***В. А. Баранов***

**Параллельный корпус славянских списков паримейника: материал и постановка задачи[[1]](#footnote-1)**

**Аннотация.** Дается общая характеристика четырех рукописей славянского паримейника, транскрипции которых размечаются для демонстрации параллельных чтений. Описываются возможности модуля параллельных корпусов системы «Манускрипт» (manuscripts.ru). Показано несоответствие модели базы данных корпуса и состава богослужебных циклов паримий и предложено решение проблемы. Сформулированы требования к дополнительным формам демонстрации материалов корпуса для сопоставления текстологических особенности рукописей.

**Ключевые слова.** Параллельный корпус, средневековые славянские рукописи, паримейник

**1. Параллельный корпус как способ демонстрации списков средневекового текста**

Исторический корпус «Манускрипт» (manuscripts.ru), содержащий транскрипции средневековых славянских кодексов и отрывков X–XV вв., предоставляет пользователям различные средства просмотра (близкие к оригинальным и транслитерированные современной кириллицей или латиницей тексты, прямые и обратные формо- и словоуказатели, однотекстовые и многотекстовые конкордансы и др.) и анализа (количественные и статистические сведения о лингвистических единицах и их сочетаемости) данных.

Одним из способов визуализации некоторых транскрипций корпуса является демонстрация соответствующих друг другу фрагментов списков одного текста, например, Евангелий[[2]](#footnote-2), майских служебных миней[[3]](#footnote-3), летописей[[4]](#footnote-4). Работа с параллельными корпусами (поиск и выбор рукописей, режимы демонстрации, фильтры) обеспечивается специализированным модулем, текстовыми данными для которого являются а) размеченные на фрагменты транскрипции, б) словари фрагментов, единицы которых выступают инвариантами соответствующих текстовых единиц, в) параметры и значения словарных единиц [Баранов и др. 2008; Баранов и др. 2010; Аникина и др. 2012a; Аникина и др. 2012b].

Интерфейс модуля позволяет выбрать списки текста, указать рукопись, по которой будут выравниваться фрагменты, указать диапазон листов этой рукописи или фрагменты для демонстрации, ввести маску искомой лингвистической единицы, определить порядок следования списков, а также форму визуализации фрагментов – *полный текст*, *инципит* <…> *эксплицит* или только *заголовки*. Порядок следования фрагментов зависит от выбранной формы: при визуализации текстов соответствующие друг другу части располагаются одна рядом с другой, при визуализации заголовков – следование фрагментов соответствует их следованию в каждом из списков, а соответствующие друг другу части соединяются линиями, что позволяет увидеть различия в расположении фрагментов в рукописях.

Несмотря на неоднократную апробацию модели базы данных, процедур поиска и демонстрации соответствующих друг другу фрагментов, подготовка нового корпуса требует доработки и настройки менеджера в соответствии с особенностями текстов.

**2. Параллельный корпус славянского паримейника**

***2.1. Рукописи***

План работ по проекту «Средневековые тексты в современном контексте (новые методы и принципы представления средневековых текстов сегодняшним пользователям)» (рук. проф. О. Ф. Жолобов) предусматривает, в частности, создание параллельного корпуса славянских списков паримейника, богослужебной книги, содержащей тексты из Ветхого и Нового Заветов, которые читаются на вечерне в дни праздников, в дни Великого Поста и Страстной недели и др.

В качестве источников для корпуса выбраны четыре древнейших славянских кодекса XII, XIII и XIV вв. – Лазаревский (РГАДА, ф. 381, № 50), Захариинский (РНБ, Q.п.I.13), Федоровский (РГАДА, ф. 381, № 60) списки, а также рукопись РГБ, Тр. 4.

Рукописи содержат паримии Рождественско-Богоявленского (Рождество, Богоявление, паримии Водосвятия), Триодного (Великий пост, Сырная неделя, Лазарева суббота, Страстная неделя, Пентекостарий), месяцесловного циклов [Алексеев 2008: 234–245]. Списки имеют утраты, набор паримий и входящих в них чтений в ряде случаев не совпадают.

***2.2. Модель. Разметка. Словарь***

Для создания параллельного корпуса необходимы разметка транскрипций на паримии и соотнесение соответствующих друг другу текстов различных списков друг с другом.

Модель базы данных позволяет создать словарь фрагментов необходимого типа, а в каждом тексте – иерархию *текст – фрагмент – символ* и установить связь между словарными и текстовыми единицами. Каждый фрагмент может иметь характеристики (параметры) и значения характеристик, позволяющие организовать поиск и визуализацию единиц с помощью веб-форм корпуса.

Использование словарных и текстовых единиц для разметки транскрипций паримейников на паримии предполагает создание такого словаря паримий, в котором каждая единица уникальна и имеет идентифицирующие ее характеристики.

***2.3. Проблема: традиция vs модель***

Традиционно в печатных изданиях греческих профитологиев (Prophetologion) паримии имеют сквозную нумерацию, при которой каждая паримия получает номер, соотносящий ее с циклом или его частью и местом в цикле (или в части цикла). Так, восемь паримий, читающихся на Рождество, получили, например, в издании [*Höeg* et al. 1939, 1980–1981: 600] номер с L1a по L1h, а паримия, читающаяся на вечерне Страстного понедельника, – L36b и т.д. При предложенной системе нумерации паримии, содержащие один и тот же текст, например, стихи с 1-го по 13-й первой главы Бытия (Быт. 1.1–13), получают различные номера (в указанном издании это номера L1a, L2a, L5b, L41b).

Использующийся в печатных изданиях принцип сквозной нумерации паримий противоречит модели базы данных корпуса, так как предполагает соотносить одну и ту же паримию, читаемую в различных циклах, с различными словарными единицами.

Кроме того, использование некоторого условного номера, пусть и соотносимого с циклом или его частью, не является интуитивно понятной «меткой» паримии, которая могла бы использоваться как при разметке корпуса (ср., например, «Быт. 1.1» Библии), так и при поиске фрагментов и их визуализации.

Индивидуальным значением паримии могла бы стать «метка», содержащая сведения о чтениях Библии, составляющих ее текст. Но в связи с тем, что в большом количестве случаев набор стихов, глав (а иногда и книг) паримии достаточно велик (например, вторая паримия на Богоявление содержит тексты Исх. 14:15–18, 21–23, 27–29, паримия на Сретение, 2 февраля – тексты Исх. 12:51–13:2, 10–12, 14–16, 22:28; Лев. 12:2–4, 6, 8; Числ. 8:16–17 и др.), то использование таких значений в качестве основной индивидуализирующей характеристики не представляется возможным.

***2.4. Решение: традиция & модель***

В настоящее время в качестве рабочего варианта выбрана комбинированная характеристика паримий словаря, которая включает а) тип фрагмента, б) индекс, указывающий на цикл (его часть) и литерный порядковый номер в цикле (его части), а также в) комментарий, содержащий сведения о книге(ах), главе(ах), стихах текстов паримии и г) тему паримии[[5]](#footnote-5):

Тип функционально-структурного раздела: паримия

Номер: P1b[[6]](#footnote-6)

Комментарий: Чис. 24:2–3, 5–9, 17–18 «Пророчество Валаама».

При такой идентификации отдельной единицей словаря становится паримия, текст которой читается на определенной службе и содержит текст, соотносимый с соответствующими стихами Библии. Списки содержат также паримии, которые читаются в разное время и состав стихов в которых идентичен: Быт. 1:1–13 (3х – на Рождество, на Богоявление, на вечерне Понедельника 1-й недели Великого поста), Иез. 43:27–44.4 (4х – на 8 сентября, 21 ноября, 25 марта, 15 августа), Ис. 61:10–62.5 (2х – на вечерне Страстной субботы и на 11 мая) и др. (всего – 28 паримий). В полутора десятках случаев пересечение частичное: Быт. 1:1–13 (паримия на Рождество и др.) и Быт. 1:1–5 (паримия на вечерне Страстной субботы), Ис. 9:5–6 (паримия на Рождество) и Ис. 8:13–9.6 (паримия на 6-м часе Понедельника 3-й недели Великого поста) и др.[[7]](#footnote-7) Указание на повторяющуюся паримию осуществляется с помощью значения, включающей несколько номеров, например: паримия Быт. 1:1–13 – на Рождество P1a = P2a = P5b, на Богоявление P2a = P1a = P5b, на Вечерне понедельника 1‑й недели Великого поста P5b = P2a = P1a.

Наличие инвариантного (принятого в качестве нормирующего) набора чтений для каждой словарной паримии позволяет описать текстологические особенности паримий в текстах. Отличия могут касаться количества входящих в паримию стихов и порядка их следования. Так, одна из паримий на Вознесение, содержащая, согласно [Алексеев 2008: 240], чтения Зах. 14:1–11, в паримейнике Тр. 4 включает чтения Зах. 14:1–4, 8–11 (110г – 111б), а в Захариинском – Зах. 14:1–3, 5, 4, 8–11 (л. 219в). Расхождения между ожидаемым и текстовым составом паримий связаны также с утратами части листов рукописей.

***2.5. Автоматизация сопоставления. Разметка на стихи***

Автоматизированный анализ текстологических особенностей списков предполагает предоставление пользователю возможности поиска и сопоставительной визуализации не только самих паримий, но и их состава, что возможно при наличии разметки как границ паримий, так и границ фрагментов, соответствующих стихам Библии, внутри них.

Помимо разметки списков на паримии, в рамках работ по проекту осуществляется разметка паримий на фрагменты, соответствующие стихам Ветхого и Нового Заветов. Работа выполняется с помощью словаря стихов Библии, связь контекстов рукописей с которыми обеспечивает поиск и демонстрацию параллельных стихов.

Особенностью процедуры разметки текстов на стихи является частое лексическое, грамматическое и синтаксическое своеобразие текстов славянских паримий по сравнению с соответствующими стихами славянской Библии. Следствием этого является сложность соотнесения в ряде случаев границ стихов Библии и границ соответствующих фрагментов паримий.

***2.6. Визуализация корпуса***

Наличие в корпусе процедур поиска и демонстрации соответствующих друг другу фрагментов разных списков, содержащих одни и те же тексты, позволяет уже на этапе разметки получить первый вариант параллельного корпуса четырех списков славянского паримейника[[8]](#footnote-8). Основным отличием визуализации транскрипций будет выравнивание списков на основе одновременно обеих разметок – разметки на паримии и вложенной разметки на стихи. Другим необходимым компонентом вывода параллельных контекстов должно стать упорядочение паримий не только по их следованию в одном из списков, но и по их следованию внутри богослужебных циклов. Третьим необходимым условием визуализации является демонстрация отсутствия паримий и/или стихов в том или ином списке.

**3. Заключение**

Любая прикладная разработка, связанная с представлением средневековых кодексов с помощью компьютерных технологий, безусловно, должна продолжать традиции печатных публикаций и одновременно обладать такими инструментами поиска и визуализации данных, которые обеспечивают многовходовость, настраиваемость, группировку, фильтрацию и другие возможности работы с материалом. При решении задач разметки данных может возникнуть ситуация противоречия между составом и структурой текста, традицией их представления и моделью, положенной в основу компьютерной системы. Предложенная система идентификации паримий позволяет, с одной стороны, сохранить связь с принятой в православной литургике их нумерацией, а с другой – однозначно идентифицировать их при работе с параллельным корпусом.

**Источники**

Паримейник (Лазаревский, или Сковородкинский), нач. XII в., РГАДА, ф. 381, оп. 1, ед. хр. 50, 126 л.

Паримейник (Захариинский), 1271 г., РНБ, Q.п.I.13, 264 л.

Паримейник (Федоровский II), втор. пол. XIII в., РГАДА, ф. 381, № 60, 109 л.

Паримейник, 2-й пол. XIV в., РГБ, Тр. 4, 142 л.

**Литература**

Алексеев А. А. (2008), Библия в богослужении. Византийско-славянский лекционарий. СПб.

Аникина Р. А., Баранов В. А. (2012a), Параллельный корпус русских летописей XIII–XV вв. в Интернете: инструментарий и методика лингвотекстологического анализа средневекового текста. *Интеллектуальные системы в производстве*, 2 (20), с. 157–162.

Аникина Р. А., Баранов В. А. (2012b), Параллельный корпус русских летописей в Интернете: цели, задачи, технологическая основа, использование. *Информационные технологии и письменное наследие: материалы IV междунар. науч. конф.*, с. 12–18.

Баранов В. А., Гнутиков Р. М. (2008), Электронное критическое издание средневекового текста: постановка задачи, основные требования и инструментальная подготовка. *Современные информационные технологии и письменное наследие: от древних текстов к электронным библиотекам: материалы междунар. науч. конф.*, с. 36–44.

Баранов В. А., Дубовцев С. В. (2010), Электронное критическое издание средневекового славянского текста: модель данных и визуализация лингвистических единиц. *Интеллектуальные системы в производстве*, 1, с. 280–287.

Höeg C., Lake S., Zuntz G., Engberg G. (1939, 1980–1981), Monumenta Musicae Byzantinae. Lectionaria. Vol. I. Prophetologium. Pars. I. Lectiones Nativitatis et Epiphaniae; Vol. I. Prophetologium. Pars. II. Lectiones anni Immobilis. Copenhagen.

**References**

Alekseev A. A. (2008), Biblija v bogosluzhenii. Vizantijsko-slavjanskij lekcionarij. Saint Petersburg.

Anikina R. A., Baranov V. A. (2012a), Parallel'nyj korpus russkih letopisej XIII–XV vv. v Internete: instrumentarij i metodika lingvotekstologicheskogo analiza srednevekovogo teksta. *Intellektual'nye sistemy v proizvodstve*, 2 (20), pp. 157–162.

Anikina R. A., Baranov V. A. (2012b), Parallel'nyj korpus russkih letopisej v Internete: celi, zadachi, tehnologicheskaja osnova, ispol'zovanie. *Informacionnye tehnologii i pis'mennoe nasledie: materialy IV mezhdunar. nauch. konf.*, pp. 12–18.

Baranov V. A., Gnutikov R. M. (2008), Jelektronnoe kriticheskoe izdanie srednevekovogo teksta: postanovka zadachi, osnovnye trebovanija i instrumental'naja podgotovka. *Sovremennye informacionnye tehnologii i pis'mennoe nasledie: ot drevnih tekstov k jelektronnym bibliotekam: materialy mezhdunar. nauch. konf.*, pp. 36–44.

Baranov V. A., Dubovcev S. V. (2010), Jelektronnoe kriticheskoe izdanie srednevekovogo slavjanskogo teksta: model' dannyh i vizualizacija lingvisticheskih edinic. *Intellektual'nye sistemy v proizvodstve*, 1, pp. 280–287.

Höeg C., Lake S., Zuntz G., Engberg G. (1939, 1980–1981), Monumenta Musicae Byzantinae. Lectionaria. Vol. I. Prophetologium. Pars. I. Lectiones Nativitatis et Epiphaniae; Vol. I. Prophetologium. Pars. II. Lectiones anni Immobilis. Copenhagen.

**Parallel Text Corpus of Slavic Manuscripts   
of Prophetologium: Material and Statement of the Problem**

V. A. Baranov

**Abstract.** The characteristics of the four manuscripts of the Slavic prophеtologium, which are marked up to demonstrate parallel readings, are presented. The capabilities of the parallel corpora module of the "Manuscript" system (manuscripts.ru) are described. The discrepancy between the model of the corpus database and the composition of the liturgical cycles of the parimias is shown, and a solution to the problem is proposed. The task of pointing out the repeating parimias and simultaneously preserving their continuous numbering was solved using a special markup. Requirements for additional forms of visualization of parallel readings for the comparison of textological features of manuscripts are formulated.

**Keywords.** Parallel text corpus, Slavic medieval manuscripts, prophetologium

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Баранов Виктор Аркадьевич**

Ижевский государственный технический университет им. М. Т. Калашникова (Россия).

**Baranov Victor Arkadievich**

Kalashnikov Izhevsk State Technical University (Russia).

***E-mail: victor.a.baranov@gmail.com***

1. Работа выполняется при поддержке Российского фонда фундаментальных исследований (РФФИ), проект № 20-512-18001. [↑](#footnote-ref-1)
2. http://manuscripts.ru/mns/cred.cred?koll=61092969&f\_type=12014 [↑](#footnote-ref-2)
3. http://manuscripts.ru/mns/cred.cred [↑](#footnote-ref-3)
4. http://manuscripts.ru/mns/cred.cred?koll=62133570&f\_type=14001 [↑](#footnote-ref-4)
5. См. [Алексеев 2008: 234–245]. [↑](#footnote-ref-5)
6. Где P – паримия, 1 – паримия на Рождество, b – 2-я паримия службы на Рождество. [↑](#footnote-ref-6)
7. Сопоставление паримий по степени совпадения входящих в них чтений имеет особое значение при анализе содержательной стороны паримий, так как «сходные по содержанию праздники имеют тождественный набор паримий» [Алексеев 2008: 167]. [↑](#footnote-ref-7)
8. http://manuscripts.ru/mns/cred.cred?koll=94052725&f\_type=12014 [↑](#footnote-ref-8)