

RU

Соотношение религиозной и военной лексики в армянском тексте:
квантитативный подход

Андреев С. Н.

Аннотация. Целью данного исследования является выявление скрытых тенденций и закономерностей в распределении лексических единиц двух тематических классов – религиозной и военной лексики – в стихотворных текстах русской литературы, посвященных Армении. Эти темы являются важными параметрами армянского текста, они реализуются в большинстве произведений об Армении и отражают важные стороны ее истории и культуры. Для анализа привлекаются стихи, созданные в период XIX-XXI вв. Научная новизна исследования состоит в том, что в нем рассматривается вопрос лексической репрезентации двух образно-тематических параметров в рамках квантитативного подхода, что позволяет объективно определить характер освещения военной и религиозной тем на различных этапах становления и развития армянского текста. Изучение соотношений лексических единиц в текстах произведений позволяет изучить как общие тенденции в восприятии этой страны в русской литературе, так и индивидуальные особенности построения поэтами художественного образа Армении. Для квантитативного анализа используются коэффициенты вариации, коэффициент Бузмана, оценка статистической значимости (хи-квадрат), евклидово расстояние. Результаты проведенного анализа позволили установить взаимосвязи двух важных для армянского текста тем и динамику их изменений во времени. Наиболее сильные изменения наблюдались между вторым (дореволюционным) и третьим (советским) периодами. Сопоставительный анализ произведений 8 поэтов, писавших про Армению, позволил провести их классификацию и определить различия в частоте использования военной и религиозной лексики.

EN

Correlation between religious and military vocabulary
in the Armenian text: A quantitative approach

Andreev S. N.

Abstract. The study aims to identify hidden trends and patterns in the distribution of lexical units of two thematic classes – religious and military vocabulary – in the poetic texts of Russian literature dedicated to Armenia. These themes represent important parameters in the Armenian text, they are realised in most works about Armenia and reflect important aspects of its history and culture. The analysis involves poems created in the XIX-XXI centuries. The scientific novelty of the study lies in the fact that it examines the issue of the lexical representation of two figurative-thematic parameters within the framework of a quantitative approach, which makes it possible to objectively determine the nature of the coverage of military and religious themes at various stages of the formation and development of the Armenian text. The study of the relations of lexical units in the texts of the works allows the researcher to examine both general trends in the perception of this country in Russian literature and the individual features of the poets' construction of Armenia's artistic image. The coefficients of variation, the Busemann coefficient, the estimation of statistical significance (chi-square) and the Euclidean distance are used for quantitative analysis. The results of the analysis made it possible to establish the interrelationships of two important themes for the Armenian text and the dynamics of their changes over time. The strongest changes were observed between the second (pre-revolutionary) and the third (Soviet) periods. A comparative analysis of the works of 8 poets who wrote about Armenia made it possible to classify them and determine differences in the frequency of military and religious vocabulary.

Введение

Анализ локальных текстов, к классу которых принадлежат анализируемые стихотворные произведения, ставит своей задачей построение модели, отражающей основные характеристики формируемого образа данного

локуса. В числе характеристик рассматриваемого локуса, релевантных для его лингвистической модели, в последнее время рассматривалась широкая парадигма параметров, среди которых представлены равноуровневые и разноаспектные признаки: тематическая отнесенность лексических единиц, образная система, синтаксические особенности определений, ключевые сочетания слов и др. Исследование этих характеристик и их соотношений позволило получить ряд интересных результатов о ведущих тенденциях в образовании паттернов армянского текста (Andreev, 2021; Карасик, 2022; Павлова, Романова, 2023; Шафранская, 2022).

В этой статье рассматривается тематическая отнесенность лексики к военной сфере и к религии – два крайне важных параметра (Бидеркесен, 2020; Бойко, 2016; Демидович, 2015; Бугаева, 2010; Гассельблат, 2015; Сергеева, 2010), которые особенно значимы для армянского текста, поскольку отражают исторические и культурные особенности этой страны: борьбу армянского народа за свою свободу и за свою религию – христианство, принятое с IV века в качестве государственной религии. Указанное определяет актуальность исследования.

В ходе исследования будут решаться следующие задачи: определение относительных частот лексики двух рассматриваемых тематических групп, установление степени вариативности репрезентации военной и религиозной лексики во времени, выявление соотношений данных тематических классов, сопоставление индивидуальных стилей в пространстве признаков двух тематических классов, определение степени сходства/различия индивидуальных стилей и средних показателей по выборке. Для анализа и решения указанных задач используются различные методы количественного анализа, соответственно: нормализация абсолютных частот, коэффициент вариации, коэффициент Бузмана, критерий хи-квадрат и евклидово расстояние.

Анализ осуществляется в рамках количественного подхода. Для выявления соотношений пространственных признаков, степени устойчивости их частоты, для сопоставления стилей авторов используются коэффициенты вариации, Бузмана, критерий хи-квадрат, евклидово расстояние.

Теоретической базой исследования служат работы известных специалистов в области количественной лингвистики Р. Г. Пиотровского (2006), Г. Я. Мартыненко (2009), Ю. А. Тулдавы (1987) и Г. Альтмана (Altmann, 2015).

Практическая значимость исследования определяется тем, что предложенная методика исследования и полученные результаты могут быть использованы в процессе преподавания теоретических дисциплин по языкознанию (лексикологии, теоретической грамматики) и литературоведению. Они также могут составить базу в практических исследованиях для сопоставительного анализа с другими локальными текстами.

Тема религии (РЛ) представлена в выборке 101 словом. Наиболее частотные по убыванию: Бог, молиться, священный, душа, вера, креститься, благословлять, ковчег, жертвоприношения, Ной, библейский, воскресить, перуны, Потоп, пророк, рай, храм, христианство. Далее идут слова, встретившиеся 1 раз: адский, алтарь, Артемида, Заратустра, знамение, Иисус, католикос и др.

Основу списков составляют существительные (77%), на глаголы и прилагательные приходится соответственно 6% и 17%: молиться, креститься, священный, библейский и др.

Военная лексика (ВН) в выборке включает 132 слова (по убыванию частотности): враг, битва, бой, войска, сражаться, воин, ополчиться, победа, булат, гибель, пасть, полк, солдаты, борьба, брань, война, кровь, мечи, отряд, пехота, погром, разить, стрела. Слова с частотой 1 составляют 37%: броня, вал, взвод, дембель, донесение, дружина, защитники, окоп, оружие, ратник, пуля и др.

Здесь существительные составляют 83%, глаголы – 13% (осаждать, разить), прилагательные – 4% (воинский, солдатский).

В качестве материала настоящего исследования используется выборка из корпуса текстов, созданного в рамках проекта 22-18-00339 «Электронный ресурс “Армянский текст русской поэзии”: репрезентация локального текста русской литературы», поддержанного РНФ. Общий объем выборки составляет 10% от корпуса, включающего произведение, полностью посвященные Армении. Основные характеристики выборки приведены в работе (Андреев, 2022), в которой рассматривались три периода развития стихотворных текстов про Армению. В настоящей работе добавлен еще один период – советский, начиная с 20-х годов XX века до 90-х, в связи с чем произведены некоторые изменения в количестве рассмотренных текстов. В целом, таким образом, в этом исследовании выделяются следующие периоды:

- 1 период (П-1): XIX в.;
- 2 период (П-2): начало XX в. (до 20-х годов);
- 3 период (П-3): 20-е – 80-е годы XX в.;
- 4 период (П-4): 90-е годы XX в. – XXI в.

В каждом из периодов были определены частоты указанных выше параметров и проведены все дальнейшие расчеты, что позволило сопоставить особенности изменений в описании пространства в стихах про Армению.

Обсуждение и результаты

Относительные частоты тематических групп

В результате подсчетов были получены данные о частотах указанных параметров в текстах, относящихся к различным периодам. В связи с тем, что периоды не равны по объему, для их сопоставления в ряде случаев необходимо использовать не абсолютные, а относительные частоты, что достигается путем деления частот на количество строк соответствующего периода. Затем полученные данные умножаются на 1000 для того, чтобы избежать больших дробей (т. н. расчет частот на 1000 строк). Другой подход к нормализации данных

для сопоставления текстов, разных по объему, был предложен Д. Лабе (Labbé, 2007) и был использован в работе (Андреев, 2022), однако оба подхода в данном случае дают в итоге одинаковые результаты, поэтому здесь взят более простой. Полученные относительные частоты представлены в Табл. 1.

В этой таблице показаны также относительные частоты для всей выборки, при этом они не равны сумме частот отдельных периодов, деленных на 4.

Таблица 1. Частоты тематических групп лексики

Тексты	Религиозная лексика	Военная лексика
Период 1	19,66	30,80
Период 2	18,65	37,97
Период 3	5,41	5,41
Период 4	20,00	10,00
Выборка	14,57	19,04

В графическом виде эти данные представлены на Рис. 1. По оси *x* отложены значения РЛ, по оси *y* – ВЛ.

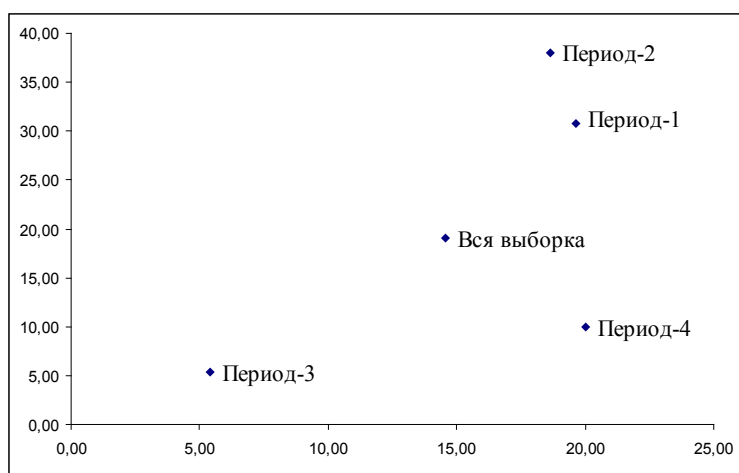


Рисунок 1. Диаграмма рассеяния периодов по относительным частотам двух тем

Как следует из диаграммы рассеяния (Рис. 1), наблюдается своеобразное противопоставление первого и второго периодов, с одной стороны, и третьего и четвертого периодов – с другой. Кроме того, если смотреть только по оси *x*, то Период-3 противопоставлен всем остальным.

Вариативность признаков во времени

Рассмотрим степень вариативности рассматриваемых параметров в армянском тексте за два века. Для этого был использован коэффициент вариации:

$$V = \frac{\sigma}{M},$$

где *V* – коэффициент вариации, σ – стандартное отклонение, *M* – средняя арифметическая. Результаты вычислений даны в Табл. 2.

Таблица 2. Коэффициент вариации двух тематических классов

Σ, M, V	Религиозная Лексика	Военная лексика
Стандартное отк.	7,04	15,79
Ср. арифметич.	15,93	21,05
Коэф. вариации	44,18	75,04

Коэффициент вариации в большинстве случаев изменяется в диапазоне 0% – 100% (иногда в случаях сверхсильной вариации признака он может и превысить 100%). Определение границ уровня вариации устанавливается опытным путем. В нашем случае, исходя из материала исследования, имеет смысл применить трехчастное деление, позволяющее выделить слабое варьирование, варьирование средней силы и сильные изменения в частоте признака по периодам. Тогда к слабому следует отнести варьирование от 0% и до 33%. Диапазон от 33% и до 66% можно рассматривать как вариацию средней силы, а все превышающие этот порог значения *V* – как показатель сильной вариации.

Если исходить из такой градации уровня вариативности, можно сделать заключение о достаточно больших изменениях в частоте слов этих тематических классов по периодам. Наиболее сильной вариативностью обладает класс военной лексики, а у РЛ вариативность хоть и ниже, но тоже достаточно велика.

Имеет смысл рассмотреть, в каких периодах происходит усиление варьирования. Если все относительные частоты по каждому тематическому классу принять за 100%, то в каждом периоде наблюдается следующая картина (Таблица 3).

Таблица 3. Процентное распределение относительных частот двух тематических классов по периодам

Период	Религия (%)	Война (%)
Период 1	30,85	36,59
Период 2	29,27	45,11
Период 3	8,49	6,42
Период 4	31,39	11,88
Всего	100	100

Данные этой таблицы отражены также графически (Рис. 2).

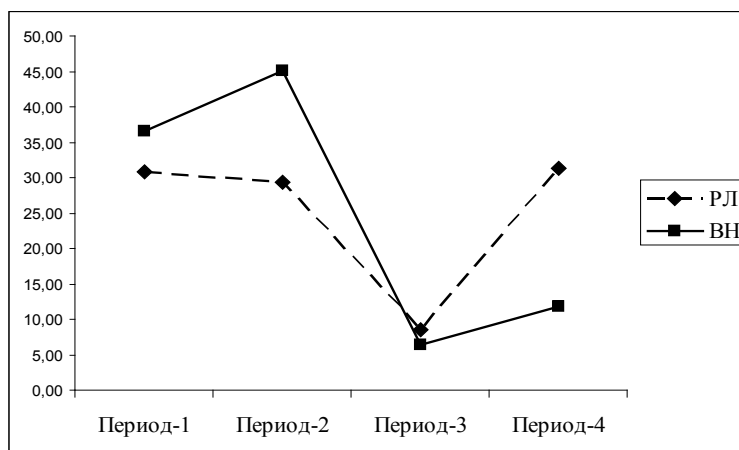


Рисунок 2. Частотное распределение двух тематических классов по периодам

Первый период отличается тем, что охватывает приблизительно треть всей лексики РЛ и даже более трети лексики ВН. Второй период несколько менее насыщен лексикой РЛ, однако в нем резко увеличивается использование военной лексики. Третий (советский) период сильно выделяется тем среди остальных, что здесь происходит резкое падение как религиозной лексики, так и военной лексики. Причем падение частоты использования лексических единиц ВН еще более сильно выражено, чем падение РЛ. В четвертом периоде максимально растет использование религиозной лексики и относительно слабо растет тема войны.

Редкое использование религиозной лексики в советский период достаточно понятно и было предсказуемо. Однако насколько оно уменьшилось по сравнению с предыдущим периодом, априорно представить достаточно сложно. Наши данные показывают, что частота РЛ упала более чем в три раза. С другой стороны, в этом периоде падение произошло и в использовании военной лексики – в 7 раз. Таким образом, военные события начала XX века в Армении и тема геноцида армянского народа стали выражаться другими тематико-мотивными способами, и тема войны в Армении заменилась темой скорби по жертвам того времени.

Если не учитывать третий период, то коэффициент вариации существенно изменяется – для ВН он становится равным 55,3%, что несколько ниже значения в Табл. 2, но все равно отражает достаточно сильное варьирование. Зато для РЛ коэффициент снижается до 3,6%, что означает крайне высокую стабильность темы. В первом, втором и четвертом периодах тема религии оказывается константной.

Соотношение признаков по периодам. Коэффициент Бузмана

Как видно из графика, использование слов данных тематических классов по-разному соотносилось между собой. Представляется возможным уточнить, как менялось это соотношение в разные периоды. Для этого можно использовать коэффициент, который получил распространение в лингвистике под названием меры Бузмана и в большинстве случаев используется как критерий определения статистики или динамики стиля (Andreev, Místecký, 2018). Однако этот коэффициент также может быть использован и с другой целью – для определения того, какой из двух параметров значимо превалирует в тексте. Коэффициент Бузмана (Andreev, Místecký, Altmann, 2018; Místecký, Altmann, 2019) имеет следующий вид:

$$B_{x/y} = \frac{x}{x + y},$$

где x – частота первого из соотносимых параметров, y – второго. Проверка осуществляется при помощи критерия хи-квадрат:

$$\chi^2 = \frac{(x - y)^2}{x + y}.$$

Мы будем говорить о значимом превалировании одного из признаков, если $\chi^2 > 3,84$ (для уровня статистической значимости $p = 0,05$ и 1 степени свободы).

Интерпретация осуществляется по критериям, предложенным в работе (Andreev, Místecký, Altmann, 2018), согласно которым в случае $0,45 \leq V_{xy} \leq 0,55$ использование сопоставляемых классов x и y сбалансировано (Б).

При $V_{xy} > 0,55$ можно говорить о преобладании первого из двух параметров, т. е. x , при $V_{xy} < 0,45$ – о преобладании второго из соотносимых параметров (y). Проверка производится при помощи критерия хи-квадрат. При $\chi^2 > 3,84$ можно говорить о значимом превалировании одного из параметров (З). В случае если $\chi^2 \leq 3,84$, речь может идти о тенденции к превалированию (Т).

В результате использования этого подхода были получены следующие результаты. Следует отметить, что здесь брались не относительные, а абсолютные частоты признаков (Табл. 4).

Таблица 4. Соотношение РЛ и ВН в стихотворных текстах об Армении

Тексты	Вр/лвн	Хи-квадрат	Интерпретация
Период 1	0,39	3,75	Т
Период 2	0,33	9,89	З
Период 3	0,50	0,00	Б
Период 4	0,67	5,00	З
Вся выборка	0,43	4,12	З

Соотношение рассматриваемых параметров в различных периодах различно. В первом имеет место тенденция к превалированию военной темы, в третьем периоде использование лексики двух тем сбалансировано. В остальных случаях различия статистически значимы. Однако четвертый и второй периоды оказываются противопоставленными и образуют оппозицию, в которой первый период более ориентирован на военную репрезентацию, второй – на религиозную.

Частоты двух параметров в индивидуальных стилях

Указанные выше данные получены на материале выборки – как целой, так и ее отдельных частей, соответствующих четырем этапам развития армянского текста. Разумеется, это некоторая усредненность, отражающая основную тенденцию, складывающуюся из индивидуальных подходов авторов к отражению истории и культуры страны. Поэтому представляется интересным сравнить, как эти тенденции соотносятся с индивидуальными особенностями описания поэтов. Для такого сопоставительного анализа были взяты 8 поэтов. В их число вошли В. Брюсов, С. Городецкий, О. Мандельштам, М. Матусовский, И. Снегова, В. Звягинцева, А. Григоров, Г. Кубатьян.

В результате подсчетов получены относительные частоты (нормализованы в расчете на 1000 слов текста), которые приводятся в Табл. 5.

Таблица 5. Относительные частоты лексики двух тем у 8 авторов

Автор	РЛ	ВН
Брюсов	7,51	10,24
Городецкий	19,81	10,34
Мандельштам	15,52	5,17
Звягинцева	4,75	3,88
Снегова	0,78	0,78
Матусовский	8,33	13,33
Григоров	6,02	7,03
Кубатьян	25,64	7,21

Евклидово расстояние частот от выборочной средней

Основываясь на полученных данных, можно сравнить индивидуальный подход каждого автора к отражению указанных тем. Для этого была определена степень сходства/различия частот двух тем у каждого автора со средним значением по выборке при помощи использования евклидова расстояния:

$$d_{(p,q)} = \sqrt{\sum_{k=1}^n (p_k - q_k)^2},$$

где p и q – точки в n -мерном пространстве.

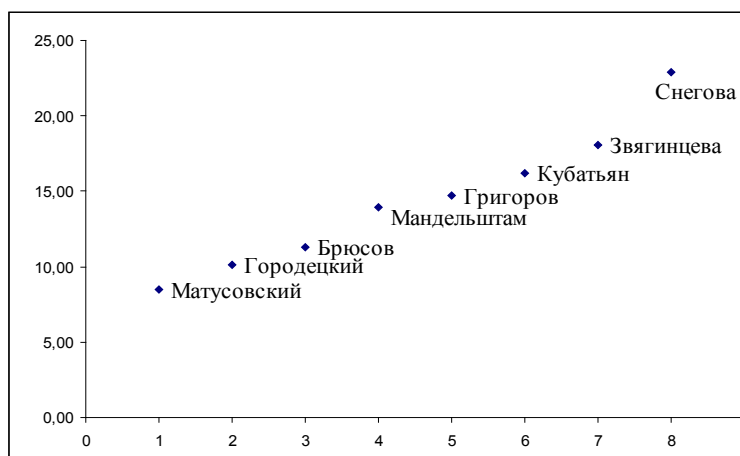
Результаты вычисления расстояний между каждым автором и средним по выборке показаны в Табл. 6. Авторы ранжированы по нарастанию расстояния (увеличению отличий) от выборочной средней.

Максимально близок к средней по выборке Матусовский, максимально отличается от среднего показателя Снегова. Увеличение расстояния, как видно на Рис. 3, нарастает постепенно.

Первые четыре автора, которые ближе всех других к средней по выборке (Матусовский, Городецкий, Брюсов и Мандельштам), имеют различные соотношения двух тем. У Матусовского наблюдается тенденция к превалированию военной лексики. У Брюсова соотношение близко к сбалансированному виду. У Городецкого имеет место тенденция к превалированию РЛ, а у Мандельштама это превалирование имеет статистически значимый характер.

Таблица 6. Евклидово расстояние частот двух тем у авторов от средней по выборке

Автор	Евклидово расстояние
Матусовский	8,46
Городецкий	10,16
Брюсов	11,29
Мандельштам	13,90
Григорьев	14,75
Кубатьян	16,20
Звягинцева	18,07
Снегова	22,89

**Рисунок 3.** Евклидово расстояние авторов в пространстве признаков РЛ и ВН от средней по выборке

Авторы, относительно дальше расположенные в пространстве признаков РЛ и ВН от средней (Григорьев, Кубатьян, Звягинцева и Снегова), по соотношению двух тем в своих произведениях более единообразны: за исключением Кубатьяна, у которого ярко выражено (статистически значимо) преимущество религиозной лексики над военной, у всех остальных темы сбалансированы.

Заключение

Определение вклада параметров религиозной и военной лексики и соотношение их частот позволяет охарактеризовать основные тенденции в построении модели данного локуса – образа Армении в стихотворных текстах на протяжении двух веков.

Вариативность использования обоих параметров достаточно значительна. При этом религиозная лексика выпадает из общей частотности только в третьем (советском) периоде, в остальных периодах ее относительная частота является константной.

Наибольшие изменения в описании военных реалий и религии происходят при переходе от второго к третьему периоду, минимальные изменения наблюдаются между периодами 1 и 2. Относительно использования военной лексики имеет место оппозиция пар периодов: 1 и 2 vs. 3 и 4.

Проведенное исследование в ряде случаев подтверждает представления о частоте использования лексических единиц двух изучаемых тем, однако позволяет уточнить и в ряде случаев найти скрытые тенденции в описании культуры Армении и ее борьбы за свободу.

Использование евклидова расстояния позволило определить индивидуальные особенности 8 поэтов, сделавших большой вклад в создание образа Армении. Максимально близко к усредненной частоте по частотам слов двух параметров оказались стихи М. Матусовского, С. Городецкого и В. Брюсова. Наиболее удалены от средних значений по выборке стихи В. Звягинцевой и особенно И. Снеговой.

Полученные результаты могут быть в дальнейшем использованы для сопоставительного анализа с другими локусами. С другой стороны, эти результаты создают базу для выявления индивидуальных различий в описании художественного мира Армении у различных авторов с указанием степени их отклонения от среднего значения в выборке (центроида) и позволяют охарактеризовать то, как каждый автор воспринимает историю этой страны и при помощи каких параметров стремится ее отразить.

Источники | References

1. Андреев С. Н. Распределение сенсорной лексики в локальных армянских текстах XIX-XXI вв. // Филологические науки. Вопросы теории и практики. 2022. Т. 15. Вып. 12.
2. Бидеркесен Д. Военная лексика как особый элемент лексической системы языка // Казанский лингвистический журнал. 2020. Т. 3. № 1.

3. Бойко Б. Л. Военная лексика в оригинале и переводе (на материале романа Э. М. Ремарка «Возвращение») // Вестник Московского университета. Серия 22 «Теория и практика перевода». 2016. № 2.
4. Бугаева И. В. Язык православной сферы: современное состояние, тенденции развития: автореф. дисс. ... д. филол. н. М., 2010.
5. Гассельблат О. А. Специфика представления лексико-семантического поля «Религия» в творчестве Иосифа Бродского в соотношении с текстами Серебряного века // Общество & человек. 2015. № 4 (10).
6. Демидович Т. В. Семантико-стилистический и лингвокультурологический аспекты изучения военной лексики (на материале художественной литературы о Великой Отечественной войне): автореф. дисс. ... к. филол. н. Волгоград, 2015.
7. Карасик В. И. Самовыражение через иную культуру: армянский текст в русской поэзии // Теория языка и межкультурная коммуникация. 2022. № 4 (47).
8. Мартыненко Г. Я. Введение в теорию числовой гармонии текста. СПб.: Изд-во Санкт-Петербургского университета, 2009.
9. Павлова Л. В., Романова И. В. Созвучие Сарьяна и Армении в восприятии русских поэтов // Известия Смоленского государственного университета. 2023. № 1 (61).
10. Пиотровский Р. Г. Лингвистическая синергетика: исходные положения, первые результаты, перспективы. СПб.: Изд-во Санкт-Петербургского университета, 2006.
11. Сергеева Е. В. Особенности употребления библеизмов в русской поэзии XX века // Лингвистические параметры художественного текста: межвуз. сб. науч. ст. Оренбург: Изд-во Оренбургского государственного педагогического университета, 2010.
12. Тулдава Ю. А. Проблемы и методы квантитативно-системного исследования лексики. Таллин: Валгус, 1987.
13. Шафранская Э. Ф. Армянский текст: стихи и проза // Филологические науки. Научные доклады высшей школы. 2022. № 6s.
14. Altmann G. Problems in Quantitative Linguistics. Lüdenscheid: RAM-Verlag, 2015. Vol. 5.
15. Andreev S., Místecký M. Activity in Czech and Russian Nineteenth-Century Sonnets: A Contrastive Study // Glottotheory. International Journal of Theoretical Linguistics. 2018. Vol. 9. Iss. 1.
16. Andreev S., Místecký M., Altmann G. Sonnets: Quantitative Inquiries. Lüdenscheid: RAM-Verlag, 2018.
17. Andreev V. Style Evolution: Space and Movement in Longfellow's Lyrical Poems // Bakhtiniana. 2021. Vol. 16 (3).
18. Labbé D. Experiments of Authorship Attribution by Intertextual Distance in English // Journal of Quantitative Linguistics. 2007. Vol. 14 (1).
19. Místecký M., Altmann G. Tense and Persons in English: Modelling Attempts // Glottometrics. 2019. Vol. 46.

Финансирование | Funding

- RU** Исследование выполнено в рамках проекта 22-18-00339 «Электронный ресурс “Армянский текст русской поэзии”: репрезентация локального текста русской литературы», поддержанного РФФИ.
- EN** The reported study was funded by the RSF as a part of project No. 22-18-00339 “The electronic resource “The Armenian text of Russian poetry”: Representation of the local text of Russian literature”.

Информация об авторах | Author information

- RU** Андреев Сергей Николаевич¹, д. филол. н., проф.
¹ Смоленский государственный университет

- EN** Andreev Sergey Nickolayevich¹, Dr
¹ Smolensk State University

¹ smol.an@mail.ru

Информация о статье | About this article

Дата поступления рукописи (received): 19.06.2023; опубликовано online (published online): 21.08.2023.

Ключевые слова (keywords): Армения; военная лексика; религиозная лексика; коэффициент вариации; евклидово расстояние; Armenia; military vocabulary; religious vocabulary; coefficient of variation; Euclidean distance.