

Министерство образования РФ
Алтайский государственный университет

Ассоциация «История и компьютер»

**КРУГ ИДЕЙ:
ЭЛЕКТРОННЫЕ РЕСУРСЫ
ИСТОРИЧЕСКОЙ ИНФОРМАТИКИ**

Труды VIII конференции Ассоциации
«История и компьютер»

Под редакцией
Л.И. Бородкина и В.Н. Владимирова

Издательство  Алтайского
университета
Москва-Барнаул 2003

ББК 63с51
К84

Отв. редакторы:

Л.И. Бородкин, В.Н. Владимиров

Редакционная коллегия:

Л.И. Бородкин, В.Н. Владимиров, С.Г. Кашенко, Д.В. Колдаков,
Э.В. Летенков, А.Н. Полевая

Редакционный совет:

Е.Н. Балыкина (Минск); д.и.н. Л.И. Бородкин (Москва);
к.и.н. В.Н. Владимиров (Барнаул); к.и.н. И.М. Гарскова (Москва);
к.и.н. С.А. Жакишева (Алматы); к.и.н. Е.В. Злобин (Москва);
д.и.н. В.В. Канищев (Тамбов); д.и.н. С.Г. Кашенко (Санкт-Петербург);
к.п.н. А.В. Кириллов (Санкт-Петербург); к.и.н. И.Н. Киселев (Москва);
к.и.н. А.А. Кобринский (Москва); д.и.н. С.И. Корниенко (Пермь);
к.и.н. Г.В. Можаяева (Томск); к.т.н. А.Ф. Оськин (Полоцк);
д.и.н. В.В. Подгаецкий (Днепропетровск); А.Н. Полевая (Москва);
к.и.н. И.Е. Сафонов (Воронеж); д.и.н. В.Н. Сидорцов (Минск);
д.и.н. Т.И. Славко (Тверь); д.и.н. А.К. Соколов (Москва);
к.и.н. И.Ф. Юшин (Москва)

Рецензенты:

д.и.н. Н.М. Арсентьев, д.т.н. Е.В. Бауман

К84 Круг идей: электронные ресурсы исторической информатики:
Труды VIII конференции Ассоциации «История и компьютер» /
Под ред. Л.И. Бородкина, В.Н. Владимирова. Москва; Бар-
наул: Изд-во Алт. ун-та, 2003. 588 с.: илл.
ISBN 5–7904–0241–0

Настоящий сборник содержит статьи участников VIII конференции Ассоциации «История и компьютер», состоявшейся в июне 2002 г. в Санкт-Петербурге.

Издание подготовлено при участии кафедры архивоведения и исторической информатики Алтайского государственного университета.

Редакционный совет благодарит Институт «Открытое общество» за финансовую поддержку данного издания.

Эмблема на обложке: В.Н. Куплевацкий

ББК 63с51

ISBN 5–7904–0241–0

© Ассоциация «История и компьютер», 2003

*В.А. Баранов, А.А. Вотинцев, Р.М. Гнутиков, О.В. Зуга,
А.Н. Миронов, С.А. Никифорова, С.В. Ощепков,
В.А. Романенко, Е.В. Рябова*

**Электронные издания
древних письменных памятников
и технология создания
полнотекстовых баз данных¹**

Известно, что текстологи различают издание текста и публикацию отдельного древнего письменного памятника (рукописи). Текст должен быть сначала изучен и только потом может быть издан². На основании этого положения Д.С. Лихачев однозначно определяет первый тип (издания критические, дипломатические) как издания научные, а вторые — как ненаучные, как «сырой материал» для дальнейших исследований (Лихачев, 491—496). Напомним, что исследователь, настаивая на таком подходе к изданию памятников древнерусской литературы, имел в виду в первую очередь их предварительную текстологическую и литературоведческую интерпретацию (Лихачев, 486, 487).

Вместе с тем совершенное издание текста с точки зрения его изученности может быть осуществлено, в частности, лишь при точном и адекватном эпохе понимании его языковых особенностей и при корректной интерпретации текста средствами современного русского языка. К сожалению, это в настоящее время не всегда достижимо в связи с несколькими причинами:

1. Недостаточной изученностью языковых особенностей древнерусского и среднерусского периодов;
2. Принципиально различными на некоторых языковых уровнях средствами выражения аналогичных смыслов в древности и в настоящее время³;
3. Невозможностью современными средствами орфографии и пунктуации передать в издании то, что читал, понимал и воспринимал читатель в древности⁴.

Одновременно с этим, как будто бы всеми признается, что публикация отдельной рукописи может быть осуществлена в том случае, если издается древнейший или наиболее важный список текста, рукопись хорошей сохранности, датированная, единичная, иллюминированная и т.п. (Лихачев, 493). При этом требования к изданию меняются: должны быть максимально сохранены графико-орфографические особенности, а литературоведческое, текстологическое изучение может быть лишь предварительным. Но при таком типе издания, ценном для лингвистов, уменьшается доступность текста для текстологов, литературоведов, историков и ученых других специальностей⁵. Вывод Д.С. Лихачева неутешителен: лингвистические, текстологические, литературоведческие издания памятников принципиально различны и несовместимы: «Сторонники общих изданий древних текстов напрасно полагают, что между языковедами, литературоведами и историками может быть достигнуто единство требований, если только историки и литературоведы согласятся пользоваться изданиями “более точными в орфографическом и графическом отношении”, чем те, какими они пользовались до сих пор. Эти различия гораздо более серьезны» (Лихачев, 479—480), и далее: «Даже если бы превосходные лингвисты и превосходные историки объединили свои усилия, — все равно при совместной подготовке текста к изданию им пришлось бы встать либо на точку зрения лингвиста и издавать для лингвистов, либо стремиться удовлетворить потребности историков... Одним словом, научные издания, как нам кажется, могут быть только целенаправленными. Лишь в очень простых случаях (короткий памятник в одном или двух списках) можно вести работу по подготовке издания, имея в виду интересы нескольких наук одновременно» (Лихачев, 481—482; см. также 498).

Дело осложняется также тем, что даже самая подробная, тщательно подготовленная публикация неизбежно содержит те или иные изменения текста, его упрощение из-за выбранного технического способа тиражирования (в современной терминологии — создается транскрипция

текста), любая публикация — это *интерпретация*⁶. И поэтому никогда нет уверенности в том, что в издании или публикации будут отражены все особенности рукописи, к которым однажды кому-либо из исследователей (дело касается в первую очередь палеографических, графических особенностей, особенностей оформления и исполнения рукописи, выбора и интерпретации разночтений) необходимо будет обратиться⁷.

Можно ли в одной публикации одновременно воспроизвести все графические, орфографические и структурные особенностей рукописи и сделать его ценным для ученых других специальностей и доступным для широкого круга читателей? Можно ли дать возможность в пределах одного издания работать с фактами рукописи с разной степени условности и точности их передачи? Ответ на эти и другие вопросы, которые появляются при желании совместить «несовместимое» в издательской деятельности, по нашему мнению, может быть найден: положительные результаты могут быть получены при использовании электронных изданий древних письменных памятников.

В настоящее время принципы создания электронных публикаций, вопросы конкретного исполнения такого рода изданий только начинают разрабатываться⁸.

По нашему мнению, принципиальными требованиями к электронному изданию древнерусских рукописей являются:

1) *Адекватность оригиналу*, то есть сохранение всех графико-орфографических и структурных особенностей памятника — вариантных и выносных букв, титл и выносных знаков, их типов, знаков оформления текста, инициалов и изображений, сохранение расположения знаков, строк и столбцов, текста на поле, передача различий в почерках и т.д.;

2) *Трансформируемость составляющих*, иначе возможность изменения графического вида текста в сторону его упрощения: автоматическое устранение графических дублетов, раскрытие титл, внесение надстрочных букв в строку, а также возможность изменения структуры рукописи: добавление пробелов и иных разделителей, упрощающих чтение и идентификацию составляющих единиц текста, подача текста не только «строка в строку», но и с учетом текстовой, содержательной структуры — выделение отдельных чтений, стихов, разделов и т.п.;

3) *Информативность*, а именно наличие такого справочного аппарата и в таком объеме, который в обычном издании зачастую невозможен: наличие указателей различных типов (форм и слов, прямых и обратных, текстовых и этимологических, сокращений и слов под титлом), наличие списков форм, употребленных в однотипных частях

текста, имеющих, например, одинаковые грамматические характеристики, а также возможность получения информации, скрыто представленной в самом тексте: количественных, статистических данных о структуре, составе и составляющих текста;

4) *Доступность*, что обуславливается электронной формой представления рукописи и справочных материалов.

Такого рода изданию должны быть присущи возможности, отличающие его от печатного:

1) Быстрый поиск необходимой информации по графическим, орфографическим, структурным, а при проведении необходимой предварительной лингвистической работы — также и по грамматическим, семантическим, тематическим и иным признакам;

2) Возможность регулярного дополнения новой информацией, как лингвистической, так и экстралингвистической;

3) Неограниченная тиражируемость с помощью электронных средств копирования;

4) Возможность подготовки бумажных эквивалентов текста и справочного аппарата (указателей, словарей, комментариев) на основе электронного издания.

Для осуществления электронного издания необходимо решить достаточно много теоретических и практических задач, в частности:

1) Создать текстовый процессор, способный передавать все палеографические, графико-орфографические и структурные особенности древнерусских текстов;

2) Спроектировать и создать базу данных, дающую возможность хранить как саму электронную копию рукописи, расчлененную на минимальные информативные составляющие, и дополнительную, неявно представленную в ней информацию, так и любую информацию исследовательского, комментирующего, ссылочного и тому подобного характера;

3) Разработать структуру представления текста и справочного аппарата, а также удобный и понятный интерфейс электронного издания, включающий, в частности, средства выбора структуры полученной информации с точки зрения расположения и упорядочивания текста и его единиц и различных форм их визуализации;

4) Провести текстологическое, палеографическое и лингвистическое исследование публикуемого текста, набор и его компьютерную обработку с целью создания адекватного электронного эквивалента рукописи и упрощенных его аналогов.

По нашему мнению, электронное издание должно включать следующие компоненты:

1) Базу данных, отражающую палеографические, графико-орфографические, структурные особенности оригинала и позволяющую работать как собственно с текстом, так и с указателями разного типа, конкордансами, выборками необходимой информации с учетом максимально возможного количества критериев поиска;

2) Палеографические, лингвистические, текстологические и иные комментарии и исследования, связанные с текстом рукописи по принципу гипертекста;

3) Электронную графическую копию текста, имеющую два варианта: копию для быстрого просмотра страниц и копию с максимально возможным разрешением кадра.

Каждая из форм представления текста и справочного материала должна удовлетворять требованиям определенного круга исследователей и читателей, а в случае необходимости позволяет обратиться как к более подробной, детальной информации о тексте и его единицах, так и к интерпретации текста, справочному аппарату, комментариям и к обобщающим исследованиям.

Среди древнейших славянских письменных памятников, таких, как Остромирово Евангелие 1056–1057 гг., Изборники Святослава 1073 г. и 1076 г. и др., давно введенных в научный оборот, хорошо известных широкому кругу любителей древней словесности, есть рукописи, не менее ценные как в научном, так и в общекультурном отношении, но менее известные. К ним относится, в частности, служебная Миней на май (Пуятинна Миней), переписанная в Древнем Новгороде в первой половине XI в. с южнославянского оригинала. Неоднократно небольшие отрывки рукописи публиковались в научных работах, ее палеографические, языковые и текстологические особенности были предметом исследования многих ученых — В.М. Маркова, В.В. Колесова, Н.А. Нечунаевой, В.А. Баранова. В 1998–2000 гг. текст полностью издан в журнале *Palaeoslavica*⁹, в 2001 г. часть памятника опубликована Л.И. Щеголевой¹⁰.

Ценность рукописи неоспорима: структура и состав Миней не имеет аналогов среди других древнейших служебных Миней, ее языковые особенности (графика, орфография, фонетика, морфология) очень архаичны по сравнению со всеми древнерусскими памятниками XI в. Следует ожидать, что дальнейшее исследование текста учеными различных специальностей, не только лингвистами и текстологами, принесет немало интересного.

В Удмуртском госуниверситете несколько лет назад началась работа по подготовке издания этой уникальной рукописи. Помимо традиционной публикации была подготовлена полнотекстовая база данных, включающая, кроме слов, словоформ и их грамматических значений, еще и несколько трансформированных видов текста, позволяющая осуществлять поиск, выборку и упорядочивание информации. Фрагмент этой базы в виде электронного издания доступен в настоящее время по адресу в Интернет <http://io.udsu.ru/ptm/>.

Собственно база данных представляет собой один из модулей создаваемой Системы обработки письменных памятников «Манускрипт» (Manuscript) (далее — Система). Разработанная модель базы данных позволяет создавать разнообразные справочные материалы, работать с текстом как с совокупностью фрагментов, добавлять другие тексты, проводить сравнение выборок по разным параметрам, а также добавлять комментирующую, справочную, исследовательскую информацию в процессе работы над текстами.

Помимо базы данных, в состав Системы в настоящее время входит специализированный редактор для набора, редактирования и занесения текстов в базу, комплекс сервисных программ, конверторы для связи с другими базами данных, средства обработки запросов, конверторы в издательские системы.

Появление подобного рода изданий ставит на повестку дня необходимость активного обсуждения принципов, подходов и технологий создания электронных текстовых ресурсов на основе древних и средневековых письменных памятников. Надеемся, что данная статья будет интересна всем, кто интересуется технологией представления исторических документов с помощью компьютера.

1. Представленное в Интернет электронное издание *Путятиной Миной* является web-интерфейсом к полнотекстовой базе данных¹¹. Каждая страница, появляющаяся в окне браузера, является не статическим файлом, хранящимся на web-сервере, а формируется всякий раз динамически на основе информации, хранящейся в базе. Эта информация представлена в виде объектов, характеризующихся различными свойствами и многочисленными связями, в терминах нашей системы объекты называются *единицами* текста. Например, в тексте, который содержит 135 листов, насчитывается ~ 1 220 000 единиц, связей между единицами ~ 1 500 000, свойств единиц ~ 3 660 000.

Этот компонент Системы позволяет по запросу пользователя получать фрагменты текста, прямые и обратные слово- и формоуказатели,

количественные указатели, конкордансы, элементы иллюстраций, осуществлять поиск словоформы по образцу и характеризующим ее грамматическим признакам.

Совершенно естественно, что авторы и разработчики не стали отходить от состава традиционного издания: электронное издание содержит те же разделы, что и печатное, — *тексты, указатели, комментарии, библиографию, иллюстрации* и т.д.¹² Мы руководствовались тем, что представление материала, близкое традиционному печатному, поможет исследователю легко пользоваться как текстами, так и указателями. В то же время пользование таким изданием значительно отличается от пользования печатным текстом. Основной принцип работы — получение выборок с помощью запросной формы. В результате обработки запроса выдаются все единицы, удовлетворяющие критериям запроса, сообщается их количество. Форма показа результатов запроса выбирается пользователем.

Запросная форма имеет два раздела: раздел примеров и web-форму ввода значений (см. рис. 1).

В разделе примеров пользователю предоставлена возможность познакомиться с фрагментами текста указателей без заполнения запроса. Раздел ввода значений содержит поля для ввода значений.

В поле «Форма результата» предоставляется возможность выбора вида текста — оригинальный (максимально приближенный к оригиналу рукописи¹³), преобразованный (текст, подготовленный по правилам дипломатического издания¹⁴), нормализованный (текст, реконструированный по правилам старославянского языка¹⁵), современный¹⁶; возможность выбора вида и типа указателя — указатель словоформ, начальных форм, обратные указатели форм и слов, количественные указатели; а также иллюстраций — фотокопии некоторых страниц рукописи, копии оригинал-макета печатного издания, эскизов графических элементов и пр. В поле «Листы от ... до ...» указываются интересующие исследователя листы и страницы рукописи. В поле «Образец поиска» вводится интересующая словоформа или ее часть. В поле «Грамматические значения» могут быть введены грамматические характеристики слов и/или словоформ (значения вводятся в любой последовательности в сокращенном написании).

На рис. 2, 3, 4 можно увидеть результаты выполнения запросов: фрагменты текста и указатели.

2. Полнотекстовая база данных, в которой хранится доступная в Интернет информация, организована в соответствии со специально

разработанной моделью. Модель позволяет хранить и описывать практически любые объекты, которые интересуют или могут заинтересовать исследователя (*единицы*) (см. рис. 5¹⁷). Так, единицами являются и могут являться собственно *текст*, каждая *текстовая словоформа*, каждая из *морфем*, *пунктуационные знаки*, *синтагмы*, *предложения*, *фрагменты*, выделенные на определенном основании, и др. В Системе отсутствуют какие-либо ограничения на список исследуемых объектов. В процессе работы пользователь может расширить этот список, например, выделив интересующую единицу, которая обладает определенными свойствами, в новый тип.

Единицы, формально представленные в тексте в виде знака или в виде набора расположенных в определенном порядке знаков, обладают координатами и размерами (высота и ширина). Кроме того, существуют единицы, не имеющие явного графического представления в тексте рукописи. Это *начальные формы словоформ*, *морфологические*, *синтаксические*, *семантические* и подобные признаки; *эквиваленты текстовых единиц* и т.д.

Любая единица может обладать уникальным набором характеристик, определяемых типом единицы. Причем на уровне модели данных заданы лишь две из них: *координаты* и *размер*. Это обусловлено тем, что, во-первых, они присущи большинству единиц, имеющих графическое представление, во-вторых, эти характеристики чаще всего встречаются в первичных запросах.

В модели реализован механизм зависимости набора допустимых характеристик от значений других характеристик (одной или нескольких, объединенных в группу). Иначе говоря, по мере описания единицы растет список допустимых характеристик. Так, возможность указать *род* словоформы появляется лишь после указания характеристики *часть речи* (значение *имя существительное* или *имя прилагательное*) либо пары характеристик — *часть речи* (значение *глагол*) и *время* (значение *прошедшее*).

В Системе предусмотрена возможность ведения различных *словарей* (рис. 6). Любая единица может являться элементом словаря. Применение словарей может значительно снизить трудоемкость описания единиц. Например, единица может наследовать часть характеристик единицы словаря *начальных форм*, а именно те характеристики, которые не указаны для нее явно. Поэтому целесообразно описывать лишь характеристики, либо отличные от характеристик элемента словаря, либо отсутствующие у словарного элемента.

Между выявленными составляющими текста (единицами) может существовать множество разнообразных *связей* (рис. 7). Возможность исследования, описания этих связей является одной из основных задач, решаемых Системой. В существующей модели множество типов рассматриваемых связей также определяется пользователем в соответствии с решаемыми задачами. Связь определяется множеством единиц (концов связи) и характером вхождения единицы в связь (типом конца связи). Связь может обладать характеристиками. Примеры связей: *связь вхождения, связь следования, связь с элементом словаря* и др.

Для описания структурного состава исследуемых объектов (единиц) введено понятие *иерархия*, определяющее правила построения связей между единицами различного типа (рис. 8).

Использование иерархий позволяет уменьшить дублирование информации, так как единицы наследуют значения характеристик единицы более высокого уровня. Например, если для единицы не указано время написания, оно определяется временем написания единицы следующего уровня, согласно иерархии, определяющей свойство *время написания* (рис. 9).

Как показало предварительное тестирование, предлагаемая модель хранения информации о древних текстах является достаточно универсальной и гибкой. В то же время для предоставления исследователю удобных и быстрых средств извлечения разнообразной информации из базы данных в некоторых случаях возникает необходимость создания дополнительных структур данных для ускоренного доступа к данным базы при выполнении наиболее часто используемых запросов.

Представленная модель требует специального структурирования данных рукописей, языка и экстралингвистических данных, использующихся при исследовании письменных памятников.

3. Как уже было сказано, в качестве составляющих исследуемого объекта (текста, рукописи, языковой системы) выступают единицы. Они представляют собой выделенные на тех или иных основаниях фрагменты текста или рукописи, а также введенные пользователем абстрактные объекты, которые связаны с выделенными и реально существующими в тексте или рукописи единицами. Основанием выделения является, в частности, наличие в тексте/рукописи единиц с идентичными структурно-функциональными признаками (свойствами).

Выделение единицы осуществляется указанием *состава (набора) составляющих ее подъединиц, границ единицы, ее типа* и принадлежности к определенной *подсети* и *иерархии* в ней. Работа с единицей предполагает

установление *связей* с единицами того же уровня, а также определение *значений ее свойств и значений свойств ее связей*.

При выделении единиц и установлении их свойств и связей пользователь имеет возможность: 1) указать иной состав (набор) подъединиц и, соответственно, иные границы выделения той же единицы, то есть дать вариант выделения единицы; 2) указать иное значение свойства, а именно дать вариантное значение; 3) указать вариантную связь; 4) указать иное значение для связи; 5) указать приоритет вариантов выделения, связей и значений исходя из 100-процентной шкалы, а также *предположительность* выделения, свойств и значений.

Между единицами существуют различного рода связи. Совокупность единиц и связей между ними представляет собой *сеть единиц*¹⁸. В сети единиц на основании свойств определенной предметной области выделяются *подсети*. Подсети имеют структуру иерархической подчиненности, иначе, являются *иерархиями*. Например, в иерархии геометрической подчиненности единицей самого нижнего уровня является «знак», а единицей самого высокого уровня — «рукопись» (рис. 13).

Группировка единиц в подсети и иерархии осуществляется 1) для указания правил построения единиц любого уровня из минимальных единиц системы — знаков — по определенным правилам; 2) указания на наличие иерархических связей между типами единиц; 3) указания правил следования значений свойств единицами той или иной иерархии; 4) научных исследований, так как группировка единиц в подсети в той или иной мере точности отражает (моделирует) структуру и систему отношений между единицами исследуемой области — речи и языка.

Важнейшим структурным условием существования между двумя единицами иерархических отношений является невозможность выхода границ более низкой единицы (подъединицы) за границы более высокой (единицы).

Важнейшим сущностным условием наличия между двумя единицами иерархических отношений и их вхождения в одну подсеть является аналогичность единиц в рамках представляемой (моделируемой) области — в понятиях текста и языка¹⁹.

Следует различать *несколько типов отношений* между единицами, входящими в одну иерархию. *Отношения полного вхождения* предусматривают отсутствие в единице более высокого уровня неопределенных фрагментов на более низком уровне (см., например, геометрическую подсеть). *Отношения неполного вхождения* предусматривают наличие в единице более высокого уровня фрагментов, которые на более низком

уровне не могут быть подвергнуты делению (см., например, словообразующий ярус языковой подсети).

Единицы одной иерархии а) могут не иметь идентичных свойств (например, многие единицы языковой подсети), б) могут иметь только идентичные свойства (например, многие пары единиц геометрической подсети) и, в) могут иметь как идентичные, так и индивидуальные свойства (например, многие пары геометрической подсети).

Отношения наследования предусматривают перенос значений идентичных свойств с более высоких единиц на более низкие единицы (см., например, свойства и значения единиц геометрической подсети). Отношения без наследования не предусматривают перенос значений свойств более высоких единиц на более низкие единицы (см., например, словообразующий ярус языковой подсети). Наследование осуществляется в том случае, если у единиц одной иерархии существуют идентичные свойства.

У единиц некоторой иерархии имеются не только идентичные свойства, что дает возможность изменить значение данного свойства у единицы более низкого уровня при наличии значения у единицы более высокого уровня (например, значения свойства *размер*), но и повторяющиеся свойства, значение которых не может быть изменено у единицы более низкого уровня при наличии его у единицы более высокого уровня (например, значение свойства *материал*). Введение повторяющихся свойств предусмотрено для случая, когда в рукописи не выделена единица более высокого уровня с данным свойством. Например, не выделена единица *тетрадь*, тогда значение свойства *материал* может быть задано у единицы *лист*.

При необходимости пользователь может ввести отсутствующую в Системе единицу, указать ее тип и свойства. Новая единица вводится или в уже имеющуюся подсеть и/или иерархию, или определяются новые подсеть и/или иерархия.

При введении новой *подсети* обязательно указываются наиболее общая (максимальная) единица и минимальная единица (для моделируемой области в качестве первой чаще всего выступают рукопись и текст, в качестве второй — знак). После этого указываются промежуточные единицы новой подсети. При введении новой единицы в уже существующую подсеть указывается место единицы в иерархии. Для этого указываются ближайшие (верхняя и нижняя) единицы. Если между этими единицами отсутствуют какие-либо иные единицы, то новая единица становится их промежуточной единицей.

Если между ближайшими верхней и нижней единицами уже существует промежуточная(ые) единица(ы), то новая единица должна составить *альтернативный участок иерархии*, не пересекающийся с уже имеющимся, иначе — *новую иерархию*.

Основанием и необходимым условием введения новой единицы между существующими является ее полное (без остатка) вхождение в ближайшую верхнюю единицу и возможность выделения в ней единиц ближайшего нижнего уровня.

Связи между любыми тремя единицами (как основными, так и вспомогательными) сети представляют собой «треугольник», в котором «длинная связь» соединяет крайние из трех единиц, а две «короткие связи» — промежуточную единицу с ее ближайшими единицами (рис. 10, 11, 12). При построении связей между единицами четырех и более уровней соблюдается правило: в смежных «треугольниках» «длинная связь» одного из них может быть одновременно лишь «нижней короткой связью» другого. При нарушении этого правила, которое может произойти на этапе маркирования текста, связи автоматически упорядочиваются.

4. Количество подсетей в Системе не ограничено и определяется требованиями пользователей. В частности, в качестве подсетей в Системе существуют геометрическая, лингвистическая подсети, подсети по времени или месту написания фрагмента, по писцу, по автору и т.д.

Приведем пример трех подсетей, построенных на основе описанных принципов.

Основными единицами **геометрической подсети** являются *рукопись*, *лист/страница*, *строка*, *знак*, вспомогательными — *тетрадь*, *блок*, *подстрока* (рис. 13).

Основные единицы геометрической подсети являются первичными, обязательными для выделения в составе рукописи и минимально достаточными при определении координат единиц рукописи. Крайними единицами подсети являются единицы *знак* и *рукопись*. В начале работы с новым текстом по умолчанию устанавливается *страница*, которая соотносится с экраном, служащим местом ввода знаков текста. Объединение страниц (листов) в *рукопись*, а также разбиение страниц (блоков) на строки могут быть произведены в любом порядке. Выделение вспомогательных единиц может быть произведено только после установления основных ближайших единиц.

Приведем примеры свойств и значений некоторых единиц геометрической подсети:

Рукопись²⁰: материал: *пергамен, бумага, воск* и др.; размер (длина): мм; размер (ширина): мм; размер (диаметр): мм...

Лист (аналог): материал (повторяющееся свойство), размер (повторяющееся свойство), тип: *основной, вклейка, наклейка, печать, обложка, форзац* и др.; соединение (для неосновных типов): *к..., до..., после...*; место соединения: мм; способ (средство) соединения: *пришито, приклеено, шнур*, др.; номер в тетради: $I...N$; номер в рукописи: $I...N$; количество поверхностей: $I...N$.

Блок: тип: *столбец основного текста, поле (правое, левое, верхнее, нижнее, угол, между столбцами, между строками...), сноска, колоннитул, пагинация*, др.; размер (длина): мм; размер (ширина): мм; размер (диаметр): мм; последовательность: *после..., до..., независимо*, др.; нумерация блоков одного типа: $I...N$; средство соединения блоков одного типа: *знак, без средства*, др.; точка соединения неосновного блока с основным: *к...*; средство соединения неосновного блока с основным: *знак, без средства*, др.; количество строк в блоке: $I...N$; координаты начала блока: мм; координаты конца блока: мм.

Знак: координаты: мм.

Сеть единиц рукописи по времени создания включает только основные единицы: *рукопись, фрагмент по времени создания, знак/изображение* (рис. 14). Наличие подсети позволяет отметить в рукописи фрагменты, созданные в иное время, чем вся рукопись в целом (например, позднейшие приписки). Для указания времени создания фрагментов рукописи обязательным является указание времени создания высшей единицы подсети — самой рукописи.

Свойства и значения единиц подсети по времени создания (рукопись): **рукопись**²¹: *вилка дат: 1056–1057, без даты; фрагмент по времени создания: вилка дат: 1 пол. XII в.*, др.

Аналогичной структурой и составом обладают подсети единиц по месту создания рукописи и по создателям, по времени, месту, авторам текста и др.

Основными единицами **языковой подсети** являются *текст, предложение-фраза, словоформа, знак* (рис. 15). Эти единицы являются первичными, обязательными для выделения в составе текста и минимально достаточными при его лингвистическом исследовании. Обязателен также порядок введения единиц *знак* (1) и *текст* (2), которые являются соответственно крайними²² единицами в языковой подсети. Порядок выделения единиц *словоформа* и *предложение-фраза* при наличии единиц *знак* и *текст* может быть любым. При выделении единиц *слово-*

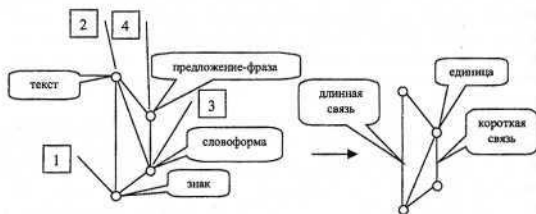


Рис. 10. Граф основных единиц языковой подсети (первичная и перестроенная структура).

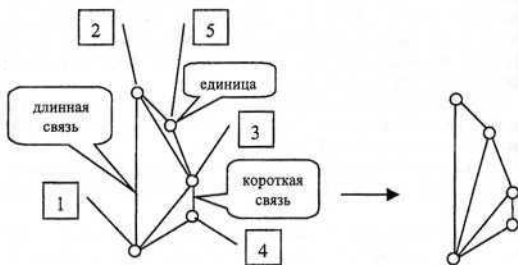


Рис. 11. Граф вспомогательных единиц языковой подсети (первичная и перестроенная структура).

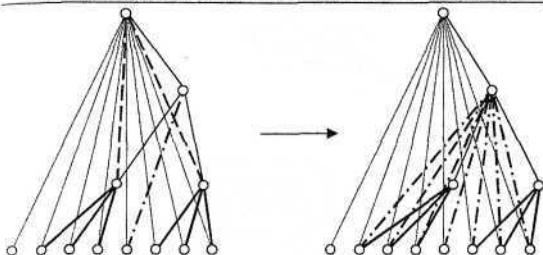


Рис. 12. Графы связей между единицами разных уровней и перестроение первичного графа в граф с основной структурой.

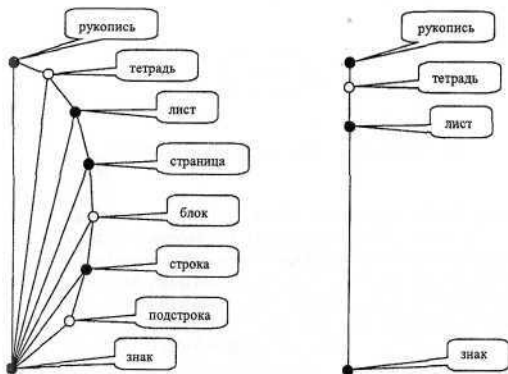


Рис. 13. Граф единиц геометрической подсети и одна из иерархий подчиненности.



Рис. 14. Граф единиц подсети по времени создания.

форма и *предложение-фраза* могут быть опущены иные составляющие текста (например, знаки оформления). Необходимым условием для выделения единиц *словоформа* и *предложение-фраза* является разделение исследуемого объекта на блоки (текст на поле, например), если в рукописи они есть, и на листы (страницы), это дает возможность получить сведения о координатах расположения единиц относительно первого листа рукописи и работать с каждым блоком отдельно.

Единица *предложение-фраза* к единице *текст* и единица *словоформа* к единице *предложение-фраза* находятся в отношениях неполного вхождения. Это значит, что при выделении *словоформы* и *предложения-фразы* могут остаться неопределенные, с точки зрения их функции в тексте, фрагменты.

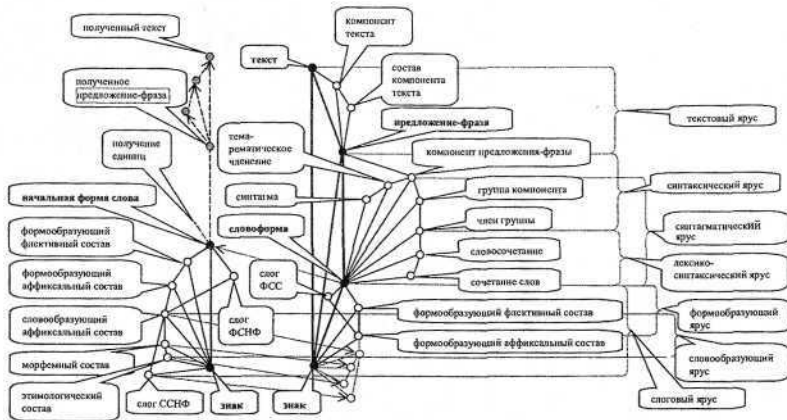


Рис. 15. Граф единиц языковой подсети с подграфом начальной формы слова²⁵.

Вспомогательные единицы языковой подсети по отношению к единицами более высокого уровня находятся в отношениях полного вхождения. При разделении на вспомогательные единицы выделяются все составляющие «родителя» с указанием их функциональной нагрузки в его составе.

Приведем свойства и значения некоторых единиц языковой подсети.

Компонент текста²⁴: значения: *заглавие, собственно текст, рубрикатор компонента, отделитель компонента...*

Состав компонента текста: значения: *сложное синтаксическое целое (ССЦ), диалог, рубрикатор группы, отделитель группы...*

Компонент предложения-фразы (ПФ)²⁵: значения: *предикативная часть, средство связи с предыдущим ПФ, средство связи компонентов ПФ, осложнитель ПФ, детерминант ПФ, разделитель предикативных частей²⁶, отделитель ПФ...*

Группа компонента: значения: *ядро группы подлежащего, ядро группы сказуемого, полупредикативная группа, средство связи групп, разделитель групп, осложнитель группы, детерминант группы...*

Член группы: значения: *подлежащее, сказуемое, определение, дополнение, обстоятельство, средство связи членов, разделитель членов...*

Формообразующий флективный состав (слово): значения: *основа грамматическая, окончание.*

Формообразующий аффиксальный состав (основа без окончания): значения: *основа формообразующая, формообразующий аффикс.*

Словообразующий состав (основа без окончания и формообразующего аффикса): значения: *основа словообразующая, словообразующий аффикс.*

Морфемный состав (часть слова без окончания, формообразующего и словообразующего аффиксов): значения: *основа непроеизводная, приставка, суффикс, постфикс, интерфикс...*

Этимологический состав (часть слова без окончания, формообразующего, словообразующего аффиксов и непроеизводной основы): значения: *корень, приставка, суффикс, постфикс, интерфикс...*

5. Понятно, что, помимо серверной части и реализованных к настоящему времени predetermined средств построения запросов для чтения информации из базы данных, Система по работе с древнерусскими текстами должна обладать удобным **клиентским интерфейсом**. Основной программой клиентской части должен стать **специализированный текстовый редактор**, имеющий собственную объектную модель, поддерживающий коллективную работу, имеющий возможность

загрузки документа и работы с ним в локальном режиме. Пользователю редактор предоставляет возможность как набора текстов и наполнения ими баз данных, так и редактирования и разметки текстов, установления связей между единицами, добавления и изменения их свойств и значений и другие возможности (рис. 16).

Отличительными особенностями редактора являются: 1) возможность набора электронной копии рукописи в адекватном оригиналу виде; 2) обеспечение работы с текстом как системой лингвистических и текстологических единиц; 3) постоянное взаимодействие с базой данных²⁷.

Работа с текстом как с системой единиц, находящихся в иерархических отношениях и объединенных в подсети, и взаимодействие редактора с базой данных диктует специфические операции, которые пользователь выполняет при работе с текстом и его объектами, и некоторые специфические режимы работы.

Ввод и редактирование текста осуществляется при помощи стандартного для текстовых процессоров экрана. При необходимости пользователь может увидеть границы активных в данный момент единиц, например словоформ, синтагм, фрагментов, написанных разными писцами, и т.д.

При сохранении фрагмента в этом режиме в базе данных сохраняется следующая информация: формируются единицы типа *знак* (по одной на каждый введенный знак); устанавливается связь введенных знаков с единицами более высокого уровня в геометрической подсети (строкой, листом, рукописью); каждой новой единице присваивается значение свойства *физические координаты*; другие операции.

Помимо собственно набора, удаления, вставки и других стандартных операций этот экран позволяет фрагментировать набранный текст (создавать единицы) и устанавливать связи выделенной единицы с ближайшими единицами иерархии. Таким образом, текст размечается на части, необходимые для его дальнейшего исследования: это могут быть заголовки, фрагменты композиционно-тематического состава (главы, стихи, погодные записи и т.п.), разделы календарно-временного состава, читаемые в определенные дни года (памяти святым, разделы служебных Евангелий), части логико-структурного состава (зачин, основная часть, отступление, заключение и под.), фрагменты, отличающиеся друг от друга жанровыми и стилевыми характеристиками, и т.п. Разметка может включать самую разнообразную информацию о рукописи и тексте — от указания дефектов и их видов и особенностей исполнения

рукописи до сведений об авторах отдельных фрагментов и времени их создания. Все это осуществляется выделением фрагмента, выбором соответствующей подсети, типа единицы и указанием на установление связи этого выделенного фрагмента с единицами иерархии, в первую очередь, с ближайшей, находящейся выше по иерархии, единицей.

Просмотр выбранной единицы как элемента подсети осуществляется в специальном режиме, который наглядно демонстрирует ее связь с соседними единицами иерархии, ее место по отношению к ним и единицам того же типа (рис. 17). Режим обеспечивает также переход между единицами иерархии, их визуализацию, редактирование, удаление и другие операции. В этом режиме осуществляется также ввод и редактирование значений единиц текста и рукописи.

Оба описанных режима (набор и редактирование, работа с иерархиями) связаны между собой возможностью позиционирования на одной и той же единице. Иначе говоря, выделение нужной единицы в одном из режимов представления рукописи позволяет найти эту же единицу в другом режиме представления.

Ввод и редактирование значений единиц осуществляется с помощью таблиц свойств единиц того или иного типа (рис. 18). Предварительная обработка текста позволяет представить пользователю те значения данной единицы, которые уже известны Системе. Пользователь может подтвердить, изменить или удалить значение.

Понятно, что эффективность работы пользователя с разрабатываемой Системой зависит от многих условий. Конечно, немаловажную роль при этом играют точность представления изучаемой области — рукописей и текстов — в виде модели баз данных, правильность структурирования лингвистической, текстологической, археографической и другой заносимой в базу или получаемой в ходе обработки текстов информации, понятность и удобность интерфейса, разнообразие предоставляемых пользователю вторичных данных — выборки, указателей, перечней, возможность сравнивать на разных основаниях имеющиеся в базе тексты, их фрагменты, единицы. Одновременно с этим для создания полноценного инструмента научных исследований, несомненно, необходимо наполнение модели текстовой информацией. Чем большее количество текстов имеется в распоряжении пользователя, тем более точными, достоверными, надежными и корректными будут его выводы. Поэтому особое внимание разработчиков уделяется обеспечению удаленного доступа к базе и дистанционной работе с ней. По сути, создаваемый специализированный редактор

и должен выполнять функции программы, обеспечивающей удаленный доступ к базе данных через Интернет. Это должно позволить ввести в Систему необходимое количество рукописей.

Наш рассказ о Системе Манускрипт, о ее компонентах, находящихся в настоящее время на разных стадиях разработки и завершенности, имеет непосредственное отношение к проблеме электронных изданий и к вопросам представления древнего рукописного наследия с помощью новых технологий.

Знакомство и работа с текстом рукописи, представленном в различной степени условности, развитый поиск по различным критериям, получение необходимых для научных исследований указателей, создание сравнительных указателей нескольких текстов или различных фрагментов одного, возможность наложения выборок (например, выявление их пересечения или разности), выявление разночтений нескольких единиц при наличии нормализованного образца — вот только небольшой перечень тех возможностей, которые могут быть реализованы с помощью Системы Манускрипт.

Приложение

Список работ по теме статьи

1. Баранов В.А. Автоматизированная обработка древнерусского текста (Цели и задачи, проблемы, пути решения) // Вестник УдГУ. № 6. Ижевск, 1994. С. 36–50.

2. Баранов В.А., Марков В.М., Князева О.А., Миронов А.Н., Никифорова С.А., Пуленков О.И. Древнерусские богослужебные тексты как источник изучения истории русского языка и культуры: подготовка к изданию Майской служебной Минееи XI в. (Путятиной) // Материалы Всероссийского конкурса научно-исследовательских проектов в области гуманитарных наук. Гуманитарная наука в России: Соросовские лауреаты. Москва, 1996. С. 96–107.

3. Баранов В.А. Заметки об особенностях состава и письма Путятиной Минееи XI в. // *Opera Slavica*, VI, 2, 1996. С. 17–28.

4. Баранов В.А. Палеографические особенности почерков Путятиной Минееи XI века // Вестник УдГУ. № 7. 1996. С. 31–64.

5. Баранов В.А. Лексикографическое описание древнерусских текстов (к вопросу о графико-орфографических и грамматических сведениях в автоматизированных словарях) // *Porovnávací opus statiky a dynamiky*

súčasného ruského jazyka z teoretického a praktického hľadiska. Zborník ruského jazyka a literatúry Filozofickej fakulty Univerzity J.A.Komenského v Bratislave. Bratislava, 1998. С. 13–17.

6. Баранов В.А., Никифорова С.А. Графико-орфографические особенности новгородских служебных миней XI-XIV вв. // Porovnávací prýs statiky a dynamiky súčasného ruského jazyka z teoretického a praktického hľadiska. Zborník ruského jazyka a literatúry Filozofickej fakulty Univerzity J.A.Komenského v Bratislave. Bratislava, 1998. С. 19–25.

7. Баранов В.А. Древнерусская лексикография: теория и практика // Opera Slavica, VIII, 2, 1998. С. 30–41.

8. Баранов В.А. К вопросу о требованиях к автоматизированным информационно-поисковым системам обработки древнерусских текстов (поиск информации) // Prouchavanje slovenskih jezika, knjizhevnosti i kultura u inoslovenskoj sredini. Beograd: Filoloski fakultet, Slavistichko drushtvo Srbije, 1998. С. 52–57.

9. Баранов В.А. К вопросу о соотношении почерков и инициалов в майской служебной Минее XI века // Тезисы докладов научно-практической конференции «Удмуртия накануне третьего тысячелетия». Ижевск, 26–27 марта 1998 г. С. 74–75.

10. Баранов В.А., Миронов А.Н., Миронов А.Л., Пуленков О.И., Рябова Е.В. Структура, состав и принципы заполнения компьютерной базы древнерусских текстов // Тезисы докладов 4-й Российской университетско-академической научно-практической конференции. Часть 3. 21–23 апреля 1999 г. Ижевск, 1999. С. 132–133.

11. Баранов В.А. Электронные издания древнерусских памятников и их соотношение с печатными // Тезисы докладов 4-й Российской университетско-академической научно-практической конференции. Часть 3. 21–23 апреля 1999 г. Ижевск, 1999. С. 133–134.

12. Баранов В.А. Отражение фонетических и фонематических изменений древнерусского языка в новгородских служебных Минеех XI-XIV вв. // Тезисы докладов 4-й Российской университетско-академической научно-практической конференции. Часть 3. 21–23 апреля 1999 г. Ижевск, 1999. С. 136–137.

13. Баранов В.А., Шарова А.В. Оформление Новгородской служебной миней XI в. (ГПБ, Соф. 202) // Тезисы докладов 4-й Российской университетско-академической научно-практической конференции. Ч. 3. 21–23 апреля 1999 г. Ижевск, 1999. С. 130–131.

14. Баранов В.А. Богослужебные тексты как источник для изучения исторической фонетики русского языка (На материале текстов XI–

XIV вв.) // *Príspevky ke slovanske filologii. Ústav slavistiky Filozofické fakulty Masarykovy univerzity. Brno, 1999. С. 9–20.*

15. Баранов В.А. Издание древнерусских памятников и некоторые вопросы лексикографии // *Príspevky k aktuálním otázkám jazykovédne rusistiky. Ústav slavistiky Filozofické fakulty Masarykovy univerzity, Brno 1999. С. 113–121.*

16. Баранов В.А., Вотинцев А.А., Зуга О.В., Миронов А.Л., Миронов А.Н., Никифорова С.А., Пуленков О.И., Романенко А.В., Рябова А.В., Шарова А.В. Древнейший русский письменный памятник — Новгородская служебная Миней за май (Путятина) в сети Интернет // Тезисы докладов Международной конференции «Интернет. Общество. Личность. ИОЛ-2000. Новые информационно-педагогические технологии», Санкт-Петербург, 28 февраля — 3 марта 2000 г. С. 91.

17. Баранов В.А., Вотинцев А.А., Зуга О.В., Миронов А.Л., Миронов А.Н., Никифорова С.А., Пуленков О.И., Романенко А.В., Рябова А.В., Шарова А.В. Электронное издание древнейшего русского письменного памятника — Новгородской служебной Миней за май (Путятиной Миней), 1-ая половина — сер. XI в. // Новые информационные ресурсы и технологии в исторических исследованиях и образовании. Сборник тезисов докладов и сообщений Всероссийской конференции. Подмоскowie, «Боровое», 6–9 апреля 2000 г. М., 2000. С. 31–32.

18. Баранов В.А., Вотинцев А.А., Зуга О.В., Миронов А.Л., Миронов А.Н., Никифорова С.А., Пуленков О.И., Романенко А.В., Рябова А.В. Полнотекстовые базы данных древнейших древнерусских рукописей // Новые информационные ресурсы и технологии в исторических исследованиях и образовании. Сборник тезисов докладов и сообщений Всероссийской конференции. Подмоскowie, «Боровое», 6–9 апреля 2000 г. М., 2000. С. 29–30.

19. Баранов В.А., Вотинцев А.А., Зуга О.В., Килина Л.Ф., Миронов А.Н., Никифорова С.А., Ощепков С.В., Романенко А.В., Рябова А.В. Адекватность филологических терминов и терминов моделирования данных в представлении лингвистических объектов // Тезисы докладов 5-й Российской университетско-академической научно-практической конференции. Ч. I. / Отв. ред. В.А.Журавлев, С.С.Савинский. Ижевск, 2001. С. 50–51.

20. Баранов В.А., Вотинцев А.А., Зуга О.В., Килина Л.Ф., Миронов А.Н., Никифорова С.А., Ощепков С.В., Романенко А.В., Рябова А.В. Текстовый редактор для обработки и исследования древних славянских текстов // Тезисы докладов 5-й Российской университетско-

академической научно-практической конференции. Ч. 1. / Отв. ред. В.А.Журавлев, С.С.Савинский. Ижевск, 2001. С. 53–54.

21. Баранов В.А., Зуга О.В., Никифорова С.А., Рябова А.В. Древнерусские письменные памятники и проектирование полнотекстовых баз данных // Материалы XIII конференции Ассоциации АИК. Санкт-Петербург, 26–29 июня 2002 г. Информационный бюллетень Ассоциации «История и компьютер». № 30. Специальный выпуск. М., 2002. С. 86–87.

22. Баранов В.А., Вотинцев А.А., Гнутиков Р.М., Зуга О.В., Мионов А.Н., Никифорова С.А., Ощепков С.В., Романенко А.В., Рябова А.В. Структура и функции информационно-поисковой системы «Манускрипт» // Материалы VIII конференции Ассоциации АИК. Санкт-Петербург, 26–29 июня 2002 г. Информационный бюллетень Ассоциации «История и компьютер». № 30. Специальный выпуск. М., 2002. С. 87–89.

23. Лихачев Д.С. Текстология. На материале русской литературы X–XVII веков. / Издание второе, перераб. и доп. Л.: Наука, 1983. 639 с.

Примечания

- ¹ Работы, о которых рассказывается в статье, осуществляются при поддержке РФФИ (гранты №№ 02-07-90318-в, 02-07-90424-в) и РГНФ (грант № 01-04-12027в).
- ² Хорошо известна мысль Д.С. Лихачева о том, что «изучение памятника должно предшествовать его изданию. Сперва изучить, а потом издать на основе этого изучения — таков принцип советской текстологии. Он диаметрально противоположен укоренившейся в старой, дореволюционной науке практике издавать памятники для их последующего изучения» (Лихачев, 486–487).
- ³ Показательно высказывание Д.С. Лихачева относительно реконструкции текстов также и в отношении выбора языковых средств, которые не свойственны языку памятника, для передачи издаваемого текста: «Невозможно смешивать в реконструкциях явления языка разных эпох, вкрапливать в язык позднейший, представляющий собой законченную систему определенной эпохи, элементы языка более раннего, соединять то и другое, соединять орфографические явления второго южнославянского влияния с более древними и т.п. Нельзя также реконструировать древний текст в формах языка более позднего. Последнее возможно только тогда, когда реконструируется не текст, а содержание исчезнувшего текста» (Лихачев, 472).
- ⁴ Например, о трудностях применения современной пунктуации для передачи древнего текста говорит Д.С. Лихачев: «Очень трудный вопрос — это

вопрос о расстановке знаков препинания. Прочтешь текст — это не только правильно разделить текст на слова, но и понять его синтаксически. Внешним выражением такого правильного синтаксического понимания древнего текста и является правильная расстановка в нем знаков препинания. Оставление текста без знаков препинания (полностью или частично) во многих случаях означает, что издатель не решается интерпретировать текст или что он его не понял. Серьезным препятствием к расстановке знаков препинания служит то обстоятельство, что современная расстановка знаков препинания приспособлена к современному же синтаксису, который сильно отличается от синтаксиса древнего. Применять же древнюю пунктуацию невозможно: она не систематична и совершенно не изучена... В ряде случаев текстолог просто не знает, как расставить знаки препинания, и либо в таком случае вовсе их не ставит, либо модернизирует текст» (Лихачев, 162–163; см. также далее).

- ⁵ См., например, замечание Д. С. Лихачева о необходимости специальных лингвистических знаний для работы с оригиналами рукописей: «Прочтение текста списка невозможно без знания двух специальных дисциплин: истории языка той эпохи или тех эпох, к которым относятся текст списка и текст самого произведения, и палеографии. Если знание последней дисциплины до известной степени признается всеми текстологами, то обстоятельное знание языка эпохи практически часто отсутствует у текстологов и порождает очень большое количество ошибок прочтения и установления текста в современных изданиях» (Лихачев, 155; далее см., например: 156, 161).
- ⁶ «В сущности, при любом использовании печатного текста древнего памятника в научных целях, необходимо проверять его теми же текстологическими приемами, что и при рукописном списке. Принципиальных различий нет, есть только различия, вызываемые спецификой самого материала. Печатные издания древнерусских памятников также должны изучаться текстологом, если возникает необходимость пользоваться их текстом, как и рукописные списки» (Лихачев, 456).
- ⁷ Требование необходимости обращения к факту рукописи, несмотря на его бесспорность (см., например, Лихачев, 1983, 170, 171, 173), в отношении работы с древним рукописным наследием не всегда соблюдается; в частности, в силу и объективных причин.
- ⁸ Оставляем в стороне вопросы, связанные с технологиями переноса на электронные носители оцифрованного изображения текста рукописей. Это лишь иная по сравнению, например, с фотокопией форма хранения изображения страниц рукописных памятников. Не касаемся здесь и технологии создания печатных изданий, созданных с использованием компьютера и его развитых возможностей набора и воспроизведения внешнего вида рукописи.
- ⁹ Путьятина Минеев на май / Подг. текста и параллели М. Ф. Мурьянова; ред., предисловие и комм. А. Б. Страхова // (1–9 мая) — *Palaeoslavica VI. Cambridge (Massachusetts)*, 1998. С. 114–208; (10–17 мая) — *Palaeoslavica VII. Cambridge*

- (Massachusetts), 1999. С. 136–217; (18–31 мая) — Palaeoslavica VIII. Cambridge (Massachusetts), 2000. С. 123–221.
- ¹⁰ Щеголева Л.И. Путятина Миняя (XI в.) в круге текстов и истолкования. 1–10 мая. М.: Территория, 2001. 496 с.
 - ¹¹ Сайт «Путятина Миняя» создан при финансовой поддержке «Института «Открытое общество» (Фонд содействия)» (грант № 1BE984).
 - ¹² Электронное издание содержит текст рукописи в различных видах, указатели слов и словоформ различного типа и вида, справочные материалы археологического, историко-филологического, комментирующего и библиографического характера, иллюстрации, помощь по работе с запросными формами, сведения об авторах и разработчиках.
 - ¹³ Текст рукописи передается лист в лист, столбец в столбец, строка в строку, буква в букву, без деления на слова. Заглавные и строчные буквы различаются. Вариантные буквы, все лигатуры и диграфы сохраняются. Надстрочные знаки, титла, небуквенные знаки в строке заменены на условные.
 - ¹⁴ Текст разделен на слова. В начале словоформ различаются заглавные и строчные буквы. Раскрытие титл, расшифровка сокращений, дополнение букв отмечается особыми знаками. Используются условные знаки восстановления и расшифровки текста, а также знаки [?] и [!].
 - ¹⁵ Текст разделен на слова. В начале словоформ различаются заглавные и строчные буквы. Отсутствуют надстрочные знаки, титла раскрыты. Отсутствуют варианты буквы. Словоформы нормализованы по правилам старославянской графики и орфографии.
 - ¹⁶ Текст дается в современной транслитерации.
 - ¹⁷ На рисунках 5–9 полная модель представления данных и ее фрагменты приведены в виде диаграмм «сущность-связь».
 - ¹⁸ Используемые понятия *сети* и *иерархии* соответствуют типичным понятиям, принятым в математической теории графов. Единицы представляют собой вершины графа.
 - ¹⁹ Используемые нами термины *уровень иерархии*, *тип единицы* и *единица* не являются аналогами языковых терминов *уровень языка*, *языковая единица* и *единица одного уровня*. С помощью первых описывается любая совокупность выделяемых в тексте или языке компонентов, обладающих идентичными свойствами (признаками), и их отношения с компонентами, которые имеют аналогичные свойства. Причем под *единицей* подразумевается конкретный компонент текста или языка; под *типом единицы* — абстрактное понятие, объединяющее конкретные единицы данного типа; под *уровнем* — понятие, объединяющее абстрактную единицу, ее конкретных представителей и противопоставленное объединению единиц другого уровня; под *иерархией* — совокупность единиц, находящихся в отношениях вложенности. Условно можно говорить о соответствии понятия *единица* понятию *конкретный представитель языковой единицы*, понятия *тип единицы* — понятиям *единица языка* и *единица речи*, понятия *ярус иерархии* — понятию *языковой*

- уровень, понятия *уровень иерархии* — совокупности соответствующих единиц языка и единиц речи.
- ²⁰ Вданном перечислении принят следующий принцип выделения: единицы — полужирным шрифтом, свойства — обычным, значения — курсивом.
- ²¹ См. предыдущую сноску.
- ²² С точки зрения моделируемых областей, в частности языковой, крайние единицы подсетей условно можно назвать высшей и низшей единицами. Это возможно хотя бы потому, что эти единицы в каждой из иерархий конкретной подсети действительно являются таковыми.
- ²³ ФСС — формообразующий состав словоформы; ФСНФ — формообразующий состав начальной формы; ССНФ — словообразующий состав начальной формы.
- ²⁴ Названия типов единиц (в частности, вспомогательных языковых единиц) являются условными и часто связаны с названиями наиболее важных в лингвистическом отношении единиц этого уровня, выделяемых в составе родителя.
- ²⁵ В понятие *компонент ПФ*, в отличие, например, от понятий *компонент простого предложения* и *компонент сложного предложения*, включаются все структурно-функциональные фрагменты предложения, которые можно выделить на основе их индивидуальной роли в его построении и оформлении, в первую очередь — в его письменной форме.
- ²⁶ Значением единиц одного уровня в ряде случаев является наименование функции той или иной единицы другого уровня. Так, разделитель частей предложения-фразы является одновременно знаком «тире», а отделитель предложения-фразы — знаком «точка» уровня *знаки языковой иерархии*.
- ²⁷ Универсальные принципы, заложенные в основу редактора, позволяют предположить, что он может быть использован не только в указанном узкоспециализированном направлении — для работы со славянскими рукописями, — но и как инструмент для обработки текстов, которые написаны на других языках и которые имеют сложную структуру и состав и обладают сложным набором свойств.