

DOI: <https://doi.org/10.15688/jvolsu2.2020.6.4>

UDC 811.161.1'04:27-247

LBC 81.411.2-03

Submitted: 30.06.2019

Accepted: 12.10.2020

**QUANTITATIVE INVESTIGATION OF THE PANTELEYMON GOSPEL DATING
FROM THE LATE 12th TO THE EARLY 13th CENTURIES
(THREE STATISTICAL EXPERIMENTS)**

Victor A. Baranov

Kalashnikov Izhevsk State Technical University, Izhevsk, Russia

Oksana V. Zuga

Udmurt State University, Izhevsk, Russia

Abstract. The work presents the results of the quantitative and statistical comparative analysis of the most frequent word forms and combinations of the Old Russian of the Panteleymon Gospel (RNB, Sof. 1). The work aims to reveal the degree of closeness of the Panteleymon Gospel to the other gospels and the medieval Slavonic texts of other genres, represented in sub-corpora of historical corpus “Manuscript: Slavic Written Heritage”. The work was carried out with the help of the special modules of statistics and *n*-grams. The comparison of the lists of single-, two- and three-component linguistic units, automatically extracted from the manuscripts, with the respective lists of several sub-corpora points to the presence of the quantitative-statistical characteristics of the linguistic components of the manuscripts which can be recognized as important. The data of the three experiments are summarized. The first experiment showed that the smallest differences of the frequency lists exist between the Panteleymon Gospel and the sub-corpus of complete aprakoses and the greatest differences between the manuscript being analyzed and the sub-corpus of short aprakoses. This makes possible to recognize that the composition of the lists, the order and the relative frequency of the forms in them are the important characteristics of the manuscript or the sub-corpus. The application of the Weirdness measure helped to extract from the Panteleymon Gospel the word forms which are supposed to be significant – those, having the highest weight within the sub-corpora of different genres (*вамь, имь, азъ, емоу, рече, аще*). It has been established that the volume and composition of contrasted sub-corpora do not influence the result, and the use of the collections of complete and short aprakoses as contrast sub-corpora helped to specify the list of such forms (*яко, къ, бо, о(т), имь, есть, аще*). The investigation of two- and three-component combinations, extracted with the help of the statistical measure T-score, gave the following results: a list of fixed combinations – invariable composition formulas (*ев[ан](г)[елие] ш(т) ма[т](о)[ея]* etc.), inherent to all gospels, was made; entire grammatical structures (*аже даль юси* etc.) were listed, as well as stable semantic complexes and their parts (*[да] любите другъ друга* etc.). Statistically important sequences having in the Panteleymon Gospel a statistical weight, which is considerably higher than in the contrast sub-corpora – *нѣсте ли чьли, имать животъ вѣчьныи* etc. have been revealed.

Key words: Old Russian manuscripts, Panteleymon Gospel, statistical methods, key words, *n*-grams.

Citation. Baranov V.A., Zuga O.V. Quantitative Investigation of the Panteleymon Gospel Dating from the Late 12th to the Early 13th Centuries (Three Statistical Experiments). *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 2. Yazykoznanie* [Science Journal of Volgograd State University. Linguistics], 2020, vol. 19, no. 6, pp. 43-57. (in Russian). DOI: <https://doi.org/10.15688/jvolsu2.2020.6.4>

УДК 811.161.1'04:27-247

ББК 81.411.2-03

Дата поступления статьи: 30.06.2019

Дата принятия статьи: 12.10.2020

**ОПЫТ КВАНТИТАТИВНОГО ИССЛЕДОВАНИЯ
ПАНТЕЛЕЙМОНОВА ЕВАНГЕЛИЯ КОНЦА XII – НАЧАЛА XIII В.
(ТРИ СТАТИСТИЧЕСКИХ ЭКСПЕРИМЕНТА)¹**

Виктор Аркадьевич Баранов

Ижевский государственный технический университет им. М.Т. Калашникова, г. Ижевск, Россия

Оксана Владимировна Зуга

Удмуртский государственный университет, г. Ижевск, Россия

Аннотация. Цель работы – выявление степени близости Пантелеймонова Евангелия (РНБ, Соф. 1) другим Евангелиям и средневековым славянским текстам других жанров, представленным в подкорпусах исторического корпуса «Манускрипт: славянское письменное наследие». Работа выполнена с помощью специализированных модулей статистики и *n*-грамм. Сопоставление перечней автоматически извлеченных их рукописи одно-, двух- и трехкомпонентных лингвистических единиц с соответствующими перечнями нескольких подкорпусов позволяет определить такие количественно-статистические характеристики лингвистических компонентов рукописей, которые могут быть признаны существенными. Обобщены данные трех экспериментов. Первый эксперимент показал, что наименьшие отличия частотных списков существуют между Пантелеймоновым Евангелием и полными апракосами, а наибольшие – между анализируемой рукописью и краткими апракосами. Это позволяет признать, что состав перечней, порядок следования и относительная частота форм в них являются существенными характеристиками рукописи или подкорпуса. В ходе второго эксперимента, проведенного с применением статистической меры Weirndness, из Пантелеймонова Евангелия извлечены словоформы, претендующие на роль значимых – имеющие максимально высокий вес на фоне разножанровых подкорпусов. Установлено, что объем и состав контрастного подкорпуса не влияют на результат; применение в качестве контрастных подкорпусов коллекций полных и кратких апракосов позволило уточнить список таких форм. В ходе третьего эксперимента (анализ двух- и трехкомпонентных сочетаний, извлеченных с помощью статистической меры T-score) был установлен перечень устойчивых сочетаний (неизменяемых композиционных формул, цельных грамматических структур, устойчивых семантических комплексов и их частей), свойственных всем Евангелиям, и выявлены статистически значимые последовательности, имеющие в Пантелеймоновом Евангелии статистический вес значительно выше, чем в контрастных подкорпусах.

Ключевые слова: древнерусские рукописи, Пантелеймоново Евангелие, статистические методы, ключевые слова, *n*-граммы.

Цитирование. Баранов В. А., Зуга О. В. Опыт квантитативного исследования Пантелеймонова Евангелия конца XII – начала XIII в. (три статистических эксперимента) // Вестник Волгоградского государственного университета. Серия 2, Языкознание. – 2020. – Т. 19, № 6. – С. 43–57. – DOI: <https://doi.org/10.15688/jvolsu2.2020.6.4>

Средневековые славянские тексты как объект и предмет исследования

Статистический анализ давно и продуктивно используется для решения различных задач в области русской филологии и лингвистики². В настоящее время необходимый материал для этого предоставляют большие текстовые коллекции и корпуса. Создание машиночитаемых копий средневековых славянских текстов, их разметка и размещение в специализированных системах хранения и обработки данных дает возможность начать статистические эксперименты и на таком материале.

Одним из интернет-ресурсов, содержащих размеченные транскрипции средневековых славянских рукописей, является текстовый корпус «Манускрипт: славянское письменное наследие», в котором размещено более 130 рукописей и отрывков X–XV вв. объемом более 3,5 млн текстовых прецедентов и который

снабжен специализированными инструментами для обработки, поиска и демонстрации лингвистических данных³.

Самой большой коллекцией корпуса является собрание Евангелий XI–XIV вв., включающее и один из ранних русских списков полного апракоса – новгородское Пантелеймоново Евангелие, не часто привлекавшее внимание лингвистов.

Общая характеристика рукописи и ее электронное издание

Пантелеймоново Евангелие (далее – ЕП) было создано в конце XII – начале XIII (?) в. [Сводный каталог..., 1984, с. 167] (другая датировка – XII в. [Марков, 2001]) и названо по имени св. Пантелеймона, изображенного на л. 224. По содержанию и композиции представляет собой полный апракос, в котором имеются евангельские чтения на субботние, воскресные и будние дни недели. Памятник

хранится в отделе рукописей Российской национальной библиотеки (РНБ), в Софийском собрании (Соф.), под № 1. Рукопись содержит 224 листа; не издавалась. Вследствие ветхости памятник недоступен широкому кругу читателей.

Интерес к ЕП не случаен: в ряду древнейших церковнославянских письменных памятников русского извода у него особое место как у одной из рукописей, в которой отразились значимые для истории русского языка особенности фонетики и грамматики [Зуга, 2009; Марков, 2001]. Сравнение текстов Остромирова и Пантелеймонова Евангелий дало возможность В.М. Маркову обнаружить яркие различия, свидетельствующие о происходивших в русском языке XII в. изменениях. Одним из достижений работы ученого стала демонстрация эффективности и результативности сопоставления рукописей, содержащих одни и те же тексты: «...даже прямолинейное сопоставление текстов приводит, как кажется, к достаточно значимым результатам, коль скоро дело касается источников, разнесенных во времени и, вместе с тем, представляющих однородный языковой материал» [Марков, 2001, с. 33].

Сказанное позволяет продолжить сопоставление ЕП и других евангельских списков с помощью иных методов, например корпусных и количественно-статистических. Дает ли их применение возможность обнаружить значимые характеристики текстов, можно понять, только проведя соответствующие эксперименты. В статье представлены результаты автоматического извлечения одно-, двух- и трехкомпонентных текстовых единиц из ЕП, сопоставление их с данными других евангельских текстов, объединенных в три подкорпуса, с целью выявления степени близости ЕП последним.

Возможность выполнить эксперименты предоставляют пользовательские сервисы корпуса «Манускрипт: славянское письменное наследие», в который включена интернет-версия электронного издания рукописи⁴. Оно позволяет познакомиться с археографической, текстологической, лингвистической, библиографической информацией о рукописи и тексте, получить сведения об их структуре и составе, просмотреть

полный текст рукописи (машиночитаемую транскрипцию, содержащую лингвистическую и аналитическую разметку) и текст дипломатического издания (преобразованный текст), построить прямые, обратные, количественные, сравнительные указатели слов и словоформ, конкордансы.

Эти данные послужили материалом нескольких статистических экспериментов для нахождения таких количественных характеристик Пантелеймонова Евангелия, которые отличают его от других рукописей корпуса. Объектом анализа стали автоматически извлеченные одно-, двух- и трехкомпонентные сочетания текстовых форм, статистическая значимость которых устанавливается с помощью сопоставления с коллекциями полных и кратких апракосов, а также коллекциями текстов других жанров⁵.

Количественный и статистический анализ Пантелеймонова Евангелия

Общие сведения об объеме данных

Объем рукописи ЕП – 68 734 текстовые формы. Базовый подкорпус русских списков Евангелий XI–XIV вв.⁶: количество рукописей – 9, отрывков – 2; объем – 522 793 текстовые формы, а также два старославянских списка – Ассеманиево Евангелие и Саввина книга; объем – 76 644 формы.

Эксперимент 1. Наиболее частотные словоформы

В работе [Баранов, 2019в] показаны существенные различия между составом и порядком следования первых десяти наиболее частотных слов в подкорпусе русских списков Евангелий и в подкорпусах рукописей других жанров, а также между их статистическими оценками.

Единство текстов Евангелий с точки зрения количественных характеристик подтверждается составом, порядком следования и относительным количеством наиболее частотных словоформ ЕП и трех подкорпусов Евангелий – русских списков полного апракоса, русских списков краткого апракоса и старославянских списков (см. таблицу 1⁷).

Таблица 1. Наиболее частотные словоформы в ЕП и подкорпусах Евангелий

Table 1. The most frequent word forms in the EP and subcorps of the Gospels

R	ЕП			Евангелия (АП)			Евангелия (АК)			Евангелия (ст.-слав.)		
	w	F	f	w	F	f	w	F	f	w	F	f
1	н	5516	0,080	н	17281	0,079	н	12711	0,077	н	5871	0,077
2	же	2344	0,034	же	9093	0,041	же	6851	0,041	же	3099	0,040
3	вз	2186	0,032	вз	6924	0,031	вз	5118	0,031	вз	1740	0,023
4	не	1498	0,022	не	4557	0,021	сѧ	3321	0,020	сѧ	1579	0,021
5	сѧ	1462	0,021	сѧ	4475	0,020	не	3319	0,020	не	1479	0,019
6	ѿ	1170	0,017	ѿ	3555	0,016	ѿ	2799	0,017	ѿко	988	0,013
7	ѿко	1031	0,015	ѿко	2959	0,013	на	1854	0,011	на	908	0,012
8	на	785	0,011	на	2462	0,011	реѹе	1743	0,010	емоуѹ	755	0,010
9	реѹе ⁸	734	0,011	реѹе	2266	0,010	емоуѹ	1618	0,010	реѹе	746	0,010
10	емоуѹ	626	0,009	емоуѹ	1941	0,009	ѿко	1605	0,010	отъ	719	0,009
11	естъ	585	0,009	его	1816	0,008	его	1408	0,008	да	586	0,008
12	да	530	0,008	естъ	1795	0,008	да	1240	0,007	его	579	0,008
13	къ	523	0,008	да	1674	0,008	естъ	1228	0,007	се	578	0,008
14	его	493	0,007	се	1437	0,007	се	1074	0,006	естъ	538	0,007
15	се	450	0,007	аше	1372	0,006	къ	1028	0,006	къ	506	0,007
16	нмъ	434	0,006	нмъ	1376	0,006	бам	934	0,006	нмъ	387	0,005
17	бамъ	428	0,006	бам	1299	0,006	съ	912	0,005	аше	385	0,005
18	аше	414	0,006	къ	1253	0,006	аше	870	0,005	бо	363	0,005
19	бо	396	0,006	съ	1195	0,005	нмъ	870	0,005	бам	372	0,005
20	съ	363	0,005	бо	1185	0,005	азъ	836	0,005	о	367	0,005
21	азъ	292	0,004	азъ	911	0,004	бо	775	0,005	азъ	348	0,005

Примечание. Евангелия (АП) – списки полного апракоса, без ЕП (объем – 220 023); Евангелия (АК) – списки краткого апракоса (объем – 166 153); Евангелия (ст.-слав.) – два старославянских списка (объем – 76 644). R – ранг словоформы; w – словоформа; F – абсолютная частота; f – относительная частота ⁹. В случае вариативности текстовых прецедентов в качестве маски использовались регулярные выражения, например: (н|г), (с|ѧ|ѿ), (ѿ|ѧ|ѿ)к(о|о|ѡ), (н|е|е)с(н|г), ([о о ѡ] т. | ѿ), (н|е|е)м(оу|ѹ|ѹ|ѹ) и под. Различно переданные в транскрипциях корпуса сѧ и же в постпозиции и не в препозиции, а также формы местоимения н в косвенных падежах (емоуѹ/емоуѹ же, его/его же) в таблице приведены суммарно.

Сопоставление первых 10 наиболее частотных форм позволяет увидеть: а) полную идентичность состава ЕП и других полных апракосов и порядка следования форм в них; б) различия полных и кратких апракосов в порядке следования форм; в) отличия в порядке следования форм в старославянских списках по сравнению с русскими списками; г) заметно различную относительную частоту союза ѿко в русских списках полных и кратких апракосов (0,015–0,013 vs 0,010) и предлога ѿ/отъ в русских и старославянских списках (0,016–0,017 vs 0,009).

Перечни форм, имеющих ранги с 11-го по 21-й: а) в ЕП и полных апракосах идентичны по составу, близки по порядку следования и по

относительному количеству (относительные значения 7 форм равны, 3 формы различаются на 0,001, 1 форма – на 0,002); б) в ЕП и в кратких апракосах также совпадают по составу, различаются порядком следования и относительной частотностью (относительные значения 2 форм равны, 7 форм различаются на 0,001, 2 форм – на 0,002); в) в кратких русских и старославянских апракосах различаются по составу (рус. съ, ст.-слав. о), близки по порядку следования и относительному количеству (относительные значения 6 форм не различаются, 3 форм отличаются на 0,001, 1 формы – на 0,002). Наименьшие различия находим в парах «ЕП – полные апракосы» и «краткие апракосы – старославянские списки», наи-

большие – в «ЕП – краткие апракосы», что понятно: ЕП является полным апракосом.

Обнаруженные схожесть и контрастность между наиболее часто встречающимися формами, имеющими ранги с 1-го по 21-й, в ЕП, подкорпусах русских списков кратких и полных Евангелий и в старославянских рукописях позволяют сделать вывод о неслучайности существующих между подкорпусами совпадений и различий и признать, что состав, порядок следования и относительная частота форм в совокупности являются существенными характеристиками документа или подкорпуса.

Эксперимент 2. Значения наиболее частотных форм в соответствии с мерой Weiridness

Таблица 2. Вес наиболее частотных словоформ в ЕП в соответствии с мерой Weiridness

Table 2. Weight of the most frequent word forms in the EP according to the Weiridness measure

R	w	F	f	R ₁	Weiridness ₁	R ₂	Weiridness ₂	R ₃	Weiridness ₃	R ₄	Weiridness ₄
1	н	5516	0,080	14	1,722	15	1,216	12	1,022	10	1,049
2	же	2344	0,034	12	1,934	14	1,269	21	0,825	21	0,827
3	въ	2186	0,032	13	1,750	13	1,356	15	1,011	12	1,032
4	не	1498	0,022	19	1,233	17	1,061	6	1,052	8	1,091
5	сѧ	1462	0,021	21	0,636	21	0,649	7	1,046	9	1,064
6	ѡ	1170	0,017	11	2,061	8	1,834	5	1,054	16	1,010
7	ѧко	1031	0,015	16	1,368	10	1,510	2	1,115	1	1,553
8	на	785	0,011	18	1,252	18	0,956	13	1,021	14	1,024
9	реуе	734	0,011	2	29,911	2	7,487	8	1,044	13	1,025
10	кмоу	626	0,009	6	14,702	5	3,188	10	1,032	18	0,935
11	ксть	585	0,009	8	4,888	9	1,814	9	1,043	5	1,152
12	да	530	0,008	9	3,618	11	1,507	14	1,013	11	1,033
13	къ	523	0,008	15	1,582	16	1,077	1	1,336	3	1,230
14	нго	493	0,007	10	2,161	12	1,382	20	0,869	19	0,846
15	се	450	0,007	7	5,945	6	2,120	17	1,002	15	1,013
16	нмъ	434	0,006	4	22,517	3	4,581	16	1,010	4	1,206
17	бамъ	428	0,006	1	174,472	1	8,409	4	1,055	7	1,108
18	аше	414	0,006	3	27,797	7	2,073	19	0,966	6	1,150
19	бо	396	0,006	17	1,359	19	0,779	3	1,070	2	1,235
20	съ	363	0,005	20	0,877	20	0,664	18	0,972	17	0,962
21	азъ	292	0,004	5	17,542	4	3,371	11	1,026	20	0,844

Примечание. R, R₁, R₂ – ранги словоформ; w – словоформа; F – абсолютная частота; f – относительная частота; Weiridness₁ – вес словоформы в ЕП при сравнении ее частотности в трех подкорпусах (списков майских служебных миней, миней на другие месяцы года и стихирарей); Weiridness₂ – вес словоформы в ЕП при сравнении ее частотности в семи подкорпусах (списков майских служебных миней, миней на другие месяцы года, стихирарей, Апостолов, Паренесисов, Псалтырей и летописей¹¹); Weiridness₃ – вес словоформы в ЕП при сравнении ее частотности в подкорпусе полных апракосов; Weiridness₄ – вес словоформы в ЕП при сравнении ее частотности в подкорпусе кратких апракосов.

Weirdness, полученный при сравнении абсолютной частоты употребления формы в ЕП с ее количеством в различных подкорпусах, а также ранги словоформы – номер по порядку в каждом из сопоставлений.

Представим эти данные в виде диаграмм (рис. 1–4). На рисунках 1 и 2 показаны значения в соответствии с $Weirdness_1$ и $Weirdness_2$, на рисунках 3 и 4 – в соответствии с $Weirdness_3$ и $Weirdness_4$.

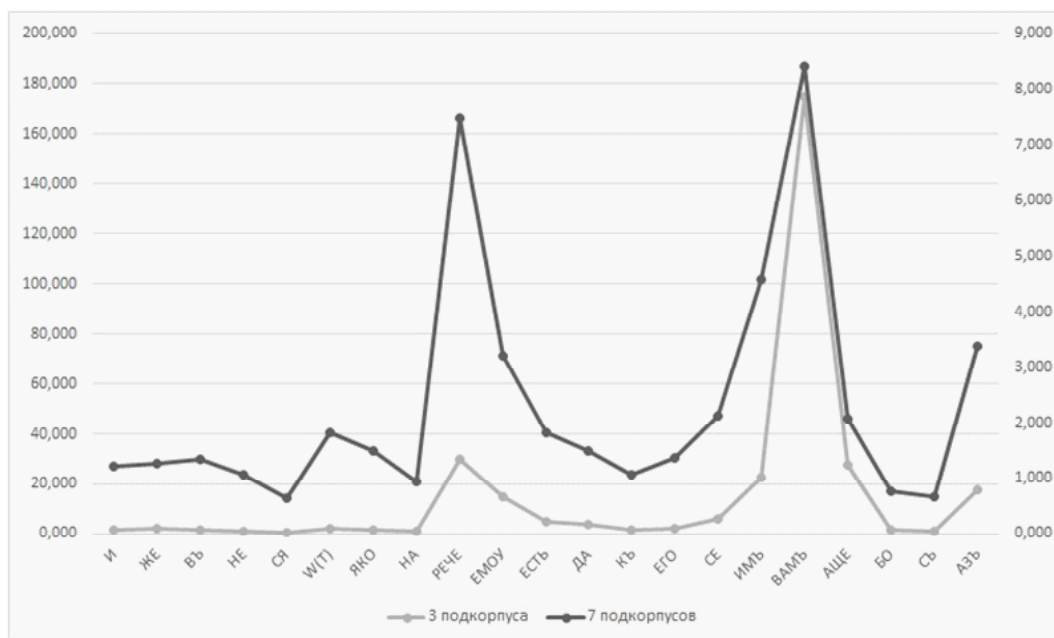


Рис. 1. Значения наиболее частотных словоформ ЕП в соответствии с мерой Weirdness при сравнении с тремя и семью разножанровыми корпусами

Fig. 1. Values of the most frequent EP word forms in accordance with the Weirdness measure when compared with three and seven different-genre corpora

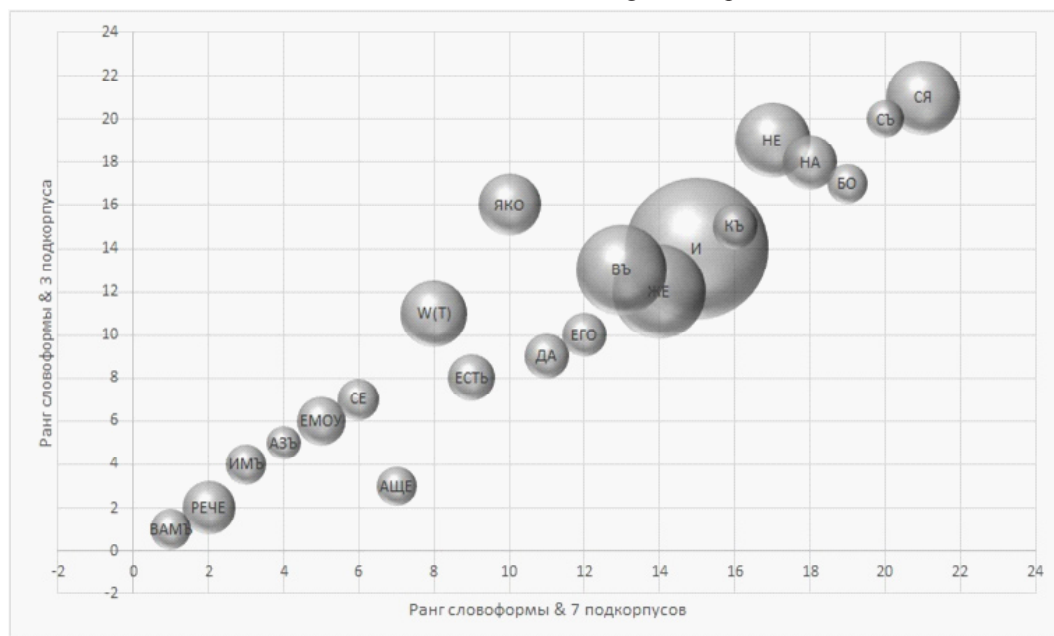


Рис. 2. Ранги наиболее частотных словоформ ЕП в соответствии с мерой Weirdness при сравнении с тремя и семью разножанровыми корпусами

Fig. 2. Ranks of the most frequent EP word forms according to the Weirdness measure when compared with three and seven different-genre corpora

На диаграммах хорошо видно, что изменение объема альтернативных подкорпусов и различия в их жанрах не приводят к принципиальному изменению значимости (веса, ранга) словоформ ЕП. Несмотря на то что величина меры Weiridness для некоторых форм может значительно отличаться при ее вычислении на основе различных альтернативных подкорпусов, степень значимости («странности») форм практически идентична (см. рис. 1: местоимения **ВАМЪ**, **АЗЪ**, союз **АЩЕ**, глагол **РЕЧЕ** и некоторые другие). Это демонстрируется также соответствием рангов одной и той же формы друг другу (см. рис. 2: местоимение **ВАМЪ** имеет ранги 1-1, глагол **РЕЧЕ** – 2-2 и т. д.).

Таким образом, данные диаграммы позволяют увидеть формы, характеризующие ЕП на фоне подкорпусов других жанров. В пределах 10 первых рангов находятся формы местоимений **ВАМЪ**, **НМЪ**, **АЗЪ**, **НМОУ**, глагол **РЕЧЕ**, союз **АЩЕ**, местоимение (союз-частица) **СЕ**.

Сравнение частотности словоформ в ЕП с их частотностью в полных и кратких апракосах демонстрирует как идентичную, так и существенно различающуюся оценку некоторых форм с помощью меры Weiridness.

Некоторая часть значений лежит в области значения 1,0 (см. рис. 3; на рис. 4 эти формы находятся на линии 0-0 – 24-24 и близко к ней), что свидетельствует об отсутствии значимых расхождений в частотности этих форм в ЕП и в двух альтернативных подкорпусах.

В то же время значения некоторых форм несколько или значительно превышают 1,0, что говорит об их большей относительной частотности в ЕП по сравнению с тем или иным подкорпусом: союз **КАКО**, предлог **КЪ**, частица **БО** и имеют максимально высокий ранг (на рисунке 4 эти формы находятся в начале осей рангов). С большой долей вероятности можно считать, что частотность этих форм является особенностью именно ЕП на фоне других списков.

Одновременно видно, что оценка некоторых форм с помощью двух альтернативных подкорпусов (кратких и полных апракосов) существенно различна и даже противоположна. Так, предлог **Ѡ** имеет низкое значение (ранг 16) на фоне кратких апракосов и высокое (ранг 5) при сравнении с полными, и наоборот: словоформа **НМЪ**, глагольная форма **КСТЬ**, союз **АЩЕ** высоко оцениваются при сравнении ЕП с краткими апракосами, но значительно ниже на фоне подкорпуса полных (см. рис. 4).

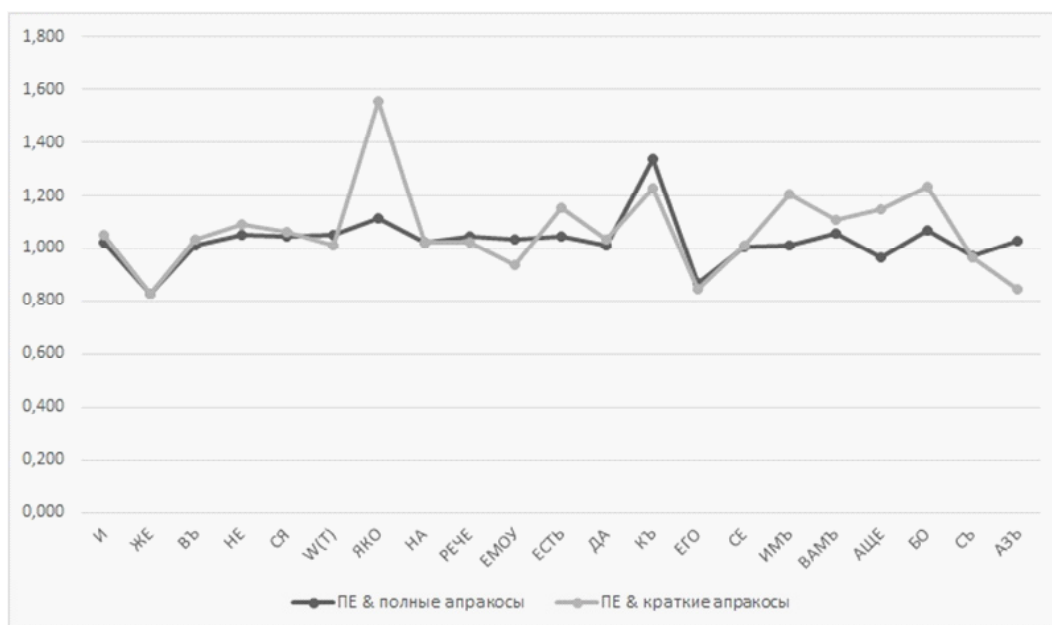


Рис. 3. Значения наиболее частотных словоформ ЕП в соответствии с мерой Weiridness при сравнении с подкорпусом полных и кратких апракосов

Fig. 3. Values of the most frequent EP word forms in accordance with the Weiridness measure when compared with the subcorpus of full and short aprakos

Таким образом, можно заключить, что союз **яко**, предлог **къ**, частица **бо**, а также предлог **ѿ** (на фоне других рукописей полного апракоса), местоимение **нмъ**, глагольная форма **ѣсть**, союз **аще** (при сравнении с краткими апракосами) являются особенностью списка ЕП.

Подобные соотношения, демонстрирующие расхождения частотности служебных слов и некоторых форм местоимений и глагола **быти** в ЕП и подкорпусах разного состава, позволяют говорить о значимых количественно-статистических тенденциях, реализующихся в превышении средней частотности форм ЕП над средней частотностью аналогичных форм в контрастных подкорпусах.

Эксперимент 3. Извлечение статистически значимых сочетаний

Как известно, мера T-score дает возможность выявить в документе сочетания, частотность совместного использования компонентов которых выше статистически ожидаемой средней величины, иначе – найти такие сочетания, компоненты которых используются друг с другом чаще, чем с другими формами в тексте. Значение меры зависит от количества

сочетаний компонентов и от количества каждого из компонентов в документе ¹².

Приведем первые 20 биграмм ЕП, русских и старославянских списков (см. табл. 3 ¹³).

Перечни близки по составу: 9 из 20 сочетаний трех выборок идентичны (**въ оно**, **глаголю вамъ**, **рече нмъ**, **рече гъ**, **онъ же**, **рече кмоу**, **глаголю кмоу**, **онн же**, **дане**), другие совпадают в двух группах (ЕП – др.-рус.: **же рече**, **съ нмъ**, **оученици кго**; ЕП – ст.-слав.: **къ немюу**, **ѿ матъѹ**, **глаголю нмъ**, **н рече**; др.-рус. – ст.-слав.: **рече же**, **же въ**).

Сама по себе подобная близость показательна и иллюстрирует жанровое единство сопоставляемых документов и подкорпусов.

Иная картина в начале перечней триграмм. Приведем первые 20 в ЕП, древнерусских и старославянских рукописях (см. табл. 4).

Состав статистически наиболее значимых триграмм так же, как и биграмм, безусловно, определен жанром и включает евангельские структурные, грамматические, семантические единства. Так, мерой высоко оценены композиционные формулы (**ѣвѣ ѿ ма**, **пама прѣбна оца**, **повелнѣ**

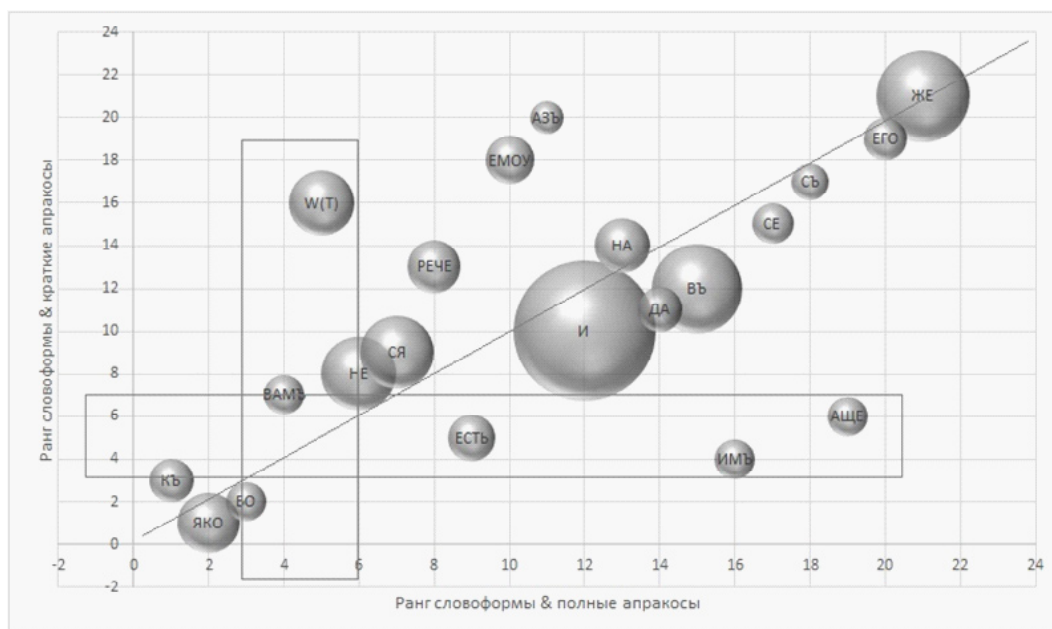


Рис. 4. Ранги наиболее частотных словоформ ЕП в соответствии с мерой Weirdness при сравнении с подкорпусом полных и кратких апракосов

Fig. 4. Ranks of the most frequent EP word forms according to the Weirdness measure when compared with the subcorpus of full and short aprakos

дне, [реуе] г̃ прнтзую сню и под.), грамматически единые структуры (нѣсте лн ѹьлн – Мф. 12: 3, Мф. 22: 31, Мф. 19: 4, Мк. 12: 2 ЕП; аже далъ еси – Ин. 17: 12 и др.), устойчивые семантические комплексы или их части (нмѣан оушн слышатн [да слышнтъ] – Мф. 11: 15, Мф. 13: 9, Мк. 4: 9, Лк. 8: 8 и др.; [да] люенте дроугъ дроуга – Ин15:17 и нек. др.).

Некоторые триграммы являются составной частью сочетаний, включающих большее количество грамматических или семантических компонентов: об онъ полъ – онъ полъ мора – Ин. 6: 22, 25 и др., донъдеже положю врагы – положю врагы твоа – твоа подъножнн ногама – подъножнн ногама твонма, Мф. 22: 44, Мк. 12: 36, Лк. 20: 43, къ овьцамъ погыбъшнмъ – овьцамъ погыбъшнмъ домоу [нздранлева] – Мф. 10: 06, Мф. 15: 24,

[реуе] г̃ прнтзую сню – Мф. 21: 33, Мф. 22: 1, Лк. 16: 1 и др.

Перечни ЕП и русского подкорпусов частично пересекаются. Так, в список 20 наиболее частотных форм входят нмѣан оушн слышатн [да слышнтъ], [да] люенте дроугъ дроуга, триграммы фраз донъдеже положю врагы твоа подъножнн ногама твонма, об онъ полъ мора / нордана и др.

Увеличение количества анализируемых триграмм, получивших высокое значение меры T-score, до 50 позволяет выявить и другие семантически цельные сочетания. Например, в ЕП и русских списках – [не трѣбують] съдравнн враѹа нъ болащнн – Лк. 5: 31; вндѣрнзѹи днннн [лежаща] – Лк. 24: 12; нмоу же нѣсмь достоннн [отрѣшнтн ремень сапогоу его] – Ин. 1: 27; в русских и старославянских списках – азъ нсмь лоза нстнннннн – Ин. 15: 1.

Таблица 3. Перечень первых 20 статистически значимых биграмм (мера T-score)

Table 3. List of the first 20 statistically significant bigrams (T-score measure)

№ п/п	ЕП			Древнерусские списки			Старославянские списки		
	Биграмма	F	T-score	n-грамма	F	T-score	Биграмма	F	T-score
1	въ оно	194	13.467	онъ же	597	23.640	ѳѡ ѡ	157	12.497
2	г̃лю вамъ	129	11.247	реуе же	621	22.021	реуе же	175	11.976
3	реуе нмъ	130	11.001	г̃лю вамъ	469	21.469	реуе г̃	122	10.724
4	реуе г̃	125	10.997	въ оно	501	21.456	онъ же	106	9.982
5	онъ же	123	10.724	оно врема	418	20.405	реуе нмоу	105	9.629
6	къ нмоу	112	10.502	реуе нмъ	431	20.007	реуе нмъ	91	9.168
7	реуе нмоу	93	9.062	же р̃е	466	19.311	г̃л̃ѡ вамъ	83	8.835
8	г̃ла нмоу	77	8.557	оууеннцн нго	329	17.895	къ нмоу	102	8.357
9	ѡ матьѳ	74	8.501	р̃е нмъ	319	17.265	же ѳѡ	74	7.887
10	же реуе	105	8.358	сконмъ оууеннкомъ	292	17.062	же бѣ	76	7.875
11	г̃ла нмъ	64	7.805	онн же	306	17.025	въ оно	63	7.758
12	ѡ лоукы	62	7.782	же бѣ	367	16.846	ѡ маѳ	59	7.662
13	н реуе	149	7.384	азъ нсмь	287	16.813	н реуе	155	7.537
14	онн же	57	7.344	да не	384	16.465	г̃ла нмоу	56	7.272
15	съ ннмь	54	7.303	въ оно	280	16.216	л̃оу г̃ла	50	7.040
16	оууеннцн нго	53	7.191	реуе нмоу	299	16.180	онн же	52	7.036
17	къ ннмъ	48	6.866	р̃е г̃	266	16.155	да не	67	6.927
18	да не	66	6.824	г̃ла нмоу	272	16.088	г̃ла нмоу	49	6.836
19	амннъ г̃лю	45	6.672	съ ннмь	261	16.019	же нѣ	56	6.740
20	въ домъ	48	6.621	ннкто же	270	16.007	г̃ла нмъ	37	5.937

Таблица 4. Перечень первых 20 статистически значимых триграмм (мера T-score)¹⁴

Table 4. List of the first 20 statistically significant trigrams (T-score measure)

№ п/п	ЕП			Древнерусские списки			Старославянские списки		
	n-грамма	F	T-score	n-грамма	F	T-score	n-грамма	F	T-score
1	нѣстѣ лн ѹьлн	7	2.402	прпѣнаа оцѣ наше	52	6.780	ѣба ѿ ма	53	3.626
2	нмать жнвотъ бѣуьнъин	8	2.327	памъ прпѣнаа оцѣ	57	6.700	гѣ прнтъѹж снѣж	12	3.324
3	нмѣян оушн слъшатн	5	2.206	стра стго мунѣка	56	6.551	по вѣлнцѣ дне	10	2.922
4	онъ полъ мора	5	2.140	нмѣян оушн слъшатн	22	4.405	амн амн глѣж	14	2.745
5	об онъ полъ	5	2.124	люентѣ дроугъ дроуга	22	4.303	об онъ полъ	8	2.698
6	гѣ прнтъѹю сню	10	2.014	стра стго сѣномунѣка	23	4.248	но лѣ ѣ	8	2.674
7	подъножнѣ ногама твонма	4	1.999	стра стхъ мунѣкъ	18	4.148	гѣ своимъ оуѹеннкомъ	10	2.633
8	вѣсею дшѣю своею	4	1.995	прнше дъшнмъ к нѣмоу	21	3.835	снмоне нонннъ люеншн	5	2.233
9	овѣцамъ погъбѣшнмъ домоу	4	1.994	ѹа нѣн сѣптѣм	52	3.757	всѣ нмѣннѣ свое	5	2.153
10	твоѣ подъножнѣ ногама	4	1.988	положю врагъи твоѣ	15	3.675	аже далъ сн	6	2.149
11	донъдеже положю врагъи	4	1.980	мъзда баша многа	15	3.579	нонннъ люеншн лн	5	2.146
12	положю врагъи твоѣ	4	1.979	снмоне нонннъ люеншн	12	3.436	гѣ прнтъѹж снѣж	5	2.114
13	люентѣ дроугъ дроуга	4	1.968	к нѣмоу нюдѣомъ	20	3.428	ѣба ѿ лоу	13	2.107
14	вѣставъ запрѣтн вѣтроу	4	1.966	а дроугокѣ паде	24	3.316	ѣба ѿ ма	7	1.982
15	вѣра твоѣ спсе	4	1.960	об онъ полъ	20	3.270	прнтъѹж снѣж подобно	4	1.974
16	кмоу же нѣсмь достоннъ	4	1.940	коупнша нмь село	11	3.259	хѣалж тебѣ въздаѣж	4	1.957
17	къ овѣцамъ погъбѣшнмъ	4	1.939	нлн нного ѹанмъ	13	3.236	вѣлнцѣ дне мѣ	4	1.934
18	домъ вашъ поустъ	4	1.924	правъи творнтѣ стъза	10	3.132	ѿ лоу гѣ	5	1.922
19	къ немѣ нюдѣомъ	6	1.903	стра стдо мунѣкоу	10	3.132	онъ полъ нордана	4	1.918
20	възвѣдъ оун свон	4	1.900	нмена баша написана	10	3.127	азъ смь лоза	5	1.916

Несмотря на ожидаемое достаточно большое количество пересечений трех выборок, состав триграмм в ЕП и в русском и старославянском подкорпусах Евангелий несколько различен. Так, среди триграмм, имеющих наибольшее статистическое значение в соответствии с мерой T-score, девять не повторяются в сопоставляемых подкорпусах: **НѢСТѢ ЛН УЪЛН, НМАТЬ ЖНВОТЪ ВЪУЪНЪИН, ГЪ ПРНТЪУЮ СНЮ, ВЪСЮ ДШЮ СВОЮ, ВЪСТАВЪ ЗАПРЪТН ВЪТРОУ, ВЪРА ТВОА СПСЕ, КЪ ОВЪЦАМЪ ПОГЪБЪШНМЪ, ДОМЪ ВАШЪ ПОУСТЪ, КЪ НЕМЪ НЮДЪОМЪ, ВЪЗВЕДЪ ОУН СВОИ**. Этот факт позволяет говорить и о специфике триграмм ЕП на фоне подкорпуса древнерусских и старославянских списков.

Заключение

Результаты трех экспериментов по сопоставлению одно-, двух- и трехкомпонентных сочетаний ЕП и других евангельских списков исторического корпуса «Манускрипт» свидетельствуют о таких количественно-статистических характеристиках лингвистических компонентов каждой из рукописей, которые могут быть признаны существенными.

Результаты первого эксперимента демонстрируют идентичность или близость порядка следования и относительного количества наиболее частотных форм ЕП и других полных апракосов при одновременных диагностируемых отличиях перечней ЕП от перечней кратких апракосов и старославянских списков: наименьшие различия обнаружены в парах «ЕП – полные апракосы» и «краткие апракосы – старославянские списки», а наибольшие – в «ЕП – краткие апракосы». Это убеждает в том, что состав перечней, порядок следования и относительная частота форм в них в совокупности являются существенными характеристиками рукописи или подкорпуса.

Второй эксперимент с использованием статистической меры Weiridness позволил установить, что состав статистически значимых форм ЕП (формы местоимений **ВАМЪ, НМЪ, АЗЪ, КМОУ**, глагол **РЕУЕ**, союз **АЩЕ**, местоимение (союз-частица) **СЕ**), выявляемых при сопоставлении с перечнями словоформ разножанровых подкорпусов, не зависит от

объема контрастного подкорпуса; а сопоставление с подкорпусами кратких и полных апракосов дало возможность выявить формы, отличающие ЕП от других списков: союз **ЯКО**, предлог **КЪ**, частицу **БО**, а также предлог **Ѡ** (на фоне других рукописей полного апракоса), местоимение **НМЪ**, глагольная форма **КСТЪ**, союз **АЩЕ** (при сравнении с краткими апракосами).

Текстовое единство Евангелий и одновременно специфику текста ЕП демонстрирует третий эксперимент, в котором, с одной стороны, из 20 биграмм, извлеченных с помощью меры T-score, половина идентична в ЕП и подкорпусах разных изводов, а триграммы с максимальным статистическим значением – это неизменяемые композиционные формулы (**ѸВЪ Ѡ МА** и др.), цельные грамматические структуры (**АЖЕ ДАДЪ ЕСИ** и др.), устойчивые семантические комплексы и их части (**[ДА] ЛЮБНТЕ ДРОУГЪ ДРОУГА** и др.), с другой стороны, практически половина биграмм и триграмм ЕП отсутствует в перечне значимых сочетаний сопоставляемых подкорпусов (**НѢСТѢ ЛН УЪЛН, НМАТЬ ЖНВОТЪ ВЪУЪНЪИН** и др.).

Полученные результаты носят как общий, так и частный характер, позволяют говорить о результативности применения статистических методов к лингвистическому материалу славянских средневековых письменных памятников, о возможности обнаружения их неизвестных количественных характеристик.

ПРИМЕЧАНИЯ

¹ Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (РФФИ) в рамках проекта «Лингвостатистический анализ однокомпонентных и многокомпонентных лексических единиц исторического корпуса «Манускрипт»» (проект № 18-012-00463).

The reported study was funded by Russian Foundation for Basic Research (RFBR), the projects “A linguistic statistical analysis of one-component and multicomponent lexical units of the historical corpus ‘Manuscript’” (project No. 18-012-00463).

² См., например: [Копотев, 2014; Литвинова, 2016; Мартыненко, 2014; 2015; Марусенко, 1990; Морозов, 1915; Прикладная..., 2016; и др.].

³ О корпусе см., например: [Баранов, 2019а; Baranov, 2018].

⁴ URL: http://www.manuscripts.ru/mns/portal.main?p1=21&p_lid=1.

⁵ Для извлечения данных и сопоставления их с данными других подкорпусов использованы специализированные модули корпуса – модуль статистики и модуль *n*-грамм (см. список источников) [Баранов, 2019б; Baranov, Gnutikov, 2019].

⁶ URL: <http://manuscripts.ru/mns/portal.main?p1=30>.

⁷ Количество форм подобрано таким образом, чтобы перечни ЕП, подкорпусов полных и кратких Евангелий максимально пересекались; большие по размеру списки дают меньшее количество совпадающих форм.

⁸ Наличие в списке, кроме служебных слов, местоимений и формы *mV|V* связано со структурно-жанровыми особенностями текстов и списков Евангелий.

⁹ Относительная частота словоформы: $f = F / N$, где *F* – абсолютное количество употреблений словоформы в документе(ах); *N* – общее количество словоформ в документе(ах) (объем).

¹⁰ $Weirdness_w = \frac{W_s/T_s}{W_g/T_g}$, где «*W_s* – частотность слова в коллекции предметной области; *T_s* – совокупная частотность слов в коллекции предметной области; *W_g* – частотность слова в контрастной коллекции; *T_g* – совокупная частотность слов в контрастной коллекции» [Клышинский, Кочеткова, 2014, с. 368].

¹¹ Состав подкорпусов см.: http://manuscripts.ru/mns/portal.main?p1=1&p_lid=1.

¹² $T\text{-score} = \frac{F(w_1, w_2) - \frac{F(w_1) \times F(w_2)}{N}}{\sqrt{F(w_1, w_2)}}$, где *w₁* – первый компонент; *w₂* – второй компонент; *F(w₁)*, *F(w₂)* – абсолютные (независимые) частоты *w₁* и *w₂*; *F(w₁, w₂)* – абсолютная частота *n*-граммы в подкорпусе; *N* – общее число словоформ в корпусе [Evert, 2004]. См. также: [Кочеткова, 2013; Пивоварова, Ягунова, 2014; Хохлова, 2008; Ягунова, Пивоварова, 2010].

¹³ В выборке представлены биграммы, между компонентами которых отсутствуют символы-разделители.

¹⁴ В выборке представлены только триграммы, между компонентами которых отсутствуют символы-разделители.

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 Куртенэ и мировая лингвистика. В 2 т. Т. 1 : Междунар. 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.

Баранов В. А., 2019в. Опыт применения количественных и статистических методов для поиска значимых слов в историческом корпусе (на материале средневековых славянских гимнографических и евангельских кодексов) // *Studia Hymnographica*. Band II / hrsg. von H. Rothe, C. Schnell. Paderborn ; München ; Wien ; Zürich : Verlag Ferdinand Schöningh. P. 149–201. (*Patristica Slavica* ; Band 24).

Зуга О. В., 2009. Из наблюдений над характером языковых различий в славянских списках Евангелия XII–XIII вв. (на материале «Притчи о блудном сыне») // *Вестник Вятского государственного гуманитарного университета*. № 3 (2). С. 40–46.

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.

Копотев М. В., 2014. Введение в корпусную лингвистику. Прага : Animedia Company. 230 с.

Кочеткова Н. А., 2013. Статистические языковые методы. Коллокации и коллигации // *Новые информационные технологии в автоматизированных системах*. № 16. С. 301–305. URL: <https://cyberleninka.ru/article/n/statisticheskie-yazykovye-metody-kollokatsii-i-kolligatsii> (дата обращения: 31.05.2020).

Литвинова Т. А., 2016. Судебная автороведческая экспертиза текста с целью установления пола его автора: проблемы и перспективы // *Современное право*. № 7. С. 111–115.

Марков В. М., 2001. Из наблюдений над языком Пантелеймонова Евангелия (XII век) // Марков В. М. Избранные работы по русскому языку. Казань : ДАС. С. 31–56.

Мартыненко Г. Я., 2014. Стилеметрия: возникновение и становление в контексте междисциплинарного взаимодействия // *Структурная и прикладная лингвистика*. № 10. С. 3–23.

Мартыненко Г. Я., 2015. Стилеметрия: возникновение и становление в контексте междисциплинарного взаимодействия. Ч. 2. Первая половина XX века: расширение междисциплинарных контактов стилеметрии // *Структурная и прикладная лингвистика*. № 11. С. 9–28.

- Марусенко М. А., 1990. Атрибуция анонимных и псевдонимных литературных произведений методами распознавания образов. Л. : Изд-во Ленингр. ун-та, 164 с.
- Морозов Н. А., 1915. Лингвистические спектры : Средство для отличия плагиатов от истинных произведений того или другого известного автора : Стилометрический этюд // Изв. Отд. рус. языка и словесности Имп. АН. СПб. Т. XX, кн. 4. С. 93–134.
- Пивоварова Л. М., Ягунова Е. В., 2014. От коллокаций к конструкциям // Русский язык: грамматика конструкций и лексико-семантические подходы. СПб. С. 568–617. (Acta Linguistica petropolitana ; т. 10, № 2). URL: <https://www.elibrary.ru/item.asp?id=23195921> (дата обращения: 31.05.2020).
- Прикладная и компьютерная лингвистика, 2016 / под ред. И. С. Николаева, О. В. Митрениной, Т. М. Ландо. М. : ЛАНАНД. 320 с.
- Сводный каталог славяно-русских рукописных книг, хранящихся в СССР: XI–XIII вв., 1984. М. : Наука. 406 с.
- Хохлова М. В., 2008. Экспериментальная проверка методов выделения коллокаций // Slavica Helsingiensia 34. Инструментарий русистики : Корпусные подходы. Хельсинки. С. 343–357. URL: <https://www.elibrary.ru/item.asp?id=26581613> (дата обращения: 31.05.2020).
- Ягунова Е. В., Пивоварова Л. М., 2010. Природа коллокаций в русском языке. Опыт автоматического извлечения и классификации на материале новостных текстов // Сб. НТИ. Серия 2. № 6. С. 30–40.
- Ahmad K., Gillam L., Tostevin L., 1999. University of Surrey participation in Trec8: Weirdness indexing for logical documents extrapolation and retrieval // Proc. of Eighth Text Retrieval Conference (Trec-8). 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 / ed. by L. Sels, J. Fuchsbaauer, V. Tomelleri, I. de Vos. P. ; Bristol : Peeters Leuven. P. 283–308.
- Baranov V.A., Gnutikov R.M., 2019. 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” / comp. and ed. by A. Miltenova, V. Baranov, H. Miklas, K. Hawkins, J. Fuchsbaauer. Sofia : Gutenberg Publishing House. P. 9–28.
- Evert S., 2004. Association Measures // Computational Approaches to Collocations. URL: <http://collocations.de/AM/index.html> (date of access: 31.05.2020).

ИСТОЧНИКИ

- ЕП–Евангелие апракос полный («Пантелеймоново Евангелие») // РНБ. Соф. 1, кон. XII – нач. XIII (?) в. 224 л.
- Коллекция славянских Евангелий корпуса «Манускрипт». URL: <http://manuscripts.ru/mns/portal.main?p1=30> (дата обращения: 31.05.2020).
- Корпус «Манускрипт: славянское письменное наследие». URL: <http://manuscripts.ru/> (дата обращения: 31.05.2020).
- Модуль статистики корпуса «Манускрипт». URL: <http://manuscripts.ru/mns/!cred2.stat> (дата обращения: 31.05.2020).
- Модуль n-грамм корпуса «Манускрипт». URL: http://manuscripts.ru/mns/cred_ngr.stat (дата обращения: 31.05.2020).
- Электронное издание Пантелеймонова Евангелия. URL: http://manuscripts.ru/mns/portal.main?p1=21&p_lid=1 (дата обращения: 31.05.2020).

REFERENCES

- Baranov V.A., 2019a. Sozдание i ispolzovanie istoricheskikh korpusov slavyanskikh pismennykh pamyatnikov [Creation and using of historical corpora of slavonic manuscripts]. *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: mezhdunar. konf.: VII Boduenovskie chteniya (Kazan. feder. un-t, 28-31 okt. 2019 g.): tr. i mater.: v 2 t.* [I.A. Baudouin de Courtenay and World Linguistics: International Conference: VII Baudouin Readings (Kazan Federal University, October 28-31, 2019): Proc. and materials: in 2 vol.]. Vol. 1, Kazan, pp. 24-30.
- Baranov V. A., 2019v. Opyt primeneniya kolichestvennykh i statisticheskikh metodov dlya poiska znachimykh slov v istoricheskom korpuse (na materiale srednevekovykh slavyanskikh gimnograficheskikh i evangelskikh kodeksov) [The experience of applying quantitative and statistical methods to search for meaningful words in the historical corpus (based on

- medieval Slavonic hymnographic and gospel manuscripts)]. *Studia Hymnographica*, vol. II, Hans Rothe and Claudia Schnell (eds.), Paderborn, Munich, Vienna, Zurich, 2019, pp. 149-201 (Patristica Slavica, vol. 24, eds. Hans Rothe, vol. 131).
- Zuga O.V., 2009. Iz nablyudeniy nad kharakterom yazykovykh raznochteniy v slavyanskikh spiskakh Evangeliiya XII–XIII vv. (na materiale «Pritchi o bludnom syne») [Observing the character of alternative version in the 12th - 13th centuries Slavic Gospels (on the material of the Parable of the Prodigal Son)]. *Vestnik Vyatskogo gosudarstvennogo gumanitarnogo universiteta* [Herald of Vyatka State University], no. 3 (2), pp. 40-46.
- Klyshinskiy Ye.S., Kochetkova N.A., 2014. Metod izvlecheniya tekhnicheskikh terminov s ispolzovaniem mery strannosti [Method for extracting technical terms using a strangeness measure]. *Novye informacionnye tehnologii v avtomatizirovannykh sistemah* [New information technologies in automated systems], no. 17, pp. 365-370.
- Kopotev M.V., 2014. *Vvedenie v korpusnuyu lingvistiku* [Introduction to Corpus Linguistics]. Prague, Animedia Company Publ. 230 p.
- Kochetkova N.A., 2013. Statisticheskie yazykovye metody. Kollokatsii i kolligatsii [Statistical language methods. Collocations and Colligations]. *Novye informacionnye tehnologii v avtomatizirovannykh sistemah* [New information technologies in automated systems], no. 16, pp. 301-305. URL: <https://cyberleninka.ru/article/n/statisticheskie-yazykovye-metody-kollokatsii-i-kolligatsii> (accessed 31 May 2020).
- Litvinova T.A., 2016. Sudebnaya avtorovedcheskaya ekspertiza teksta s tselyu ustanovleniya pola ego avtora: problemy i perspektivy [Forensic authorship examination towards gender attribution: challenges and prospects]. *Sovremennoe pravo*, no. 7, pp. 111-115.
- Markov V.M., 2001. Iz nablyudeniy nad yazykom Panteleymonova Evangeliiya (XII vek) [From observations of the language of the Panteleimon Gospel (12th century)]. Markov V.M. *Izbrannye raboty po russkomu yazyku* [Selected works on the Russian language], Kazan, Publishing House “DAS”, pp. 31-56.
- Martynenko G.Ya., 2014. Stilemetriya: vzniknovenie i stanovlenie v kontekste mezhdistsiplinarnogo vzaimodeystviya [Stylemetry: emergence and evolution in context of interdisciplinary interaction]. *Strukturnaya i prikladnaya lingvistika* [Structural and Applied Linguistics], no. 10, pp. 3-23.
- Martynenko G.Ya., 2015. Stilemetriya: vzniknovenie i stanovlenie v kontekste mezhdistsiplinarnogo vzaimodeystviya. Chast 2. Pervaya polovina XX veka: rasshirenie mezhdistsiplinarnykh kontaktov stilemetrii [Stylemetry: emergence and formation in the context of interdisciplinary interaction. Part 2. The First Half of the 20th century: the Expansion of interdisciplinary Contacts]. *Strukturnaya i prikladnaya lingvistika* [Structural and Applied Linguistics], no. 11, pp. 9-28.
- Marusenko M.A., 1990. *Atributsiya anonimnykh i psevdonimnykh literaturnykh proizvedeniy metodami raspoznavaniya obrazov* [Attribution of anonymous and pseudonymous literary works by pattern recognition methods]. Leningrad, Izdatelstvo Leningradskogo universiteta. 164 p.
- Morozov N.A., 1915. Lingvisticheskie spektry: Sredstvo dlya otlichiya plagiatov ot istinnykh proizvedeniy togo ili drugogo izvestnogo avtora: Stilemetricheskii etyud [Linguistic spectra: A tool for distinguishing plagiarism from the true works of one or another well-known author: Stylemetric sketch]. *Izvestiya Otdeleniya russkogo yazyka i slovesnosti Imperatorskoy Akademii nauk* [Proc. of the Department of Russian Language and Literature of the Imperial Academy of Sciences], St. Petersburg, vol. XX, no. 4, pp. 93-134.
- Pivovarova L.M., Yagunova E.V., 2014. Ot kollokatsiy k konstruktsiyam [From collocations to syntax constructs]. *Russkiy yazyk: grammatika konstruktsiy i leksiko-semanticheskie podkhody* [Russian language: grammar of constructions and lexical-semantic approaches], St. Petersburg, pp. 568-617. (Acta Linguistica petropolitana, vol. 10, no. 2). URL: <https://www.elibrary.ru/item.asp?id=23195921> (accessed 31 May 2020).
- Prikladnaya i kompyuternaya lingvistika*, 2016. [Applied and Computational Linguistics] Nikolaeva I.S., Mitreninoy O.V., Lando T.M. (eds.). Moscow, LANAND. 320 p.
- Svodnyy katalog slavyano-russkikh rukopisnykh knig, khranyashchikhsya v SSSR: XI–XIII vv.* [The consolidated catalog of Slavic-Russian manuscript books stored in the USSR: 11th-13th centuries], 1984. Moscow, Nauka Publ. 406 p.
- Hohlova M.V., 2008. Eksperimentalnaya proverka metodov vydeleniya kollokatsiy [Experimental verification of collocation allocation methods]. *Slavica Helsingiensia 34. Instrumentariy rusistiki: Korpusnye podhody* [Instrumentation of Russian Studies: Corpus Approaches],

- Helsinki, pp. 343-357. URL: <https://www.elibrary.ru/item.asp?id=26581613> (accessed 31 May 2020).
- Yagunova E.V., Pivovarov L.M., 2010. Priroda kollokatsiy v russkom yazyke. Opyt avtomaticheskogo izvlecheniya i klassifikatsii na materiale novostnykh tekstov [The nature of collocations in Russian. The experience of automatic extraction and classification on the material of news texts]. *Sbornik nauchno-tekhnicheskoy informacii* [Collection of scientific and technical information], ser. 2, no. 6, Moscow, pp. 30-40.
- Ahmad K., Gillam L., Tostevin L., 1999. University of Surrey participation in Trec8: Weirdness indexing for logical documents extrapolation and retrieval. *Proc. of Eighth Text Retrieval Conference (Trec-8)*, Gaithersburg, pp. 717-724.
- Baranov V., 2018. A Text Corpus of Medieval Manuscripts as a Goal and a Tool for Linguistic Research. Lara Sels, Jürgen Fuchsbaier, Vittorio Tomelleri and Ilse de Vos (eds.). *Editing Mediaeval Texts from a Different Angle: Slavonic and Multilingual Traditions*. Paris, Bristol, Ct, Peeters Leuven, pp. 283-308.
- Baranov V.A., Gnutikov R.M., 2019. The statistics and n-gram modules of the historical corpus "Manuscript". A. Miltenova, V. Baranov, H. Miklas, K. Hawkins, J. Fuchsbaier (eds.). *Digital and Analytical Approaches to the Written Heritage*, proc. of the 7th international conference El'Manuscript "Textual Heritage and Information Technologies", Sofia, pp. 9-28.
- Evert S., 2004. Association Measures. *Computational Approaches to Collocations*. URL: <http://collocations.de/AM/index.html> (accessed 31 May 2020).

SOURCES

- Evangelie aprakos polnyy («Panteleymonovo Evangelie») [The Gospel of Aprakos complete ("Panteleymonovo Evangelie")]. *RNB*, Sof. 1, end. 12th - beg. 13th century(?), 224 f.
- Kollektsiya slavyanskikh Evangeliy korpusa «Manuskript»* [Collection of Slavic Gospels of the corpus "Manuscript"]. URL: <http://manuscripts.ru/mns/portal.main?p1=30> (accessed 31.05.2020).
- Korpus «Manuskript: slavyanskoe pismennoe nasledie»* [Corpus "Manuscript: Slavonic written heritage"]. URL: <http://manuscripts.ru/> (accessed 31.05.2020).
- Modul statistiki korpusa «Manuskript»* [Statistics module of the corpus "Manuscript"]. URL: <http://manuscripts.ru/mns/!cred2.stat> (accessed 31.05.2020).
- Modul n-gramm korpusa «Manuskript»* [N-gramm module of the corpus "Manuscript"]. URL: http://manuscripts.ru/mns/cred_ngr.stat (accessed 31.05.2020).
- Elektronnoe izdanie Panteleymonova Evangeliya* [Electronic edition of Panteleimon's Gospel]. URL: http://manuscripts.ru/mns/portal.main?p1=21&p_lid=1 (accessed 31.05.2020).

Information About the Authors

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, victor.a.baranov@gmail.com, <https://orcid.org/0000-0003-1730-6359>

Oksana V. Zuga, Candidate of Sciences (Philology), Associate Professor, Department of the Russian Language, Theoretical and Applied Linguistics, Udmurt State University, Universitetskaya St, 1, 426034 Izhevsk, Russia, ozuga@rambler.ru, <https://orcid.org/0000-0002-2675-4818>

Информация об авторах

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, victor.a.baranov@gmail.com, <https://orcid.org/0000-0003-1730-6359>

Оксана Владимировна Зуга, кандидат филологических наук, доцент кафедры русского языка, теоретической и прикладной лингвистики, Удмуртский государственный университет, ул. Университетская, 1, 426034 г. Ижевск, Россия, ozuga@rambler.ru, <https://orcid.org/0000-0002-2675-4818>