

Использование формата TEI для публикации и анализа списков произведений вопросно-ответного жанра¹

М. Г. Бабалык, А. Г. Варфоломеев, А. В. Пигин
Петрозаводский государственный университет

The article deals with the problems of textological study and publication of Russian medieval written documents in the form of questions and answers. The “Conversation of the Three Holy Hierarchs” spread in a great number of manuscript copies is cited as an example of such a text. The authors of the article suggest using the TEI format and provide a description of a possible option for TEI marking.

Использование формата TEI для публикации и анализа литературных памятников имеет давние традиции в литературоведении [Vanhouste, 2004]. Появившись в 80-е годы XX в., этот формат сначала был основан на принципах SGML, затем, начиная с 4-й версии (TEI P4), стал XML-форматом. С 2007 г. существует и продолжает совершенствоваться новая версия — TEI P5, учитывающая тенденции развития XML-технологий [Wittern et al, 2009]. Традиционно TEI имеет модульную структуру — кроме набора базовых элементов и атрибутов, имеется также много дополнительных наборов, ориентированных на конкретные жанры или задачи. На сайте TEI-Консорциума (www.tei-c.org) перечислено более сотни проектов, использующих TEI для различных литературоведческих и лингвистических целей, в частности, для представления корпусов письменной и разговорной речи на различных языках, электронных публикаций бумажных изданий, писем и рукописей, а также средневековых грамот.

Однако TEI позволяет ставить и другие задачи на основе таких жанров средневековой письменности, которые еще не при-

¹ Статья написана в рамках проекта (№ 10-04-12118в), поддержанного грантом Российского гуманитарного научного фонда.

влекались в подобных проектах. Так, существует обширная литература вопросно-ответного жанра (диалога), требующая текстологического изучения с помощью компьютерных технологий.

Жанр диалога (беседы, в том числе вопросно-ответной) является очень древним в мировой литературе. В древнерусской оригинальной (непереводной) литературе эта форма впервые была использована в Речи Философа в Повести временных лет. В такой форме написаны и многие апокрифы: Вопросы Иоанна Богослова Господу на горе Фаворской, Вопросы Иоанна Богослова Аврааму на Елеонской горе, Луцидариус, Вопросы о Тивериадском море и другие. В XVIII–XX вв. вопросно-ответная форма активно использовалась старообрядческими книжниками в полемических целях.

Текстологическое изучение вопросно-ответных произведений имеет свою специфику, которая определяется в первую очередь высокой степенью вариативности этих текстов. Разные списки одного вопросно-ответного памятника могут различаться как количеством вопросно-ответных пар, так и их последовательностью. Для сравнения списков друг с другом и для построения вероятных стемм списков необходимо применять компьютерные технологии. Более того, необходимо организовать работу сетевого научного сообщества и виртуальной среды для исследования списков, так как они находятся в самых разных рукописных хранилищах. Такие задачи требуют использования общепринятого, универсального формата для представления текстов, который был бы удобен для многократного текстологического анализа с помощью различных методик. Поэтому для нашего исследования мы выбрали TEI.

Объектом исследования стал популярный на Руси апокриф — Беседа трех святителей. Считается, что Беседа возникла на византийской почве приблизительно в V–VI вв. [Мочульский, 1893]. Ранняя история этого текста достаточно хорошо изучена [Мочульский, 1893; Nachtigal, 1901–1902; Бучилина, 1994; Miltenova, 2004]. Вошедшая в русскую письменность, Беседа стала активно переписываться русскими книжниками, которые вносили в нее новый материал. Именно поэтому в науке высказывалось мнение, что русские списки Беседы следует изучать отдельно от греческих, латинских и южно-славянских.

В переводе на славянский язык Беседа известна уже в XI в. [Рождественская, 1980]. Самые ранние русские списки датируются XV в., до этого времени на Руси Беседа известна лишь в выписках [Рождественская, 2003].

Памятник построен в форме вопросов и ответов, изложенных от имени трех святителей-богословов IV в. Василия Великого, Григория Богослова и Иоанна Златоуста. Такая атрибуция является ложной. Использование этих имен в заглавии апокрифа должно было сделать произведение авторитетным и достоверным. Основное содержание апокрифа — библейская история, изложенная подчас с элементами еретической (богомильской) интерпретации.

На данный момент мы изучили около 80 списков Беседы XV–XX вв. из хранилищ Москвы, Санкт-Петербурга, Нижнего Новгорода, Вологды, Петрозаводска, Киева. Многие списки вышли из-под пера старообрядческих книжников. Встречаются в нашей подборке и списки из монастырских собраний.

Число вопросов и ответов в списках колеблется от 1-й вопросно-ответной пары до 130 (в большинстве списков — около 20–50). Иногда в состав одной и той же рукописи входят сразу несколько списков Беседы. Фрагменты из Беседы встречаются также в составе более крупных вопросно-ответных компиляций.

Учитывая особенности списков Беседы, мы предлагаем следующий вариант TEI-разметки. Основная структурная единица — пара «вопрос-ответ» — является содержимым элемента `<p>` с атрибутами `@n` (номер) и `@type` (идентификатор типа, присваиваемый исследователем). Внутри `<p>` помещаются элементы `<div type="q">` и `<div type="a">` для текстов вопроса и ответа соответственно. Последовательность вопросно-ответных пар составляет содержание элемента `<body>`. Кроме этого, список Беседы обычно имеет название, которое помещается в элемент `<head>`, находящийся внутри элемента `<front>`. Пара из элементов `<front>` и `<body>` составляет содержание элемента `<text>`, представляющего собственно текст списка. Переход текста на новый лист отмечается пустым элементом `<pb/>` с атрибутами `@n` (номер листа) и `@type` со значениями “a” или “r” (аверс или реверс). Метаинформация об источнике помещается отдельно в элементе `<teiHeader>`.

Если в одной рукописи встречается несколько списков Беседы, приходится использовать элемент `<group>`, позволяющий группировать тексты в коллекции, при этом элемент `<teiHeader>` относится ко всей группе.

Сравнение списков для построения стемм в проектируемой информационной системе (виртуальной среде исследований) предполагается нами следующим образом. Во-первых, тексты списков, размеченные в соответствии с вышеизложенными правилами, будут преобразованы в последовательности типов вопросно-ответных пар, то есть представлены фактически строками символов, в которых однотипные пары обозначены одинаковыми символами. Таким образом, сравнение текстов по набору и последовательности вопросно-ответных пар можно будет производить с помощью алгоритмов сравнения строк. Во-вторых, различие в текстах однотипных вопросов и ответов также может быть легко учтено, так как исходные тексты будут присутствовать в системе наряду с формализованными представлениями списков благодаря гибкости XML-формата.

Список литературы

- Бучилина, 1994 — Бучилина Е. А. Апокриф «Беседа трех святителей» как памятник средневековой русской литературы. Автореф. дис. ... канд. филол. наук. М., 1994.
- Мочульский, 1893 — Мочульский В. Н. Следы народной Библии в славянской и древнерусской письменности. Одесса, 1893.
- Рождественская, 1980 — Рождественская М. В. Беседа трех святителей (комментарии к тексту) // Памятники литературы Древней Руси. XII в. М., 1980. С. 649.
- Рождественская, 2003 — Рождественская М. В. Новгородская берестяная грамота № 10 — «апокрифическая загадка»? // Берестяные грамоты: 50 лет открытия и изучения. Материалы междунар. конф. (Великий Новгород, 24–27 сентября 2001). М., 2003. С. 310–320.
- Miltenova, 2004 — Miltenova A. Erotapokriseis. Съчиненията от кратки въпроси и отговори в старобългарската литература. София, 2004.
- Nachtigal, 1901–1902 — Nachtigal R. Ein Beitrag zu den Forschungen über die sogenannte «Беседа трех святителей» (Gespräch dreier Heiligen) // Archiv für slavische Philologie. Berlin, 1901. Bd. 23. S. 1–95; 1902. Bd. 24. S. 321–408.
- Vanhoutte, 2004 — Vanhoutte E. An Introduction to the TEI and the TEI Consortium // Literary and Linguistic Computing. 2004. Vol. 19. No. 1. P. 9–16.
- Wittem at al, 2009 — Wittem C., Ciula A., Tuohy C. The making of TEI P5 // Literary and Linguistic Computing. 2009. Vol. 24. No. 3. P. 281–296.