

¹Е. А. Мишина, ²А. А. Пичхадзе

^{1,2}Институт русского языка им. В.В. Виноградова РАН
(Москва, Россия)

¹kmishina@mail.ru, ²rusyaz@yandex.ru

ДРЕВНЕРУССКИЙ ПОДКОРПУС НАЦИОНАЛЬНОГО КОРПУСА РУССКОГО ЯЗЫКА¹

Статья посвящена древнерусскому подкорпусу Национального корпуса русского языка, включающему оригинальные древнерусские памятники (летописи, жития и др.) и древнерусские переводы с греческого. Корпус содержит около полумиллиона словоупотреблений. В статье описывается грамматическая разметка подкорпуса и способы построения запросов с учетом специфики древнерусских текстов. Обращается внимание на возможность построения синтаксических запросов в подкорпусе. Приводятся и анализируются данные некоторых конкретных запросов, демонстрирующих: 1) употребление конструкции «личная форма глагола **быти** + причастие» (типа **бѣ ловы дѣя**), 2) отсутствие форм действительного причастия настоящего времени и имперфекта от перфективных глаголов **дати** и **поустити** при отсутствии отрицания, 3) порядок слов в группе «личная глагольная форма — прямое дополнение — косвенное адресатное дополнение, выраженное местоимением **и (ѣмоу, еи, имъ)**»).

Ключевые слова: текстовые корпуса, древнерусский язык, грамматическая разметка, морфологические и синтаксические запросы.

Древнерусский подкорпус в составе Национального корпуса рус-

¹ Работа над статьей и над корпусом осуществлялась при поддержке Программы фундаментальных исследований Президиума РАН «Корпусная лингвистика» и РГНФ (проекты №14-04-12035 и № 15-04-12050).

ского языка (НКРЯ) изначально задумывался как ресурс по древнеславянским книжным¹ текстам восточнославянского происхождения. Проблема отбора памятников, которые нужно включать в ресурс, предназначенный для того, чтобы давать представление о древнерусском языке, всегда стояла остро и для составителей словарей языка древнерусского и старорусского периода. В отечественной лексикографии она решалась по-разному: И. И. Срезневский и составители «Словаря русского языка XI–XVII вв.», во-первых, считали допустимым использовать любые источники, даже южнославянского происхождения, если они были переписаны на Руси, а во-вторых, использовали памятники, сохранившиеся в поздних списках; составители «Словаря древнерусского языка (XI–XIV вв.)» пытались ограничить круг памятников теми, где встречаются лексические русизмы и которые сохранились в списках не позже рубежа XIV–XV вв. В результате в одни словари попала южнославянская лексика, не усвоенная древнерусским языком даже в его церковнославянском варианте, а в другие словари не попала лексика, представленная в памятниках, заведомо возникших на Руси в ранний период, но сохранившихся в поздних списках. При создании древнерусского подкорпуса НКРЯ был выбран средний путь: привлечение памятников с восточнославянскими элементами в лексике, возникших в период XI–XIV вв., даже если они сохранились только в поздних списках².

Разумеется, в подкорпус включались прежде всего оригинальные древнерусские тексты: на данный момент туда вошли древнерусские летописи и сочинения Кирилла Туровского. Сам по себе отбор оригинальных памятников не создает проблем, все они по возможности должны быть отражены в подкорпусе. Однако при отборе конкретных текстов возникают вопросы аутентичности и оценки источников, как при всяком обращении к средневековым памятникам. Выбор составителей подкорпуса опирался на новейшие источниковедческие исследования. С учетом работ последних лет было принято решение вводить в подкорпус «Повесть временных лет» не

¹ Материал берестяных грамот представлен в особом подкорпусе НКРЯ, поскольку их язык обладает специфическими особенностями, которые обуславливают иной формат разметки и поиска.

² О каждом памятнике дается краткая информация относительно времени создания или перевода. Она появляется в окне, всплывающем при выделении мышью названия памятника в результатах запроса.

по древнейшему Лаврентьевскому списку 1377 г., который сравнительно хорошо изучен и издан со словоуказателем, а по Ипатьевскому списку начала XV в., не менее важному для восстановления первоначального текста памятника (см. [Гиппиус 2002; 2007–2008]). Это, впрочем, не исключает возможности в дальнейшем присоединить и Лаврентьевский список — открытость подкорпуса позволяет объединять в его составе разные версии одного и того же текста. В ходе работы приходилось предпринимать специальные текстологические и лингвистические исследования, на основе которых, например, произведено разграничение Галицкой и Волынской летописей в составе Галицко-Волынского свода [Юрьева 2013], выполнена текстологическая группировка списков повествовательных сочинений Кирилла Туровского и отобраны наиболее представительные рукописи [Баранкова, Макеева 2013].

Что касается переводных церковнославянских памятников, то они отбирались на основании специального исследования переводов с восточнославянскими элементами в лексике [Пичхадзе 2011], которое позволило отделить небольшую группу переводных памятников, переведенную древнерусскими переводчиками, от более обширной группы, которая была переведена, по всей вероятности, южнославянскими переводчиками, работавшими на Руси. На настоящем этапе в подкорпус включена только первая небольшая группа текстов, принадлежащих восточнославянским переводчикам. В перспективе, при расширении подкорпуса, в него могут быть введены и тексты второй группы.

Единственный памятник южнославянского происхождения, представленный на сегодняшний день в подкорпусе, — Изборник 1076 г. Его присутствие в подкорпусе оправдано в силу того, что он подвергся ощутимому редактированию на русской почве и в некоторых отношениях (например, по количеству форм с восточнославянским полногласием) является уникальным для своего времени источником данных по истории именно древнерусского языка [Мушинская, Мишина, Гольшенко 2009]. Представляется правильным и впредь вводить в древнерусский подкорпус памятники, отредактированные на Руси и впитавшие на восточнославянской почве заметное число языковых русизмов, прежде всего лексических.

На сегодняшний день древнерусский подкорпус, содержащий около полумиллиона словоупотреблений, представляет пользова-

телю книжные тексты XI–XIV вв., в наименьшей степени затронутые южнославянским влиянием и максимально отражающие восточнославянские языковые особенности. Грамматическая разметка текстов производилась вручную (с использованием прецедентных разборов) и затем редактировалась, что обеспечивает довольно высокую степень её надежности.

Тексты, входящие в древнерусский подкорпус, приводятся с сохранением древней оригинальной орфографии рукописей (в том числе и те, которые цитируются по изданиям), в некоторых из памятников (например, в «Истории Иудейской войны Иосифа Флавия», «Александрии») при разметке сохранены знаки акцентуации. В связи с этим для удобства пользователя в древнерусском подкорпусе НКРЯ разработана трёхступенчатая система поиска («точная», «упрощенная» и «модернизированная»), позволяющая в соответствии с той или иной исследовательской задачей учитывать разные орфографические режимы. В режиме «точной» орфографии можно искать словоформу или фрагмент словоформы с учетом оригинальной орфографии (буква в букву). В перспективе такой режим должен обеспечить возможность поиска нетривиальных орфографических приемов (например, написания «и десятичного» без точки — **і** — в заимствованиях). В режиме «упрощенной» орфографии можно строить запросы с учетом орфографических вариантов в одной и той же позиции: на виртуальной клавиатуре выделяются, например, узкие варианты букв (**ѣ**, **о**) и **и** восьмеричное, и отдельно — их орфографические варианты, которые могут писаться в определенных позициях (например, **ѣ=к**, **о=ѡ=ѡ** после букв гласных или в начале слова, **і=ї=і** на конце строки и т. п.). Наконец, можно строить запрос в «модернизированной» орфографии, в которой приравниваются не только орфографические варианты, но и некоторые буквы, обозначающие разные фонемы: **ѣ** (и все его орфографические варианты) и **ѣ**, **и** (и все его орфографические варианты) и **ѡ**. Этот режим удобен для пользователей, ориентирующихся на современный облик слова, а также для запросов по словам, орфографический образ которых может варьироваться (например, слов с приставками **пре-/прѣ-**).

Для поиска по леммам наиболее удобными являются режимы, учитывающие упрощенную и модернизированную орфографию, поскольку леммы в общем для всех памятников словаре (всплываю-

шем при наборе леммы) представлены в унифицированной древнерусской орфографии¹.

Обновление и редактирование пополняемого сводного словаря текстов, включенных в древнерусский подкорпус НКРЯ, а также организация эффективного поиска по леммам, учитывающего, например, такие параметры, как разграничение по омонимам, представляет собой отдельную непростую задачу. При создании древнерусского подкорпуса была предпринята попытка её решить. Дополнительные трудности заключаются в том, что, во-первых, в некоторых памятниках представлен только один из омонимов, а во-вторых, в том, что оформление омонимов в разных текстах может отличаться (омонимы в словарях к отдельным памятникам могут получить отличающиеся толкования и т. д.); тем не менее все расхождения по отдельным текстам должны в конечном итоге быть унифицированы в сводном словаре. Создание эффективных алгоритмов автоматизации этого процесса — во многом задача будущего. На данный момент для разграничения омонимов при поиске конкретной леммы целесообразно указывать её частеречную характеристику (например, **и** «союз» и **и** «местоимение»). Если омонимы имеют одинаковые грамматические характеристики (например, **лоуќкъ** ‘оружие’ и **лоуќкъ** ‘лука седла’), то в словаре, который всплывает при наборе леммы, помимо недифференцированной леммы **лоуќкъ**, появляются еще две леммы **лоуќкъ**, при которых указаны их значения. Соответственно, можно осуществлять поиск по каждому омониму отдельно. Однако для некоторых памятников, где омонимы не разграничены (например, «История Иудейской войны»), поиск оказывается возможным только по недифференцированной лемме; в этом случае пользователю придется внимательно просматривать все найденные контексты и разграничивать омонимы самостоятельно. Подсказкой при определении значения омонимов (как и остальных слов) является греческое соответствие, если памятник переводной: оно указано в окне, всплывающем при выделении мышью соответствующей словоформы в результатах запроса.

Разбор древнерусского текста — работа, не вполне поддающаяся стандартизации. Разметка в определенной мере отражает взгляды исследователя, который её осуществляет, на предмет исследования,

¹ Она отражает состояние до падения и прояснения редуцированных.

на то, какой набор признаков необходим, полезен или актуален для научного анализа. Научная трактовка грамматических явлений, отраженных в древнерусских текстах, постоянно меняется, иногда (об этом ниже) непосредственно в процессе разметки. Так, в последнее время появились работы [Петрухин, Сичинава 2006; 2008], посвященные так называемому «русскому плюсквамперфекту», в которых было показано, что эта аналитическая форма, состоящая из *л*-причастия и перфекта в роли связки, представляет собой сверхсложное прошедшее со значением «аннулированного результата», т. е. имеет свою семантику, отличную от значений других перфектных форм, и обозначает действие, результат которого был отменен последующим ходом событий. Помета «сверхсложное прошедшее» была использована при разметке текстов подкорпуса, и теперь эта форма может быть легко найдена при помощи запроса.

Поскольку потребность добавлять релевантные пометы возникает при изучении древних текстов постоянно, система разметки разрабатывалась таким образом, чтобы исследователь мог легко вводить новые признаки по собственному усмотрению. В результате древнерусский подкорпус несет отпечаток индивидуальных подходов к анализу текста. Разные памятники, входящие в подкорпус, могут быть размечены как с разной степенью подробности, так и с разным наполнением одной и той же категории. Разумеется, это не касается базовых признаков и категорий, интерпретация которых однозначна.

Расхождения в грамматической разметке разных памятников проявляются, например, в разборе имен собственных. Как и все остальные леммы, они получают частеречные характеристики: личные имена — характеристику «личное», топонимы — «топоним», отчества и этнонимы определяются как «существительные», а обозначения жены или вдовы по мужу — как «прилагательные». В ряде памятников частеречные пометы отчеств, этнонимов и обозначений жены или вдовы по мужу исчерпываются базовыми («существительное» или «прилагательное»), в других памятниках одновременно с одной из этих обязательных помет дополнительно указывается другая, более детальная: «отчество», «этноним», «по мужу». Это означает, что при запросе на «существительное» во всех текстах подкорпуса будут найдены отчества и этнонимы, а при запросе «прилагательное» — наименования по мужу, но для некоторых памятников

есть возможность построить и более точный запрос на «отчество», «этноним», «по мужу». Расхождения в разметке вызваны прежде всего тематикой памятников: во многих церковнославянских текстах не встречаются ни отчества, ни наименования жены по мужу, поэтому эти пометы используются при разборе летописей и отсутствуют среди грамматических помет остальных памятников. Однако в случае с этнонимами расхождение объясняется исключительно предпочтениями исследователей: одни из них считают целесообразным выделять этнонимы специальной пометой, а другие — нет. В подкорпусе этнонимы размечены в летописях, «Александрии» и сочинениях Кирилла Туровского, а в остальных памятниках они имеют только помету «существительное».

Есть объективные причины, которые приводят к расхождениям в грамматической разметке. Они связаны с неоднородностью явлений, попадающих в традиционной грамматике под одну категорию. Например, категория презенса по-разному трактуется исследователями: одни включают в нее формы с референцией к настоящему и будущему (так размечена Новгородская I летопись), другие разделяют её на две категории — настоящего и настоящего-будущего времени. Это относится и к аналитическим формам так называемого будущего I и будущего II, принадлежащих к числу недостаточно изученных и неоднозначно трактуемых категорий в древнерусском (ср. в последнее время новые интерпретации в [Юрьева 2010; Козлов 2014; Пенькова 2014]). Применительно к будущему I, состоящему из личной формы глаголов **начати**, **почати**, **имѣти**, **мощи**, **хотѣти** в настоящем времени и инфинитива, во многих случаях проблематично уже само объединение этих двух глагольных форм: они могут разбираться и как несвязанные формы в свободной синтаксической конструкции. При разборе переводных памятников исследователь может опираться на данные оригинала: в греческом форма будущего времени образуется синтетически, и если в соответствии с ней в переводе употребляется сочетание **начати**, **почати**, **имѣти**, **мощи**, **хотѣти** с инфинитивом, есть основания рассматривать сочетание как аналитическую форму будущего I. Однако при разборе оригинальных текстов исследователь вынужден исходить только из своих собственных представлений о будущем I и о семантике конкретного контекста, и четкие критерии здесь отсутствуют. В результате сочетания **начати**, **почати**, **имѣти**, **мощи**, **хотѣти** с инфинитивом

в сходных контекстах могут быть размечены, в зависимости от точки зрения исследователя, либо как «будущее I», либо как свободные словосочетания. Трудность разбора таких конструкций связана еще и с тем, что в них входят глаголы разной семантики: сочетания с **начати** и **почати** легче трактовать как временные формы, чем сочетания с модальными глаголами, а сочетания с разными модальными глаголами могут развивать разную семантику, как это было показано применительно к конструкциям **хотѣти** и **имѣти** с инфинитивом [Козлов 2014]. Хотя на данный момент в подкорпусе используется традиционная помета «будущее I», не исключено, что в ближайшее время новые исследования позволят заменить её на несколько помет, более адекватно характеризующих явления, обобщенные под этой рубрикой.

Что же касается будущего II, представляющего собой сочетание *л*-причастия с презенсом от основы *буд-*, то дискуссия по поводу семантики и грамматического статуса этой формы продолжается уже полтора столетия: одни исследователи считают её временной формой, другие — формой наклонения (см. обзор мнений в [Пенькова 2014]). Работа над подкорпусом неожиданно подвела черту в этом споре: при разметке «Повести об Акире Премудром» обнаружилась форма повелительного наклонения **вѹди възалъ** ‘твоя взяла, ты победил (букв.: будь взявший)’, что исключает трактовку будущего II как наклонения, хотя допускает модальные оттенки у этой формы, вообще присущие многим аналитическим перфективным формам. С другой стороны, при разметке «Жития Андрея Юродивого» были отмечены формы будущего II с референцией к прошедшему времени [Там же]. Эти факты показывают, что форма будущего II обладала очень разнообразной семантикой (подробнее см. указанную работу Я. А. Пеньковой), и объединение их под одной пометой «будущее II» выглядит достаточно условным. Можно предположить, что в ближайшее время формы, сейчас объединенные пометой «будущее II», по итогам дальнейших исследований получают какие-то дополнительные дифференцирующие характеристики.

Палеослависты обращаются к корпусам в первую очередь за лексической и морфологической информацией. Однако древнерусский подкорпус позволяет довольно успешно строить синтаксические запросы. В подкорпусе отмечается управление предлогов, принадлежность предиката целевому и косвенно-побудительному

придаточному предложению (глагольная форма получает помету «в значении долженствования»), а также специфическое синтаксическое употребление словоформ: имен прилагательных и причастий — в роли существительных или предикативов, некоторых падежных форм существительных — в роли наречий и т. п. Это дает возможность получать точные результаты в случаях, когда, например, необходимо построить запрос не только на существительные и местоимения, но и остальные субстантивы, т. е. прилагательные и причастия в роли существительных. В частеречных пометах предусмотрены две характеристики для местоимений: «местоимение-существительное» — для местоимений, имеющих субстантивный статус, и «местоимение-прилагательное» — для местоимений, выполняющих роль определения. Отмечаются некоторые предложно-падежные сочетания.

Некоторые синтаксические запросы не дают такого точного результата, как запросы по лексике и морфологии, и требуют ручного редактирования. Тем не менее они могут быть полезны для получения предварительного чернового материала, из которого нужно исключить не относящиеся к делу примеры. Так, неплохой результат дают запросы на дательный самостоятельный (запрос на причастие в дательном падеже).

Большое преимущество корпусов состоит в том, что они позволяют вести поиск по нескольким словам и/или формам одновременно (формы могут быть охарактеризованы только грамматически, без указания лемм). Это хорошее подспорье не только для поиска цитат и фразеологизмов, но и для синтаксического анализа — например, для получения сведений об управлении глаголов. В древнерусском подкорпусе, указав две соответствующие формы, нетрудно отыскать конструкции «**как** + инфинитив» со значением следствия, «личная форма глагола **быти** + причастие», «3 л. ед. ч. глагола **быти** + инфинитив» со значением долженствования, конструкции с отрицанием «**не** + глагольная форма» и др.

Результаты синтаксических запросов предоставляют иногда неожиданную информацию, которая ждет своего исследователя. Так, например, запрос на конструкцию «личная форма глагола **быти** + причастие» (типа **бѣ бо ловы дѣла шлѣгъ**) показывает 9 примеров из «Жития Андрея Юродивого», 5 примеров из «Александрии», 5 примеров из «Истории Иудейской войны», один пример из Избор-

ника 1076 г., более двух десятков примеров из Повести временных лет, 8 примеров из Киевской летописи, 8 примеров из совсем небольшой по объему Волынской летописи и 4 примера из Галицкой летописи. В «Пчеле», Новгородской I летописи и Суздальской летописи надежных примеров нет. Конструкция «личная форма глагола **быти** + причастие» считается специфически книжной, калькирующей соответствующую греческую конструкцию, распространенную в библейских текстах [Живов 2012: 235; Петрухин 2012: 239–243]. Она чрезвычайно частотна в русской агиографии — например в «Житии Феодосия Печерского» [Живов 2012: 236–238]. Между тем данные подкорпуса демонстрируют распределение, не коррелирующее напрямую со степенью книжности текста. Летописи используют указанную конструкцию интенсивнее, чем древнерусские церковнославянские переводные памятники: в одной только Повести временных лет она встречается чаще, чем во всех церковнославянских текстах подкорпуса вместе взятых, а из летописей её очень активно употребляет наименее книжная по языку Волынская летопись — в отличие от заметно ориентированной на церковнославянский стандарт Галицкой летописи. Этот факт требует специального исследования. Разумеется, летописи склонны к использованию клише, и рассматриваемая конструкция может принадлежать к их числу. Тем не менее её распространенность обращает на себя внимание. Возможно, конструкция «личная форма глагола **быти** + причастие» не была просто заимствованием и имела опору в восточнославянском языковом узусе. Если же признать за ней исключительно книжный статус, полученные данные, вероятно, можно интерпретировать как указание на независимость усвоения церковнославянских элементов летописями и церковнославянской древнерусской письменностью. Данные подкорпуса информативны еще в одном отношении, поскольку демонстрируют, что использование рассматриваемой конструкции обусловлено типом дискурса: она употребляется в нарративе (исторических повествованиях, агиографии) и почти полностью отсутствует в гномических текстах (Изборнике 1076 г., «Пчеле»).

Интересный материал дают и результаты запроса на «**яко** + инфинитив». Наряду с ожидаемыми конструкциями следствия (**быша громи велици и блисканїи млънїа, яко всемоу миру подвижатиса** 'были громы великие и блистания молний, так что весь мир при-

шел в движение'), представленными как в церковнославянских, так и летописных источниках, запрос показывает множество примеров омонимичной конструкции при глаголах со значением 'давать клятву, обещание; провозвещать; договориться': **крестъ цѣловати, клатиса, съдоумати яко/како** + инф., ср. в Новгородской I летописи: **сдоумаша <яко> изгонити князя своего всеволода** 'договорились изгнать князя своего Всеволода'; и **такъ оувѣчаста . такъ нѣ воквати на ц(с)ръградъ** 'и так договорились не воевать против Царьграда'; в Киевской летописи: **цѣлоуи крѣтъ яко имѣти братью въ любовь** 'целуй крест дружить с братьями'. Эта конструкция фиксируется в русских летописях, «Александрии» (83.9) и в «Истории Иудейской войны», вообще близкой по языку к летописям. Греческого аналога она не имеет и в церковнославянских текстах не употребляется: там используется либо инфинитивное дополнение без союза, либо придаточное предложение со сказуемым в финитной форме, вводимое союзом **яко**. Судя по всему, **яко** + инф. при глаголах клятвы и обещания — специфическая древнерусская книжная конструкция, генезис которой еще предстоит выяснить (см. о её функционировании [Власова 2014: 189–194]).

Иногда полезными оказываются запросы с исключением слов или грамматических характеристик. Так, например, одним из актуальных вопросов в исследовании формирования категории вида в древнерусском языке является способность перфективных глаголов образовывать формы имперфекта и причастий настоящего времени. Запрос на форму действительного причастия настоящего времени от глагола **дати** в подкорпусе выдает 19 примеров из 7 памятников, а на форму имперфекта от глагола **дати** — 18 примеров из 6 памятников. Однако эти же запросы с исключением отрицательной частицы **нѣ** показывают, что без отрицания формы имперфекта и действительных причастий настоящего времени от глагола **дати** не встречаются ни разу. Аналогичный запрос на формы имперфектов от глагола **пѹстити** и его приставочных производных с исключением отрицательной частицы **нѣ** показывает, что формы имперфекта с формантом **-стѣ-** (**пѹстѣхуѹ, допѹстѣхуѹ, припѹстѣхуѹ, роспѹстѣста**) — в отличие от форм имперфектов с формантом **-ща-** (**пѹщѣхуѹ, испѹщѣше, воспѹщѣше, ѿпѹщѣше**) — встречаются только с отрицанием (как и малочисленные имперфекты от других приставочных глаголов совершенного вида). Полученные дан-

ные могут свидетельствовать в пользу того, чтобы считать глаголы **дати** и **пѹстити** глаголами совершенного вида уже в ранний древнерусский период [Мишина 2012].

В известной мере подкорпус можно использовать для изучения порядка слов. Несмотря на неточность результатов, запросы дают в некоторых случаях картину, которая даже с учетом неточностей в целом позволяет составить адекватное представление о проблеме. В [Пичхадзе 2012: 185] на материале древнерусской «Пчелы» была отмечена тенденция, ставшая закономерностью в современном русском языке: дополнение, выраженное местоимением, предшествует дополнению, выраженному существительным, если только последнее не участвует в контрастном противопоставлении [Рогова 1962: 47]. Вот как по данным подкорпуса распределяются словопорядки в группе «личная глагольная форма (V) — прямое дополнение (O) — косвенное адресатное дополнение (A), выраженное местоимением и (**ѣмоу, еи, имъ**)» в русских летописях:

	VAO	OAV	VOA	OVA	AVO	AOV
Повесть временных лет	26	1	7	2	1	1
Киевская летопись	47	9	11	5	3	4
Новгородская I летопись	11	2	0	0	0	0
Галицкая летопись	12	3	0	3	2	0
Волынская летопись	14	1	3	3	2	0
Суздальская летопись	37	2	10	2	2	0

Приведенные в таблице данные подтверждают наличие тенденции к препозиции дополнения, выраженного местоимением, по отношению к дополнению, выраженному субстантивом, уже в древнерусском языке (сравнительно большие цифры третьей колонки объясняются частотностью в летописях конструкции **нарекоша имъ ѣмоу/еи X**, которая употребляется с порядком VOA). Помимо этого, полученные данные лишней раз демонстрируют, насколько предпочтительнее для древнерусского книжного нарратива была препозиция глагольной формы по отношению к прямому и косвенному адресатному дополнениям по сравнению с постпозицией.

Древнерусский подкорпус моложе многих других подкорпусов НКРЯ, и практика его использования еще невелика. Он, безусловно,

нуждается в усовершенствованиях, в частности в унификации параметров разметки для всех текстов, в оптимизации словаря лемм, в расширении метаданных, в развитии поисковых возможностей. Ресурс для этого имеется, поскольку потенциал осуществленной разметки памятников шире того, что представлено на данный момент в древнерусском подкорпусе. Так, может быть добавлен поиск по греческим леммам и их славянским соответствиям; по альтернативным разборам словоформ, которые допускают неоднозначную трактовку; по ошибочным и испорченным словоформам и др. Тем не менее уже сейчас подкорпус позволяет получать разнообразную информацию, а главное, открывает неожиданные аспекты в изучении древнерусской письменности и ставит перед исследователями нетривиальные вопросы.

Литература

Баранкова Г. С., Макеева И. И. Повествовательные произведения Кирилла Туровского и проблемы русского литературного языка // Письменность, литература, фольклор славянских народов. История славистики. XV Международный съезд славистов. Минск, 20–27 августа, 2013 г. Доклады российской делегации. М., 2013. С. 52–80.

Власова Е. А. Инфинитив и сослагательное наклонение в косвенной речи в русских летописях XI–XVI вв. // Русский язык в научном освещении. 2014. № 1 (27). С. 185–205.

Гиппиус А. А. О критике текста и новом переводе-реконструкции Повести временных лет // Russian Linguistics. 2002. Vol. 26. № 1. С. 63–126.

Гиппиус А. А. К проблеме редакций Повести временных лет // Славяноведение. 2007. № 5. С. 20–44; 2008. № 2. С. 3–24.

Живов В. М. Об употреблении одной книжной конструкции в памятниках восточнославянской письменности: <глагол **быти** с причастием наст. времени> // I. Podtergera (Hrsg.): Schnittpunkt Slavistik. Ost und West im wissenschaftlichen Dialog. Festgabe für Helmut Keipert zum 70. Geburtstag. Teil 3: Vom Wort zum Text. Göttingen, 2012. S. 235–249.

Козлов А. А. К грамматической семантике старославянских конструкций **хотѣти/ имѣти** с инфинитивом // Русский язык в научном освещении. 2014. № 1 (27). С. 122–149.

Мишина Е. А. «Ситуация напрасного ожидания» и отрицание // Русский язык в научном освещении. 2012. № 2 (24). С. 219–241.

Мушинская, Мишина, Гольштенко 2009 — Изборник 1076 г. / Изд. подгот. М. М. Мушинская, Е. А. Мишина, В. С. Гольштенко. Т. I–II. М., 2009.

Пенькова Я. А. К вопросу о семантике так называемого будущего сложного II в древнерусском языке (на материале «Жития Андрея Юродивого») // Русский язык в научном освещении. 2014. № 1 (27). С. 150–184.

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 П. В., Сичинава Д. В. Еще раз о восточнославянском сверхсложном прошедшем, плюсквамперфекте и современных диалектных конструкциях // Русский язык в научном освещении. 2008. № 1 (15). С. 224–258.

Петрухин П. В. К проблеме реконструкции и перевода Повести временных лет // Русский язык в научном освещении. 2012. № 1 (23). С. 232–249.

Пичхадзе А. А. Переводческая деятельность в домонгольской Руси: лингвистический аспект. М., 2011.

Пичхадзе А. А. Порядок слов в сочетаниях «глагол + прямое и адресатное дополнение» в древнерусской «Пчеле» // I. Podtergera (Hrsg.): Schnittpunkt Slavistik. Ost und West im wissenschaftlichen Dialog. Festgabe für Helmut Keipert zum 70. Geburtstag. Teil 3: Vom Wort zum Text. Göttingen, 2012. С. 181–194.

Рогова К. А. Место прямого дополнения в простом повествовательном предложении // Ученые записки Ленинградского гос. университета. № 302. Серия филологических наук. Вып. 61. Исследования по грамматике русского языка. III. Л., 1962. С. 40–49.

Юрьева И. С. Особенности древнерусских инфинитивных сочетаний с глаголом **начати** // Русский язык в научном освещении. 2010. № 2 (20). С. 270–286.

Юрьева И. С. Некоторые особенности синтаксиса, морфологии и лексики так называемой Галицко-Волынской летописи // Лингвистическое источниковедение и история русского языка. <2012–2013>. М., 2013. С. 135–151.

¹E. A. Mishina, ²A. A. Pichkhadze
^{1,2}Vinogradov Russian Language Institute
of the Russian Academy of Sciences
(Moscow, Russia)
¹kmishina@mail.ru, ²rusyaz@yandex.ru

OLD RUSSIAN SUBCORPUS OF THE RUSSIAN NATIONAL CORPUS

The paper deals with the Old Russian subcorpus of the Russian National Corpus that comprises original Old Russian texts (chronicles, hagiography etc.) and Old Russian translations from Greek. The subcorpus contains approximately 500 000 words. The paper describes the grammatical tagging used in the subcorpus and possible ways of constructing complex queries taking into account specific features of Old Russian texts and with special emphasis on possibility to construct syntactic queries. The results of some queries are presented and analyzed in the article: 1) the usage of the construction “finite form of the verb *byti* + participle” (e. g. *bě lovy děja*), 2) the lack of active praesens participle and imperfect forms of the perfective verbs *dati* and *pustiti* except the contexts with negation, 3) word orders in the group “finite verb + direct object + indirect object represented by the pronoun **и** (ѣмѹ, еи, имѣ)”.

Key words: text corpora, Old Russian, grammatical tagging, morphological and syntactic queries.

References

Barankova G. S., Makeeva I. I. [Narrative works of Cyril of Turov and problems of the Russian literary language]. *Pis'mennost', literatura, fol'klor slavyanskikh narodov. Istoriya slavistiki. XV Mezhdunarodnyi s'ezd slavistov. Minsk, 2013. Doklady rossiiskoi delegatsii* [Writing, literature, and folklore of the Slavic peoples. The history of Slavic studies. XV International Congress of Slavists. Minsk, 2013 Reports of the Russian delegation]. Moscow, 2013, pp. 52–80. (In Russ.)

Vlasova E. A. [The use of infinitive and subjunctive forms in modal patterns of indirect speech in Russian chronicles of the XI—XVI

centuries]. *Russkii yazyk v nauchnom osveshchenii*, 2014, no. 1 (27), pp. 185–205. (In Russ.)

Gippius A. A. [On text criticism and a new reconstructive translation of the Primary Chronicle]. *Russian Linguistics*, 2002, vol. 26, no. 1, pp. 63–126. (In Russ.)

Gippius A. A. [On the problem of redactions of the Primary Chronicle]. *Slavyanovedenie*, 2007, no. 5, pp. 20–44; 2008, no. 2, pp. 3–24. (In Russ.)

Zhivov V. M. [About the use of one bookish construction in East Slavic manuscripts: <verb byti with present participle>]. *I. Podtergera (Hrsg.): Schnittpunkt Slavistik. Ost und West im wissenschaftlichen Dialog. Festgabe für Helmut Keipert zum 70. Geburtstag. Teil 3: Vom Wort zum Text. Göttingen*, 2012, pp. 235–249. (In Russ.)

Kozlov A. A. [On the semantics of old Church Slavonic *xotěti/ iměti* + infinitive constructions]. *Russkii yazyk v nauchnom osveshchenii*, 2014, no. 1 (27), pp. 122–149. (In Russ.)

Mishina E. A. ["Situation of futile expectations" and negation]. *Russkii yazyk v nauchnom osveshchenii*, 2012, no. 2 (24), pp. 219–241. (In Russ.)

Mushinskaya, Mishina, Golyschenko 2009 — *Izbornik 1076 g. M. M. Mushinskaya, E. A. Mishina, V. S. Golyschenko (eds.). V. I–II. Moscow*, 2009.

Pen'kova Ya. A. [Towards the problem of the so called «compound Future II» semantics in the Old Russian language (on the material of The Life of St. Andrew the Fool)]. *Russkij jazyk v nauchnom osveshchenii*, 2014, no. 1 (27), pp. 150–184. (In Russ.)

Petrukhin P. V., Sichinava D. V. ["Russian pluperfect" in typological perspective]. *Verenitsa liter. K 60-letiyu V. M. Zhivova*. [Collection of articles for the 60th anniversary of V. M. Zhivov]. Moscow, 2006, pp. 193–214. (In Russ.)

Petrukhin P. V., Sichinava D. V. [Once again about East Slavic complicated past, the pluperfect and modern dialectal constructions]. *Russkii yazyk v nauchnom osveshchenii*, 2008, no. 1 (15), pp. 224–258. (In Russ.)

Petrukhin P. V. [On the reconstruction and translation of the Russian Primary Chronicle]. *Russkii yazyk v nauchnom osveshchenii*, 2012, no. 1 (23), pp. 232–249. (In Russ.)

Pichkhadze A. A. *Perevodcheskaya deyatel'nost' v domongol'skoi*

Rusi: lingvistikheskii aspekt [Translational activity in pre-Mongol Rus': the linguistic aspect]. Moscow, 2011.

Pichkhadze A. A. [Word order in "verb + direct and indirect object" combination in Old Russian "Pchela"]. I. Podtergera (Hrsg.): *Schnittpunkt Slavistik. Ost und West im wissenschaftlichen Dialog. Festgabe für Helmut Keipert zum 70. Geburtstag. Teil 3: Vom Wort zum Text*. Göttingen, 2012, pp. 181–194. (In Russ)

Rogova K. A. [The place of direct object in a simple declarative sentence]. *Uchenye zapiski Leningradskogo gos. universiteta. № 302. Seriya filologicheskikh nauk. Vyp. 61. Issledovaniya po grammatike russkogo yazyka. III* [Scientific notes of the Leningrad State University. No. 302. Series in Philology. Vol. 61. Studies on the grammar of the Russian language. III]. Leningrad, 1962, pp. 40–49. (In Russ)

Yur'eva I. S. [The peculiarities of Old Russian nachati + infinitive constructions]. *Russkii yazyk v nauchnom osveshchenii*, 2010 no. 2 (20), pp. 270–286. (In Russ)

Yur'eva I. S. [Some peculiarities of syntax, morphology and vocabulary of the so-called Galician–Volhynian Chronicle]. *Lingvisticheskoe istochnikovedenie i istoriya russkogo yazyka, <2012–2013>*, Moscow, 2013, pp. 135–151. (In Russ.)