

Научная статья
УДК 81.32
EDN AWTVNX
DOI 10.17150/2713-1734.2025.7(2).224-247



А.В. Боровский

*Байкальский государственный университет,
г. Иркутск, Российская Федерация*

Л.А. Михалёва

*Байкальский государственный университет,
г. Иркутск, Российская Федерация*

Ф.Э. Мосоркин

*Байкальский государственный университет,
г. Иркутск, Российская Федерация*

Метод матрицы парных мер как инструмент сравнения лингвистических множеств

Аннотация. В статье представлена разработка программного обеспечения «Помощник лингвиста», предназначенного для поиска соответствий между словами двух списков с использованием метода матрицы парных мер. Программа попарно сравнивает слова из двух списков, проводит статистические расчёты парных метрик, строит графики для выборок случайных величин, проводит сортировку матрицы парных мер и выдает сборник соответствий между словами двух лексических множеств.

Впервые особое внимание уделено применению данного ПО для выявления скрытых лексических связей между топонимами Иркутской области и старорусским языком, существовавшим 500 лет назад.

В работе рассматриваются основные тенденции лексических изменений, включая замену и утрату букв, перестановку слогов, инверсию слов, трансформацию древнерусских корней, а также влияние национальных языков на фонетическую адаптацию старорусских слов. Также выделены закономерности взаимодействия языков в условиях культурных контактов. Статья подчёркивает перспективы дальнейших исследований, включая расширение словарной базы, использование разнообразных парных метрик, а также углубленный анализ морфологических и фонетических закономерностей эволюции старых слов. Результаты работы могут существенно обогатить понимание исторической лексики и топонимики Восточной Сибири, а также расширить возможности для сравнительных лингвистических исследований.

Ключевые слова. Системный анализ лингвистических списков, историко-математическая лингвистика, парные метрики, метод матрицы парных мер, статистические расчёты парных метрик, старорусский язык, историческая лексика, нерусскоязычные топонимы Иркутской области.

Информация о статье. Дата поступления: 2 апреля 2025 г.; дата принятия к публикации: 11 июня 2025 г.; дата онлайн-размещения: 8 июля 2025 г.

Original article

A.V. Borovsky

*Baikal State University,
Irkutsk, Russian Federation*

L.A. Mikhaleva

*Baikal State University,
Irkutsk, Russian Federation*

F.E. Mosorkin

*Baikal State University,
Irkutsk, Russian Federation*

The Method of the Matrix of Paired Measures as a Tool for Comparing Linguistic Sets

Abstract. The article presents the development of the Linguist's Assistant software, designed to find correspondences between words in two lists using the matrix of paired measures method for comparing words and statistical calculations for numerical values of paired metrics. The program plots random variables, sorts a matrix of paired measures, and provides a dictionary of correspondences between words of two lexical sets. For the first time, special attention is paid to the use of this software to identify hidden lexical connections between non-Russian-speaking toponyms of the Irkutsk region and the Old Russian language that existed 500 years ago.

The paper examines the main trends of lexical changes, including the replacement and loss of letters, the rearrangement of syllables, the inversion of words, the transformation of Old Russian roots, as well as the influence of the Buryat language on the phonetic adaptation of Old Russian words. The patterns of interaction of languages in the context of cultural contacts are also highlighted. The article highlights the prospects for further research, including the expansion of the vocabulary, the use of a variety of paired metrics, as well as an in-depth analysis of the morphological and phonetic patterns of the evolution of old words. The results of the work can significantly enrich the understanding of the historical vocabulary and toponymy of Eastern Siberia, as well as expand opportunities for comparative linguistic research.

Keywords. System analysis of linguistic lists, historical and mathematical linguistics, paired metrics, matrix method of paired measures, statistical calculations of paired metrics, Old Russian language, historical vocabulary, non-Russian toponyms of the Irkutsk region.

Article info. Received 2 April, 2025; Accepted 11 June, 2025; Available online 8 July, 2025.

Введение

Топонимика является важной частью исторического и культурного наследия региона. Исследование происхождения географических названий помогает выявить миграционные процессы, влияние различных языковых пластов и особенности языкового взаимодействия. Изучение топонимов с утерянным смысловым значением относится к задачам историко-математической лингвистики, основы которой изложены, например, в работах [1–6].

Актуальность исследования связана с необходимостью углублённого анализа исторической лексики и её отражения в топо-

нимах региона. Топонимы несут в себе важную информацию о языковых контактах, этнических процессах и культурных изменениях, происходивших на данной территории. Их исследование позволяет глубже понять историю миграционных потоков и связи между различными языковыми группами, что делает изучение данного вопроса актуальным не только с точки зрения истории, но и для современной лингвистики и культурологии.

Добавим, что прогресс в изучении топонимов с утерянным смысловым значением обеспечивает применение методологии системного анализа и разнообразных математических моделей и методов.

В Иркутской области имеется большой пласт топонимики, который никак не интерпретируется на современном русском языке. Поскольку на территории области проживают три языковые группы населения - русские, буряты и эвенки, то лингвисты называют условно эту группу топонимов «бурятско-эвенкийской». С севера к Иркутской области примыкает территория проживания якутского народа, с запада территория енисейских остяков (лингвисты их часто называют кетами), которые также оставили следы в нерусскоязычной областной топонимике. Имеются многочисленные примеры, когда буряты и эвенки не могут интерпретировать нерусскоязычные топонимы и отнести их к бурятскому, либо эвенкийскому языкам. Это обстоятельство говорит о необходимости дополнительного изучения нерусскоязычных топонимов Иркутской области.

Известно, что русское население начало заселять Восточную Сибирь с начала 17-го века и, возможно, даже со второй половины XVI в. Например, село Оёк под Иркутском существовало уже в 1632 г., поселение Качуг на реке Лене известно с 1621 г., город Киренск был образован в 1630 г. как Никольский погост, село Урик — с 1673 г. Лингвистические исследования истории языков показывают, что язык является живой субстанцией. За сто лет примерно 10 % разговорных слов выходят из употребления и заменяются новыми или заимствованными словами. Значит за 400 лет с момента появления носителей русского языка в Восточной Сибири следует ожидать вывода из русского языка до 40% слов. Поэтому в топонимике должен существовать пласт старорусских слов, которые вышли из употребления и смысл их уже забыт. Примерами могут являться слова: *Оёк=скакать, подсказывать* (примером является сохранившееся выражение в современном русском языке «ёкнуло сердце»); *Качуг=петля реки* (соответствиями в русском языке являются слова «кочка» или «качка»); *Карлук=клей из рыбьих костей*; *Куреть=кривое, изогнутое* (село находится на склоне холма) и т.д. В связи с этим представляет большой интерес сравнить группу нерусскоязычных топонимов Иркутской области со старорусским языком, на котором разговаривали русские переселенцы в Восточной Сибири в 16-17-м веках.

В данной работе для анализа топонимов Иркутской области и их соответствий в старорусском языке применяется метод матрицы парных мер (МПМ), который ранее использовался для поиска соответствий между русским и бурятским языками [7].

Изучение соответствий между санскритом и русским языком проводилось в работе [8]. Соответствия между санскритом и нерусскоязычными топонимами Иркутской области найдены в работе [9] методом перебора слов с применением парной метрики Рэтклиффа-Обершелпа [10]. Новизна данной работы заключается в применении метода МПМ для автоматизированного поиска соответствий между современным звучанием топонимов и лексикой старорусского языка. Данный метод представляет собой количественный подход, позволяющий выявить скрытые связи между лексическими единицами и выявить закономерности, которые могли бы остаться незамеченными при использовании традиционных методов сравнительно-исторического анализа [11]. В отличие от классического этимологического исследования корневой системы слов [12], где исследователь субъективно подбирает возможные соответствия, МПМ учитывает совокупность фонетических и морфологических параметров, что делает результаты более объективными и воспроизводимыми.

Метод МПМ является математическим помощником историков и лингвистов. Он находит несколько наиболее близких соответствий для топонима с утерянным смысловым значением и словами языка, с которым производится сравнение. Окончательное правильное соответствие можно определить только анализируя старые тексты, которые раскроют смысловое значение слова.

Метод МПМ может значительно упростить и ускорить анализ, позволяя систематизировать данные и выявлять закономерности, недоступные при традиционном сравнительном анализе. Благодаря автоматизированному подбору соответствий, исследователи могут оперативно анализировать большие объемы информации, что открывает новые возможности для изучения происхождения топонимов. Кроме того, применение данного метода может послужить основой для дальнейших исследований, направленных на выявление глубинных связей между старорусской лексикой и современными географическими названиями, а также их взаимосвязи с другими языковыми пластами, такими как бурятский, эвенкийский, якутский, кетский языки.

Предшествующие исследования включали работы по топонимике Сибири [13–15], анализам заимствований в бурятском языке [7], а также использование санскрита как возможного источника топонимов [8; 9]. Отличие данной работы в применении количественных методов анализа, а именно — метода матрицы парных мер. Данный подход открывает новые перспективы в изучении топонимики региона.

Модели и методы

В работе применялся метод матрицы парных мер (МПМ), который относится к теоретико-множественным методам сравнения лингвистических множеств [16]. Данный метод заключается в следующем. Берутся два лингвистических множества, которые представляются в виде списков слов $A = \{a_i\}, i = 1, 2 \dots I$ и $B = \{b_j\}, j = 1, 2 \dots J$. Далее рассчитывается матрица $\rho_{ij}, i = 1, 2 \dots I; j = 1, 2 \dots J$, элементами которой являются парные метрики, описывающие близость двух слов. Примером может служить метрика Рэтклиффа-Обершелпа (РО), которая описывает долю совпадающих букв в двух словах

$$\rho_{ij} = \frac{2k_{ij}}{|a_i| + |b_j|}. \quad (1)$$

В формуле (1) k_{ij} — число совпадающих букв в каждом из двух слов с номерами i и j , $|a_i|, |b_j|$ — условное обозначение длины слов, т.е. количество букв в каждом слове. Величина ρ_{ij} лежит в пределах $0 \leq \rho_{ij} \leq 1$. Если в двух словах нет одинаковых букв, то $\rho_{ij} = 0$. Если все буквы в двух словах совпадают, то $\rho_{ij} = 1$. Поскольку, значение парной метрики лежит в интервале $[0, 1]$, иногда в литературе говорят о «вероятности» совпадения двух слов. Мы считаем применение термина «вероятность» в данном случае некорректным. Будем говорить просто о значении парной метрики.

После получения матрицы парных мер ρ_{ij} выполняются следующие операции над матрицей:

1. Вычисление среднего значения парных мер для всех элементов матрицы

$$\langle \rho \rangle = \frac{1}{I \cdot J} \sum_{i=1; j=1}^{I; J} \rho_{ij}. \quad (2)$$

2. Нахождение максимальных значений мер по строкам матрицы

$$\rho_i^{\max} = \max_j \rho_{ij}. \quad (3)$$

$i = \text{Const}$

3. Нахождение максимальных значений мер по столбцам матрицы

$$\rho_j^{\max} = \max_i \rho_{ij}. \quad (4)$$

$j = \text{Const}$

4. Нахождение среднего значения от максимальных мер по строкам матрицы

$$\langle \rho_i^{max} \rangle = \frac{1}{I} \sum_{i=1}^I \rho_i^{max}. \quad (5)$$

5. Нахождение среднего значения от максимальных мер по столбцам матрицы

$$\langle \rho_j^{max} \rangle = \frac{1}{J} \sum_{j=1}^J \rho_j^{max}. \quad (6)$$

6. Вычисление дисперсий для всех средних величин

$$D(\rho_{ij}) = \frac{1}{I \cdot J} \sum_{i=1}^I \sum_{j=1}^J (\rho_{ij} - \langle \rho \rangle)^2. \quad (7)$$

$$D(\rho_i^{max}) = \frac{1}{I} \sum_{i=1}^I (\rho_i^{max} - \langle \rho_i^{max} \rangle)^2. \quad (8)$$

$$D(\rho_j^{max}) = \frac{1}{J} \sum_{j=1}^J (\rho_j^{max} - \langle \rho_j^{max} \rangle)^2. \quad (9)$$

Стандартное отклонение находится как корень из дисперсии

$$\sigma = \sqrt{D}.$$

7. Сортировка строк матрицы по убыванию с запоминанием исходных номеров

$$\rho_{iksort} = \{\rho_{ik(i,j)}\}_{j=1, \dots, J; k=1, \dots, J}. \quad (10)$$

Здесь метрика ρ_{ij} в результате сортировки в строке i с места j переходит на место k .

8. Вычисление средних значений мер по столбцам от отсортированной матрицы

$$\bar{\rho}_k = \frac{1}{I} \sum_{i=1}^I \rho_{iksort}. \quad (11)$$

9. Построение графика средних значений мер по столбцам от отсортированной матрицы.

10. Выдача словаря на базе нескольких первых столбцов отсортированной матрицы. Практика показала, что достаточно рассматривать первые 10 соответствий в отсортированной матрице $\rho_i = (\rho_{[i, 1]}, \rho_{[i, 2]}, \dots, \rho_{[i, 10]})$. Словарь тогда выглядит следующим обра-

зом: a_i — слово из первого списка, b_j — слово из второго списка, ρ_j — мера соответствия между этими словами, c_j — перевод слова b_j на русский язык.

$$a_i \quad b_{j(k=1)} \quad \rho_{j(k=1)} \quad c_{j(k=1)} \quad b_{j(k=2)} \quad \rho_{j(k=2)} \quad c_{j(k=2)} \quad \dots \quad b_{j(k=10)} \quad \rho_{j(k=10)} \quad c_{j(k=10)}$$

Здесь индекс j является функцией k . Индекс $i = 1, 2 \dots I$ пробегает все топонимы.

Разработка программного обеспечения

В последние годы автоматизация лингвистических исследований набирает всё большую популярность. В рамках данного исследования была разработана программа для сравнения двух лингвистических множеств с использованием метода матрицы парных мер (МПМ). Основной задачей программы является автоматизированный поиск соответствий между словами двух множеств, что позволяет значительно ускорить процесс лингвистического анализа и повысить его точность. Программа позволяет учесть фонетические изменения звуков, морфологические преобразования слов, вскрыть закономерности исторического развития языка, выявить латентные связи, которые могут быть незаметны при традиционных методах анализа. Программа представлена в виде оконного приложения. Для начала рассмотрим инструменты разработки.

Основным языком программирования для реализации данной программы выбран Python¹. Для создания каркаса программы использована библиотека PyQt, которая часто применяется для разработки приложений с графическим интерфейсом. Чтобы работать с импортами списков слов и экспортом результатов применяется библиотека `orenpuxl`, так как все входные данные записываются в Excel-файлах, которые должен предоставить пользователь. Для работы с Word-файлами и экспорта отчётов использована библиотека `docx`.

Основные требования к приложению приведены в виде диаграммы *use case* на рис. 1.

Опишем процесс работы с программой. Сначала пользователь должен загрузить необходимые списки слов, которые будут сравниваться. После этого все слова будут отображены и пронумерованы в поле «Список А» и «Список В». Потом нужно выбрать на интерфейсе программы необходимую модель преобразования слов: без преобразования D0, с переводом согласных в консонантные классы и гласных в отдельный класс гласных звуков D1, с переводом только согласных в консонантные классы и исключением из рассмотрения гласных звуков D2. Затем нажимается кнопка «обработать слова». Необходимая статистика будет отображена в поле «результаты»:

¹ Язык программирования Питон. URL: <https://docs.python.org/3.12/>.

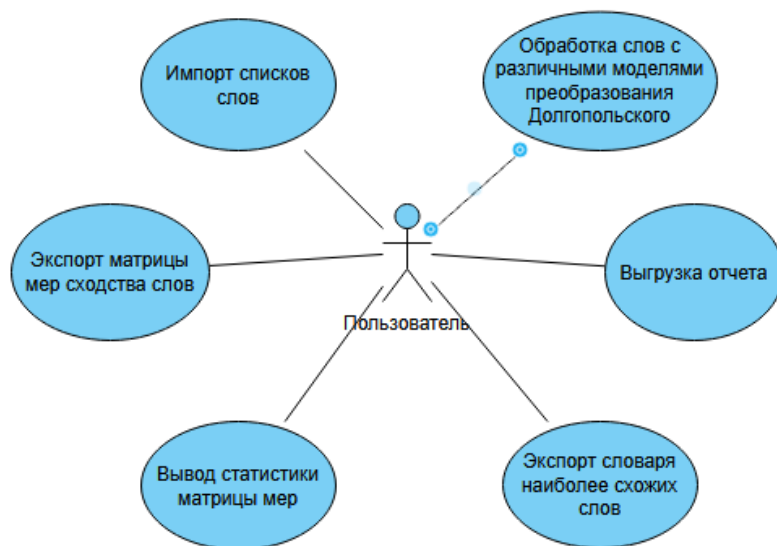


Рис. 1. Диаграмма работы с программным обеспечением (use case)

- Среднее арифметическое по всем элементам матрицы мер;
- Среднее арифметическое от максимумов мер по строкам;
- Среднее арифметическое от максимумов мер по столбцам;
- Дисперсия по всем элементам матрицы мер;
- Среднеквадратичное отклонение по всем элементам матрицы мер;
- Дисперсия от максимумов мер по строкам;
- Среднеквадратичное отклонение от максимумов мер по строкам;
- Дисперсия от максимумов мер по столбцам;
- Среднеквадратичное отклонение от максимумов мер по столбцам.

После обработки слов у пользователя есть возможность экспортировать матрицу парных мер сходств слов по указанной модели в виде Excel-файла. Также можно экспортировать словарь, в котором приведены по 10 наиболее схожих слов для каждого слова из первого списка. Таким образом, приложение обрабатывает списки слов, необходимые результаты выводит в окне «результаты» или их можно экспортировать в Excel-файл или Word-файл.

Подготовка исходных списков слов

Для анализа использовались:

1. Список населенных пунктов Иркутской области.
2. Список гор Иркутской области.
3. Список рек Иркутской области.

4. Старорусский словарь, включающий слова, использовавшиеся до XV в.

Списки 1–3 были составлены на основе географических справочников и карт.

Словарь старорусского языка² был очищен от омонимов, повторяющихся форм и редких слов с неустановленной этимологией. Подготовка данных также включала нормализацию словоформ, устранение некоторых архаизмов, не имеющих современных аналогов, и применение морфологического анализа для выявления корневых форм слов. Например, старорусское слово «гость» могло означать как «гость», так и «купец». В очищенном словаре остаётся только одно значение, наиболее релевантное в данном контексте. В старорусском языке существовали варианты одного слова с различными окончаниями, например, «братъ» и «брат», которые в нормализованном словаре сведены к одной форме. Были исключены редкие слова с неустановленной этимологией, к примеру слово «гълухъ», встречающееся в редких источниках с неясным значением. Форма «хожаху» (прошедшее время множественного числа от «ходить») заменяется на современный эквивалент ходили. Слово «вег» (устаревшее обозначение пути) было исключено, так как оно не имеет точного современного аналога. Разные формы слова, такие как «рекоша», «рече», «рекл» (разные формы глагола «говорить»), приводятся к единой корневой форме «ректи» (или её современному аналогу «сказать»).

Определение чувствительности метода МПМ

В данном пункте исследована чувствительность метода матрицы парных мер. Под чувствительностью будем понимать относительное изменение какого-либо среднего значения $\langle \rho \rangle$ матрицы мер при замене одного слова на новое.

$$\varepsilon = \frac{\Delta \langle \rho \rangle}{\langle \rho \rangle} \quad (12)$$

Был взят список населённых пунктов Иркутской области из 567 слов и словарь старорусского языка на 2442 слова. Были заменены 10 слов в словаре старорусского языка на новые, которых ранее не было в словаре. Абсолютные изменения вычисленных средних показаны в табл. 1. Метод МПМ чувствует появление даже одного нового слова. При этом отклонения, указанные в табл. 1, уменьшаются примерно в 10 раз. Чувствительность метода на изменение одного слова достигает $2 \cdot 10^{-5}$.

² Старорусские слова. URL: <chrome-extension://efaidnbmnnnnibpcajpcgglefindmkaj/https://barley-malt.ru/wp-content/uploads/2014/08/staroruss-27082014.pdf> (дата обращения: 02.04.2025).

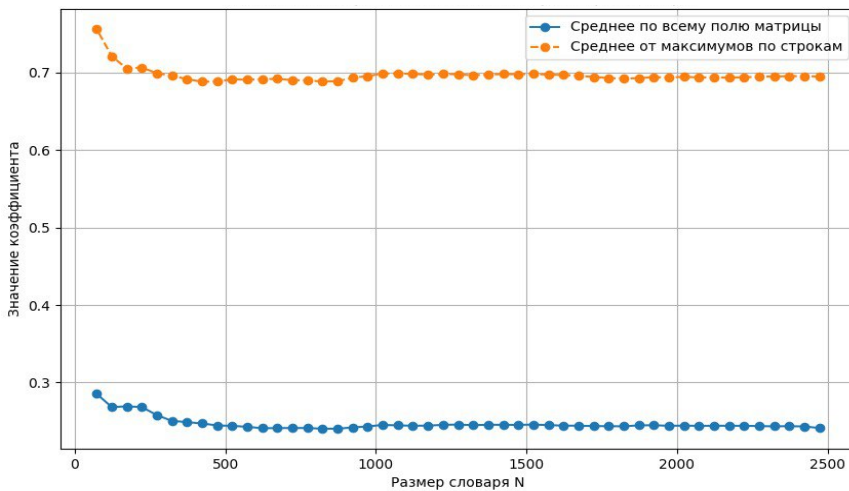
Таблица 1

Определение чувствительности метода МПМ

Параметр	$\langle \rho \rangle$	$\Delta \langle \rho \rangle$	ε
Среднее по всему полю матрицы $\langle \rho \rangle$	0.241122	0,000051	0,000212
Средние от максимумов по строкам $\langle \rho_i^{max} \rangle$	0.694414	0,000214	0,000308
Средние от максимумов по столбцам $\langle \rho_j^{max} \rangle$	0.776584	0,000392	0,000505

Зависимость средних величин от длины словаря

Средние значения, которые вычисляются на основе матрицы парных мер, являются статистическими характеристиками сравниваемых языковых множеств. В связи с этим представляет интерес изучение зависимости средних величин от длины словаря. Был выбран тот же вариант исходных данных для матрицы мер, что и при изучении чувствительности МПМ. Такое сравнение представлено на рис. 2.

**Рис. 2. Зависимость средних значений от размера словаря**

Имеется также другое соображение для изучения указанной зависимости. Увеличение количества слов в словаре приводило к расширению возможностей лингвистического анализа топонимов, но усложняло расчёты. Поэтому были проведены тестовые запуски программы с различными объемами словарных данных для поиска оптимального баланса между скоростью и возможностями анализа. Рисунок 2 показывает, что статистические характеристики практически не зависят от длины словаря, начиная с 300 слов в словаре.

Результаты

Результаты статистических расчётов с использованием формул (1–9) совпадения лингвистических множеств представлены в табл. 2.

Таблица 2

**Статистические характеристики совпадения
лингвистических множеств: группы нерусскоязычных топонимов
Иркутской области — словарь старорусского языка**

Параметр	Населённые пункты	горы	реки
D0			
$\langle \rho \rangle$	0.241 ± 0.153	0.245 ± 0.150	0.236 ± 0.152
$\langle \rho_i^{max} \rangle$	0.776 ± 0.073	0.760 ± 0.087	0.776 ± 0.075
$\langle \rho_j^{max} \rangle$	0.694 ± 0.082	0.659 ± 0.086	0.616 ± 0.093
D1			
$\langle \rho \rangle$	$0.569 \pm 0,144$	$0.577 \pm 0,141$	$0.556 \pm 0,143$
$\langle \rho_i^{max} \rangle$	$0.976 \pm 0,036$	$0.971 \pm 0,038$	$0.975 \pm 0,038$
$\langle \rho_j^{max} \rangle$	$0.924 \pm 0,064$	$0.905 \pm 0,067$	$0.874 \pm 0,074$
D2			
$\langle \rho \rangle$	$0.385 \pm 0,224$	0.401 ± 0.217	$0.351 \pm 0,227$
$\langle \rho_i^{max} \rangle$	0.984 ± 0.068	0.979 ± 0.070	$0.981 \pm 0,094$
$\langle \rho_j^{max} \rangle$	0.929 ± 0.085	0.898 ± 0.101	$0.861 \pm 0,106$

Изучение задачи проводилось в рамках следующих трёх моделей.

Модель D0. Брались перечни топонимов, населённых пунктов (567 шт.), гор (290 шт.) и рек (130 шт.), и словарь старорусского языка 500-летней давности³ на 2442 слова. Проводились расчёты матрицы парных мер согласно формуле (1). Каждый топоним рассматривался с каждым словом словаря. Общее количество посчитанных парных мер составило величину $(567 + 290 + 130) \times \times 2442 = 2\,410\,254$. Парная мера (1) учитывает долю совпадающих букв в двух словах. При этом порядок следования букв не учитывается. Такой подход не является ни плохим, ни хорошим он позволяет выбрать из двух с половиной миллионов пар слов, 10 тыс. слов с близким буквенным составом и не более того.

Модель D1. Брались те же перечни топонимов и старорусский словарь. Однако, слова предварительно обрабатывались. Буквы в словах переводились в классы звуков [17], разработанные известным лингвистом А.Б. Долгопольским. Давно установлено, что при эволюции языков во времени некоторые звуки легко переходят друг в друга, а некоторые группы звуков могут испытать му-

³ Старорусские слова. URL: <chrome-extension://efaidnbmninnibpcapjpcgglefindmkaj/https://barley-malt.ru/wp-content/uploads/2014/08/staroruss-27082014.pdf> (дата обращения: 02.04.2025).

тацию с меньшей вероятностью. Классы звуков А.Б. Долгопольского представлены в табл. 3.

Таблица 3

Классы согласных для русского языка

	Класс согласных	Буквы русского языка
1	Р-класс	П, Б, В, Ф
2	Т-класс	Т, Д
3	С-класс	С, З, Ц, Ч, Ш, Щ, Ч
4	М-класс	М
5	Н-класс	Н
6	Р-класс	Р, Л
7	К-класс	К, Г, Х
8	Н-класс (нулевой класс)	Все гласные, включая Ё, Ъ, Ю, Я

В методе предполагается, что используя предложенную систему обозначений классов (Р, Т, S, М, N, R, К, Н) можно закодировать любые сравниваемые реальные словоформы. Одинаковые классы для двух слов сравниваемых языков указывают на большую вероятность того, что слова имеют родственное происхождение в отличие от словоформ, кодировки классов для которых существенно отличаются.

Приближение D2. В этом приближении исключён из рассмотрения класс гласных звуков, который кодируется символом Н. Гласные звуки в любых языках подвижны, что приводит к возникновению диалектов. Для выявления общих слов в различных языках лингвисты часто используют костяк согласных. Если костяки согласных в двух словах совпадают, то вероятность того, что эти слова родственные повышается.

Статистические характеристики. Элемент матрицы $\rho_{ij} = m$, $m \leq n$ представляет собой правильную рациональную дробь $0 \leq \rho_{ij} \leq 1$, где m и n — целые числа. Строки и столбцы матрицы представляет собой конечные дискретные выборки случайных правильных рациональных дробей. Существование элементов матрицы в виде рациональных чисел в рассматриваемой задаче какой-то особой роли не играет. Дробь в компьютере рассматриваются в десятичной записи. Выборки обладают среднеарифметическими значениями, дисперсиями и стандартными отклонениями.

Величина $\langle \rho \rangle$ есть среднее арифметическое от всех элементов матрицы мер. Указаны также доверительные интервалы. Напомним, что запись $\langle \rho \rangle = 0.241 \pm 0.153$ означает, что величина ρ с вероятностью 68 % попадает в интервал значений $[0.241 - 0.153, 0.241 + 0.153]$ в предположении, что случайная величина ρ распределена по нормальному закону. Закон распределения случайных величин, в данной работе не изучался. Поэтому стандартные отклонения указаны в рамках гипотезы о нормальном распределении.

Величины $\langle \rho_i^{max} \rangle$, $\langle \rho_j^{max} \rangle$ есть средние значения от максимальных парных мер по строкам и столбцам матрицы мер. График максимального сходства ρ_i^{max} для модели D0 представлен на рис. 3. Рассматривается матрица мер: населенные пункты — словарь старорусского языка.

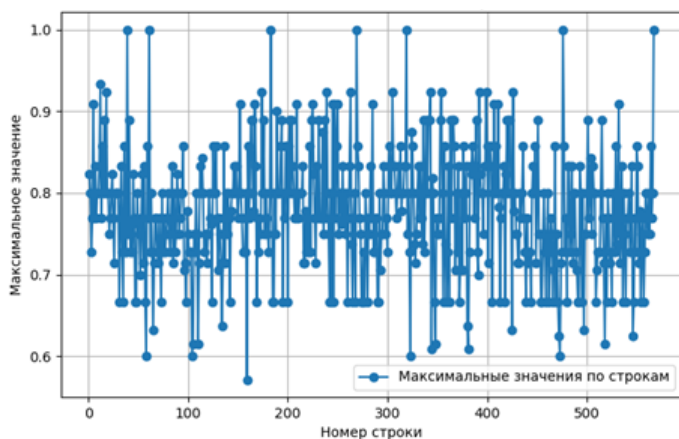


Рис. 3. График максимальных значений парных мер по строкам как функция номера строки

Вернёмся к обсуждению результатов, представленных в табл. 2.

1. Значения средних от мер, максимальных по строкам, всегда больше, чем значения средних от мер, максимальных по столбцам $\langle \rho_i^{max} \rangle > \langle \rho_j^{max} \rangle$. Причина заключается в прямоугольной форме матрицы мер. В самом деле если бы в словаре старорусского языка были все топонимы, то среднее от максимальных мер по строкам равнялось бы единице, а среднее от максимальных по столбцам было бы меньше единицы, так как нашлись бы столбцы, где отсутствуют пересечения слов.

2. Значения средних $\langle \rho \rangle$, $\langle \rho_i^{max} \rangle$, $\langle \rho_j^{max} \rangle$ для трех групп топонимов — населённых пунктов, гор и рек достаточно близки. Например, в модели D0:

Параметр	Населённые пункты	Горы	Реки
$\langle \rho \rangle$	0.241 ± 0.153	0.245 ± 0.150	0.236 ± 0.152

Данное обстоятельство указывает на языковую близость всех трёх групп названий. Можно также предполагать, что для всех трёх групп названия были даны примерно в один и тот же временной период. Иначе слова бы видоизменились и средние значения бы разошлись по величине.

1. Сравнение средних для моделей D0 и D1, D2 показывает их увеличение для моделей D1, D2. Такое увеличение указывает на то, что современное звучание топонимов разошлось со временем с произношением слов старорусского языка.

2. Сравнение средних для моделей D1 и D2 показывает, что средние по строкам и столбцам в моделях D1 и D2 примерно совпадают, а средние по всем элементам матрицы парных мер значительно отличаются. Такое отличие связано с исключением класса гласных букв в модели D2. Учёт гласных букв увеличивает разнообразие слов, что приводит к уменьшению средней доли совпадающих букв в словах.

3. Обращают на себя внимание в представленных расчётах высокие значения (близость к единице) средних от максимальных по строкам и столбцам особенно в моделях D1 и D2. Это указывает на наличие в списках топонимов и словаре старорусского языка слов с близким звучанием или с близким составом классов А.Б. Долгопольского.

Отсортируем каждую строку матрицы парных мер по убыванию значений метрик на основе формулы (10). Вычислим затем средние значения для каждого столбца отсортированной матрицы, формула (11). График таких средних в зависимости от номера столбца (порядковый номер слов в словаре) представлен на рис. 4.

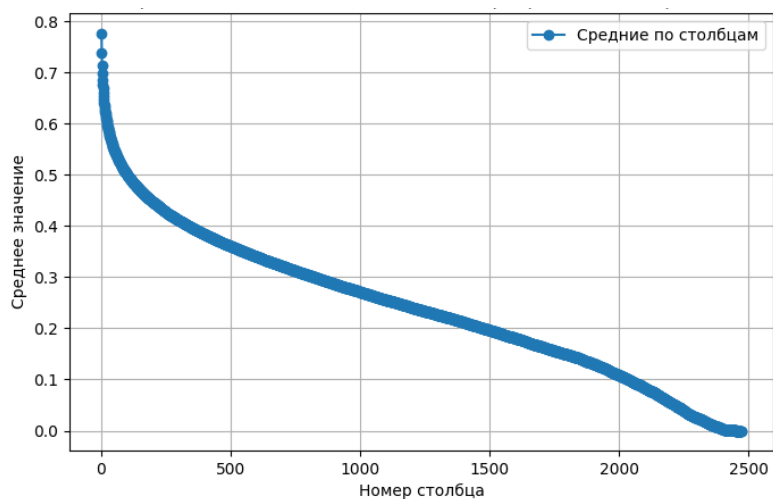


Рис. 4. График среднего значения метрики по столбцам для отсортированной матрицы как функции номера столбца. Модель D0. Населённые пункты - словарь старорусского языка

На рис. 4 можно выделить три области. В первой области $0 < k < 50$ при $k \rightarrow 0$ значения мер резко возрастают. Это область возможного заимствования топонимов из старорусского языка.

В области $50 < k < 2000$ имеет место близкая к линейной зависимость средних мер от номера k . Слова любого языка состоят из слогов, которых в русском языке порядка 200. Слога повторяются в различных словах и составляют общую часть несовпадающих по смыслу слов. Вторая область указывает на слоговую структуру слов.

Третья область $2000 < k < 2442$ состоит из практически несовпадающих слов, для которых доля совпадающих звуков в словах меньше 0.1.

Одним из результатов работы программы является «Словарь соответствий». К каждому топониму подыскивается по 10 наиболее близких слов старорусского языка на основе формулы (12). Критерием подбора слов является парная мера близости слов Рэтклиффа-Обершелпа (РО), которая равна доле совпадающих звуков (букв) в двух словах.

Таблица 4

Фрагмент «Словаря соответствий»

Яр	яр	1.00	пря	0.80	зря	0.80
Моты	мыто	1.00	ометы	0.89	поматы	0.80
Барда	брада	1.00	бразда	0.91	рада	0.89
Бадар	брада	1.00	бразда	0.91	рада	0.89
Ут	ту	1.00	кут	0.80	тук	0.80
Коты	коты	1.00	ток	0.86	толк	0.75
Зима	зима	1.00	взаим	0.89	измена	0.80
Алтарик	параклит	0.93	алкати	0.92	крата	0.83
Одинск	сходник	0.92	искони	0.83	дискос	0.83
Колтоши	толико	0.92	волотки	0.86	точило	0.77
Тайтура	татаур	0.92	оратай	0.77	заутра	0.77
Катарма	камара	0.92	матка	0.83	крата	0.83
Закора	проказа	0.92	рака	0.80	зрак	0.80
Никилей	еликий	0.92	зелейник	0.80	елеонский	0.75
Тремина	матери	0.92	метание	0.86	мнети	0.83
Маракан	камара	0.92	кампан	0.77	рукама	0.77
Солзан	соблазн	0.92	назола	0.83	сало	0.80
Ангарск	красна	0.92	анфракс	0.86	агарян	0.77
Ерема	мережа	0.91	раменье	0.83	аер	0.75
Тарма	аромат	0.91	матка	0.80	крата	0.80
Аксай	стайка	0.91	фаска	0.80	касиа	0.80
Курма	рукама	0.91	крома	0.80	ручка	0.80
Добчур	обруч	0.91	удоб	0.80	бердо	0.73
Исеть	сиеть	0.91	сеть	0.89	истее	0.80
Тарай	оратай	0.91	врата	0.80	крата	0.80
Карам	камара	0.91	рукама	0.91	рака	0.89
Келора	рекла	0.91	кораблец	0.86	стрекало	0.86

Окончание табл. 4

Квиток	кивот	0.91	свиток	0.83	волотки	0.77
Челяба	лечба	0.91	обаче	0.73	блаже	0.73
Кеуль	кудель	0.91	купель	0.91	скудель	0.83
Игжиновка	кандиловжиг	0.90	иноковати	0.78	виклина	0.75
Ката	матка	0.89	крата	0.89	казати	0.80

Число 10 является условным. Его можно изменять. Численное моделирование показывает, что число 10 является достаточным для успешного поиска большинства соответствий слов. «Словарь соответствий» для нерусскоязычных топонимов Иркутской области занимает объём порядка 100 стр. Мы его естественно не можем привести в короткой статье. Однако, планируем его издание в электронном виде после дополнительных исследований.

Отметим, что программное обеспечение, раскрытое в данном исследовании, является всего лишь помощником лингвиста или историка, оно позволяет кардинально сузить объём поиска соответствий между топонимами и словарями языка. Для Иркутской области вместо 2,5 млн вариантов программа выдаёт ограниченное поле в 10 тысяч вариантов для дальнейшего поиска. Окончательное соответствие может быть установлено только при изучении старых текстов с установлением смыслового значения тех или иных старых слов. Также при увеличении размера словаря может снижаться точность оценки сходства из-за увеличения количества близких слов и языковой вариативности. Кроме того, эффективность метода зависит от качества предварительной обработки словаря.

В результате работы программы было получено множество соответствий между топонимами Иркутской области и старорусскими словами. Выявлены основные тенденции:

1. Значительное количество топонимов демонстрирует фонетические изменения, типичные для старорусского языка:

Замена *г* на *в* (характерно для старорусских диалектов)

Голымба → *Волымба* (трансформация названий сибирских рек)

Редукция гласных в безударных слогах

Ангарь → *Ангара* (исчезновение редуцированных гласных ъ, ь)

Оглушение конечных согласных

Иркутъ → *Иркут* (современный Иркутск)

Метатеза (перестановка звуков)

Берёзовъ → *Берёзово* (как в диалектных названиях деревень)

Адаптация бурятских и эвенкийских названий через старорусские фонетические нормы

Тулын → *Тулун*

Илга → *Илгая*

2. Обнаружены примеры утраты и замены букв, что может свидетельствовать об адаптации слов к местному произношению:

Утрата звуков (редукция)

Иркутъ → Иркутск (исчезновение конечного мягкого «ъ», характерного для древнерусского языка)

Ангарь → Ангара (утрата редуцированного «ъ» в конце слова)

Ленес → Лена (возможная адаптация эвенкийского названия)

Замена звуков (фонетическая адаптация)

Олекма → Олёкма (замена «е» на «ё», упрощение произношения)

Тукулан → Тукалан (упрощение согласного сочетания)

3. Выявлены случаи перестановки слогов, характерные для исторической трансформации топонимов, которые могут быть объяснены дислолией речи людей, обладавших ограниченной грамотностью:

Перестановка слогов в русских топонимах

Братково → Братск (упрощение сложного сочетания, характерное для просторечия)

Метатеза в заимствованных топонимах

Тугур → Гутур (эвенкийское название с изменённым порядком слогов)

Коченёво → Коченева (вариант, появившийся из-за народного произношения)

Улус-Ботогол → Ботогол-Улус (смена порядка слов в названии бурятских селений)

Изменения в сложных названиях из-за особенностей устного произношения

Вилюйск → Вилуй (исключение сложного окончания)

4. Зафиксированы закономерности изменения суффиксов и окончаний, указывающие на влияние диалектных особенностей и языковых контактов:

Замена древнерусских окончаний на более типичные для сибирских говоров

Иркутъ → Иркутск (добавление -ск вместо исчезнувшего мягкого знака)

Якутъ → Якутск (переход к -ск, типичному для названий городов)

Адаптация бурятских и эвенкийских названий через русские словообразовательные модели

Баргузин → Баргузинск (добавление русского суффикса -ск)

Тунка → Тункинский (образование прилагательной формы с -ский)

Замена редких суффиксов на более привычные

Ленес → Лена (упрощение окончания, возможно под влиянием русских аналогов)

Анаул → **Анадырь** (изменение окончания на -ырь, характерное для северных названий)

Влияние диалектов на топонимические окончания

Шилка → **Шилкин** (переход к притяжательной форме, характерной для сибирских диалектов)

Витим → **Витимский** (добавление русскоязычного суффикса при освоении территории).

Метод матрицы парных мер позволяет учитывать различные виды фонетических и графических изменений. В сравнении с предыдущими исследованиями [7; 9], использовавшими санскрит и бурятский язык, метод МПМ выявил ряд новых соответствий, ранее не зафиксированных. Применение данного метода позволило обнаружить следующие тенденции:

5. Некоторые топонимы происходят от старорусских слов с потерей конечных согласных (Култукъ → Култук, Иркутъ → Иркут).

6. Встречаются случаи перестановки букв и слогов, что указывает на влияние фонетических законов (Тугур → Гутур).

7. Обнаружены примеры адаптации старорусских корней под бурятскую фонетику (Турка → Турхэ).

8. Прослеживаются закономерности лексического заимствования, характерные для миграционных процессов средневекового периода (Олекма → Олёкма - русское произношение эвенкийского названия реки).

Приведём примеры некоторых соответствий:

1. Топоним «Урик» — слово «уречь» (ограждать, защищать); слог ур сохранился к примеру в военном крике «ура»; имеются также слова со слогом ур — «урочище», «ургал» — закрытое, защищённое от солнца место. Гал, гало — солнце, отсюда же «уголь».

2. Топоним «Оса» — слово «оса» (защитное укрепление); как известно, осы строят из специального состава крепкие ульи, которые являются для них защитными укреплениями.

3. Топоним «Ангара» — возможная связь со словом «анг» (изгиб, петля, что может описывать извилистый характер реки). Другая интерпретация «анга» — горы, холмы, сопки, холмистая местность, чему не противоречит первая интерпретация. Слово «анга» однокоренное с современным словом «нога» — как изгиб человеческого тела. В обратном прочтении возникают однокоренные слова — гон, гнать, погоня и огонь, который гонит обитателей леса. Слово «анга» явно имеет очень древнее происхождение, при помощи которого люди описывали окружающую действительность. Слово анга закрепилось в русском языке в словах ангар — сооружение, похожее на гору; ангина — бугры в горле. В Крыму имеется перевал Ангарский — т.е. горный. В Иркутской области гора Ангасолка — т.е. гора с выходами соли.

Выводы

Разработано программное обеспечение «Помощник лингвиста». Оно предназначено для поиска соответствий между словами двух списков. ПО использует метод парных мер, статистические расчёты, сортировку матрицы парных мер, выполняет построение графиков выборок случайных величин и соответствующих средних значений, выдаёт словарь соответствий между словами двух списков (лингвистических множеств).

Модели D0–D2 выбраны как отражающие разные уровни эволюции слов во времени: отсутствие изменений (D0) возникновение диалектов (D1), расхождения диалектов и трансформация слов (D2), ~~глубокая эволюция слов (D3)~~. Это позволяет оценить вклад разных аспектов эволюции слов в итоговое сходство.

Для исследования эволюции слов на глубину в 400 лет важны модели D0 и D1. В отдельных случаях для учёта национальной адаптации слов требуется привлекать модель D2.

Разработанное ПО показало свою эффективность в выявлении скрытых лексических связей между нерусскоязычными топонимами Иркутской области и старорусским языком 500 летней давности. Топонимика Восточной Сибири и, в частности, Иркутской области, представляет собой сложную лингвистическую систему. На территории происходило взаимодействие нескольких народов, русских, бурят, эвенков, якутов, енисейских остяков. Каждый народ оставил свою группу топонимов. Осложняет дело то, что языки меняются во времени, смысл части топонимов уже забыт, слова выведены из употребления. Описанная в данной работе компьютерная программа реализует методологию системного анализа применительно к изучению топонимики региона.

Основные тенденции включают:

1. Утрата букв и перестановка слогов в процессе адаптации слов.

В процессе адаптации слов в разных языковых системах часто происходят изменения, связанные с утратой букв или перестановкой слогов, что является частью процесса упрощения или эволюции фонетики слов. Программное обеспечение выявляет такие изменения, что позволяет понять, как слова адаптировались к местной фонетической системе. Это может быть связано с различиями в произношении, недостаточной грамотностью населения в средневековый период или природой языковых контактов.

2. Инверсия слов.

Интересной особенностью, выявленной при анализе, является явление инверсии — чтение слов слева направо и справа налево, что может быть связано с особенностями записи слов в разных языковых традициях. Например, в определенных случаях топонимы могли быть адаптированы или зафиксированы в противоположном порядке из-за тюркской (арабской) практики письменности.

3. Трансформация древнерусских корней в измененную форму ПО также помогает отслеживать изменения в древнерусских корнях, их трансформацию в изменённые формы. Это может включать как фонетические изменения, так и морфологические преобразования, например, добавление или удаление суффиксов, изменение окончания слов.

4. Влияние бурятского языка на фонетическую адаптацию старорусских слов.

Значительное влияние бурятского языка на фонетическое восприятие и адаптацию старорусских слов также было выявлено с помощью этого ПО. Топонимы и другие слова, заимствованные из русского языка, подвергались изменениям в соответствии с особенностями бурятского произношения, что проявляется в изменении звуков и структуры слов. Это свидетельствует о глубоком языковом взаимодействии в историческом контексте.

5. Закономерности взаимодействия языков на стыке культур. Программное обеспечение позволяет выявить закономерности, которые происходят на стыке культур и языков. Например, взаимодействие старорусского языка с местными языками Сибири привело к взаимным заимствованиям, которые изменяли фонетику, грамматику и лексику как в русском, так и в национальных языках.

Перспективы дальнейших исследований включают расширение словарной базы и углублённого понимания лексических и фонетических изменений. Необходимо расширять словарную базу, включив в неё более широкий спектр топонимов, лексем и различных словарных единиц, что позволит исследовать их эволюцию в ещё большем объёме и в разных контекстах. Расширение метрик, используемых для сравнения слов, откроет новые возможности для более точной оценки изменений и точности соответствий. Включение метрик, адаптированных к специфике различных языков (например, тюркских или других региональных языков), позволит выявить закономерности и скрытые связи между языками, что обогатит лексикографический и фонетический анализ.

Одним из важных направлений дальнейших исследований является расширение спектра языков для сравнительного анализа, включая тюркские языки. Это поможет глубже понять закономерности взаимодействия сибирских языков, их влияние на русский язык и наоборот.

Важным шагом в дальнейшем исследовании станет углубленный анализ изменений в морфологии и фонетике старорусских слов, что позволит выявить не только поверхностные изменения, но и более глубокие трансформации, происходившие в истории языка. Анализ таких изменений может существенно расширить наше понимание не только о конкретных топонимах, но и об исторической лексике в целом. Будущие исследования могут учи-

тивать также социальную и культурную динамику, которая влияет на формирование лексических и топонимических изменений. Углубление в исторические процессы миграции, взаимодействия народов и изменения в культурной идентичности откроет новые перспективы для понимания эволюции языка.

Разработанное программное обеспечение представляет собой эффективный инструмент для лингвистических исследований, который способен значительно улучшить понимание эволюционных процессов, происходивших в языке. Особое внимание стоит уделить его применению для анализа старорусских и местных топонимов, выявляя скрытые связи и закономерности их эволюции. Перспективы дальнейших исследований в этом направлении могут не только углубить знания о лексике и топонимике Восточной Сибири, но и в целом обогатить изучение исторического языка и культурных процессов в данной области.

Список использованной литературы

1. Абдрахманова Р.Дж. Лингвистическая типология : учеб. пособие / Р.Дж. Абдрахманова. — Бишкек : Изд-во КРСУ, 2014. — 108 с.
2. Алпатов В.М. История лингвистических учений : учебник и практикум / В.М. Алпатов, С.А. Крылов. — 5-е изд., перераб. и доп. — Москва : Юрайт, 2024. — 452 с.
3. Батура Т.В. Математическая лингвистика и автоматическая обработка текстов на естественном языке : учеб. пособие / Т.В. Батура. — Новосибирск : Изд-во НГУ, 2016. — 166 с. — EDN WOMTNJ.
4. Плунгян В. Почему языки такие разные. Популярная лингвистика / В. Плунгян. — Москва : АСТ-пресс книга, 2010. — URL: http://royallib.com/book/plungyan_vladimir/pochemu_yaziki_takie_raznie_populyarnaya_lingvistika.html.
5. Пиотровский Р.Г. Математическая лингвистика / Р.Г. Пиотровский, К.Б. Бектаев, А.А. Пиотровская. — Москва : Высш. шк., 1977. — 383 с. — EDN ZDHKKT.
6. Lobovikov V.O. Artificial Intelligence and an Almost Unknown Aspect of Mathematical Linguistics / V.O. Lobovikov. — DOI 10.17506/18179568_2022_19_3_170. — EDN KOOCKW // Discourse-P. — 2022. — Vol. 19, no. 3. — P. 170–184.
7. Боровский А.В. Изучение связи между русским и бурятским языками методом матрицы мер близости между консонантными классами слов / А.В. Боровский, В.В. Братищенко, Е.Е. Раковская. — DOI 10.17150/2713-1734.2023.5(1).19-33. — EDN CEMRYO // System Analysis & Mathematical Modeling. — 2023. — Т. 5, № 1. — С. 19–33.
8. Аспекты санскритско-славянской этимологии на примере русского языка: тематические группы / Д.В. Денисов, М.Ю. Журавлев, Н.А. Трибунская [и др.]. — Самара : Слово, 2021. — 352 с. — EDN HGBLOM.
9. Боровский А.В. Изучение корреляции между топонимами Иркутской области и словами санскрита / А.В. Боровский, Е.Е. Раковская, И.А. Картеева. — DOI 10.17150/2500-2759.2022.32(3).610-622. — EDN UFJGAV // Известия Байкальского государственного университета. — 2022. — Т. 32, № 3. — С. 610–622.
10. Ratcliff J.W. Pattern-matching-the gestalt approach / J.W. Ratcliff, D.E. Metzner // Dr Dobbs Journal. — 1988. — Vol. 13, no. 7. — P. 46.
11. Мейе А. Сравнительный метод в историческом языкознании / А. Мейе ; пер. с фр. А.В. Дилигенской. — 2. изд., испр. и доп. — Москва : УРСС, 2004. — 101 с. — EDN QQTRGP.

12. Омельченко В.В. Методические аспекты комплексного анализа корневой системы и образно-смыслового содержания слов разных языков / В.В. Омельченко. — EDN UNITLT // Вестник Российского университета дружбы народов. Серия: Теория языка. Семиотика. Семантика. — 2015. — № 3. — С. 148–160.

13. Мельхеев М.Н. Географические названия Восточной Сибири. Иркутской и Читинской области / М.Н. Мельхеев. — [Иркутск] : Вост.-Сиб. кн. изд-во, 1969. — 121 с.

14. Ахадов Э. Тюркская топонимика Восточной Сибири (Бурятия, Иркутская область) / Э. Ахадов // ВКонтакте. — 2025. — 3 янв. — URL <https://vk.com/@dolma60-turkskaya-toponimika-vostochnoi-sibiri-buryatiya-irkutskaya>.

15. Гурулёв С.А. Географические названия Иркутской области : топонимический словарь / С.А. Гурулёв. — Иркутск : Изд-во Ин-та географии им. В.Б. Сочавы СОРАН, 2015. — 575 с.

16. Боровский А.В. Применение вычислительных методов в лексикостатистических исследованиях топонимов Иркутской области / А.В. Боровский, Е.Е. Раковская. — EDN GURKJV // Ляпуновские чтения 2024 : материалы 40-й междунар. конф., Иркутск, 02–06 дек. 2024 г. — Иркутск, 2024. — С. 30–32.

17. Долгопольский А.Б. Гипотеза древнейшего родства языковых семей Северной Евразии с вероятностной точки зрения / А.Б. Долгопольский // Вопросы языкознания. — 1964. — № 2. — С. 64–69.

References

1. Abdrakhmanova R.Dzh. Linguistic typology. Bishkek, Kyrgyz-Russian Slavic University Publ., 2014. 108 p.

2. Alpatov V.M., Krylov S.A. History of linguistic teachings. 5th ed. Moscow, Yurait Publ., 2024. 452 p.

3. Batura T.V. Mathematical linguistics and automatic processing of natural language texts. Novosibirsk State University Publ., 2016. 166 p. EDN: WOMTNJ.

4. Plungyan V. Why Languages Are So Different. Popular Linguistics. Moscow, AST-press kniga Publ., 2010. Available at: http://royallib.com/book/plungyan_vladimir/pochemu_yaziki_takie_raznie_populyarnaya_lingvistika.html.

5. Piotrovskii R.G., Bektaev K.B., Piotrovskaya A.A. *Mathematical linguistics*. Moscow, Vysshaya Shkola Publ., 1977. 383 p. EDN: ZDHKKT.

6. Lobovikov V.O. Artificial Intelligence and an Almost Unknown Aspect of Mathematical Linguistics. *Discourse-P*, 2022, vol. 19, no. 3, pp. 170–184. EDN: KOOCKW. DOI: 10.17506/18179568_2022_19_3_170.

7. Borovsky A.V., Bratishchenko V.V., Rakovskaya E.E. The Study of the Relationship between the Russian and Buryat Languages Using the Matrix of Proximity Measures between Consonant Word Classes. *System Analysis & Mathematical Modeling*, 2023, vol. 5, no. 1, pp. 19–33. (In Russian). EDN: CEMRYO. DOI: 10.17150/2713-1734.2023.5(1).19-33.

8. Denisov D.V., Zhuravlev M.Yu., Tribunskaya N.A., Medvedeva N.Yu., Zhuravleva T.D., Litvyakova T.A., Martynova A.B. *Aspects of Sanskrit-Slavic etymology on the example of the Russian language: thematic groups*. Samara, Slovo Publ., 2021. 352 p. EDN: HGBLOM.

9. Borovsky A.V., Rakovskaya E.E., Karteeva I.A. Study of the Correlation between Toponyms of the Irkutsk Region and Sanskrit Words. *Izvestiya Baikal'skogo gosudarstvennogo universiteta = Bulletin of Baikal State University*, 2022, vol. 32, no. 3, pp. 610–622. (In Russian). EDN: UFJGAV. DOI: 10.17150/2500-2759.2022.32(3).610-622.

10. Ratcliff J.W., Metzener D.E. Pattern-Matching-the Gestalt Approach. *Dr Dobbs Journal*, 1988, vol. 13, no. 7, pp. 46.

11. Meillet A. *La méthode comparative en linguistique historique*. Honoré Champion, 1925. 132 p. (Russ. ed.: Meillet A. *The Comparative Method in Historical Linguistics*. 2nd ed. Moscow, UPSS Publ., 2004. 101 p.). EDN: QQTRGP.

12. Omelchenko V.V. Methodological Aspects of Complex Analysis of Root System and Image and Sense Content of Words of Various Languages. *Vestnik RUDN. Seriya: Teoriya yazyka. Semiotika. Semantika* = *RUDN Journal of Language Studies, Semiotics and Semantics*, 2015, no. 3, pp. 148–160. (In Russian). EDN: UHITLT.

13. Melkheev M.N. *Geographical names of Eastern Siberia. Irkutsk and Chita regions*. Irkutsk, Vostochno-Sibirskoe knizhnoe izdatel'stvo Publ., 1969. 121 p.

14. Akhadow E. Turkic toponymy of Eastern Siberia (Buryatia, Irkutsk region). *Vkontakte*, 2025, January 3. Available at: <https://vk.com/@dolma60-turkskaya-toponimika-vostochnoi-sibiri-buryatiya-irkutskaya> (In Russian).

15. Gurulev S.A. Geographical names of the Irkutsk region. Irkutsk, 2015. 575 p.

16. Borovskii A.V., Rakovskaya E.E. Application of computational methods in lexicostatistical studies of toponyms of the Irkutsk region. In *Lyapunov readings 2024. Proceedings of the 40th International Conference*, Irkutsk, 02-06 December, 2024. Irkutsk, 2024, pp. 30–32. (In Russian). EDN: GURKJV.

17. Dolgopol'skii A.B. The hypothesis of the most ancient relationship of the language families of Northern Eurasia from a probabilistic point of view. *Voprosy jazykoznanija* = *Topics in the Study of Language*, 1964, no. 2, pp. 64–69. (In Russian).

Информация об авторах

Боровский Андрей Викторович — доктор физико-математических наук, профессор, кафедра математических методов и цифровых технологий, Байкальский государственный университет, г. Иркутск, Российская Федерация, e-mail: andrei-borovskii@mail.ru.

Михалёва Людмила Анатольевна — аспирант, кафедра математических методов и цифровых технологий, Байкальский государственный университет, г. Иркутск, Российская Федерация, e-mail: mihalova.ludmila@mail.ru.

Мосоркин Федот Эдуардович — аспирант, кафедра математических методов и цифровых технологий, Байкальский государственный университет, г. Иркутск, Российская Федерация, e-mail: mosorkin@bk.ru.

Information about the Authors

Andrey V. Borovsky — D.Sc. in Physics and Mathematics, Professor, Department of Mathematical Methods and Digital Technologies, Baikal State University, Irkutsk, Russian Federation, e-mail: andrei-borovskii@mail.ru.

Lyudmila A. Mikhaleva — PhD Student, Department of Mathematical Methods and Digital Technologies, Baikal State University, Irkutsk, Russian Federation, e-mail: mihalova.ludmila@mail.ru.

Fedot E. Mosorkin — PhD Student, Department of Mathematical Methods and Digital Technologies, Baikal State University, Irkutsk, Russian Federation, e-mail: mosorkin@bk.ru.

Вклад авторов

А.В. Боровскому принадлежат идея разработки и постановка задачи;
Л.А. Михалёвой принадлежат подготовка словаря старорусского языка, получение и анализ результатов;

Ф.Э. Мосоркину принадлежит разработка программного обеспечения.

Contribution of the Authors

The authors contributed equally to this article. The authors declare no conflicts of interests.

Для цитирования

Боровский А.В. Метод матрицы парных мер как инструмент сравнения лингвистических множеств / А.В. Боровский, Л.А. Михалёва, Ф.Э. Мосоркин. — DOI 10.17150/2713-1734.2025.7(2).224-247. — EDN AWTVNX // System Analysis & Mathematical Modeling. — 2025. — Т. 7, № 2. — С. 224–247.

For Citation

Borovsky A.V., Mikhaleva L.A., Mosorkin F.E. The Method of the Matrix of Paired Measures as a Tool for Comparing Linguistic Sets. *System Analysis & Mathematical Modeling*, 2025, vol. 7, no. 2, pp. 224–247. (In Russian). EDN: AWTVNX. DOI: 10.17150/2713-1734.2025.7(2).224-247.