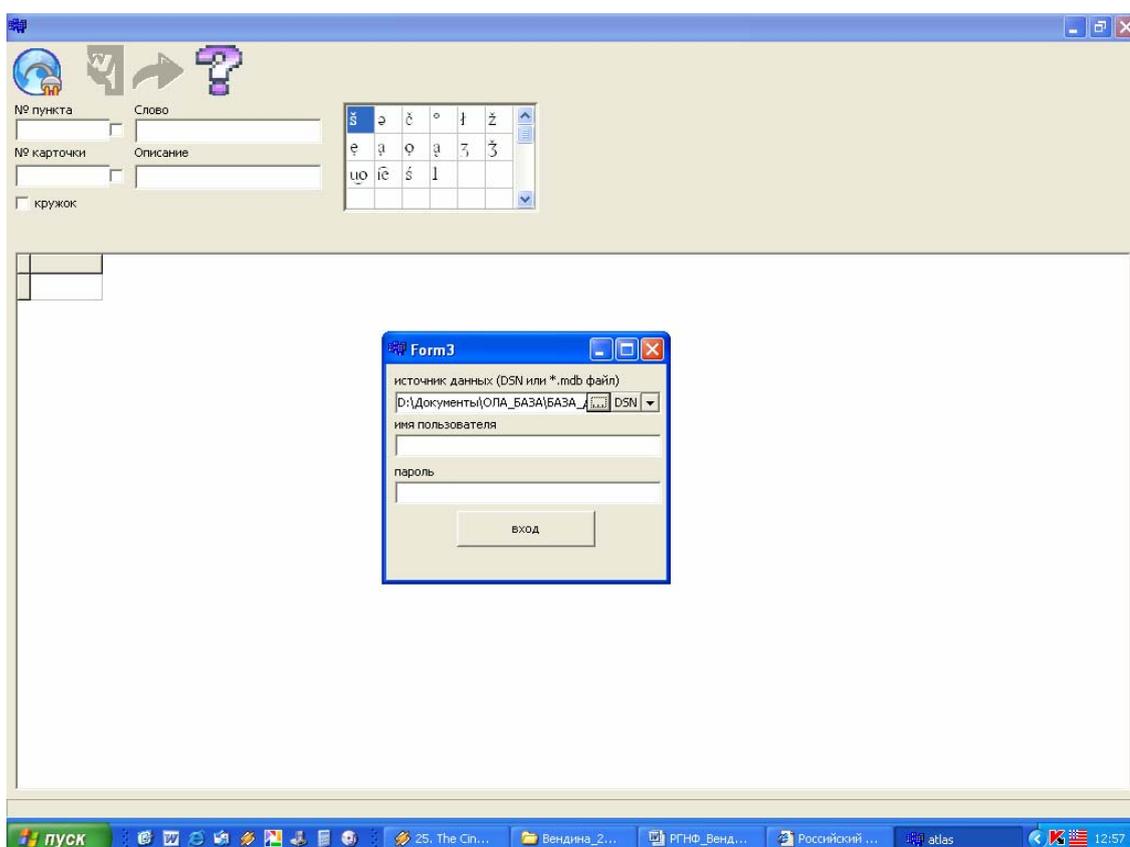


Рис. 1. Интерфейс программы *Atlas Project* (соединение с базой).

Основные характеристики программы *Atlas Project* (см. ниже рис. 2-4):

- Основные операции с базой данных: ввод, исправление (редактирование) и удаление записей;
- Поддержка программой *unicode* кодировки для отображения большого количества спецсимволов;
- Автоматическая проверка уникальности записи, проверка индекса на «снятие» и «необследованность»;

- Перенос данных из одной базы в другую, параллельный ввод данных разными людьми с последующим соединением всех данных в конечную (единую) базу;
- Горячие клавиши на основные операции: F1 - вывод справки, CTRL+I – вставка записи, CTRL+U – обновление записи, CTRL+Space – заполнение поля данными из текущей записи в таблице, CTRL+S – пропуск текущего пункта в режиме быстрого ввода записей, CTRL+D – удаление выделенных записей в таблице, CTRL+A – вызов консоли управления текущей базой;
- Автозамена обычных символов на спецсимволы при вводе записи (см. ниже рисунки 2-4);
- Перенос и конвертация данных ОЛА в документ MS Word, то есть генерация готового индекса из базы данных с автоматическим созданием указанного номера и сигнатуры индекса (например, L 1845, M 30);
- Автоматическое слияние «соседних» пунктов с одинаковым значением (525 z'v'er', 526 z'v'er' = 525-526 z'v'er');
- Изменение названия столбца, его длины, добавление новых автозамен (спецсимволов), а также настройка документа MS Word;

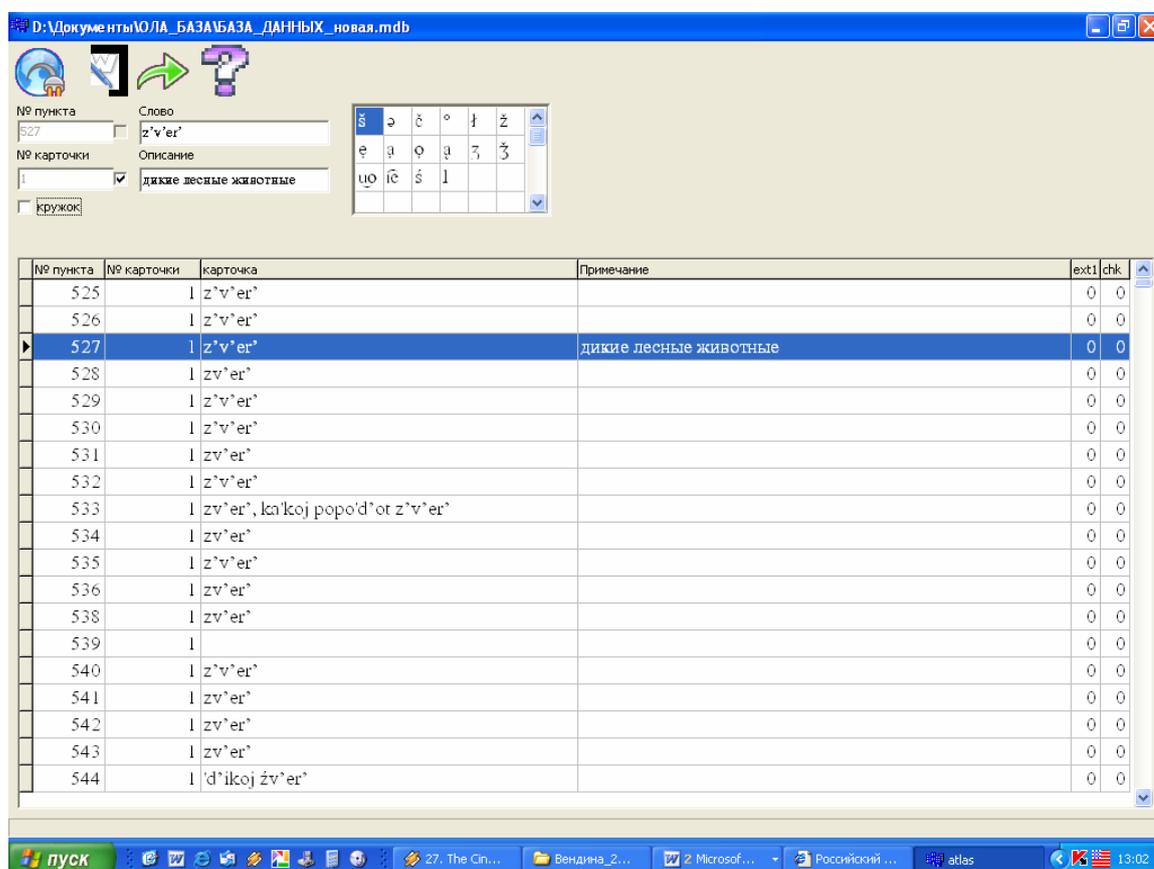


Рис. 2. Интерфейс программы Atlas Project (база данных)

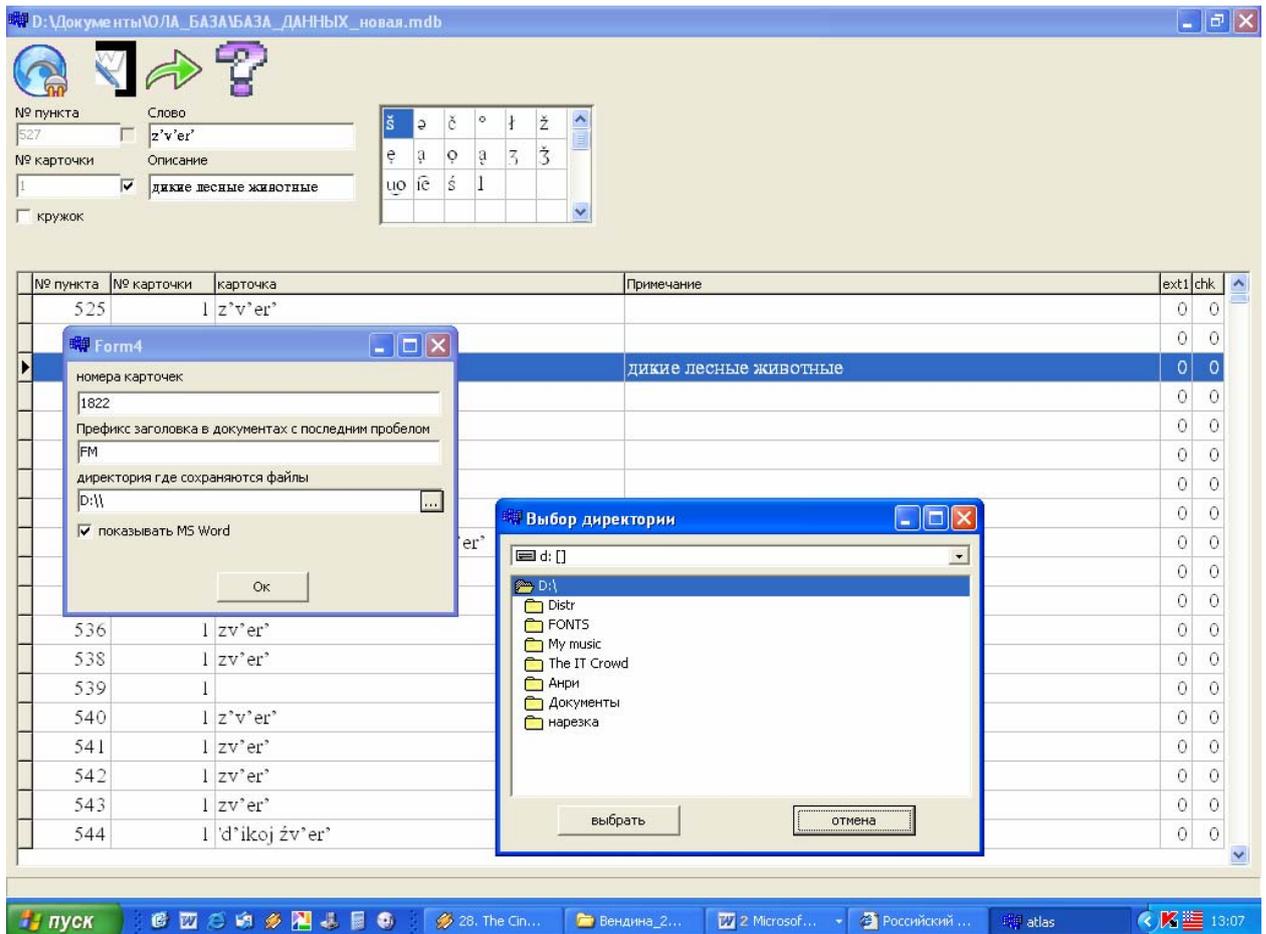


Рис. 3. Интерфейс программы *Atlas Project* (перенос и конвертирование данных ОЛА в документ MS Word, создание готового сводного индекса).

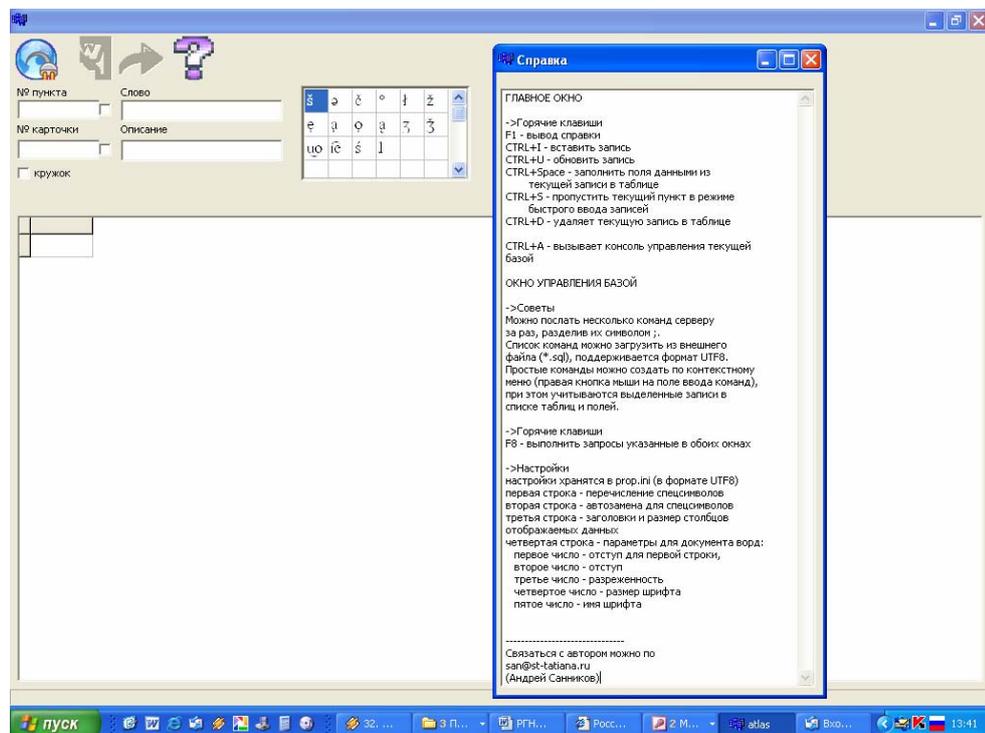


Рис. 4. Интерфейс программы *Atlas Project* (справка)

В ближайшем будущем через Интернет будет организован ограниченный удаленный доступ участников проекта к базе данных ОЛА для координации работы, что позволит максимально быстро просматривать и редактировать диалектные материалы, введенные в базу.

Программа *Atlas Project* может быть использована при создании *единой* базы данных ОЛА (почти 3 млн. карточек), в перспективе доступной для открытого использования внутри научного сообщества. База данных обеспечит возможность в соответствии с современными технологиями наиболее оперативной обработки материалов ОЛА. Проект рассчитан на научную аудиторию всех славянских (и не только) стран, которые занимаются диалектологией, лингвогеографией и сравнительным языкознанием.

ЛИТЕРАТУРА

Вопросник 1965 – Вопросник Общеславянского лингвистического атласа. М., 1965.

Жиго 1996 – П. Жиго. Опыт использования ЭВМ в обработке материалов ОЛА // ОЛА. Материалы и исследования. М., 1996.

Израелян 2006 – А.Г. Израелян. Использование реляционных баз данных в обработке материалов ОЛА // ОЛА. Материалы и исследования 2003-2005. М., 2006.