



DOI: <https://doi.org/10.15688/jvolsu2.2023.6.1>

UDC 811.163'04
LBC 81.416-03

Submitted: 19.07.2023
Accepted: 09.10.2023

CYRIL-METHODIAN AND EASTERN BULGARIAN WORDS IN THE MANUSCRIPTS OF THE 10th – 15th CENTURIES (TEXT CORPUS STUDY)¹

Victor A. Baranov

Kalashnikov Izhevsk State Technical University, Izhevsk, Russia;
Kazan Federal University, Kazan, Russia

Abstract. The correlation of statistical characteristics of the so-called Cyrillo-Methodian and Eastern Bulgarian words in groups of texts characterized by different textological and(or) codicological meanings is presented: Glagolitic – Cyrillic, service – non-service, archaic – Eastern Bulgarian subcorpora. Synonymous pairs of *vrētishche – vlasēnitsa* ‘rough (horsehair) clothes’; *zhrtva – trēba* ‘sacrifice’; *radi – dēlya* ‘because of, due to, on account of, for’; *тъкътo – тъchiyu* ‘only, just, merely’; *vrat'nikъ – vratar* ‘gatekeeper, doorkeeper’; *outro – zautra* ‘(early) in the morning’; *yako – aky* ‘how, as, like’; *aminъ – parvo* ‘rightly’; *aromatъ – vonya* ‘(fragrant) spices’; *iyudēi – zhidъ* ‘Jew’ are analyzed. The method of comparing the statistical meaning of the word observed in the subcorpora with the expected meaning is applied. The statistics measures Log-Likelihood, TF*ICTF and Weirdness were used. The components of synonymic pairs were extracted from subcorpora and evaluated using the historical corpus statistics module. Comparison of the statistical preference of the components of synonymic pairs in different subcorpora made it possible (a) to confirm the known confinement of each of the components to archaic and Eastern Bulgarian texts opposed to each other, (b) to show a different ratio of the components of pairs in different subcorpora, and also (c) to draw conclusions about the dependence of the preference of components on the lexical and lexical-derivational characteristics of lexemes.

Key words: Cyrillic-Methodian words, Eastern Bulgarian words, synonymic pairs, linguistic statistics, text corpus.

Citation. Baranov V.A. Cyril-Methodian and Eastern Bulgarian Words in the Manuscripts of the 10th – 15th Centuries (Text Corpus Study). *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 2. Yazykoznanie* [Science Journal of Volgograd State University. Linguistics], 2023, vol. 22, no. 6, pp. 5-20. (in Russian). DOI: <https://doi.org/10.15688/jvolsu2.2023.6.1>

УДК 811.163'04
ББК 81.416-03

Дата поступления статьи: 19.07.2023
Дата принятия статьи: 09.10.2023

КИРИЛЛО-МЕФОДИЕВСКАЯ И ВОСТОЧНОБОЛГАРСКАЯ ЛЕКСИКА В РУКОПИСЯХ X–XV ВВ. (КОРПУСНОЕ ИССЛЕДОВАНИЕ)¹

Виктор Аркадьевич Баранов

Ижевский государственный технический университет им. М.Т. Калашникова, г. Ижевск, Россия;
Казанский (Приволжский) федеральный университет, г. Казань, Россия

Аннотация. Представлено соотношение статистических характеристик так называемой кирилло-мефодиевской и восточноболгарской лексики в группах текстов, характеризующихся различными текстологическими и / или кодикологическими значениями: глаголический – кириллический, служебный – неслужеб-

ный, архаичный – восточноболгарский. Рассмотрены синонимические пары *врѣтище – власѣница*; *жрътва – трѣба*; *ради – дѣля*; *тъкъмо – тъчию*; *вратъникъ – вратарь*; *оутро – заутра*; *яко – акы*; *аминь – право*; *ароматъ – воня*; *шюдѣи – жидь*. Применен метод сопоставления наблюдаемого в подкорпусе статистического значения слова с ожидаемым. В качестве статистик использованы меры Log-Likelihood, TF*ICF и Weirness. Компоненты синонимических пар извлечены из подкорпусов и оценены с помощью модуля статистики исторического корпуса. Сопоставление статистической предпочтительности компонентов синонимических пар в разных подкорпусах позволило подтвердить известную приуроченность каждого из компонентов противопоставленным друг другу архаичным и преславским текстам, показать различия в соотношении компонентов пар в разных подкорпусах и сделать выводы о зависимости предпочтительности компонентов от лексических и лексико-словообразовательных характеристик лексем.

Ключевые слова: кирилло-мефодиевская лексика, восточноболгарская лексика, синонимические пары, лингвистическая статистика, текстовый корпус.

Цитирование. Баранов В. А. Кирилло-мефодиевская и восточноболгарская лексика в рукописях X–XV вв. (корпусное исследование) // Вестник Волгоградского государственного университета. Серия 2, Языкознание. – 2023. – Т. 22, № 6. – С. 5–20. – DOI: <https://doi.org/10.15688/jvolsu2.2023.6.1>

Введение

Традиционно в историко-лингвистических исследованиях при анализе материала письменных источников используют количественные сведения об анализируемых фактах, которые приводятся в абсолютных или относительных величинах или описательно: «присутствует – отсутствует», «больше – меньше», «чаще – реже», «значительно чаще – значительно реже», «преобладает – единично» и под. Однако подобного рода количественные сопоставления не учитывают, например, существенного фактора случайности отсутствия либо единичной представленности языковой единицы в некотором документе (группе документов) или отсутствия сведений о значимости (незначимости) обнаруженного различия в частоте употребления одной и той же единицы в нескольких документах. Чрезвычайно критичным отсутствие сведений о значимости (незначимости) выявленных различий становится в том случае, когда нет возможности проверить выводы на большом объеме текстового материала: например, количество рукописей начала славянской письменности, которые сохранились до настоящего времени, крайне ограничено.

Вопрос о кирилло-мефодиевских и преславских лексических элементах в древнеболгарской письменности

Одним из наиболее активно обсуждающихся на протяжении более чем ста пятидесяти лет вопросов из области кирилло-мефодиевской проблематики является вопрос о

лексической маркированности кирилло-мефодиевских текстов, с одной стороны, и преславских – с другой [Димитрова, 2016; Добрев, 1979; Желязкова, 2016а; 2016б; Илиев, 2016; Милтенов, 2008; 2020; Новикова, 2013; Славова, 1989; 2013; Станков, 2018; Шафарик, 1860/1861; Jagić, 1913; Voss, 1996; и др.] (историографию вопроса см., например, в: [Станков, 2018, с. 121–140]). Изучение древнеболгарских глаголических и кириллических рукописей показало, что переводчики, редакторы, правщики, переписчики использовали во многом разные графико-орфографические и языковые средства для создания текста. Отчетливые различия обнаружены между наиболее древними (кирилло-мефодиевскими / охридско-моравскими) или восходящими к ним письменными памятниками и рукописями, созданными в период Первого Болгарского царства. Наиболее ярко эти различия проявляются в лексике: при наличии в древнеболгарских рукописях синонимов в конкретном списке наблюдается предпочтение одних лексем другим. В работах представлены перечни таких синонимических рядов, извлеченных из различных текстов и рукописей (см., например: [Славова, 1989, с. 25–117; Станков, 2018, с. 141–153; Voss, 1996, S. 100–103]), определен круг памятников, которые ориентированы на различные переводческие и редакторские традиции, и соответственно в которых предпочитается кирилло-мефодиевский или преславский компонент синонимического ряда, выявлены греческие соответствия славянских лексем, предложены различные гипотезы о историко-культурных и языковых связях и отно-

шениях двух болгарских книжных центров в период второй половины IX – начала XI века.

Исследователями по-разному понимается природа лексических различий: от утверждений о сознательно, последовательно и системно осуществляемой в Преславской книжной школе редакционной работе над древнейшими переводами и работе над новыми (см., например: [Милтенев, 2008, с. 43; Славова, 1989, с. 19]) до категорического отрицания существования особых преславских лексем: «Вывод ясен, преславской редакции как таковой нет. Это означает, что нет и преславской лексики» [Станков, 2018, с. 137], «...“преславская” лексика просто не существует и что так называемые “преславизмы” не могут быть лексическими маркерами в поисках текстов преславского происхождения» [Станков, 2018, с. 139].

Болгарский исследователь Татьяна Славова указывает на наличие трех типов соотношений в синонимических рядах: оба компонента являются славянскими, славянское слово соответствует непереуведенному греческому, компоненты являются словообразовательными синонимами [Славова, 1989, с. 117], Явор Милтенев к этому ряду добавляет компоненты-словосочетания [Милтенев, 2008, с. 42].

В работах последних лет медиевисты приходят к выводу о динамичности и неоднородности возникновения, существования явления, которое традиционно называют языковыми преславизмами, подчеркивают, что основой переводческого и редакторского подхода к тексту в Восточной Болгарии являлся выбор предпочитаемых языковых средств из имеющихся – лексических синонимов и грамматических вариантов, последовательное использование которых делает их маркированными [Милтенев, 2020, с. 60, 63, 65].

Другое направление исследований – выявление текстологических различий между первоначальными и преславскими переводами: исследователями показано, что в Восточной Болгарии были созданы особого типа тексты – толковые и четьи, расширившие корпус первоначальных служебных, указываются дошедшие до наших дней рукописи, сохраняющие эту редакцию [Милтенев, 2008, с. 42; Славова, 1989, с. 120–121; Станков, 2018, с. 139]. При этом ученые все чаще указывают на необходимость более активного использования лексического

критерия для установления близости рукописей и текстов, созданных в рамках одной переводческой или редакторской традиции [Милтенев, 2008, с. 48]. В работах А.А. Пичхадзе показано, что отход от традиционного, «атомистического» подхода и применение сопоставительных приемов исследования лексических особенностей группы рукописей открывает новые перспективы в изучении письменного наследия [Пичхадзе, 2011, с. 17, 53, 54].

Кирилло-мефодиевские и восточноболгарские языковые особенности в письменных памятниках Руси

Известно, что начало церковнославянской книжности на Руси было положено письменными памятниками, переписанными с южноболгарских оригиналов: «Подавляющее большинство произведений переводной древнеславянской письменности возникло у южных славян. Оригинальная письменность восточных славян на церковнославянском языке с самого начала в той или иной степени подражала южнославянским образцам, и проникновение в текст специфически восточнославянских языковых элементов допускалось в минимальной степени» [Пичхадзе, 2011, с. 7].

Это наследие неоднородно. Оно включает рукописи, восходящие к кирилло-мефодиевским переводам богослужебных текстов, свидетельствующих «о преемстве книжных элементов языковой нормы, формирование которой началось одновременно с первыми переводами» [Максимович, 2000, с. 73], и рукописи Преславской книжной школы – ее писцы осуществили редактуру первоначальных переводов и сделали новые переводы [Максимович, 2000, с. 75].

Исследователями отмечается, что при возможности выбрать из нескольких синонимичных лексем русский книжник ориентируется на восточноболгарский образец: «Однако по всем этим позициям словоупотребление преславских памятников совпадает в древнерусском узусом» [Пичхадзе, 2011, с. 54]; и далее: «Восточноболгарская книжность, чрезвычайно популярная на Руси, своим авторитетом поддерживала включение в текст древнерусских элементов, чуждых кирилло-мефодиевскому лексикону» [Пичхадзе, 2011, с. 56].

Таким образом, разнообразие существующих выводов о соотношении кирилло-мефодиевского и симеоновского наследия, различные интерпретации одних и тех же текстологических и лингвистических данных свидетельствуют о сложности изучения истории ранней южнославянской письменной традиции и ее противоречивой судьбы на Руси.

Неоднократно указывалось, что для понимания соотношения западно- и восточноболгарских языковых особенностей, их судьбы в восточнославянской книжной традиции недостаточно исследований отдельных рукописей или использования созданных к настоящему времени лексикографических работ: необходим анализ широкого круга рукописей, анализ лексики в ее реальном, текстовом употреблении (с учетом сведений о происхождении и истории текстов), и не только с учетом ее наличия в тех или иных списках, но и отсутствия [Милтенев, 2020, с. 67, 68, 70].

Цель работы – выявить посредством количественно-статистического анализа соотношение представленности характеризующихся различными текстологическими и/или кодикологическими значениями кирилло-мефодиевских и восточноболгарских лексем в подкорпусах исторического корпуса «Манускрипт».

Материал и методы

Исходным является положение о том, что частота использования компонента синонимического ряда зависит от текстологической истории рукописей, входящих в подкорпус. Поисковым – предположение о том, что количественно-статистическое соотношение между компонентами синонимической пары в группе текстов (подкорпусе), близких своими текстологическими и кодикологическими характеристиками, носит неслучайный характер, зависит от нескольких характеристик компонентов лексической пары, в том числе лингвистических. Оно может быть в различных подкорпусах как типичным, так и индивидуальным для конкретной пары лексем.

Основными вопросами анализа являются:

– существует ли корреляция между количественными и статистическими характеристикам слова, входящего в синонимический

ряд, и текстологическими и/или кодикологическими характеристиками текстов;

– каковы количественно-статистические отношения компонентов синонимической пары в подкорпусе;

– существует ли зависимость отношения компонентов синонимической пары в подкорпусах от их лингвистических или иных характеристик.

Синонимические пары

Синонимические пары выбраны из перечней, имеющих в работах [Пичхадзе, 2011, с. 23, 54; Славова, 1989, с. 25–117; Станков, 2018, с. 146, 149, 153; Voss, 1996, S. 100–103], так, чтобы были представлены различные виды соотношения компонентов: разноосновные славянские (*врѣтище* – *власѣница*; *жрътва* – *трѣба*; *ради* – *дѣля*; *тѣкъмо* – *тѣчию*), одноосновные славянские (*вратъникъ* – *вратарь*; *оутро* – *заутра*; *яко* – *акы*), греческий – славянский (*аминь* – *право*; *аромать* – *воня*), заимствованный – заимствованный (*шодѣи* – *жидѣ*).

Методы

Одним из методов корпусного анализа является сопоставление лингвистических данных, извлеченных из подкорпусов, противопоставленных друг другу по какой-либо характеристике – времени создания, жанру, автору и т. д. Обнаружение между подкорпусами различий в таком случае интерпретируется как существование различий в языке разного времени, жанров или авторов и т. п.

Основным приемом стала статистическая оценка встречаемости лексемы в подкорпусах в сопоставлении со статистической величиной частотности слова в контрольном корпусе (модуль статистики – <https://manuscripts.ru/mns/!cred2.stat>). В качестве статистических мер оценки отклонения встречаемости от ожидаемой использованы статистики Log-Likelihood [Ляшевская, Шаров, 2009; Rayson, Garside, 2000], TF*ICF (Term Frequency * Inverse Collection Term Frequency) [Kwok, 1995; Robertson, 2004; Roelleke, 2013; Roelleke, Wang, 2006; Salton, Yang, 1973; Sparck, 1972; Wu et al., 2008], Weiridness [Бессмерт-

ный и др., 2017; Бессмертный, Юй Чуцяо, Ма Пенюй, 2016; Клышинский, Кочеткова, 2014; Ahmad, Gillam, Tostevin, 1999; Gillam, Tariq, Ahmad, 2005]. Статическое значение лексемы сопоставляется с аналогичным синонимичного ему слова. Соотношение статистических значений двух лексем сопоставляется с их соотношением в нескольких подкорпусах.

Известно, что методы TF*ICTF и Weiridness присваивают слову значение тем выше, чем чаще слово встречается в анализируемом подкорпусе и реже – в контрастном корпусе [Бессмертный и др., 2017, с. 83]. Методы считаются эффективным при работе с небольшими коллекциями для извлечения часто встречающихся терминов [Бессмертный, Юй Чуцяо, Ма Пенюй, 2016, с. 1097, 1098]. Относительно Weiridness экспериментальным путем установлено, что для «обычных слов формула странности возвращает значения, близкие к 1, а для терминов – значения, намного превышающие 1, так как в этом случае знаменатель формулы близок к 0» [Бессмертный и др., 2017, с. 83]. В метрике LL используются не только относительные, но и абсолютные частоты слов [Ляшевская, Шаров, 2009, с. VIII], значения меры LL чувствительны как к редко, так и часто встречающимся словам. Для метрики также известно значение, превышение которого дает возможность отнести слово к лексемам, значимым для подкорпуса, – 15.31 [Ляшевская, Шаров, 2009, с. VIII].

Каждая из мер имеет свои диапазоны значений. Наиболее высокие значения назначает статистика LL – десятки, а то и сотни условных единиц, значения Weiridness колеблются около 1,0, величины TF*ICTF существенно меньше 1,0. Для приведения значений к сопоставимым величинам используется нормализация данных с помощью логарифмирования².

Текстовый материал

В качестве контрольного (контрастного) корпуса выступает коллекция рукописей корпуса «Манускрипт». Подкорпуса сформированы из текстов, характеризующихся аналогичными текстологическими и/или кодикологическими свойствами, противопоставленными по этим свойствам другим рукописям корпуса. Первым основанием для противопоставления выбрана

графика списка – глаголический / кириллический алфавит (далее – Глаг³, Кир⁴).

Вторым – служебный / неслужебный тип кириллического текста (далее – КирСл⁵, КирНСл⁶). Особо из четырех текстов XI в., восходящих к восточноболгарской переводческой традиции, сформирован отдельный подкорпус (далее – ВБ⁷).

Третий – кирилло-мефодиевский / преславский: подкорпус, включающий Христинопольский апостол XII в. (содержащий «древнюю редакцию Апостола, осложненную преславскими чтениями» [Пентковская, 2009, с. 19]) и глаголическую Псалтырь Димитрия XI в., (далее – ПсАп-ЗБ⁸), и подкорпус, в который входят Толстовский апостол конца XIV в. («яркий представитель преславской редакции») [Новак, 2015, с. 232]) и Чудовская псалтырь XI в. – представитель симеоновской книжной школы (далее – ПсАп-ВБ⁹).

Инструментарий

Информационно-аналитическая система «Манускрипт» в настоящее время является одной из немногих систем хранения, обработки и демонстрации машиночитаемых транскрипций средневековых славянских письменных памятников. Корпус содержит транскрипции славянских текстов X–XV вв. общим объемом более 3,5 млн словоупотреблений, имеет мета, аналитическую и лингвистическую разметку, снабжен корпусным менеджером, обеспечивающим подготовку транскрипций, лемматизацию текстов, формирование подкорпусов, выборку лингвистических единиц и их демонстрацию (о корпусе см., например: [Baranov, 2018; Баранов, 2019а]).

Модуль статистики¹⁰ предназначен: а) для демонстрации распределения единиц корпуса (символов, текстовых форм, лемм) в рукописи или нескольких рукописях; б) для нахождения количественных и/или статистических характеристик (соответствия или несоответствия ожидаемым статистическим характеристикам) слов в подкорпусе (подкорпусах). Модуль позволяет на основе метаданных транскрипций сформировать подкорпус, сохранить его для многократного исполь-

зования, указать тип анализируемой единицы и ее маску, выбрать контрольный (контрастный) корпус, указать меру статистической оценки. Результат выводится на экран в виде таблицы со сведениями о количестве и статистическом значении лингвистических единиц в подкорпусах [Baranov, Gnutikov, 2018; Баранов, 2019б].

Возможности модуля статистики сформировать подкорпусы на основе выборок необходимых документов и извлечь из них с помощью маски регулярных выражений все текстовые формы анализируемых слов и сведения об их количественных и статистических характеристиках позволили получить информацию о соответствии или несоответствии (значимости) частотности слов их средней частоте в контрастном (контрольном) корпусе, роль которого выполнила выборка, включающая практически все транскрипции корпуса.

Результаты. Количественные и статистические данные

Синонимический ряд жертва – трѣба

Маски регулярных выражений: $\wedge ж[ьѣ]? р[ьѣ]?т.?в.?[\wedge] \$ - \wedge трѣб([ауьѣ])о[юи]ах[ъ]? ам[ьѣ]? \$$

Слово *жертва* зафиксировано во всех подкорпусах (табл. 1, 2). Относительное количество во всех подкорпусах, кроме ПсАп-ВБ, в котором оно значительно ниже, примерно одинаковое. Слово *трѣба* отсутствует в глаголическом подкорпусе (см. табл. 3, 4). Максимальное относительное количество – в подкорпусе четырех восточноболгарских рукописей; минимальное – в подкорпусе кириллических служебных.

Количественные соотношения соответствуют текстологическим свойствам списков: слово *жертва* не характерно для подкорпуса рукописей, восходящих к ЗБ редакции, слово *трѣба* – отсутствует в подкорпусе глаголических рукописей и часто используется в рукописях, созданных в Восточной Болгарии.

Статистическая активность слова *жертва* по-разному оценивается мерами: LL выделяет слово в КирНСл, ВБ, ПсАп-ВБ, Weirdness – в КирСл, ВБ, ПсАп-ЗБ. Статистическая активность слова *трѣба* также различается: LL – в КирСл, КирНСл, ВБ, ПсАп-ВБ, Weirdness – в ВБ и ПсАп-ВБ. При этом оценка статистиками в двух последних подкорпусах совпадает, а в LL выше, чем в КирСл и КирНСл.

Средние логарифмированных статистических метрик (см. рис. 1) показывают су-

Таблица 1. Количественные и статистические значения слова *жертва*

Table 1. Quantitative and statistical values of the word *zhrtva* ‘sacrifice’

Подкорпус	Все рукописи	Глаг	Кир	КирСл	КирНСл	ВБ	ПсАп-ВБ	ПсАп-ЗБ
Объем подкорпуса	2986201	97138	2889063	1583956	1329303	262429	103908	99577
Объем выборки	629	21	608	389	197	95	4	34
Среднее	0,02106	0,02162	0,02104	0,02456	0,01482	0,03620	0,00385	0,03414
LL	–	2,01357	2,00270	7,48186	21,84683	23,06930	23,73787	8,49934
TF*ICTF	–	0,00079	0,00077	0,00090	0,00054	0,00133	0,00014	0,00126
Weirdness	–	1,02473	0,99753	1,16408	0,70246	1,71590	0,18247	1,61845

Таблица 2. Логарифмированные количественные и статистические значения слова *жертва*

Table 2. Logarithmic quantitative and statistical values of the word *zhrtva* ‘sacrifice’

Подкорпус	Глаг	Кир	КирСл	КирНСл	ВБ	ПсАп-ВБ	ПсАп-ЗБ
ln(LL)	0,69991	0,69450	2,01248	3,08406	3,13850	3,16707	2,13999
ln(TF*ICTF)	-7,14348	-7,16912	-7,01312	-7,52394	-6,62258	-8,87387	-6,67664
ln(Weirdness)	0,02443	-0,00247	0,15193	-0,35317	0,53994	-1,70117	0,48147
ln(средн.)	-3,83420	-3,86110	-3,70669	-4,21179	-3,31869	-5,55980	-3,37716

Таблица 3. Количественные и статистические значения слова *трѣба*

Table 3. Quantitative and statistical values of the word *trēba* ‘sacrifice’

Подкорпус	Все рукописи	Глагол	Кир	КирСл	КирНСл	ВБ	ПсАп-ВБ	ПсАп-ЗБ
Объем подкорпуса	2986201	97138	2889063	1583956	1329303	262429	103908	99577
Объем выборки	279	0	279	78	219	105	41	13
Среднее	0,00934	0,00000	0,00966	0,00492	0,01647	0,04001	0,03946	0,01306
LL	–	0,00000	2,12333	30,27670	39,68686	126,35074	54,06898	3,23649
TF*ICTF	–	0,00000	0,00039	0,00020	0,00066	0,00161	0,00159	0,00053
Weirdness	–	0,00000	1,02993	0,52518	1,75704	4,26717	4,20821	1,39234

Таблица 4. Логарифмированные количественные и статистические значения слова *трѣба*

Table 4. Logarithmic quantitative and statistical values of the word *trēba* ‘sacrifice’

Подкорпус	Глагол	Кир	КирСл	КирНСл	ВБ	ПсАп-ВБ	ПсАп-ЗБ
ln(LL)	–	0,75299	3,41038	3,68102	4,83906	3,99026	1,17449
ln(TF*ICTF)	–	-7,84936	-8,51719	-7,32327	-6,43152	-6,44402	-7,54263
ln(Weirdness)	–	0,02949	-0,64401	0,56363	1,45095	1,43704	0,33099
ln(средн.)	–	-4,64006	-5,31356	-4,10592	-3,21861	-3,23252	-4,33857

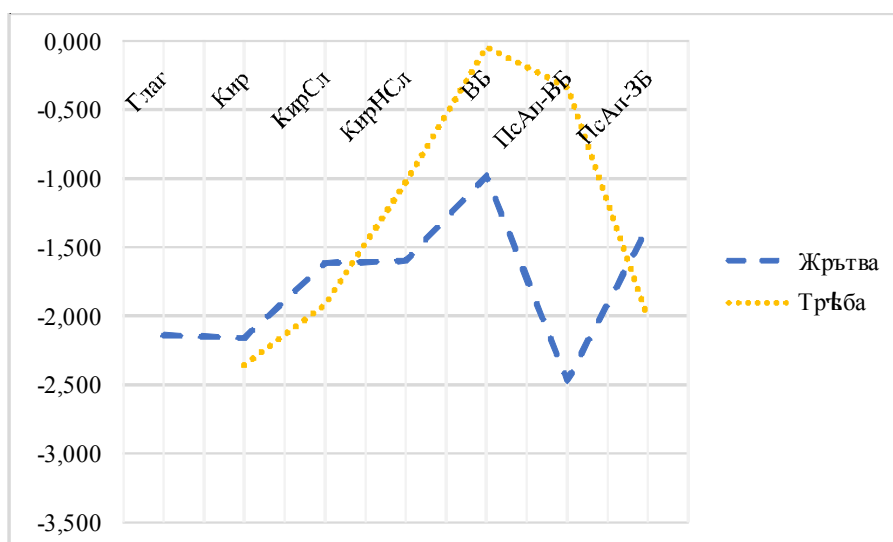


Рис. 1. Жрътва – *трѣба*. Средние логарифмических значений ln(LL), ln(TF*ICTF), ln(Weirdness)

Fig. 1. *Zhrъtva* – *trēba* ‘sacrifice’. Mean of logarithmic values ln(LL), ln(TF*ICTF), ln(Weirdness)

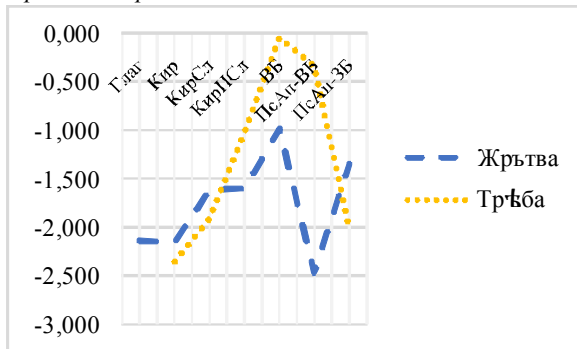
ществленные отличия между статистической активностью слов как между собой, так и между подкорпусами: а) в Глагол и ПсАп-ЗБ преобладает *жрътва*, в ВБ, ПсАп-ВБ – *трѣба*, в КирСл – *жрътва*, в КирНСл – *трѣба*; б) при большей статистической активности *жрътва* максимальный контраст обнаруживается в ПсАп-ЗБ, минимальный – в Кир и КирСл; в) при большей активности *трѣба* – максимальная контрастность в ПсАп-ВБ, меньшая – в ВБ, еще меньшая – в КирНСл.

Аналогично были исследованы и все другие пары лексем. Приведем итоговые диаграммы.

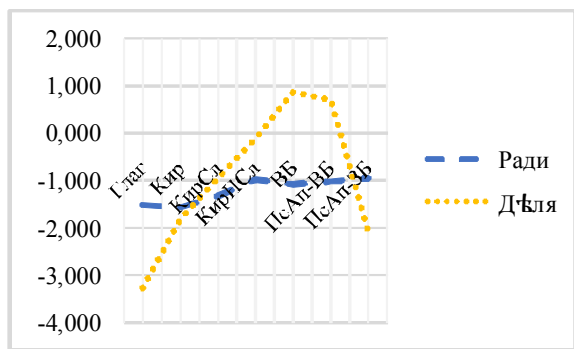
Обсуждение

Использованная в работе методика сопоставления статистических значений слов в текстовых подкорпусах позволяет увидеть ряд закономерностей в соотношении западноболгарских и восточноболгарских компонентов синонимического ряда (см. рис. 2).

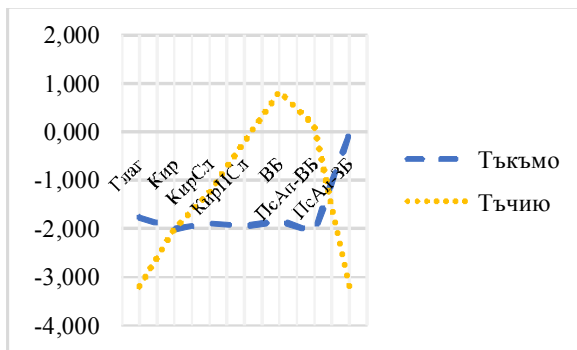
Жрътва – трѣба



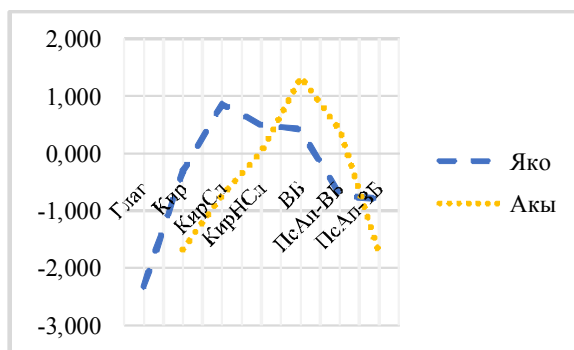
Ради – дѣля



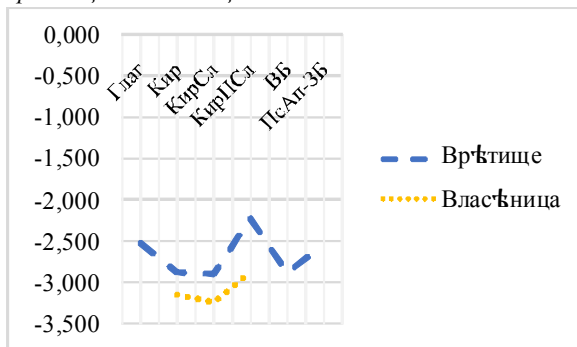
Тѣкъмо – тѣчию



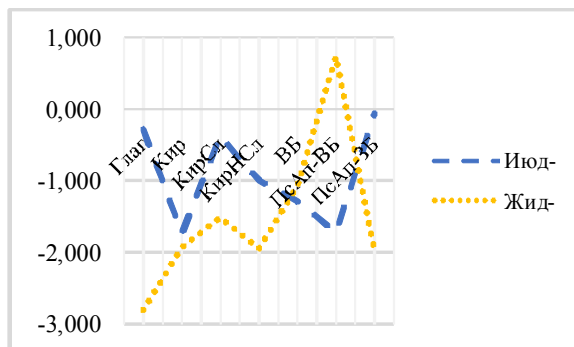
Яко – акы



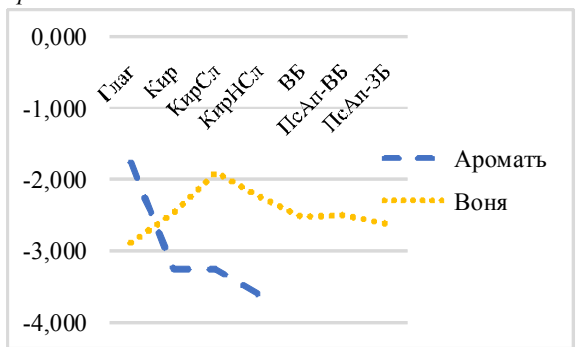
Врѣтище – власѣница



Июдѣи – жидѣ



Ароматъ – воня



Аминѣ – право

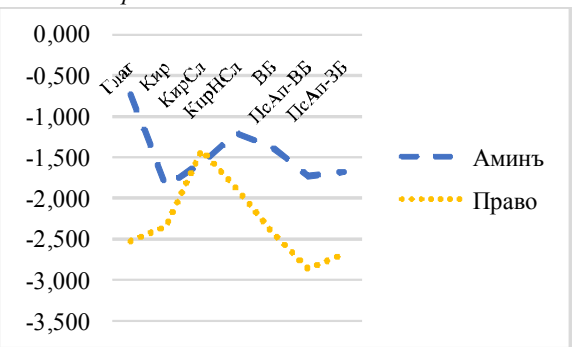


Рис. 2. Средние логарифмических значений десяти пар синонимов (см. также с. 13)

Fig. 2. Average logarithmic values of ten pairs of synonyms (see also p. 13)

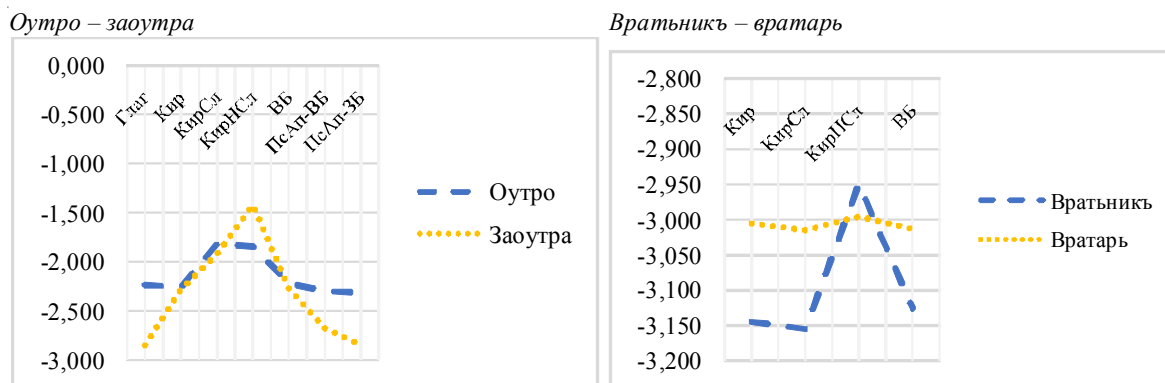


Рис. 2. Окончание
Fig. 2. Final part

1. Во всех синонимических рядах статистическая активность западноболгарского компонента выше (или он единственно возможен) в подкорпусе глаголических текстов и подкорпусе ПсАп-ЗП, включающем списки, восходящие к кирилло-мефодиевской традиции. По этой особенности подкорпус глаголических текстов отчетливо противопоставлен подкорпусу, включающему все кириллические рукописи, а подкорпус ПсАп-ЗП – подкорпусу ПсАп-ВБ, содержащему списки, восходящие к восточноболгарским рукописям.

2. Отчетливая предпочтительность восточноболгарских компонентов обнаруживается в подкорпусах ВБ и ПсАп-ВБ, включающих тексты, восходящие к Преславской традиции. О преимущественном тяготении восточноболгарского компонента именно восточноболгарской книжной школе свидетельствует и последовательная его большая активность в подкорпусе неслужебных рукописей КиНСл по сравнению с текстами подкорпуса более архаичных служебных списков.

Обнаруженная корреляция членов ряда и текстологических характеристик рукописей характеризует лишь славянские разноосновные синонимические ряды *жрътва – трѣба*, *ради – дѣля*, *тъкъмо – тъчию*, *яко – акы*. В рядах, членами которых являются заимствованные слова и одноосновные словообразовательные синонимы, отношения несколько иные.

3. В парах с заимствованными словами греческий компонент в глаголическом подкорпусе и в ПсАп-ЗБ также предпочтительнее, чем негреческий, ср. пары *шодѣи – жидь*, *аромать – воня*, *аминь – право* (*аромать* в

ПсАп-ЗБ отсутствует). В паре *аминь – право* в КирСл, КирНСл, ВБ, ПсАп-ВБ нет отчетливого предпочтения *право* – статистически более активным является западноболгарское *аминь*. В паре *шодѣи – жидь* картина аналогичная, за исключением преобладания в ПсАп-ВБ слова *жидь*. Вторым отличием этой группы является иное соотношение статистической активности западноболгарского компонента между КирСл и КирНСл, с одной стороны, и ВБ и ПсАп-ВБ – с другой: если в славянских парах (*жрътва – трѣба* и др.) первый компонент был более предпочтителен в ВБ и ПсАп-ВБ, чем в КирСл и КирНСл, то в парах греческим компонентом *шодѣи – жидь*, *аминь – право* и *аромать – воня* ситуация обратная. Особенностью соотношений компонентов в этих трех рядах является, в отличие от славянских рядов, и их существенная индивидуальность по отношению друг к другу.

4. Словообразовательные одноосновные синонимы, наряду с похожим в обоих случаях последовательно более высоким статистическим значением восточноболгарского лексического компонента в трех подкорпусах – КирСл, КирНСл и ВБ, – достаточно индивидуальны: имеют статистическое равенство обоих компонентов и несколько более высокое значение восточноболгарского в паре *оутро – заоутра*, с одной стороны, и более высокие значения восточнославянского *вратарь* в Кир, КирСл и ВБ при преобладании западноболгарского *вратъникъ* в КирНСл – с другой.

5. Пары, в которых один из компонентов в одном или нескольких подкорпусах отсут-

ствуется, демонстрируют те же тенденции и соотношения компонентов синонимического ряда, что и в тех, где компоненты встречаются во всех подкорпусах: в паре *врѣтище* – *власѣница* активность второго компонента выше в подкорпусе КирНСл рукокопий, в паре *аромать* – *воня* слово *аромать* статистически более активно в подкорпусе служебных списков КирСл, а *воня* наименее предпочтительна в ПсАп-ЗБ, чем в КирНСл, ВБ, ПсАп-ВБ; в паре *вратъникъ* – *вратарь* второй компонент имеет стабильно высокую статистическую активность во всех кириллических подкорпусах.

Обнаруженные закономерности демонстрируют, с одной стороны, типичность соотношения компонентов всех рассмотренных синонимических рядов в том или ином подкорпусе, с другой стороны – индивидуальность соотношения синонимов. Важным представляется существование связи между соотношением синонимов и их лексическими и лексико-словообразовательными характеристиками: заметно отличаются соотношения в группах разноосновных и одноосновных синонимов, в славянская пара синонимов и паре, одним из компонентов которой является греческое по происхождению слово.

Продемонстрированная корреляция между статистическими соотношениями западноболгарского и восточноболгарского компонентов синонимического ряда и текстологическими и кодикологическими характеристиками текстов (глаголический – кириллический, кирилло-мефодиевский – восточноболгарский) ожидаема, но одновременно является дополнительной достаточно надежной характеристикой текста / рукописи или группы текстов / рукописей наряду с традиционно используемыми критериями, основанными на соотношениях «есть – нет» и на соотношениях «больше – меньше» / «чаще – реже» в абсолютном или относительном выражении.

Наличие такой корреляции позволяет с доверием относиться и к различиям в соотношении синонимических рядов, компоненты которых характеризуются отличающимися лексическими или лексико-словообразовательными свойствами, а кроме того, увидеть различия в степени устойчивости таких соотношений – последовательную, типовую про-

тивопоставленность западноболгарской и восточноболгарской разноосновных славянских лексем ряда в текстологически и кодикологически противопоставленных текстах и индивидуальную конфигурацию соотношений в паре одноосновных словообразовательных синонимов или в паре, компонентом которой является греческое заимствование.

Полученные результаты позволяют сделать несколько выводов относительно вопроса о маркированности / немаркированности преславской лексики и о представленности кирилло-мефодиевских и восточноболгарских лексем в рукописях, созданных на Руси.

1. Аналогичность, а в ряде случаев и идентичность соотношений компонентов синонимических пар как в случаях представленности обоих во всех подкорпусах, так и в случаях отсутствия одного из них в некоторых из подкорпусов позволяют сделать вывод о том, что отсутствие компонента синонимической пары в подкорпусе является одним из двух аналогичных вариантов его соотношения с другим компонентом: компонент отсутствует, или его статистическая активность ниже или значительно ниже альтернативного компонента.

2. Не имеющая исключений противопоставленность компонентов исследованных синонимических рядов в глаголических и архаичных рукописях подкорпуса ПсАп-ЗБ, с одной стороны, и в текстах подкорпусов ВБ, ПсАп-ВБ – с другой, свидетельствует об устойчивых лексических предпочтениях этих текстологически противопоставленных групп текстов при использовании в них и того и другого компонентов. Статистическая предпочтительность восточноболгарских компонентов может отличаться в подкорпусах (ср. их соотношение в КирСл и КирНСл), выступая своеобразным маркером неслужебных рукописей.

3. Есть основания говорить о том, что предпочтительность того или иного компонента синонимического ряда зависит не только от того, в рамках какой писцовой школы он был переведен, создавался или редактировался, но и от собственно языковых характеристик слов пары: разноосновные славянские компоненты последовательно противопоставляются в текстологичес-

ки контрастных друг другу подкорпусов (текстах и рукописях); одноосновные славянские компоненты при аналогичном противопоставлении двух архаических подкорпусов (Глаг и ПсАп-ЗБ) не имеют подобного последовательного противопоставления в ряду кириллических; греческий и негреческий компоненты ряда конкурируют друг с другом, причем первый в ряду кириллических подкорпусов не уступает, а то и превосходит в статистической активности негреческий.

4. В восточнославянских списках, которые и составляют большую часть проанализированного материала, зафиксированы как западноболгарские, так и восточноболгарские компоненты синонимических рядов. Предпочтительность того или иного компонента зависит как от текстологических и кодикологических характеристик кириллических рукописей, так и от собственно языковых свойств лексем. При общей предпочтительности восточноболгарского компонента в ряде синонимических рядов наблюдается большая статистическая активность западноболгарского компонента, например, в парах *яко* – *акы*, *аминь* – *право*, *вратьникъ* – *вратарь*.

Заключение

Исследование доказало существование корреляции между текстами, характеризующимися различными текстологическими и/или кодикологическими свойствами, и соотношением в них компонентов синонимического ряда, а также наличие связи между соотношением компонентов ряда в различных подкорпусах и лексическими характеристиками синонимов.

Значимым результатом является доказательство результативности использования корпусных и статистических методов и приемов для анализа лингвистических данных средневековых славянских текстов и рукописей, в частности, для решения вопросов в рамках кирилло-мефодиевской проблематики.

Безусловно, полученные результаты должны быть подтверждены и уточнены в ходе анализа и других синонимических кирилло-мефодиевских – восточноболгарских лексических рядов.

ПРИМЕЧАНИЯ

¹ Работа выполнена при финансовой поддержке РФФ, проект № 20-18-00206 «Дистрибутивно-квантитативный анализ семантических изменений на основе больших диахронических корпусов».

The study is given a financial support by The Russian Science Foundation, the research project no. 20-18-00206 entitled “Distributional-quantitative analysis of semantic changes based on large diachronic text corpora”.

² «Процесс преобразования сырых частотностей с целью устранения зависимости от размера текста, а также приведения различных величин к единому виду в лингвистической информатике называется нормализацией. В качестве средства нормализации широко применяется логарифмирование по основанию 2» [Яцко, 2014, с. 3].

³ Глаголические рукописи и отрывки корпуса «Манускрипт». URL: <http://manuscripts.ru/mns/portal.main?p1=41>; <http://manuscripts.ru/mns/portal.main?p1=55>.

Всего 9 рукописей и отрывков, объем – 97 138 текстовых форм.

⁴ Кириллические служебные и неслужебные рукописи и отрывки. URL: http://manuscripts.ru/mns/portal.main?p1=1&p_lid=1.

Всего 74 рукописей и отрывков, объем – 2 889 063 текстовые формы.

⁵ Кириллические служебные рукописи и отрывки. URL: <http://manuscripts.ru/mns/portal.main?p1=30>; http://manuscripts.ru/mns/portal.main?p1=10&p_lid=1; <http://manuscripts.ru/mns/portal.main?p1=39>; <http://manuscripts.ru/mns/portal.main?p1=22>; <http://manuscripts.ru/mns/portal.main?p1=58>; <http://manuscripts.ru/mns/portal.main?p1=60>; http://manuscripts.ru/mns/portal.main?p1=8&p_lid=1; <http://manuscripts.ru/mns/portal.main?p1=72>.

Всего 52 рукописи и отрывка, объем – 1 583 956 текстовых форм: служебные (краткие и полные) Евангелия (11), служебные минеи (17), стихирари минейные (5), триоди (7), паримейники (5), апостолы (2), псалтыри (3), служебники (2).

⁶ Кириллические неслужебные рукописи и отрывки. URL: http://manuscripts.ru/mns/portal.main?p1=12&p_lid=1; <http://manuscripts.ru/mns/portal.main?p1=63>; http://manuscripts.ru/mns/portal.main?p_text=44386403; http://manuscripts.ru/mns/portal.main?p1=8&p_lid=1.

Всего 22 рукописи и отрывка, объем – 1 329 303 текстовые формы: летописи (5), сборники (4), Патерик синайский (1), поучения Ефрема Сирина (3), Слова Григория Богослова (2), Пандекты Антиоха Черноризца (1), отрывки (6).

⁷ Неслужебные рукописи и отрывки, восходящие к восточноболгарской письменной традиции. URL: [http://manuscripts.ru/mns/portal.main?p1=8&p_lid=1\\$](http://manuscripts.ru/mns/portal.main?p1=8&p_lid=1$); http://manuscripts.ru/mns/main?p_text=85179364.

Всего 4 рукописи, объем – 262 429 текстовых форм:

1. Златоструй («Златоструй А.Ф. Бычкова»), XI в. // РНБ. Q.п.1.74. 4 л.

2. Изборник («Изборник Святослава / Симеона»), 1073 г. // ГИМ. Син. 1043 (Син. 31-д). 266 л.

3. Минея четья на март («Супрасльская рукопись»), сер. XI в. // Национальная библиотека Варшавы. Собр. Зав. № 21; Университетская библиотека Люблины. Cod. Кор. 2; РНБ. Q.п.72. 285 л.

4. Слова (13) Григория Богослова, XI в. // РНБ. Q.п.1.16. 377 л.

⁸ Псалтырь и Апостол, восходящие к архаичной традиции. URL: <http://manuscripts.ru/mns/portal.main?p1=61>; <http://manuscripts.ru/mns/portal.main?p1=42>.

Всего 2 рукописи, объем – 99 577 текстовых форм:

1. Христинопольский апостол, XII в. // Национальная библиотека Украины. ИР. Ф. VIII. № 3. Л. 1–8; Библиотека Чарторыйских, Краков, Польша. Л. 9–12; Львовский исторический музей. ОР. № 37. Л. 13–303. 303 л.

2. Псалтырь Димитрия Синайского, посл. четв. XI в., Синай/Египет // Монастырь св. Екатерины. Cod. Sin. slav. 3/N. 145 л.

⁹ Псалтырь и Апостол, восходящие к восточноболгарской традиции. URL: <http://manuscripts.ru/mns/portal.main?p1=64>; http://manuscripts.ru/mns/main?p_text=44342283.

Всего 2 рукописи, объем – 103 908 текстовых форм:

1. Толстовский апостол, XIV в. // РНБ. Q.п.1.5. 93 л.

2. Псалтырь толковая Феодорита Кирского («Чудовская псалтырь»), XI в. // ГИМ. Чуд. 7. 176 л.

¹⁰ URL: http://manuscripts.ru/mns!/cred2.stat#tabs|Tabs_Group_name:Tab_1_name

СПИСОК ЛИТЕРАТУРЫ

Баранов В. А., 2019а. Создание и использование исторических корпусов славянских письменных памятников // Scripta & e-Scripta. Vol. 19. С. 33–57.

Баранов В. А., 2019б. Модуль статистики исторического корпуса «Манускрипт»: функции и демонстрация данных. 2 // И. А. Бодуэн де Куртене и мировая лингвистика. В 2 т. Т. 1 : Международный конф. : VII Бодуэновские чтения (Казан. федер. ун-т, 28–31 окт. 2019 г.) : тр. и материалы. Казань : Изд-во Казан. ун-та. С. 24–30.

Бессмертный И. А., Нугуманова А. Б., Мансурова М. Е., Байбурин Е. М., 2017. Метод контрастного извлечения редких терминов из текстов на естественном языке // Научно-технический вестник информационных технологий, механики и оптики. Т. 17, № 1. С. 81–91. URL: <http://ntv.ifmo.ru/file/article/16383.pdf>

Бессмертный И. А., Юй Чуцяо, Ма Пенной, 2016. Статистический метод извлечения терминов из китайских текстов без сегментации фраз // Научно-технический вестник информационных технологий, механики и оптики. Т. 16, № 6. С. 1096–1102. URL: <http://ntv.ifmo.ru/file/article/16157.pdf>

Димитрова А., 2016. Златоструят в преводаческата дейност на старобългарските книжовници. София: Авалон. 456 с.

Добрев И., 1979. Текстът на Добромировото евангелие и втората редакция на старобългарските богослужебни книги // Български език. Т. 29, № 1. С. 9–21.

Желязкова В., 2016а. Паримейные чтения в четьх списках Книги Исход // Studia Ceranea. Т. 6. С. 225–240.

Желязкова В., 2016б. Книга Исход в южнославянских списках XV–XVI вв. // Studia Slavistica. Vol. 13. С. 243–256.

Илиев И., 2016. Тълкуванието на Книга на пророк Даниил от Иполит Римски в старобългарската литература : автореферат на дисертация за присъждане на образователната и научна степен «доктор». София. 67 с.

Клышинский Э. С., Кочеткова Н. А., 2014. Метод извлечения технических терминов с использованием меры странности // Новые информационные технологии в автоматизированных системах. № 17. С. 365–370. URL: https://elibrary.ru/download/elibrary_21527004_14693581.pdf

Ляшевская О. Н., Шаров С. А., 2009. Введение к частотному словарю современного русского языка // Частотный словарь современного русского языка (на материалах Национального корпуса русского языка). М. : Азбуковник. С. V–XXII. URL: <http://dict.ruslang.ru/freq.pdf>

Максимович К. А., 2000. Славянизмы современного русского языка и кирилло-мефодиевское наследие // Folia Slavistica: Пале Михайловне Цейтлин. М. : [б. и.]. С. 72–84.

Милтенев Я., 2008. Лексический критерий как способ атрибуции преславских текстов (на материале славянского перевода Диалогов Псевдо-Кесария) // Славяноведение. № 5. С. 41–49.

Милтенев Я., 2020. Преславските лексикални маркери. 1. Опит за въведение // Palaeobulgarica. Vol. 44, № 2. С. 54–79.

- Новак М. О., 2015. Лексика древнеславянского перевода апостола: в поисках образа и смысла // *Перевод как средство взаимодействия культур*. № 1. С. 230–240.
- Новикова А. С., 2013. К вопросу об истории создания евангельского текста Чудовской рукописи // *Славянские языки и литературы в синхронии и диахронии* : материалы Междунар. научн. конф. М. : [б. и.]. С. 268–273.
- Пентковская Т. В., 2009. Восточнославянские и южнославянские переводы богослужебных книг XIII–XIV вв. : Чудовская и афонская редакции Нового Завета и Иерусалимский Типикон : дис. ... д-ра филол. наук. М. 497 с.
- Пичхадзе А. А., 2011. Переводческая деятельность в домонгольской Руси: лингвистический аспект. М. : Рукопис. памятники Древ. Руси. 408 с.
- Славова Т., 1989. Преславска редакция на Кирило-Методиеви старобългарски евангелски превод // *Кирило-Методиеви студии*. Кн. 6. София : Кирило-Методиевски научен център при БАН. С. 15–129.
- Славова Т., 2013. Славянският превод на Посланието на патриарх Фотий до княз Борис-Михаил. София : Св. Климент Охридски. 344 с.
- Станков Р., 2018. Проблема «преславской» и «кирилло-мефодиевской» лексики в древнеболгарском переводе Хроники Георгия Амартола // *Преславска книжовна школа*. Т. 18. С. 121–158.
- Шафарик П., 1860/1861. О происхождении и родине глаголитизма // *Чтения в Императорском Обществе Истории и Древностей Российских при Московском Университете*. № 4. Материалы Славянские. III. М. : Унив. тип. С. 29–30.
- Яцко В. А., 2014. Компьютерная лингвистика или лингвистическая информатика // *Научно-техническая информация*. Серия 2, Информационные процессы и системы. № 5. С. 1–10.
- Ahmad K., Gillam L., Tostevin L., 1999. University of Surrey Participation in TREC8: Weirdness Indexing for Logical Document Extrapolation and Retrieval (Wilder) // *Proc. 8th Text Retrieval Conference TREC*. Gaithersburg : [s. n.]. P. 717–724.
- Baranov V., 2018. A Text Corpus of Medieval Manuscripts as a Goal and a Tool for Linguistic Research // *Editing Mediaeval Texts from a Different Angle: Slavonic and Multilingual Traditions*. Leuven ; P. ; Bristol : Peeters. P. 283–308.
- Baranov V. A., Gnutikov R. M., 2018. The Statistics and n-Gram Modules of the Historical Corpus “Manuscript” // *Digital and Analytical Approaches to the Written Heritage* : Proceedings of the 7th International Conference El‘Manuscript “Textual Heritage and Information Technologies”. Sofia : Gutenberg. P. 9–28.
- Gillam L., Tariq M., Ahmad K., 2005. Terminology and the Construction of Ontology // *Terminology*, vol. 11, iss 1. P. 55–81.
- Jagić V., 1913. *Entstehungsgeschichte der kirchenslavischen Sprache*. Berlin : Weidmann. 540 p.
- Kwok K. L., 1995. A Network Approach to Probabilistic Information Retrieval // *ACM Transactions on Information Systems*. Vol. 13, № 3. P. 324–353.
- Rayson P., Garside R., 2000. Comparing Corpora Using Frequency Profiling // *Proceedings of the Comparing Corpora Workshop at ACL 2000*. Hong Kong. P. 1–6. URL: http://ucrel.lancs.ac.uk/people/paul/publications/rg_acl2000.pdf
- Robertson S., 2004. Understanding Inverse Document Frequency: On Theoretical Arguments for Idf // *Journal of Documentation*. № 60. P. 503–520. URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.438.2284&rep=rep1&type=pdf>
- Roelleke T., 2013. *Information Retrieval Models: Foundations and Relationships*. Cham : Springer. 141 p.
- Roelleke T., Wang J., 2006. A Parallel Derivation of Probabilistic Information Retrieval Models // *Proceedings of the 29th Annual ACM SIGIR Conference on Research and Development in Information Retrieval*. Seattle ; N. Y. : ACM. P. 107–114.
- Salton G., Yang C. S., 1973. On the Specification of Term Values in Automatic Indexing // *Journal of Documentation*. Vol. 29. P. 351–372.
- Sparck J. K., 1972. A Statistical Interpretation of Term Specificity and Its Application in Retrieval // *Journal of Documentation*. Vol. 28. P. 11–21.
- Voss Chr., 1996. Die Vertretung von lexikalischen textologischen Dubletten der Dichotomie Ochrid-Prslav in kirchenslavischen Abschriften der Paränese Ephraims des Syrers. Ein Beitrag zur Datierung der altbulgarischen Erstübersetzung // *Anzeiger für slavischen Philologie*. Bd. 24. S. 95–127.
- Wu H. C., Luk R. W. P., Wong K. F., Kwok K. L., 2008. Interpreting TF-IDF Term Weights as Making Relevance Decisions // *ACM Transactions on Information Systems*. Vol. 26, № 3. Art. 13. URL: https://www.scss.tcd.ie/khurshid.ahmad/Research/Sentiments/tfidf_relevance.pdf

REFERENCES

Baranov V.A., 2019a. Sozdanie i ispolzovanie istoricheskikh korpusov slavyanskikh

- pismennykh pamyatnikov [Creation and Use of Historical Corpora of Slavic Written Monuments]. *Scripta & e-Scripta*, vol. 19, pp. 33-57.
- Baranov V.A., 2019b. Modul statistiki istoricheskogo korpusa «Manuskript»: funktsii i demonstratsiya dannykh. 2 [Statistics Module of the Historical Corpus “Manuscript”: Functions and Data Demonstration. 2]. *I. A. Boduen de Kurtene i mirovaya lingvistika. V 2 t. T. 1: Mezhdunar. konf.: VII Boduenovskie chteniya (Kazan. feder. un-t, 28–31 okt. 2019 g.): tr. i materialy* [I.A. Baudouin de Courtenay and World Linguistics. In 2 Vols. Vol. 1. International Conference: The 7th Baudouin Readings (Kazan Federal University, October 28–31, 2019). Works and Proceedings]. Kazan, Izd-vo Kazan. un-ta, pp. 24-30.
- Bessmertny I.A., Nugumanova A.B., Mansurova M.E., Bayburin E.M., 2017. Metod kontrastnogo izvlecheniya redkikh terminov iz tekstov na estestvennom yazyke [Method of Rare Term Contrastive Extraction from Natural Language Texts]. *Nauchno-tehnicheskii vestnik informatsionnykh tekhnologiy, mekhaniki i optiki* [Scientific and Technical Journal of Information Technologies, Mechanics and Optics], vol. 17, no. 1, pp. 81-91. URL: <http://ntv.ifmo.ru/file/article/16383.pdf>
- Bessmertny I.A., Yuy Chutsyao, Ma Penyuy, 2016. Statisticheskii metod izvlecheniya terminov iz kitayskikh tekstov bez segmentatsii fraz [Statistical Method for Extracting Terms from Chinese Texts Without Phrase Segmentation]. *Nauchno-tehnicheskii vestnik informatsionnykh tekhnologiy, mekhaniki i optiki* [Scientific and Technical Bulletin of Information Technologies, Mechanics and Optics], vol. 16, no. 6, pp. 1096-1102. URL: <http://ntv.ifmo.ru/file/article/16157.pdf>
- Dimitrova A., 2016. *Zlatostruyat v prevodacheskata deynost na staroblgarskite knizhovnitsi* [The Zlatostruy in the Translation Activity of Old Bulgarian Scribes]. Sofia, Avalon Publ. 456 p.
- Dobrev I., 1979. Tekst na Dobromirovoto evangelie i vtorata redaktsiya na staroblgarskite bogosluzhebni knigi [The Text of the Dobromirov’s Gospel and the Second Edition of the Old Bulgarian Liturgical Books]. *Blgarski ezik* [Bulgarian Language], vol. 29, no. 1, pp. 9-21.
- Zhelyazkova V., 2016a. Parimeynye chteniya v chetikh spiskakh Knigi Iskhod [Parimea Readings in the Manuscripts of the Book of Exodus]. *Studia Ceranea*, vol. 6, pp. 225-240.
- Zhelyazkova V., 2016b. Kniga Iskhod v yuzhnoslavianskikh spiskakh XV–XVI vv. [The Book of Exodus in the South Slavic Copies of the 15th – 16th Centuries]. *Studi Slavistici*, vol. 13, pp. 243-256.
- Iliev I., 2016. *Tlkuvanieto na Kniga na prorok Daniil ot Ipolit Rimski v staroblgarskata literatura: avtoreferat na disertatsiya za priszhdane na obrazovatelnata i nauchna stepen «doctor»* [The Interpretation of the Book of Prophet Daniel by Hippolytus of Rome in Old Bulgarian Literature. Dr. philol. sci. diss.]. Sofia. 67 p.
- Klyshinskiy E.S., Kochetkova N.A., 2014. Metod izvlecheniya tekhnicheskikh terminov s ispolzovaniem mery strannosti [Method for Extracting Technical Terms Using a Weirdness Measure]. *Novye informatsionnye tekhnologii v avtomatizirovannykh sistemakh* [New Information Technologies in Automated Systems], no. 17, pp. 365-370. URL: https://elibrary.ru/download/elibrary_21527004_14693581.pdf
- Lyashevskaya O.N., Sharov S.A., 2009. Vvedenie k chastotnomu slovaryu sovremennogo russkogo yazyka [Introduction to the Frequency Dictionary of the Modern Russian Language]. *Chastotnyy slovar sovremennogo russkogo yazyka (na materialakh Natsionalnogo korpusa russkogo yazyka)* [Frequency Dictionary of the Modern Russian Language (Based on the Materials of the National Corpus of the Russian Language)]. Moscow, Azbukovnik Publ., pp. V-XXII. URL: <http://dict.ruslang.ru/freq.pdf>
- Maksimovich K.A., 2000. Slavyanizmy sovremennogo russkogo yazyka i kirillo-mefodievske nasledie [Slavicisms of the Modern Russian Language and Cyril and Methodius Legacy]. *Folia Slavistica: Rale Mikhaylovne Tseytlin* [Folia Slavistica: For Ralya Tseytlin]. Moscow, s.n., pp. 72-84.
- Miltenov Ya., 2008. Leksicheskiy kriteriy kak sposob atributsii preslavskikh tekstov (na materiale slavyanskogo perevoda Dialogov Psevdo-Kesariya) [Lexical Criterion as a Way of Attributing Preslav Texts (Based on the Slavic Translation of the Dialogues of Pseudo-Caesarea)]. *Slavyanovedenie* [Slavic Studies], no. 5, pp. 41-49.
- Miltenov Ya., 2020. Preslavskite leksikalni markeri. 1. Opit za vvedenie [Preslav Lexical Markers. 1. Introduction Attempt]. *Palaeobulgarica*, vol. 44, no. 2, pp. 54-79.
- Novak M.O., 2015. Leksika drevneslavianskogo perevoda apostola: v poiskakh obraza i smysla [Vocabulary of the Old Slavonic Translation of the Apostle: In Search of Image and Meaning]. *Perevod kak sredstvo vzaimodeystviya kultur*

- [Translation as a Means of Interaction Between Cultures], no. 1, pp. 230-240.
- Novikova A.S., 2013. K voprosu ob istorii sozdaniya evangelskogo teksta Chudovskoy rukopisi [To the Question of the History of the Creation of the Gospel Text of the Chudovskaya Manuscript]. *Slavyanskije yazyki i literatury v sinkhronii i diakhronii: materialy Mezhdunar. nauchn. konf.* [Slavic Languages and Literatures in Synchrony and Diachrony. Proceedings of the International Scientific Conference]. Moscow, s.n., pp. 268-273.
- Pentkovskaya T.V., 2009. *Vostochnoslavjanskije i yuzhnoslavjanskije perevody bogoslužebnykh knig XIII–XIV vv.: Chudovskaja i afonskaja redaktsii Novogo Zaveta i Ierusalimskij Tipikon: dis... d-ra filol. nauk* [East Slavic and South Slavic Translations of Liturgical Books of the 13th – 14th Centuries. Chudovskaya and Athos Editions of the New Testament and the Jerusalem Typicon. Dr. philol. sci. diss.]. Moscow, 497 p.
- Pichkhadze A.A., 2011. *Perevodčeskaja deyatelnost v domongolskoy Rusi: lingvističeskij aspekt* [Translation Activity in Pre-Mongol Rus: Linguistic Aspect]. Moscow, Rukopis. pamyatniki Drev. Rusi. 408 p.
- Slavova T., 1989. Preslavskaja redaktsija na Kirilo-Methodieviya staroblgarski evangelski prevod [Preslav Edition of the Cyril-Methodiev Old Bulgarian Gospel Translation]. *Kirilo-Methodievi studii. Kn. 6* [Cyril and Methodius Studies. Book 6]. Sofia, Kirilo-Methodievski nauchen tsentr pri BAN, pp. 15-129.
- Slavova T., 2013. *Slavyanskiyat prevod na Poslanieto na patriarkh Fotiy do knyaz Boris-Mikhail* [The Slavic Translation of the Message of Patriarch Photius to Prince Boris-Mikhail]. Sofia, Sv. Kliment Okhridski. 344 p.
- Stankov R., 2018. Problema «preslavskoy» i «kirillo-mefodievskoy» leksiki v drevnebolgarskom perevode Khroniki Georgiya Amartola [The Problem of “Preslav” and “Cyrillic-Methodiev” Lexicons in Ancient Bulgarian Translation Georgy Amartol’s Chronicles]. *Preslavskaja knizhovna shkola* [Preslav Literary School], vol. 18, pp. 121-158.
- Shafarik P., 1860/1861. O proiskhozhdenii i rodine glagolizma [On the Origins and Homeland of Glagoliticism]. *Chteniya v Imperatorskom Obshchestve Istorii i Drevnostey Rossiyskikh pri Moskovskom Universitete. № 4. Materialy Slavyanskiye, III* [Readings at the Imperial Society of Russian History and Antiquities at Moscow University. No. 4. Slavic Proceedings. 3]. Moscow, Univ. tip., pp. 29-30.
- Yatsko V.A., 2014. *Kompyuternaya lingvistika ili lingvisticheskaya informatika* [Computational Linguistics or Linguistic Informatics]. *Nauchno-tekhnicheskaya informatsiya. Seriya 2. Informatsionnye protsessy i sistemy* [Scientific and Technical Information. Series 2. Information Processes and Systems], no. 5, pp. 1-10.
- Ahmad K., Gillam L., Tostevin L., 1999. University of Surrey Participation in TREC8: Weirdness Indexing for Logical Document Extrapolation and Retrieval (Wilder). *Proc. 8th Text Retrieval Conference TREC*. Gaithersburg, s.n., pp. 717-724.
- Baranov V., 2018. A Text Corpus of Medieval Manuscripts as a Goal and a Tool for Linguistic Research. *Editing Mediaeval Texts from a Different Angle: Slavonic and Multilingual Traditions*. Leuven, Paris, Bristol, Peeters, pp. 283-308.
- Baranov V.A., Gnutikov R.M., 2018. The Statistics and n-Gram Modules of the Historical Corpus “Manuscript”. *Digital and Analytical Approaches to the Written Heritage: Proceedings of the 7th International Conference El’Manuscript “Textual Heritage and Information Technologies”*. Sofia, Gutenberg Publ., pp. 9-28.
- Gillam L., Tariq M., Ahmad K., 2005. Terminology and the Construction of Ontology. *Terminology*, vol. 11, iss. 1, pp. 55-81.
- Jagić V., 1913. *Entstehungsgeschichte der kirchenslavischen Sprache*. Berlin, Weidmann. 540 p.
- Kwok K.L., 1995. A Network Approach to Probabilistic Information Retrieval. *ACM Transactions on Information Systems*, vol. 13, no. 3, pp. 324-353.
- Rayson P., Garside R., 2000. Comparing Corpora Using Frequency Profiling. *Proceedings of the Comparing Corpora Workshop at ACL 2000*. Hong Kong, pp. 1-6. URL: http://ucrel.lancs.ac.uk/people/paul/publications/rg_acl2000.pdf
- Robertson S., 2004. Understanding Inverse Document Frequency: On Theoretical Arguments for Idf. *Journal of Documentation*, no. 60, pp. 503-520. URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.438.2284&rep=rep1&type=pdf>
- Roelleke T., 2013. *Information Retrieval Models: Foundations and Relationships*. Cham, Springer. 141 p.
- Roelleke T., Wang J., 2006. A Parallel Derivation of Probabilistic Information Retrieval Models. *Proceedings of the 29th Annual ACM SIGIR Conference on Research and Development in Information Retrieval*. Seattle, New York, ACM, pp. 107-114.
- Salton G., Yang C.S., 1973. On the Specification of Term Values in Automatic Indexing. *Journal of Documentation*, vol. 29, pp. 351-372.

Sparck J.K., 1972. A Statistical Interpretation of Term Specificity and Its Application in Retrieval. *Journal of Documentation*, vol. 28, pp. 11-21.

Voss Chr., 1996. Die Vertretung von lexikalischen textologischen Dubletten der Dichotomie Ochrid-Preslav in kirchenslavischen Abschriften der Paränesis Ephreaims des Syrsers. Ein Beitrag zur Datierung der altbulgarischen Erstübersetzung.

Anzeiger für slavischen Philologie, Bd. 24, S. 95-127.

Wu H.C., Luk R.W.P., Wong K.F., Kwok K.L., 2008. Interpreting TF-IDF Term Weights as Making Relevance Decisions. *ACM Transactions on Information Systems*, vol. 26, no. 3, art. 13. URL: https://www.scss.tcd.ie/khurshid.ahmad/Research/Sentiments/tfidf_relevance.pdf

Information About the Author

Victor A. Baranov, Doctor of Sciences (Philology), Professor, Head of the Department of Linguistics, Kalashnikov Izhevsk State Technical University, Studencheskaya St, 7, 426069 Izhevsk, Russia; Project Executor, Kazan Federal University, Kremlevskaya St, 18, 420008 Kazan, Russia, victor.a.baranov@gmail.com, <https://orcid.org/0000-0003-1730-6359>

Информация об авторе

Виктор Аркадьевич Баранов, доктор филологических наук, профессор, заведующий кафедрой лингвистики, Ижевский государственный технический университет им. М.Т. Калашникова, ул. Студенческая, 7, 426069 г. Ижевск, Россия; исполнитель проекта, Казанский (Приволжский) федеральный университет, ул. Кремлевская, 18, 420008 г. Казань, Россия, victor.a.baranov@gmail.com, <https://orcid.org/0000-0003-1730-6359>