

Министерство образования и науки Российской Федерации
Псковский государственный университет

Филологический факультет

**РУССКИЙ ЯЗЫК И ЛИТЕРАТУРА
В ПОЛИКУЛЬТУРНОМ
КОММУНИКАТИВНОМ
ПРОСТРАНСТВЕ**

Материалы Международной научной конференции
(26–28 апреля 2012 года, Псков)
Часть I



Псков
Псковский государственный университет
2012

УДК 801; 802
ББК 81.2 Рус
P894

Рекомендовано к изданию кафедрами
русского языка и литературы
Псковского государственного университета

P894 **Русский язык и литература в поликультурном коммуникативном пространстве** : Материалы Международной научной конференции (26–28 апреля 2012 года, Псков): в 2-х частях. Часть 1 / Отв. ред. Н. В. Большакова. Редакционная коллегия: В. К. Андреев, Н. Л. Вершинина, Л. Я. Костючук, И. В. Мотеюнайте, Т. Г. Никитина, Н. И. Яковлева. — Псков : Псковский государственный университет, 2012. — 344 с.
ISBN 978-5-91116-166-8

В сборнике представлены материалы докладов научной конференции, посвящённой фундаментальным и научно-практическим проблемам состояния и роли русского языка и литературы в современном поликультурном пространстве, истории и перспективам развития разных форм коммуникации.

В исследованиях учёных России и зарубежных стран (Белоруссии, Индонезии, Латвии, Литвы, Польши, Украины, Чехии, Эстонии) представлен опыт разработки традиционно сложившихся и новейших научных направлений современной филологии.

На страницах сборника получил освещение комплекс вопросов, связанных с филологической регионалистикой, гуманитарными проблемами региональных особенностей историко-культурных и языковых процессов.

Книга адресована специалистам разных областей гуманитарного знания, преподавателям вузов и учителям школ, а также аспирантам, магистрантам, студентам высших учебных заведений.

УДК 801; 82
ББК 81.2 Рус

ISBN 978-5-91116-166-8

© Коллектив авторов Псковского
государственного университета, 2012
© Псковский государственный университет, 2012

Литература

1. Букринская И. А., Кармакова О. Е. Структура и жанровые особенности диалектного текста // Материалы и исследования по русской диалектологии. Кн. 3 (9). М. : Наука, 2008. С. 414–427.
2. Гольдин В. Е. Теоретические проблемы коммуникативной диалектологии: Дисс. в виде научн. докл. ... докт. филол. наук. Саратов, 1997.
3. Касаткин Л. Л. Исследование говоров русских старообрядцев в Институте русского языка им. В. В. Виноградова РАН // Русские старообрядцы: язык, культура, история: Сб. статей к XIV Междунар. съезду славистов / Отв. ред. Л. Л. Касаткин. М. : Языки славянских культур, 2008. С. 573–608.
4. Мораховская О. Н. Особенности функционально-стилевой дифференциации современной диалектной лексики // Русская диалектология: Учебник для студ. филол. фак. высш. учеб. заведений / Под ред. Л. Л. Касаткина. М. : Издательский центр «Академия», 2005. С. 221–224.
5. Радиостанция «Эхо Москвы» 14.02.2012. Передача «Своими глазами», тема «Как соотечественнику вернуться в Россию?» // URL: <http://echo.msk.ru/programs/svoi-glaza/858465-echo/>, URL: <http://echo.msk.ru/programs/svoi-glaza/858465-echo/#video>.
6. Ровнова О. Г. «Полиглоты поневоле»: языковая ситуация в старообрядческих общинах Южной Америки // Staroodrządowcy za granicą // Pod redakcją M. Głuszkowskiego i St. Grzybowskiego. Toruń, Wydawnictwo naukowe Uniwersytetu Mikołaja Kopernika, 2010. S. 137–157.

И. С. Кукушкина

Конвертация «Псковского областного словаря» в формат электронного словаря на базе DWS LINGVO CONTENT

На протяжении многих лет Межкафедральным словарным кабинетом (МСК) имени проф. Б. А. Ларина Санкт-Петербургского государственного университета и кафедрой русского языка Псковского государственного университета ведётся работа по составлению «Псковского областного словаря с историческими данными». За это время вышло 23 выпуска (1967–2012) словаря, и картотека,

содержащая около двух миллионов карточек с контекстами, непрерывно пополняется новыми материалами, которые собираются ежегодными диалектологическими экспедициями. ПОС — это словарь полного типа, включающий как диалектную, так и общерусскую лексику и фразеологию, а также обширный исторический материал.

В 2011 г. МСК имени проф. Б. А. Ларина совместно с компанией АВВУУ (известные многим словари АВВУУ Lingvo) начал новый проект по переводу «Псковского областного словаря с историческими данными» (а также его Картотеки) в электронный вид. Цель данного проекта — дальнейшее сохранение материалов словаря, обеспечение лексикографов возможностью использовать профессиональное программное обеспечение для составления словарных статей, а также предоставление доступа читателям и исследователям к электронной версии словаря.

В настоящее время для создания и редактирования словарей в электронном формате применяются различные программные средства, так называемые Dictionary writing systems (DWS). К ним можно отнести и обычные текстовые редакторы вроде MS Word или редакторы баз данных, такие как MS Access, которые дают возможность готовить тексты словарей в электронном виде.

Несколько лет назад в Межкафедральном словарном кабинете был задуман и частично реализован проект по созданию электронной базы данных псковских говоров, в котором как раз использовалась система управления базами данных MS Access. К сожалению, её возможности довольно ограничены, и практика показала, что в ней трудно работать с большими массивами лексикографических данных. Но, строго говоря, настоящими «системами написания словарей» (DWS) являются программы, специально разработанные для создания конкретного словаря (обычно их используют издательства словарей и академические учреждения, научные группы), или профессиональное, достаточно универсальное программное обеспечение (ПО) для лексикографов, которое поставляется компаниями-разработчиками ПО в области лексикографии.

В данном случае требовалось создать не только готовый электронный словарь из существующих выпусков, но и удобную рабочую среду, в которой составители словаря могут работать,

продолжая подготовку следующих выпусков. В 2011 г. у МСК имени проф. Б. А. Ларина появилась возможность поработать с профессиональным программным обеспечением для лексикографов, разработанным компанией АВВУУ и предназначенным для подготовки словарей разного типа. Система LingvoContent (LC) позволяет, в частности, импортировать материалы словаря в структурированную базу данных на SQL-сервере с последующей конвертацией в формат LingvoXML. После завершения конвертации словарь можно опубликовать в электронном виде на интернет-сайте или превратить в электронное приложение для компьютеров и мобильных устройств.

Для того чтобы приспособить универсальную среду LingvoContent для диалектного Псковского словаря, необходимо было прежде всего выделить конкретные зоны словарной статьи ПОС и соотнести их с элементами структуры LC. Структура словарной статьи Псковского словаря очень своеобразна и сложна; по крайней мере, по сравнению со стандартными словарями, которые создавались ранее с помощью этой программной среды. Одной из важнейших задач было перенести всю глубину семантической структуры слова в базу данных, не потеряв значимой лексикографической информации, детально разработанной в словаре.

Для специалистов компании АВВУУ была подготовлена упрощённая иерархическая схема словарной статьи ПОС, опираясь на которую они могли ориентироваться в её уровнях и найти им соответствие в LC. На согласование Методики конвертации словаря ушло полгода. В полной мере вся сложность задачи — адаптировать LingvoContent для диалектного словаря с историческими данными и с такой сложной структурой словарной статьи — стала ясна только в процессе работы.

Основой Методики стали две инструкции — вводная статья к словарю, составленная на основе «Инструкции Псковского областного словаря» Б. А. Ларина и опубликованная в первом выпуске ПОС в 1967 г. (ПОС 1: 6–16), и обновлённая Инструкция (ПОС 15: 5–51), в которую были внесены значительные изменения и дополнения, накопившиеся за прошедшие 36 лет составления словаря. За годы работы над Псковским областным словарём многое изменилось в подаче материала. В данном проекте требовалось учесть различия

нных и поздних выпусков и привести их к общей форме. Так, например, было унифицировано лексикографическое представление глагольных форм: в первых выпусках они выводились как отдельная часть словарной статьи и повторяли структуру основной части, сейчас же записи диалектной речи, фиксирующие такие формы, приводятся вместе с остальным иллюстративным материалом при соответствующих значениях.

Кроме того, параллельно разрабатывалась Методика конвертации Картоотеки ПОС. Она в основном опирается на Методику конвертации Словаря, но некоторые отличия информации, содержащейся в картотечной записи, повлияли на структуру словарной статьи ПОС в LC и потребовали более пристального внимания к процессу конвертации, т. к. материал карточек практически не обработан лексикографически. Кроме того, в Картоотеке записи паспортизированы подробнее, чем в словаре, и для этой информации также необходимо было предусмотреть место в базе данных.

В результате совместными усилиями нам удалось создать приемлемый вариант переноса лексикографической информации из печатной версии словаря в формат электронной базы данных. В конечном варианте словарная статья в LC содержит заголовок (heading) и «тело» словарной статьи (content). В зону заголовка выносятся заголовочные слова статьи ПОС, при этом каждое оформляется отдельным тегом. Благодаря этому в случае, если заголовочных слов несколько (*вашище*, а, с. и *вашица*, и, ж.), в словнике электронного словаря отображаются все заголовочные слова статьи, а не только первое из них. «Тело» словарной статьи вмещает всю остальную информацию и делится на несколько уровней (грамматические варианты, значения, подзначения), каждый из которых может содержать от одного до нескольких структурных элементов. Это позволяет получить полноценную словарную статью электронного словаря, соответствующую исходной «бумажной» версии.

После утверждения Методики можно было приступить к конвертации уже опубликованных выпусков «Псковского областного словаря с историческими данными» в электронный формат. Для этого в МСК имени проф. Б. А. Ларина текст первых выпусков ПОС был отсканирован, тщательно вычитан и специально отформатирован для последующей автоматической обработки. После завершения обработки материалов они были

переданы в АВВУУ, и специалисты компании провели теговую разметку текстов, а затем начался процесс импорта размеченного текста в среду LingvoContent, проверка и исправление ошибок обработки. Готовую базу данных можно экспортировать в формат словаря АВВУУ Lingvo и подключить его к оболочке Lingvo, получив полноценный электронный словарь с удобным интерфейсом и возможностью полнотекстового поиска.

Однако проект рассчитан не только на превращение существующего словаря в электронное приложение, но и на дальнейшее составление. Для этого предназначено рабочее место лексикографа в среде LC.

Главное окно этого приложения состоит из четырёх частей: словник и структура выбранной статьи в левой половине *окна рабочей среды*; окно просмотра конечного результата и теговый редактор в правой. Словник позволяет легко перемещаться между словарными статьями; кроме того, в этой части отображается информация о статусе словарной статьи и пользователях, работавших с ней. Основная работа лексикографа или редактора словаря ведётся в окне тегового редактора: здесь вводится текст и производятся все изменения словарной статьи. Информация распределяется по определённым зонам (например, зона заголовка, толкования, примеров, устойчивых сочетаний и др.), размечается в соответствии с уровнями словарной статьи и типом лексикографической информации. Осуществляя такую разметку, лексикограф расставляет теги, для чего используются кнопки специальных панелей инструментов. Это панель зон, которая позволяет добавлять зоны на тот или иной уровень статьи, и панель разметки, с помощью которой можно структурировать информацию внутри зоны (например, выделять пометы, иллюстративный материал и т. п.).

Такая форма представления данных удобна для написания словарей. В то же время для исследователей и лексикографов открываются широкие возможности по поиску и отбору уже введённого в базу материала. Разнообразные инструменты поиска по тексту и фильтрации словарных статей по различным параметрам дают возможность создавать новые словари на основе материала уже существующих, работать с несколькими словарями из одного проекта.

Используя фильтры, мы можем делать выборку лексики, отбирать информацию по грамматическим, стилистическим,

географическим параметрам и многим другим. Например, можно выбрать словарные статьи, описывающие ономастическую лексику: для этого используется поиск по содержимому поля *дефиниция*. Выбрав все толкования, содержащие слова *кличка* или *прозвище* (они входят в типовые формулировки «кличка животного / кошки / коровы...» или «прозвище человека / женщины / мужчины...» в ПОС), мы можем создать специальную выборку зоонимов или антропонимов псковской лексики.

Отобранные таким образом словарные статьи можно использовать для создания новых словарей. Так, отфильтровав весь корпус словаря по помете «тильда» (~), которая в ПОС маркирует топонимическую лексику, получим все словарные статьи, содержащие названия местных мелких географических объектов (лесов, полей, лугов, речек, болот, озер, островов и т. п.). Например, в рабочей электронной версии второго выпуска ПОС получим 54 вхождения данного типа. Далее результаты поиска можно сохранить как самостоятельный словарь в текущем проекте. Конечно, такой материал потребует проверки и хотя бы минимального редактирования словаря, но эту работу невозможно сравнить с тем, сколько времени и сил потребовалось бы нам при работе с бумажной версией словаря.

Сокращения

LC — LingvoContent; Лингво-Контент.

Литература

1. Инструкция «Псковского областного словаря с историческими данными» (2-я редакция) // Псковский областной словарь с историческими данными. Вып. 15. СПб. : СПбГУ, 2004. С. 5–51.
2. Состав словаря // Псковский областной словарь с историческими данными. Вып. 1. Л. : ЛГУ, 1967. С. 6–16.