

МИНИСТЕРСТВО НА ОБРАЗОВАНИЕТО И НАУКАТА НА РЕПУБЛИКА БЪЛГАРИЯ
КИРИЛО-МЕТОДИЕВСКИ НАУЧЕН ЦЕНТЪР ПРИ БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ
ИЖЕВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ИМ. М. Т. КАЛАШНИКОВА
НАУЧНОЕ СООБЩЕСТВО “ПИСЬМЕННОЕ НАСЛЕДИЕ”
DIGITAL MEDIEVALIST SCHOLARLY COMMUNITY
ФОНДАЦИЯ „УСТОЙЧИВО РАЗВИТИЕ НА БЪЛГАРИЯ“

**Писменото наследство
и информационните технологии**

El’Manuscript–2014

Материали от V международна научна конференция
Варна, 15–20 септември 2014 г.

София · Ижевск
2014

Сборникът е издаден с финансовата подкрепа на Министерството на образованието и науката на Република България по процедура за подкрепа на международни научни форуми.

Отговорни редактори: проф. дфн В. А. Баранов
 доц. д-р В. Желязкова
 д-р А. М. Лаврентъев

Редактори: Нели Ганчева, Веселка Желязкова (български текст)
 О. В. Зуга, В. А. Баранов (руски текст)
 Кевин Хокинс (Kevin Hawkins) (английски текст)

Писменото наследство и информационните технологии [Текст] : материали от V международна науч. конф. (Варна, 15–20 септември 2014 г.) / отг. ред. В. А. Баранов, В. Желязкова, А. М. Лаврентъев. — София ; Ижевск, 2014. — 448 с.

Сборникът съдържа материали от конференция, посветена на разработването и създаването на съвременни средства за съхраняване, описване, обработка, анализ и публикуване на ръкописни и старопечатни книжовни паметници и исторически извори, а също и на въпросите за подготвянето на електронни ресурси в областта на хуманитаристиката и тяхното използване в научните изследвания и преподаването.

© Кирило-Методиевски научен център — БАН, 2014
© Ижевский государственный технический университет
им. М. Т. Калашникова, 2014
© Авторски колектив, 2014
© Лилия Тошкова — графичен дизайн на корицата, 2014

ISBN 978–954–9787–25–2

Исторические корпуса Национального корпуса русского языка как инструмент диахронических исследований грамматики

Д. В. Сичинава

Корпусная лингвистика, древнерусский язык, среднерусский (старорусский) язык, разметка корпуса, грамматика, плюсквамперфект, сравнительная степень

The Historical Corpora within the Russian National Corpus as a Diachronic Grammar Research Tool

Dmitry Sichinava

In this paper the possibilities of the Russian National Corpus as a diachronic research instrument are presented. Currently the corpus includes deeply annotated Old East Slavic texts (manuscripts and birchbark letters) as well as Middle Russian texts still lacking grammatical information. It is shown that usage of both corpora yields new results in studying rare forms such as the pluperfect and prefixed comparatives; examples of these results are given.

В Национальном корпусе русского языка (НКРЯ) в рамках программы Президиума РАН “Корпусная лингвистика” и предшествующих проектов с начала 2010-х годов ведется деятельность по разработке серии исторических корпусов. Отчасти эта деятельность связана с конвертацией для нужд Корпуса баз данных, используемых и в других проектах. Это относится прежде всего к древнерусским ресурсам — текстам древнерусского корпуса XI–XIV вв. и отдельного корпуса берестяных грамот. Тексты первого размечаются еще с середины 2000-х годов в рамках баз данных Отдела лингвистического источниковедения Института русского языка им. В. В. Виноградова РАН (доступ к базам по некоторым древнерусским летописям открыт на сайте <http://www.lrc-lib.ru/index.php%3fid=5>, руководитель А. А. Пичхадзе), второго — для сайта gramoty.ru (морфологическая разметка на нем в настоящее время недоступна; руководитель А. А. Гиппиус), в дальнейшем эти тексты становятся доступны и для НКРЯ (интерфейс между этими ресурсами обеспечивают Т. А. Архангельский и А. И. Зобнин).

Тексты древнерусского корпуса XI–XIV вв. (сейчас доступны 14 текстов: основные летописи, “Пчела”, “История Иудейской войны”, “Чудеса Николы”, Изборник Святослава 1076 г., “Александрия”, три произведения Кирилла Туровского; объём 443 тысячи слов) получают пословную морфологическую разметку со снятием омонимии. Во многих отношениях принципы этой разметки отличаются от принципов разметки основного корпуса НКРЯ, например, отдельные теги получают аналитические формы. Если в основном корпусе НКРЯ сослагательное

наклонение — это просто сочетание формы прошедшего времени и частицы *бы*, то в древнерусском корпусе формы вроде плюсквамперфекта или аналитические конструкции со значением будущего получают дополнительную помету, по которой их можно искать. Есть в этой разметке и определенные элементы параллельного корпуса (так, для памятников, переводных с греческого, в составе разметки указывается греческая лемма и словоформа оригинала), и элементы стандартной филологической адресации текста, для современных текстов не принятые (прямо в пословной разметке указывается лист рукописи и номер строки). Также пословно, со снятием омонимии, разобраны и берестяные грамоты XI–XV вв.; аппарат разметки здесь усовершенствован в связи с большим количеством фрагментарно сохранившихся слов, а также мест, трактуемых лишь предположительно (соответственно для некоторых отрезков даны альтернативные версии, в том числе неоднозначные морфологически); есть пометы “лишнее”, “порча”, “зачеркнуто” и др., которые, конечно, могут использоваться и в книжных памятниках.

Иначе устроен старорусский (среднерусский) корпус, включающий тексты XV–XVII вв. и гораздо больший по объёму (5 млн словоупотреблений). Морфологическая разметка этих текстов проектом предусмотрена, однако ее выполнение как вручную, так и автоматически наталкивается на известные сложности, связанные с огромным объёмом текстов и их орфографической и языковой пестротой. Кроме того, ряд текстов старорусского корпуса берется по изданиям, подготовленным историками и литературоведами; орфография рукописей в них существенно упрощена. В настоящее время поиск по этим текстам ведется в режиме “поиска точных форм”, с возможностью задания подстроки.

Исторические корпуса НКРЯ дают возможность фронтального исследования грамматических феноменов, мало затронутых в предшествующей литературе, в частности, из-за малого количества выявленных примеров. Например, возможен поиск по древнерусскому плюсквамперфекту через задание в грамматической разметке пометы “*rqref*”. В настоящее время корпус насчитывает 1379 примеров этой формы, но, поскольку помета даётся на обеих частях конструкции (а также на возвратном *сж*, если оно есть), то данную цифру следует поделить примерно на 2. Данная конструкция встречается по разу на 800 слов корпуса. При помощи этого поиска можно выявить, помимо стандартного церковнославянского плюсквамперфекта с *блше/бъ*, сравнительно редкие в письменных памятниках древнерусского периода и мало освещённые в древнерусских грамматиках народные “сверхсложные” формы со вспомогательным *был-*, давшие начало современной конструкции типа *пошел было*; ср. [Петрухин, Сичинава 2006]: *фильть же рече ꙗко истинуо добромьслень еси другъ и ташль сѧ еси бы(л), но оуже пращю тѧ, а храни дружбаю ѿселъ со мною* (Пчела). Сплошной анализ плюсквамперфектных форм при подготовке вошедших в корпус баз данных по русским летописям дал новые интересные результаты [Петрухин 2008], недоступные без фронтальной обработки всего корпуса примеров. Так, выявлены свойственные

этой форме с типологической точки зрения значения ‘недостигнутый результат’, ‘результатив в прошедшем’, а также специфические дискурсивные функции.

Несмотря на ограниченность поиска по грамматике в старорусском корпусе, поиск по подстроке с участием символа “*” (например, *л был*, *ла был* и т. п., в том числе с обратным порядком элементов) приносит большое количество примеров старорусского (среднерусского) плюсквамперфекта, в том числе — с нестандартными функциями. Так, удалось установить древность сочетания плюсквамперфекта с модальными глаголами, имеющего аналоги в польском и украинском, а в современном русском — ненормативные *могло было, *следовало было; ср. [Сичинава 2012]:

(1) Божий же священникъ, отвѣща, рече ему: “Сим ли хвалишися, оканне, многобожие вводя и многи боги нарицаеши, имже **было подобало** паче постыдѣтися, по реченому: ‘Да постыдятся вси кланяющиися истуканным, хвалящиися о идолѣх своих?’” Великие Минеи Четвы митрополита Макария. 26 апреля. Слово о житии и учении Стефана Пермского (1530–1554).

Выявлена нестандартная форма элемента *было* в языке XVII в. — *балъ* (ср. известное по памятникам *бола* и т. п.; иные трактовки текста, как кажется, затруднительны):

(2) *А земля у нас была балъ* (к словоформе *балъ* специальное примечание издателей: “так в ркп.”) и *приторгована толка в споре* Д. Пестров сестре Екатерине Калистратовне (конец XVII – начало XVIII в.); Грамотки 1969, № 67.

Пассаж можно трактовать как ‘А мы уже и приобрели было землю, но она спорная’. Если наша интерпретация верна, то здесь одновременно представлены два раритетных явления, характерных для исследуемого периода: ещё один вариант энклитики *было* — *балъ* [бъл], с редукцией второго гласного до нуля, и сочетании конструкции с глаголом-связкой *быть* (ср. известные в XVIII в. примеры типа *было намерен был, был принуждён было*).

Исследование русского компаратива с *по-* (об этих формах в древнерусском см. [Крысько 2006]) показывает, что в Корпусе есть примеры, появившиеся раньше, чем первый пример данного слова, включенный в Словарь XI–XVII вв. (хотя сам памятник при этом обычно входит в базу картотеки): так, *побол(ь)ши* отмечено в Домострое (до 1560 г.) и в “Послании в Кирилло-Белозерский монастырь” Ивана Грозного (1573 г.) при первом примере в Словаре с датой 1578 г.; на *полегче* есть пример из “Чина свадебного” XVI в., а в Словаре — только XVII в. Форма *посильнѣе* в Словаре отсутствует, однако представлена в Корпусе:

(3) Да и Баитерек о том ему добре ж говорил: не соромся деи князь Смаиль, пиши “государем”, немцы деи **посилнее** тебя, да все деи у них государь города поимал. Посольская книга по связям Московского государства с Ногайской Ордой. Кн. 5-я (1557–1561 гг.).

Отмечена также форма *повѣдомос(т)нѣе* ‘содержательнее, информативнее’ (ср. в Словаре — только *повѣдомѣе* ‘осведомленное’):

(4) Извол(ь) втписат(ь) ко мне обо всем поведомоснее против моего пис(ь)ма.
Грамотка Я. П. Непейцыну от неустановленного лица, XVII в.

Таким образом, обращение к древнерусскому и старорусскому корпусам НКРЯ (даже без снятой омонимии в последнем), несмотря на сравнительно небольшой объём, уже является весьма действенным методом получения новых примеров, фронтального просмотра материала, пополнения исторических грамматик и (для лексико-грамматических феноменов) словарей.

Литература

- Грамотки 1969 — *Грамотки XVII – начала XVIII века* / Изд. подгот. Н. И. Торбасова, Н. П. Панкратова; под ред. С. И. Коткова. М., 1969.
- Крысько 2006 — *Крысько В. Б.* Наречия, образованные из предложных сочетаний с формой ВП ед. числа сред. рода сравнительной степени // В. Б. Крысько (ред.) Историческая грамматика древнерусского языка. Т. III. Прилагательные. М., 2006. С. 359–374.
- Петрухин, Сичинава 2006 — *Петрухин П. В., Сичинава Д. В.* “Русский плюсквамперфект” в типологической перспективе // Вереница литер: К 60-летию В. М. Живова. М., 2006. С. 193–214.
- Петрухин 2008 — *Петрухин П. В.* Дискурсивные функции древнерусского книжного плюсквамперфекта (на материале Киевской и Галицко-Волынской летописей) // В. А. Плуныян, В. Ю. Гусев, А. Ю. Урманчиева (ред.) Исследования по теории грамматики. Вып. 4: Дискурсивные категории. М.: Индрик, 2008. С. 213–240.
- Сичинава 2012 — *Сичинава Д. В.* Частицы было и бывало в русском языке XVIII века // Русский язык в научном освещении. 2012. № 2 (24). С. 257–284.