

Компьютерная лингвистика и интеллектуальные технологии:
по материалам международной конференции «Диалог 2020»

Москва, 17–20 июня 2020 г.

БАЗЫ ДАННЫХ НЕКНИЖНОЙ ПИСЬМЕННОСТИ, ИСТОРИЧЕСКИЙ КОРПУС, СЛОВОУКАЗАТЕЛЬ: ИНТЕГРАЦИЯ¹

Сичинава Д. В.

ИРЯ РАН / ВШЭ, Москва, Россия

Дышкант А. Н.

Независимый исследователь

В докладе представлена работа над двумя базами данных не книжной древнерусской письменности — базами берестяных грамот и эпитафий. Ставится задача взаимосвязи информации о памятниках археографического и собственно лингвистического характера, а именно, возможности одновременного пополнения базы данных и лингвистического корпуса с возможностью грамматического и лексического поиска. Онлайн-база данных «Древнерусские берестяные грамоты» и раздел Исторического подкорпуса НКРЯ, посвященный берестяным грамотам, нуждаются в интеграции. Для этого предусмотрены, в частности, создание онлайн-рабочего места морфологического разметки текстов, привязанного к записям базы данных, возможность экспорта XML-баз с морфологической разметкой для размещения в составе НКРЯ, возможность автоматического генерирования прямого и обратного словоуказателей к базе на основании размеченного корпуса, взаимная проверка составленных ранее вручную словоуказателей к книге А. А. Зализняка «Древненовгородский диалект» (готовится новое пополненное издание этой книги) и разметки корпуса. Создание базы данных по древнерусской эпитафике призвано преодолеть существующую разрозненность публикаций и исследований в этой области, объединив накопленный наукой материал в рамках единого электронного ресурса. В программном отношении база данных эпитафики устроена аналогично базе данных по берестяным грамотам, что обеспечивает возможность составления размеченного корпуса древнерусской эпитафики.

Ключевые слова: берестяные грамоты; эпитафика; базы данных; корпус; словоуказатель

DOI: 10.28995/NNNN-NNNN-2020-19-1-11

¹ Работа выполнена в Институте славяноведения РАН при поддержке гранта Российского научного фонда (проект №19-18-00352).

INTEGRATING A HISTORICAL CORPUS, A WORD INDEX, AND A DATABASE OF OLD EAST SLAVIC VERNACULAR WRITING

Sitchinava D. V.

IRL RAS / HSE, Moscow, Russia

Dyshkant A. N.

Independent researcher

The paper presents work on two databases of vernacular Old East Slavic writing, viz. the databases of birchbark letters and epigraphy. The task is to link the information of archaeological/historical and linguistic character, namely, the possibility of simultaneous expansion and updating of the database and of a linguistic corpus that enables grammatical and lexical search. The online database *Old East Slavic birch bark letters* and the section of the historical subcorpus of the Russian National corpus containing birch bark letters call for an integration. This includes, in particular, creating an online workstation for the morphological markup of texts linked to the database entries, the possibility of exporting XML-databases with morphological markup to include them in the RNC, the possibility of automatic generation of forward and backward word indices to the database on the basis of the marked corpus, mutual verification of previously compiled manually indices to the book by A. A. Zalizniak *The Old Novgorod Dialect* (a new updated edition of this book is being prepared) and markup of the corpus. Creation of a new database on the Old East Slavic epigraphy is designed to overcome the fragmented state of the publications and research in the field, combining the material accumulated by the scholars into a single electronic resource. From the point of view of programming, the architecture of the epigraphy database is analogous to the one of the birch bark database, which makes it possible to create an annotated corpus of the Old East Slavic epigraphy.

Key words: birchbark letters; epigraphy; databases of archeological findings; corpora; word indices

Древнерусская не книжная письменность — как берестяные грамоты, так и эпиграфика, то есть надписи на стенах церквей, камне, ювелирных изделиях и других носителях — это постоянно пополняющийся новыми находками и подверженный новым переинтерпретациям и дискуссиям класс письменных памятников средневековой Восточной Европы. Уже давно общепризнана важнейшая роль не книжной письменности в исследовании древнерусского языка.

В докладе представлены две базы данных, посвященных памятникам не книжной древнерусской письменности. Над этими базами в настоящее время ведётся работа коллектива эпиграфистов, лингвистов и историков (А. А. Гиппиус, С. М. Михеев, М. М. Дробышева, А. А. Фитискина, К. Д. Покровская, С. А. Лашук,

а также соавторы данной статьи). В ходе проекта по созданию и усовершенствованию этих баз данных ставится задача интегрирования информации о памятниках археографического и археологического характера (тип, место и время находки, контекст, фотографии, прорисы) и собственно лингвистического характера, а именно, возможности одновременного пополнения базы данных и лингвистического корпуса (составной части НКРЯ) с возможностью грамматического и лексического поиска. К настоящему времени функционирует база данных «Древнерусские берестяные грамоты» (gramoty.ru), создан работающий закрытый прототип базы данных по древнерусской эпиграфике. К концу 2021 года планируется опубликовать базу данных по древнерусской эпиграфике, осуществить интеграцию обеих баз с НКРЯ, создать механизм автоматического генерирования словоуказателя.

1. База данных «Древнерусские берестяные грамоты»

В 2007 году была опубликована онлайн-база данных «Древнерусские берестяные грамоты» (gramoty.ru), охватывающая все берестяные грамоты, известные по состоянию на 2004 г. База включает сведения о номере и месте находки каждой грамоты, томе публикации в издании «Новгородские грамоты на бересте» (Novgorod letters on birchbark 1951–2015), сохранности текста. Кроме того, указывается тип документа (письмо, долговой список, упражнение в письме и т. п.) и приводится текст грамоты (в том числе варианты со словоделением и без такового), а для большинства текстов также перевод, фотографии и/или прорисы. В базу интегрирован текст второго издания книги А. А. Зализняка «Древненовгородский диалект» [Zalizniak 2004], ссылки на соответствующие фрагменты книги доступны из записей базы, посвящённых конкретным грамотам. В 2017 году стала доступной новая версия базы, разработанная одним из соавторов настоящей статьи А. Н. Дышкантом под общим руководством А. А. Гиппиуса, включающая новые грамоты, найденные в 2006–2017 гг. и учитывающая накопившиеся за это время поправки к первой тысяче берестяных грамот (с интеграцией посвящённой этим поправкам работы [Gippius, Zalizniak 2015] в структуру базы данных), а также расширенную библиографию.

На настоящем этапе ведётся работа над доработкой структуры базы данных, связанной с археологической информацией о находке. Это добавление информации о грамотах, найденных при нестандартных обстоятельствах, уточнение данных о находке некоторых грамот, обновление данных об отношении отдельных грамот к средневековым усадьбам, уточнение сведений об отношении грамот к квадратам, уточнение информации о метрике некоторых раскопов (имеющиеся пласты, их ширина, нумерация).

Кроме того, в настоящее время нуждается в усовершенствовании доступ к существующей исследовательской литературе о берестяных грамотах; например, в нее необходимо импортировать (с удобной навигацией непосредственно от записей о грамотах) уже присутствующие на сайте в разделе «Библиотека» статьи многотомного издания «Новгородские грамоты на бересте» (Novgorod letters on birchbark 1951–2015), а также обеспечить переход

от библиографических ссылок в записях о грамотах к их расшифровке в библиографическом списке (или к полному тексту публикаций, если он размещен на сайте).

Перечисленные выше усовершенствования предполагается поместить в общий доступ в 2021 г. Подробнее о базе данных берестяных грамот в рамках настоящего проекта см. [Раздел 4](#).

2. База данных по древнерусской эпиграфике

Создание базы данных по древнерусской эпиграфике призвано преодолеть существующую разрозненность публикаций и исследований в этой области, объединив накопленный наукой материал в рамках единого электронного ресурса. Эту базу планируется опубликовать онлайн и в дальнейшем пополнять; разработан её работающий прототип, который пополняется надписями различного типа.

К настоящему времени в базу добавлено 78 надписей, каждая из которых перекрестно классифицирована по разным критериям: по типу и категории носителя (икона, крест, литургический предмет, церковные двери, бытовой предмет, стена и т. д.), по способу нанесения (граффито, дипинто, золотая наводка, резьба по камню и т. д.), по состоянию сохранности (целая, с небольшими утратами, с утратами, фрагмент, малый фрагмент), по используемому алфавиту и др. Классифицирующие признаки являются лишь частью общей структуры информации о надписи, которая включает в себя три вида сведений:

- 1) Сведения о носителе. Сюда относится: указание на происхождение носителя (город, местность, здание, раскоп), его тип, материал, инвентарный номер, место хранения.
- 2) Сведения об истории изучения надписи, представленные в виде объектов, условно называемых «интерпретациями», которые соотносятся с надписями по связи «многие к одному», т. е. одна надпись может содержать отсылку к нескольким различным «интерпретациям», каждая из которых существует в контексте того или иного конкретного исследования и показывает, как его автор описывал изучаемую надпись. Сюда относится: расположение на носителе, способ нанесения, состояние сохранности, алфавит, текст, транслитерация, перевод, изображение и прориси, содержательная категория, содержание, дата в тексте, стратиграфическая, внестратиграфическая и историческая датировки.
- 3) Т. н. «актуальные сведения», т. е. данные, отражающие позицию коллектива, работающего над описываемой базой данных. Структура «актуальных сведений» концептуально повторяет структуру «интерпретации», однако для каждого поля (будь то текст, содержательная категория или что-либо другое) допускает фиксацию двух видов сведений: во-первых, отсылок к актуальным, по мнению авторов, интерпретациям и, во-вторых, оригинальной авторской интерпретации.

Такое построение базы данных позволяет объединить в ней сведения о разных видах древнерусской эпиграфики: надписях-граффити на стенах храмов, надписях на произведениях церковного искусства, монументальных надписях на каменных крестах, надписях на бытовых предметах и др., фиксируя при этом как актуальные по мнению авторов данные, так и альтернативные, в т. ч. конкурирующие, интерпретации тех или иных надписей.

С точки зрения архитектуры в обеих базах, — на сайте gramoty.ru и разрабатываемом сайте эпиграфики, — используется система управления базы данных MySQL, «поверх» которой работает сам сайт. Ввиду использования реляционного движка базы данных, внутреннее устройство базы максимально приближено к «нормальному», т. е. информация о каждом описываемом объекте хранится в соответствующей таблице и не дублируется (это помогает в том числе и поддерживать консистентность данных в базе). Предполагается также создание интерфейса, который будет предоставлять возможность экспорта данных из базы в различных форматах.

3. Интеграция с НКРЯ

Онлайн-база данных «Древнерусские берестяные грамоты» (gramoty.ru) и раздел Исторического подкорпуса Национального корпуса русского языка (НКРЯ; ruscorpora.ru/search-birchbark.html), посвященный берестяным грамотам, нуждаются в интеграции (при сохранении автономного существования). Соответствующий подкорпус, размещенный в ряду с остальными историческими корпусами НКРЯ на сайте ruscorpora.ru (ср. литературу о древне- и среднерусском корпусе: [Pichkhadze 2005], [Lyashevskaya 2016]), доступен онлайн с 2014 г. Он включает в себя морфологически размеченные со снятой лексико-грамматической омонимией (размечено также толкование ряда лексем, информация о фрагментарно сохранённой словоформе, неоднозначности толкования и проч.) и возможностью лексико-грамматического поиска тексты грамот, включенных во второе издание «Древненовгородского диалекта» [Zalizniak 2004], с дальнейшими дополнениями вплоть до находок 2012 г. После этой даты исправления и дополнения в корпус не вносились. Таким образом, сосуществуют две рассинхронизированные базы данных по берестяным текстам, каждая со своими достоинствами и недостатками.

База gramoty.ru, включающая большую археологическую и библиографическую информацию, а также фотографии и прориси, не содержит морфологической разметки и лексико-грамматического поиска. В этом отношении она примыкает к существующим базам данных, посвящённым фиксации археологических находок, как в основном лишённых текста (ср. базу данных НовГУ им. Ярослава Мудрого, [Toporova et al. 2017]), так и комплексов, представляющих собой аналогичные берестяным грамотам массивы письменных документов (ср. базу Виндоландских табличек [Vindolanda tablets online], где ведётся только полнотекстовый поиск без возможности выдачи всех словоформ одной лексемы). Тексты берестяных грамот с лексико-морфологической разметкой более уместны в составе более крупного исторического корпуса данного языка.

В то же время база данных gramoty.ru содержит важную информацию о графическом представлении текста (выносные буквы, вписанные между строк слова), которые при морфологической разметке корпуса не сохранились.

После 2012 г., когда последний раз обновлялся корпус берестяных грамот в составе НКРЯ, при раскопках обнаружено более 100 новых грамот, в том числе значительные как по объёму, так и по лингвистическому и историческому интересу тексты. В корпусе также отсутствуют грамоты, не включённые во второе издание «Древненовгородского диалекта», по большей части мелкие фрагменты. В соответствии с этим же принципом, в корпус не входили и аналогичные фрагменты и приравненные к ним «незначительные» грамоты, найденные начиная с 2005 г. и не включённые в предварительные публикации в журнале «Вопросы языкознания».

Объём этих «неучтённых» текстов действительно невелик — по подсчётам в [Zalizniak 2004: 706], не более 2% от общего количества словоупотреблений². Тем не менее даже среди этих документов есть содержащие небезыңтересные лингвистически или исторически вхождения словоформ и словосочетаний. Даже независимо от ценности этих текстов лингвистический корпус такого ограниченного массива исторических источников должен быть *полным*; это, несомненно, осознавал и А. А. Зализняк, подготовленные которым лингвистические словоуказатели (как в составе [Zalizniak 1995, 2004], так и даже и в составе подготовленной им книги [Zalizniak to appear], охватывающей по замыслу автора только тексты, найденные после 2004 г.) охватывали 100% известных на тот момент грамот со славянским текстом. Естественной является идея в дальнейшем составлять словоуказатели к берестяным грамотам (полу)автоматически на основе электронного корпуса (см. подробнее ниже, [раздел 3](#)).

Далее, в корпусе не учтены выявленные после 2004 г. поправки к текстам берестяных грамот, затрагивающие как буквенный состав грамот, так и их интерпретацию. Некоторые из этих поправок переводят плохо интерпретируемые документы и/или мелкие фрагменты в разряд безусловно заслуживающих монографического описания; наиболее выдающийся успех — анализ новгородской грамоты № 88, в которой издатели не прочли ни одного слова и которая оказывается интереснейшим документом из примерно 80 букв, представляющим собой, по видимому, послание князя к двум посадникам: [Gippius, Zalizniak 2015: 201–205].

Впрочем, последнее обновление базы данных произошло в конце 2017 г., а новые поправки (затрагивающие как буквенный состав, так и интерпретацию, в том числе грамматическую) появляются постоянно (ср., например, посвящённую грамматической реинтерпретации ряда словоформ в грамоте № 1083 работу [Krys'ko 2020], или меняющую синтаксическую структуру и перевод грамоты № 724 интерпретацию А. Журавель, изложенную в [Petrukhin 2019]). Выход из этой ситуации такой: поправки следует одновременно вносить как в базу, так и в корпус по мере их выявления.

² Не говоря уже о том, что не все документы, пронумерованные публикаторами как грамоты, содержат с уверенностью выделяемые слова или даже буквы, например, грамоты № 396, 399, 401, аттестуемые в базе gramoty.ru как «имитация письма».

Данный процесс поддерживается участниками команды лингвистов-разработчиков НКРЯ (в которую входит один из соавторов настоящей работы Д. В. Сичинава).

В процессе интеграции базы данных «Древнерусские берестяные грамоты» и НКРЯ предусмотрены, в частности,

- 1) создание онлайн-рабочего места морфологического разметчика берестяных грамот / эпитафических текстов, привязанного к текстам базы данных,
- 2) возможность экспорта XML-баз с морфологической разметкой для размещения в составе НКРЯ, возможность автоматического генерирования прямого и обратного словоуказателей к базе на основании размеченного корпуса,
- 3) взаимная проверка составленных ранее вручную словоуказателей к книге А. А. Зализняка «Древненовгородский диалект» (см. следующий раздел) и разметки корпуса.

Тексты грамот хранятся в ныне доступной на сайте версии базы gramoty.ru в собственном формате (не XML), который позволяет выделять выносные буквы, приписки между строк и другие особенности написания. Существование морфологической разметки этот формат не предполагает и не поддерживает. При том, что база данных включает не представленные в корпусе документы, стоит задача создания алгоритма, который сохранил бы информацию, представленную в обоих источниках.

Первым шагом такого алгоритма является приведение внутреннего формата gramoty.ru к виду XML-разметки.

Второй шаг — сопоставление текстов в базе gramoty.ru и НКРЯ. На выходе этого шага два списка: те грамоты, буквенный состав или словоделение которых изменились, а следовательно существующая для них в НКРЯ разметка не актуальна и они должны быть размечены заново; и те грамоты, тексты которых остались прежними — для них будет взята разметка из НКРЯ (что не исключает дальнейшей корректировки их разметки и/или перевода).

Третьим шагом становится адаптация компонентов сайта, взаимодействующих с текстом, к новому XML-формату.

Итогом применения этого алгоритма станет база данных, в которой содержится:

- a. наиболее полный список грамот;
- b. наиболее актуальные прочтения грамот;
- c. сохраненная морфологическая разметка (где это возможно);
- d. сохраненные сведения о визуальной составляющей текста (выносные надстрочные и подстрочные буквы, слова, вписанные в межстрочное пространство).

В настоящее время разметка древнерусского и «берестяного» корпусов НКРЯ осуществляется полуавтоматически в системе Morphy, созданной в ИРЯ РАН [Arkhangelsky, Pichkhadze, Mishina 2014], изначально на базе

грамматического указателя к [Zalizniak 2004], конвертированного в электронную базу данных, а затем тексты вновь найденных грамот добавляются в существующую базу данных и размечаются.

У этой схемы есть принципиальный недостаток: любая поправка в тексте грамоты, а также любое пополнение списка текстов (которые случаются регулярно), влекут за собой изменение как разметки, так и самого тела корпуса. В итоге мы имеем схему, при которой каждая минимальная поправка в тексте грамоты требует активных действий со стороны сотрудников НКРЯ (которые, к тому же, должны быть оповещены о каждом таком изменении). Фактически это приводит к остановке совместной работы над корпусом и рассинхронизации, которая сейчас и наблюдается.

Создаваемая нами система призвана избавить нас от описанного недостатка: во-первых, в соответствии с ней вся ответственность за разметку берестяного и эпиграфического корпуса переходит к исследователям, которые непосредственно ими занимаются (что будет означать постоянную актуальность разметки); во-вторых, процесс реимпорта текстов в состав НКРЯ становится полностью автоматическим, что означает отсутствие необходимости вмешательства в него со стороны сотрудников НКРЯ.

Отдельной задачей является создание рабочего места разметчика, т. е. интерфейса, пользуясь которым разметчик может создавать XML-разметку двух типов: морфологическую и визуальную/структурную (к последней относится деление на строки, выносные буквы, вписанный текст). Рабочее место представляет собой веб-редактор со средствами ввода, которые помечают фрагменты текста необходимой морфологической и визуальной XML-разметкой. При этом в рабочее место планируется интегрировать лучшие решения, принятые в используемой для работы над базами данных древнерусского корпуса системе Morphu, в том числе модули, поддерживающих задание иерархических структур разного уровня, запоминание разборов, порождение словоуказателей и др.

В программном отношении эпиграфическая база данных устроена аналогично имеющейся базе данных по берестяным грамотам, что обеспечит возможность аналогичного составления размеченного корпуса древнерусской эпиграфики в рамках Национального корпуса русского языка (а также автоматического генерирования словоуказателей, см. ниже). Подобная перспектива представляет большое значение для грамматического и лексикографического описания древнерусского языка, поскольку заметную часть корпуса восточнославянских текстов XI–XIV вв. представляют именно надписи, в том числе происходящие из регионов, где иные виды некнижной письменности документированы недостаточно (Киев, Полоцкая земля и др.).

4. Корпус, словоуказатель и грамматическое описание

Основополагающая для «берестологических» исследований во всем мире работа А. А. Зализняка «Древненовгородский диалект» выдержала два издания [Zalizniak 1995], [Zalizniak 2004]. Эта книга посвящена грамматическому

описанию диалекта, на котором написаны берестяные грамоты Новгорода, Старой Руссы, Торжка и Пскова, а также лингвистическому комментарию к тексту практически всех известных на тот момент берестяных грамот, уже не только новгородских (кроме неславянских и мелких фрагментов). «Древненовгородский диалект» сопровождается полными словоуказателями ко всему корпусу берестяных грамот.

В последние годы А. А. Зализняк готовил книгу «Древненовгородский диалект. Дополнение» [Zalizniak to appear], включающую тексты найденных за 2015–2017 гг. берестяных грамот, поправки и новые интерпретации к ранее известным текстам, а также добавления и исправления к грамматическому очерку, сформулированные на базе этих новых текстов и интерпретаций. Работа над «Дополнением» была оборвана смертью Зализняка в декабре 2017 г. Автор не успел завершить ряд разделов грамматического очерка, привести описание некоторых берестяных грамот к формату, принятому в ДНД, полностью выверить прямой и обратный словоуказатели ко всему корпусу берестяных грамот, включенные в книгу.

В настоящее время книга А. А. Зализняка готовится к посмертной публикации (в том числе с пополнением текстами берестяных грамот, найденных после смерти автора). Прямой и обратный словоуказатели к берестяным грамотам, статистические данные, а также перечни форм в грамматическом очерке дополнительно сверяются с данными полного электронного морфологически размеченного корпуса берестяных грамот, полученного на основании интегрированной с корпусом базы данных. Всякое слово, для которого в разметке корпуса выделена хотя бы одна лемма, может быть помещено в автоматически генерируемый словоуказатель.

Например, на основании полной текстовой базы новых находок (без использования морфологического поиска) грамматическое описание в составе [Zalizniak to appear] пополнено при подготовке к печати новыми примерами, первоначально пропущенными автором: на вторую палатализацию — пример *целю* (описка вместо *цълую*, грамота 1061), на переход *t>kl* — *клицевъ* (грамота 964), шесть примеров на отпадение конечных гласных и др. Уточнена также статистика адресных форм, приводимых А. А. Зализняком [Zalizniak 2004: 37]. Из морфологии уточнены данные об окончаниях женского склонения в блоке Якима. Корпус помогает выявить также не учтённые в последней версии словоуказателя вхождения лексем, например, частицы *ли* в грамоте из Старой Руссы № 39 (в составе надписанного над строкой *ли то продале*). При этом подобных пропусков и неточностей в составленном вручную словоуказателе Зализняка сравнительно немного, и они иногда дополнительно содержательно мотивированы (например, неуверенностью в наличии лексемы в тексте).

В дальнейшем этот механизм может быть использован для подготовки дальнейших дополнений к «Древненовгородскому диалекту» и полуавтоматического пополнения словоуказателя (возможно, только в электронном формате, с регулярными обновлениями).

5. Заключение

Создаваемые и совершенствуемые в рамках этого проекта базы данных, посвящённые памятникам не книжной письменности, призваны стать удобной доступной платформой для всех, интересующихся этими памятниками — лингвистов, историков, археологов. Эти базы интегрируют и накапливают информацию, содержащуюся в разных публикациях источников, а также библиографические указатели для дальнейшего обращения к существующим интерпретациям. Параллельно с этой работой ведётся расширение и разметка корпуса берестяных грамот, который станет, таким образом, охватывать полностью весь фонд известных на сегодняшний день документов. Кроме того, станет возможным создание с той же технологией и корпуса древнерусской эпиграфики. Важное достижение проекта — в дальнейшем все новые находки не книжных текстов и уточнения к ним будут одновременно появляться в базе данных и в поиске НКРЯ, и все исправления и пополнения будут вноситься каждый раз централизованно.

К настоящему времени реализован находящийся в закрытом доступе прототип разрабатываемого ресурса (сайта и лежащей в его основе базы данных); реализован интерфейс администрирования базы, в котором выделены роли администратора и редактора. обработано более ста надписей, 78 из которых прошли процедуру импорта и были добавлены в базу. К концу 2020-го года планируется вывести сайт в открытый доступ, постепенно пополняя его уже описанными в исследованиях надписями (публикации Б. А. Рыбакова, А. А. Медынцева, С. А. Высоцкого, публикации в томах «Новгородских грамот на бересте»). К концу 2021 года планируется реализовать механизм экспорта базы данных, а также провести интеграцию с НКРЯ.

References

1. *Arkhangelsky T. A., Pichkhadze A. A., Mishina E. A.* (2014). A system for digital grammatical markup of Old East Slavic and Church Slavonic. [Sistema elektronoj grammaticheskoj razmetki drevnerusskix i cerkovnoslavjanskix tekstov]. *Palaeobulgarica*, Vol. 38, № 4, 21–37.
2. *Gippius A. A., Zalizniak A. A.* (2015). Addenda and corrigenda to the interpretations of the birchbark letters published previously [Popravki i zamehanija k chteniju ranee opublikovannyx berestjanyx gramot], V. L. Janin, A. A. Zalizniak, A. A. Gippius. Birchbark letters from the 2001–2014 excavations [Novgorodskie gramoty na bereste (iz raskopok 2001–2014 gg.)]. Vol. 12, Indrik, Moscow, 2015, pp. 196–275.
3. *Krys'ko V. B.* (2020). D'voji chepi: On the history of Slavic collective numerals [D'voji chepi: K istorii slavjanskix sobiratel'nyx chislitel'nyx], *Issues in linguistics [Voprosy jazykoznanija]*, 2020, № 4, pp. 55–69.
4. *Lyashevskaya O. N., Gavrilova T. S., Shalganova T. A.* (2016). Creating a lexical-grammatical base for the Middle Russian RNC corpus [Sozdanie leksiko-grammaticheskoj bazy dlja starorussskogo korpusa NKРЯ], El'Manuscript–2016.

- Rašytinis palikimas ir skaitmeninė technologijos: VI tarptautinė mokslinė konferencija, Vilnius, 2016 m. rugpjūčio 22–28 d. [El’Manuscript–2016. Written Heritage and Digital Technologies: VI International Scientific Conference, Vilnius, 2016. August 22–28], Lietuvos mokslo taryba, Vilnius, pp. 229–234.
5. *Novgorod letters on birchbark* [Novgorodskie gramoty na bereste], Moscow, Vol. [1]–12, 1951–2015.
 6. *Petrukhin P. V.* (2019). Where and When did the Novgorodian Savva Gather the Tribute? [Gde i kogda sobiral dan’ novgorodec Savva?], *Slověne*. Vol. 8, № 1, pp. 55–108
 7. *Pichkhadze A. A.* (2005). A Corpus of Old East Slavic translations of the 11–12 centuries and researching the translated literature of Old Rus’ [Korpus drevnerusskix perevodov XI–XII vv. i izuchenie perevodnoj knizhnosti Drevnej Rusi], *Russian National Corpus: 2003–2005* [Nacional’nyj korpus russkogo jazyka: 2003–2005], Indrik, Moscow, 251–262.
 8. *Toropova E. V. et al.* (2017). Medieval antiquities of the Novgorod land: an electronic database of archeological findings. Novgorod Yaroslav the Wise University [Srednevekovye drevnosti Novgorodskoj zemli: elektronnaja baza dannyx arxeologicheskix naxodok. NovGU im. Jaroslava Mudrogo]. Available at: <http://www.novsu.ru/archeology/db/>.
 9. *Vindolanda tablets online*. Available at: <http://vindolanda.csad.ox.ac.uk/>.
 10. *Zalizniak A. A.* (1995). Old Novgorod dialect [Drevnenovgorodskij dialekt], *Jazyki russkoj kul’tury*, Moscow, 1995.
 11. *Zalizniak A. A.* (2004). Old Novgorod dialect [Drevnenovgorodskij dialekt]. 2nd edition. *Jazyki russkoj kul’tury*, Moscow, 2004.
 12. *Zalizniak A. A.* (to appear) — Old Novgorod dialect. Addenda [Drevnenovgorodskij dialekt: Dopolnenie.] (to appear).