

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФИЛОЛОГИЧЕСКИЙ ФАКУЛЬТЕТ ТГУ
СОВЕТ МОЛОДЫХ УЧЕНЫХ

**ТРАДИЦИИ И ИННОВАЦИИ
В ФИЛОЛОГИИ XXI ВЕКА:
ВЗГЛЯД МОЛОДЫХ УЧЕНЫХ**

**Материалы Всероссийской молодежной конференции
23–25 августа 2012 г.**



ИЗДАТЕЛЬСТВО ТОМСКОГО УНИВЕРСИТЕТА
2012

9. Фарыно Е. Введение в литературоведение. – СПб., 2004. – С. 364–378.
 10. Топоров В.Н. Петербург и «Петербургский текст русской литературы» // Топоров В.Н. Миф. Ритуал. Символ. Образ: Исследования в области мифопоэтического: Избранное. – М., 1995. – С. 259 – 367.
 11. Топоров В.Н. Пространство и текст // Текст: семантика и структура. Под ред. Т.В. Цивьян. – М., 1983.
 12. Есин А.Б. Время и пространство. // Введение в литературоведение Литературное произведение: общие понятия и термины / под ред. Л.В. Чернец. – М., 2000. – С. 47–62.

Примечания

¹ В нашем случае идея Лихачева несколько видоизменилась – реальный и концептуальный уровни пространства не разнесены по отдельным произведениям в зависимости от жанра, а наоборот, находятся в пределах одного произведения.

² Булгакова А.А. Топика в литературном процессе. – Гродно, 2008.

³ Данилова И.Ф. Страшная месть: из комментария к повести А. Ремизова «Крестовые сестры». // Алексей Ремизов. Исследования и материалы / под ред. А.М. Грачевой и А.д'Амелия. – СПб., 2003. – С. 114.

⁴ Павлов А.В. Многомерность провинциализма // Филологический дискурс. Вестник филологического факультета Тюменского государственного университета. – 2004. – Вып.4. – С. 51 – 60.

⁵ Данилевский А.А. О дореволюционных «романах» А.М. Ремизова // Ремизов А.М. Избранное. – Л., 1991.

⁶ Возникает явная и уже достаточно изученная аллюзия на «комнаты без форточек» в «Преступлении и наказании» Ф.М. Достоевского.

ТОМСКИЙ ДИАЛЕКТНЫЙ КОРПУС КАК НОВАЯ ВЕТВЬ ИЗУЧЕНИЯ ОСОБЕННОСТЕЙ ГОВОРОВ СРЕДНЕГО ПРИОБЬЯ (НА ПРИМЕРЕ МОРФОЛОГИЧЕСКОЙ РАЗМЕТКИ)

Н.А. Живаго, ТГУ, студент
 Научный руководитель – Е.А. Юрина

В настоящее время междисциплинарный подход в науке является одним из наиболее востребованных. Так, взаимодействие дисциплин корпусная лингвистика и диалектология несомненно подразумевает большие перспективы в развитии обеих наук.

Рассмотрим понятие корпусная лингвистика. В.П. Захаров определяет корпусную лингвистику как «раздел компьютерной лингвистики, занимающийся разработкой общих принципов построения и использования лингвистических корпусов с использованием компьютерных технологий»¹.

Роль корпусной лингвистики в диалектологии неоспорима. Прежние способы архивирования диалектных материалов в виде рукописных источников текстов (тетради, карточные архивы) устаревают и являются менее эффективными. Формирование диалектных корпусов одновременно разрешает многие задачи, возникающие у исследователей. Это связано не только с вопросом о хранении массива текстов, существующих в единичном экземпляре, но и со сложностью лингвистического анализа текстового материала.

С 2010 г. формируется Томский диалектный корпус, организованный на базе текстов говоров Среднего Приобья. Архив собран студентами историко-филологического (позже филологического) факультета Томского государственного университета под руководством лидеров Томской диалектной школы О.Н. Блиновой, В.В. Палагиной и др. методом интервьюирования информантов-диалектоносителей говоров Среднего Приобья. Также архив содержит картотеки диалектных словарей более чем на 1 000 000 единиц².

Для создания корпуса исследователи поставили несколько крупных задач:

1. Перевод рукописных текстов в электронный формат.
2. Первичная обработка диалектных текстов. Подразумевает за собой приведение всех собранных материалов к единому графическому стандарту.

3. Разметка текстов: метаразметка, морфологическая разметка. Осуществляется при помощи компьютерной платформы Стилеанализатор V 1.60³.

Одним из признаков корпуса является разметка. «Разметка (tagging) заключается в приписывании текстам и их компонентам специальных меток (tag): внешних, экстралингвистических (сведения об авторе и сведения о тексте и др. Это кодирование информации имеет название «метаразметка»), структурных (глава, абзац, предложение, словоформа) и собственно лингвистических, описывающих лексические, грамматические и прочие характеристики элементов текста»⁴.

Морфологическая разметка входит в разряд лингвистических типов разметки и подразумевает за собой частеречное аннотирование и признаки грамматических категорий. Морфологический анализ в дальнейшем является основой для семантического и синтаксического анализа. Рассмотрим модель реализации морфологической разметки в ТДК.

За каждым корпусом текстов, располагающим морфологическим анализом, закреплены морфологические стандарты, которые подробно описывают способ аннотирования текста, параметры анализа лексических единиц, представляют все группы помет, приписываемых слову в тексте.

Морфологический стандарт Томского диалектного корпуса представляет собой характеристику групп грамматических помет лексических элементов диалектного текста и несет исчерпывающую информацию о морфологических формах слов. При морфологическом аннотировании диалектных единиц эксперты опираются в основном на морфологическую модель анализа диалектных единиц, представленную О.Н. Киселевой⁵, в характеристике общерусских слов на морфологическую модель, представленную в «Грамматическом словаре русского языка» А.А. Зализняком⁶.

Морфологическая характеристика представлена следующими группами помет:

1. Лексема, которой принадлежит словоформа.
2. Постоянные признаки слова.
3. Непостоянные признаки слова.
4. Информация о нестандартности слова.

Рассмотрим пример размеченного текста.

Родила `с {родиться|*родилась|глагол, сов, действ, возвр, неперех|изъяв, прош, 3л, ед, ж}

я {я|мест, личн|1л, ед, им}

в {в|предл}

городе {сущ, нариц, неодуш, конкр, м.р|ед, п}

Барнауле {сущ, собств, собств, неодуш, конкр, м|ед, п}

Морфологический стандарт Томского диалектного корпуса имеет некоторые различия со стандартом Саратовского диалектного корпуса и диалектного подкорпуса Национально го корпуса русского языка. Проанализируем отражение слова помидор в форме ед. числа, в винительном падеже в данных корпусах. В НКРЯ данное слово будет выглядеть следующим образом: **помидору** {помидора*=S,жен,неод=ед,вин=dial morph_gend}, в СарДК **помидору** {помидора(помидор)=S,жен,неод=ед,вин=*}, в ТДК **помидору** {помидор|*помидора|сущ, жен, неодуш|ед, в}. Во многом различия имеют внешний характер. Так, в СарДК аннотация написана кириллицей, в НКРЯ латиницей, знаки границы между постоянными и непостоянными признаками обозначены символом = в обоих корпусах. В ТДК границы обозначены символом прямой слеш, по сравнению с другими корпусами имеются различия в оформлении грамматических признаков. Морфологический стандарт ТДК содержит более детальную разработку аннотирования, что позволит в дальнейшем подвергать анализу более точные результаты поиска.

Осуществление морфологической разметки в ТДК является одной из самых необходимых операций, позволяющей изучать динамику процессов изменения грамматических особенностей языка, проводить анализ грамматических характеристик.

Примечания

¹ Захаров В.П. Корпусная лингвистика: учеб.-метод. пособие. – СПб., 2005.– С. 3.

² Юрина Е.А. Актуальные задачи компьютеризации диалектного архива говоров Среднего Приобья и пути их решения // Компьютерная лингвистика: научное направление и учебная дисциплина [Текст]: сб. науч. ст. Вып. 1. / М-во образования РБ, ГГУ им. Ф. Скорины. – Гомель: ГГУ им. Ф. Скорины, 2010. – С. 171.

³ Кубарев А.И. Развитие функциональности систем «СтилеАнализатор», направление и содержание специализированных корпусов и алгоритмов классификации на основе деревьев решений: диплом. работа. Томск: Том. гос. ун-т, 2011. – 27 с.

⁴ Захаров В.П. Указ соч. – С. 8.

⁵ Русские говоры Среднего Приобья / О.И. Блинова, О.И. Гордеева, М.Н. Янценецкая и др.; ред. В.В. Палагина. – Томск, Изд-во Том. ун-та, 1989. Ч. 2. – 324 с.

⁶ Зализняк А.А. Грамматический словарь русского языка. – М.: Русский язык, 1977. – 879 с.