

нях увидеть этнокультурные или идиокультурные смыслы, аккумулярованные в тексте / гипертексте.

Полученные результаты мы представляем в подготовленном сборнике научных трудов "Сопоставительная и кросскультурная лингвофольклористика".

---

<sup>1</sup> Степанова И. А. Жанровое своеобразие лексики русской исторической песни: Автореф. дис... канд. филол. наук. Орел, 2004.

<sup>2</sup> Разинькова О. Н. Общая характеристика словаря русских баллад // Лингвофольклористика. Курск, 2004. Вып. 8.

<sup>3</sup> Завалишина К. Г. Человек телесный в языке фольклора русского, английского и немецкого этноса // Сопоставительная лингвофольклористика: Сб. ст. Курск, 2003. Вып. 1.

<sup>4</sup> Онежские былины: В 3 т. / Записаны А. Ф. Гильфердингом летом 1871 г. Изд. 2-е. СПб., 1894-1900. Т. 1, № 51. Ст. 200.

<sup>5</sup> Там же.

<sup>6</sup> Там же. Т. 1, № 8. Ст. 41.

**В. А. Баранов, С. А. Никифорова, Л. Ф. Клима,  
О. В. Зуга, Е. В. Рябова**

**ЛИНГВИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ  
КАК ЭТАП ИССЛЕДОВАНИЯ ДРЕВНЕГО ТЕКСТА  
С ПРИМЕНЕНИЕМ КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ**

Современная компьютерная лингвистика предоставляет ученому широкие возможности в изучении текстов: компьютерные технологии позволяют увидеть конкретное явление на фоне множества аналогичных фактов языка, представленных во множестве текстов и контекстов как в синхронии, так и в диахронии. Если на современном материале автоматизированные исследования проводятся достаточно давно (с 70-х гг. XX в.), то обработка ранних русских памятников письменности развивается лишь в последнее десятилетие. Связано это, безусловно, и с особым типом представления древнего текста (это рукопись вплоть до XVI-XVII вв.), и с минимальной стандартизованностью языковой системы древнейшего времени, а значит, с высокой степенью вариативности единиц на всех языковых уровнях — от графики до синтаксиса. Именно этими причинами объясняется то, что до сегодняшнего дня не создано стройной и гибкой систе-

мы обработки лингвистического материала древнерусского и среднерусского периодов истории русского языка с применением компьютерных технологий.

В лаборатории по автоматизации филологических исследований Удмуртского государственного университета (рук. д-р филол. наук В. А. Баранов) творческим коллективом лингвистов и программистов разрабатывается информационно-поисковая система (далее ИПС) "Манускрипт", предназначенная для ввода текста любого исторического периода и любой графической сложности, для автоматизированной обработки данных в соответствии как с уже определенными, так и с потенциальными параметрами лингвистического запроса. Эта система призвана обеспечить не только максимально точное представление текста в Интернете или подготовку его качественного бумажного издания, но и дать возможность пользователю Интернета выполнять собственные научные работы с помощью запросных форм в сети, вносить собственные материалы и пополнять и корректировать базу данных, издателю же традиционных печатных публикаций – печатать не только текст рукописи, но и создавать указатели самых разных типов (от элементарных прямых, обратных и количественных формуказателей до многофункциональных словарей различного типа).

Ввод в ИПС множества разножанровых и разноструктурных текстов древнерусского и среднерусского периода позволяет исследователю подготовить на основе большого числа форм единые словари, увидеть общую динамику и направление развития языковой системы, что часто затруднительно или вообще невозможно при проведении выборки вручную.

Сложность создания такой системы состоит в необходимости координировать работу ученых разных областей знания: в первую очередь лингвистов и программистов. Специфика работы по созданию компьютерных программ требует максимально точного представления программиста о конечном результате работы, требования же лингвистов при исследовании памятника зачастую возрастают по мере изучения текста. Именно поэтому перед созданием программы требуется значительная собственно лингвистическая обработка данных текстов для разработки максимально универсальной, гибкой, комплексной базы данных с прогнозированием потенциальных запросов.

Группа лингвистов ЛАФИ УдГУ, участвующая в проекте "Университеты России – гуманитарные науки", провела сегодня такую работу<sup>1</sup>.

На материале древнейших русских рукописей (в работу сегодня введены следующие тексты: новгородские служебные минеи XI–XIV вв., Евангелия XI–XIV вв., Житие Александра Невского в списках XV–XVI вв., летопись по Лаврентьевскому списку 1377 г. и нек. др.) разработана модель представления лингвистической информации в ИПС “Манускрипт”, которая включает лексические и грамматические параметры языковых единиц (не только слов, но и более крупных конструкторов – аналитической грамматической формы, устойчивого словосочетания, синтагмы и т.д.) с учетом как синхронического, так и диахронического уровней исследования языка (модель применима при анализе древнерусского – в модели помечено \* – и среднерусского текстов, записи диалектного текста, современного текста – в модели помечено +).

1. В основу морфологической модели положен частеречный принцип как наиболее универсальный и в достаточной степени бесспорный. Выделены следующие части речи: имя существительное, субстантиват, прилагательное, числительное (счетное слово для древности), местоимение, именная синкрета, наречие, глагол, причастие, деепричастие, предикатив, предлог, союз, частица.

Конечно, заявленная бесспорность частеречной классификации относительна, о чем говорит и выделение в модели ИПС особых частей речи, не зафиксированных сегодня в традиционных классификациях, таких как субстантиват и именная синкрета (при иной точке зрения пользователя ИПС существует возможность выбора, корректировки и сопоставления грамматических помет).

Именная синкрета понимается авторами ИПС как “первичное нерасчлененное имя [которое. – Авт.] следует признать а) семантической синкретой, функционирующей в синтагме в форме, достаточно определенной точки зрения отношений с другими членами синтагмы, б) синкретичным словом, имеющим в зависимости от роли в синтагме различные функциональные формы, которые следует квалифицировать как грамматические, категориальные омонимы”<sup>2</sup>. Это часть речи древнейшего периода развития русского языка. Введение данной грамматической категории позволяет в ряде случаев корректно определить частеречную характеристику слов, имеющих синкретичное грамматическое значение. Субстантиватом же принято считать существительное адъективного склонения: такие имена чрезвычайно широко распространены как в древности, так и в современном русском языке. Термин “субстантиват” позволяет избежать, с одной стороны,

полного, без учета грамматической специфики, включения таких имен в сферу имени существительного, с другой – смешения их с именем прилагательным (особенно в древнейших текстах, где параллельно существуют субстантиваты и синонимичные им словосочетания с омонимичным прилагательным типа гръшьньи<sup>3</sup> – гръшьньи мужь (ср. существительное гръшникъ)).

Приведем пример подобной грамматической интерпретации форм субстантивата и именной синкреты в текстах, введенных в ИПС “Манускрипт”. В старославянских и древнерусских памятниках отмечается ряд грамматических форм, ярко проявляющих синкретичность семантики как лексической, так и грамматической – это притяжательные имена (типа божи, философь), характеризующие имена в звательной форме (типа прѣмУдре, славне, достохвальне), некоторые суффиксальные отыменные образования (типа блажьнь, безоумьнь). Эти формы могут квалифицироваться как существительные или как прилагательные. ИПС позволяет определять формы по-разному, в соответствии с научными взглядами исследователя: так, в контексте ПМ<sup>4</sup> – льстию въсхотѣ оукрасити мчтель твоея крѣпости / тврьдню и жесточение прѣблажене исидоре / на свое безаконие (64) форма прѣблажене может быть оценена как звательная форма либо прилагательного, либо существительного, либо субстантивата, либо именной синкреты в зависимости от теоретических предположений.

На наш взгляд, “в пределах именного сочетания вокативы [например. – Авт.] сохраняют свой семантико-грамматический синкретизм, что наиболее отчетливо проявляется, во-первых, при их синтаксической и семантико-понятийной самостоятельности в контекстах с независимым использованием вокативов в синтагмах и законченных текстах [высказываниях. – Авт.], во-вторых, при соединении их сочинительным союзом “и” с существительным в пределах одной синтагмы<sup>5</sup>. Таким образом, формы типа блажене, славне и т.д., долгое время считавшиеся или прилагательными в звательной форме, или существительными, могут рассматриваться как именная синкрета – т.е. форма синкретичная и по семантике, и по грамматическим свойствам. Число таких форм в ранних древнерусских текстах велико. Например: ПМ – радоуисА бракоуеискоуs’на (54); и на поуть исправи воля бжия оума моЕго (154);

Приведем примеры субстантиватов в текстах XII–XV вв.:

ПЕ – и видѦи мѦ видить пославышаго мѦ (25 об.); азъ О(т) вьшьникъ Есмь вы О(т) всего мира Есте ((16 об.); вѣроуяи въ нь не боудеть осоужень (7);

ЖАНХV — соуди ги *обидАщи(м)* мА и возбрани *боржщмса* со мною (157 об.), ЖАНХVI — и всь(д) на конь иде противу *ратны(х)* (907).

2. Сложной проблемой при компьютерной обработке данных текста, разделенного на слова (в широком понимании термина), становится квалификация и описание компонентов членных составных лексических единиц, составных лексико-грамматических номинативных единиц, аналитических грамматических форм, фразеологизмов, так как каждая из составляющих таких сочетаний понимается с точки зрения машинного подхода как относительно самостоятельный элемент текста, с лингвистической же точки зрения эти слова не самостоятельны и должны функционально и в смысловом отношении оцениваться в структуре нечленного сочетания. В ИПС введена специальная система связей между такими компонентами единого наименования, например:

*Составные лексико-грамматические номинативные единицы*

*Номинативная составная единица*

*ОНИМЫ*

*Антропонимы*

Личное имя                      Компонент личного имени

*Топонимы*                        Компонент топонима = географическое наименование

Гидроним                        Компонент гидронима

ПЕ — и прьдаша и *понтъскуоумоу пилатоу игемоноу* (179 об., 193); *къ низъхоженоу горь елеоньсть* и посла два оученикъ (120); гла Емоу *симонь петрь* ги (176); *келевы наречены фады* (38 об.);

ЖАНХV — до *горь араратьскы(х)* (160), ЖАНХVI — именем *пелтоуси* а въ кршеніи *филить* (907 об.); иде противу кнзА *александра ярославичА* (908);

ЛП — *ивань творимъ* (52); на *кляцинь озерь* (4), тако *Ольбегь ратиборичь* приима лукъ свои (76).

3. При описании грамматических и лексико-грамматических свойств слова учитываются как признаки, характерные для него в древнерусский период и утраченные по мере приближения к современному языковому состоянию (категория лица и тип исторической основы существительных, двойственное число изменяемых по данной категории слов и т.д.), так и свойственные слову сегодня (соответственно — категория одушевленности, современный тип склонения, *singularia* или *pluralia tantum*, вид глагола и т.д.), например:

*Изменяемость слова:*

склоняемое

*Количество числовых парадигм*

полная числовая парадигма

+Pluralia tantum

+Singularia tantum

\**Основа*

\*а

\*ја

\*о и т.д.

+*Склонение*

несклоняемое

4. Система позволяет учитывать и различные модификации слова в зависимости от позиции в тексте (морфологические — в значении другой части речи, иной грамматической формы, синтаксические — использование слова в несвойственной синтаксической функции, семантические — в значении единицы с другим статусом), например:

*Модификация*

Морфологическая модификация

*Использование слова в значении другой части речи*

в значении существительного

муж. рода.

жен. рода

ср. рода

в значении прилагательного

в значении числительного (счетного слова)

в значении местоимения и т.д.

Семантическая модификация

*Использование слова / словоформы / сочетания*

в значении единицы с другим статусом

в значении личного имени

в значении патронима

в значении прозвища

в значении географического наименования

в значении этнонима

ПМ — въ мирѣ д(ь)во б(огороди)це вьса съхрани (35 об.);  
с(ы)нь зеведьовъ иОань (38);

ПЕ — отьпоущоу вамь црА жидовьска (145);

ЖАН — ЖАНХV — обоноу стра(ну) морА варАжьскаго и до

*великаго риму* (160); *ЖАНХVI* – о пелгоуси *ижерани* (907 об.); *король части римьскыя* О(т) полунощныя страны (157); *isa нав-внв* (159 об.);

ЛП – *Оць же бѣ има малъкъ любечанинъ* (21), *хамови же яса полуденная страна* егпеть (1 об.).

5. Синтаксические свойства и значения языковой единицы (слово и больше слова) указываются в ИПС в соответствии с традиционным членением высказывания на члены предложения / не члены предложения, определяется тип и вид конструкции, тип синтаксической связи между компонентами высказывания, средства связи.

6. Особое внимание уделяется в ИПС семантике слова и его эквивалентов, так как одной из целей создания масштабной базы данных является лексикографирование большого количества текстов. В базу введены следующие параметры семантического описания слова: значение, семантическое и формальное соответствия оригинала и перевода (большая часть древнейших текстов – переводы с греческого на церковнославянский), контекст, возможность комментирования семантических и грамматических аномалий, различные типы лексических групп (антропоним, зооним, название легиона, арготизм и т.д.), лексические связи (омоним, антоним, синоним). Приведем пример одной из таких зон:

*Семантика слова*

*Значение*

*Значение*

*Соответствие между оригиналом и переводом*

*Семантическое соответствие*

*Не соответствует*

*Соответствует*

*Грамматическое соответствие*

*Не соответствует*

*Обратные субъектно-объектные отношения*

*Соответствует*

*Отсутствие эквивалента*

*Отсутствие формы*

*Отсутствие части сложного слова*

*Контекст*

*Поясняющий контекст*

*Комментарии*

*Аномалия*

ПМ — доушегоубьнь (33) / 'губящий души, пагубный' / ψυχοφθόρος<sup>6</sup> / похотил арьсениЕ д(оу)шегоубьно ражжениЕ / доушепагоубьнь, прокоудьнь, доушеврьдливьи, доушеврьдньи;

ЖАН — ЖАНХV — зьлохьтрь (156) / 'хитрый, коварный' / — / въ зьлохьтрьоу дшо не внидеть прьмдтрь / зьлохьтрьнь, хьтрь; ЖАНХVI — благовьрьнь (898) / 'набожный, благочестивый, преданный истинной вере, православный' / ευσεβής, ευσεβων, της ευσεβείας / сеи блговьрьньи и блгородньи бгомь преоудобрен'ньи и хваламь достоины великии кнзь александръ яровлавичь / благочьствивъ, благочьствьнь, доброчьствивъ, доброчьствьнь, доброговьинъ, говьинъ, чьствьнь.

В статье представлены элементы лингвистической модели базы данных, разработанной для изучения языкового материала разных исторических эпох. Очевидно, что достижение конечной цели такой работы впереди: следует сделать компьютерные исследования не просто прикладными (для разного рода подсчетов, для создания конкордансов и их представления в автоматически построенных графиках и схемах), но и собственно аналитическими. И возможно это лишь при подробном описании языковых единиц и их характеристик, при моделировании морфологической, синтаксической и семантической областей и при разработке и создании многофункциональных программ для работы с древними славянскими рукописями.

<sup>1</sup> Универсальная структура базы данных ИПС "Манускрипт" сегодня позволяет конвертировать в нее данные из любого формата.

<sup>2</sup> Баранов В. А. Формирование определительных категорий в истории русского языка. — Казань, 2003. С. 31.

<sup>3</sup> Для ИПС разработана система шрифтов, предусматривающих ввод текста любой сложности, однако в данной статье по техническим причинам древние тексты даны в упрощенном графическом виде.

<sup>4</sup> Здесь и далее приняты следующие сокращения названий древних текстов: ПМ — Путятина Миная — служебная Миная на май, XI в., 135 л. РНБ, Соф. 202; Новгородская служебная миная на май (Путятина миная). XI век: Тексты, исследования, указатели / Подг. В. А. Баранов, В. М. Марков. Ижевск, 2003. См. также: <http://io.udsu.ru/ptm/>.

ПЕ — Пантелеймоново Евангелие — Евангелие апракос полный, XII в., 224 л. РНБ, Соф. 1.

ЖАНХV — Житие Александра Невского в составе Псковской II летописи 1486 г., лл. 156–162 об.; ЖАНХVI — Житие Александра Невского в составе Лицевого летописного свода XVI в., тексты и миниатюры, факсимильное воспроизведение. СПб., 1992. 898 об. — 940 об. + 26 с.

ЛЛ — Летопись по Лаврентьевскому списку, 1377 г.

<sup>5</sup> Баранов В. А. Указ. соч. С. 72–73.

<sup>6</sup> В греческих написаниях в статье отражена не вся диакритика.