

Анна Андреевна БАХТЕРЕВА

кандидат филологических наук, старший научный сотрудник топонимической лаборатории, Уральский федеральный университет (620000, Екатеринбург, пр. Ленина, 51); старший научный сотрудник отдела этимологии и ономастики, Институт русского языка им. В. В. Виноградова РАН (121019, Москва, ул. Волхонка, 18/2)
E-mail: toponimist@yandex.ru
<https://orcid.org/0000-0003-3493-3282>

Михаил Олегович БАХТЕРЕВ

младший научный сотрудник отдела вычислительных систем, Институт математики и механики УрО РАН (620000, Екатеринбург, ул. Софьи Ковалевской, 16)
E-mail: m.bakhterev@yandex.ru
<https://orcid.org/0000-0001-8016-9946>

Методы геопространственной идентификации топонимических данных и результаты их применения

Аннотация

В статье изучаются методы геопространственной идентификации топонимических данных. Рассматривается история сбора, хранения и обработки топонимических данных в Уральском университете. Обсуждается структура существующей топонимической базы данных, а также особенности паспортизации полевого материала и его последующего соотнесения с географическими координатами. Далее сравниваются два метода геопространственной идентификации языковых данных: метод привязки вручную (на примере ГИС «Топонимия Карелии»), успешно применяемый в Институте языка, литературы и истории КарНЦ РАН, и метод автоматической привязки, разработанный для баз данных по территории Белозерья и Костромской области, созданных в Уральском федеральном университете. Преимущества автоматической геопространственной идентификации связаны с возможностью обработки больших объемов данных (десятки тысяч топонимов), что вместе с возможностями QGIS позволяет исследователю быстро получать представление о наличии или отсутствии ареальных закономерностей. Удалось привязать 46 тыс. записей. Метод позволяет реконструировать общую пространственную структуру языковых данных, но его недостатком является невозможность получить точное представление о местоположении самих называемых объектов (что иногда оказывается важным при исследовании топонимических закономерностей). Другая проблема связана с необходимостью ручной подготовки данных для последующей автоматической обработки, прежде всего требуется удаление дублирующих записей. Перспективным представляется использование инструментов Natural Language Processing для автоматической обработки текстов географических привязок, написанных на естественном языке, что может снизить трудозатраты на обработку баз данных и приблизить качество автоматической геопространственной идентификации к качеству ручной привязки. Совершенствование метода

также предполагает создание отдельных сценариев для работы с линейными (прежде всего реками и дорогами) и площадными (крупными болотами и озерами) объектами на карте.

Ключевые слова

топонимия; топонимические базы данных; ареальная лингвистика; картографирование; Белозерье; Костромская область; Свердловская область

Благодарности

Авторская работа А. А. Бахтеревой выполнена за счет гранта Российского научного фонда № 23-18-00439-П «Ономастикон и лингвокультурная история Европейской России», <https://rscf.ru/project/23-18-00439-П/>

Для цитирования

Бахтерева А. А., Бахтерев М. О. Методы геопространственной идентификации топонимических данных и результаты их применения // Вопросы ономастики. 2026. Т. 23, № 2. С. 105–120. https://doi.org/10.15826/vopr_onom.2026.23.2.023

Рукопись поступила в редакцию 10.06.2026

Рукопись принята к печати 31.08.2026

Anna Andreevna BAKHTEREVA

PhD, Senior Researcher, Toponymic Laboratory, Ural Federal University (51, Lenin Ave., 620000 Ekaterinburg, Russia); Senior Researcher, Department of Etymology and Onomastics, V. V. Vinogradov Russian Language Institute of the RAS (18/2, Volkhonka St., 121019 Moscow, Russia)

Email: toponimist@yandex.ru

<https://orcid.org/0000-0003-3493-3282>

Mikhail Olegovich BAKHTEREV

PhD, Junior Researcher, Department of Computer Systems, Institute of Mathematics and Mechanics, Ural Branch of the RAS (16, Sofia Kovalevskaya St., 620000 Ekaterinburg, Russia)

E-mail: m.bakhterev@yandex.ru

<https://orcid.org/0000-0001-8016-9946>

Methods of Geospatial Referencing of Toponymic Data and Results of Their Application

Abstract

The article examines methods of geospatial referencing of toponymic data. The history of collecting, storing, and processing toponymic data at Ural University is reviewed. The structure of the existing toponymic database is discussed, along with the specifics of documenting field materials and subsequently correlating them with geographical coordinates. The article then considers two methods of geospatial referencing of linguistic data: manual referencing, illustrated by the GIS Toponymy of Karelia, which has been successfully implemented at the Institute of Linguistics, Literature and History of the Karelian Research Centre of the Russian Academy of Sciences; and automatic referencing, developed for databases of the toponymy of Belozerye and the Kostroma Region at Ural Federal University. The main advantage of the proposed automatic geospatial referencing method is its ability to process large volumes of data (tens of thousands of toponyms), which can subsequently be visualized using QGIS-based solutions, allowing researchers to quickly assess the presence or absence

of areal patterns. Using this method, 46,000 records were successfully georeferenced. The method makes it possible to reconstruct the spatial distribution of linguistic data, but not the precise locations of individual objects, which may sometimes be important for the study of toponymic patterns. Another limitation is that the method requires manual preliminary cleaning of the collected field data; in particular, the removal of duplicate records is essential. The application of Natural Language Processing tools to the automatic processing of geographical references expressed in natural language may improve the proposed geospatial referencing method. The use of such tools could reduce the labor costs of data preprocessing and narrow the gap between automatic and manual referencing. Further improvements may include the development of additional scenarios for working with linear objects (primarily rivers and roads) and areal objects (large swamps and lakes) on the map.

Keywords

toponymy; toponymic databases; areal linguistic; mapping; Belozerye; Kostroma Region; Sverdlovsk Region

Acknowledgements

The work authored by A. A. Bakhtereva was prepared as a part of the project supported by the Russian Science Foundation (grant number 23-18-00439-II, <https://rscf.ru/en/project/23-18-00439-II/>).

For citation

Bakhtereva, A. A., & Bakhterev, M. O. (2026). Methods of Geospatial Referencing of Toponymic Data and Results of Their Application. *Voprosy onomastiki*, 23(2), 105–120. https://doi.org/10.15826/vopr_onom.2026.23.2.023

Received on 10 June 2026

Accepted on 31 August 2026

Введение

По мере накопления как ономастического, так и апеллятивного диалектного материала исследователи довольно быстро начали осознавать необходимость картографирования собранного массива данных. Такой подход позволял точнее паспортизировать собранный материал, что повышало возможности последующей проверки, уточнения и поиска нужных данных. Кроме этого, стало понятно, что карта сама по себе является очень сильным инструментом: изучение распространения тех или иных языковых явлений позволяет делать выводы об истории заселения, миграционных потоках, языковых контактах, характере взаимодействия этнолокальных групп и пр. Начинает развиваться линвогеография. Уже в 1930-е гг. началась подготовка к созданию крупнейшего бумажного издания такого рода — «Диалектологического атласа русского языка» [ДАРЯ]. В то время, при отсутствии современных возможностей автоматизации и машинного поиска, а также возможностей электронного автоматического картографирования, процесс подготовки атласа был весьма трудоемким. Работа велась отделом диалектологии и лингвогеографии Института русского языка РАН, три выпуска ДАРЯ были изданы в 1986–1996 гг.

(всего 318 карт). Не так давно усилиями сотрудников Института русского языка им. В. В. Виноградова РАН и НИУ ВШЭ был создан электронный ресурс¹ по материалам этого атласа.

Почти в то же время, когда осуществлялся сбор материала для ДАРЯ, происходил и сбор материала для крупного международного проекта — «Общеславянского лингвистического атласа» (ОЛА). С 1965 г. издается сборник «Общеславянский лингвистический атлас. Материалы и исследования» (к 2011 г. появилось 27 выпусков, далее материалы ОЛА издаются в рамках серии «Труды Института русского языка»). База данных русскоязычных материалов ОЛА, включающая около 1 млн словоупотреблений, размещена на специальном сайте². С этого же времени ведется работа над еще более крупным международным проектом — «Atlas Linguarum Europae», с 1975 по 2015 г. издано несколько выпусков, в том числе девять выпусков с картами и комментариями к ним.

Параллельно велись работы по созданию атласов языков народов РФ. Начатые еще в советское время (материалы для большинства проектов собирались еще в 1960–1990-е гг.), они получили новый импульс с развитием компьютерных технологий, существенно ускоривших обработку и систематизацию собранного материала. Можно упомянуть по крайней мере два таких издания: «Диалектологический атлас башкирского языка» [ДАБЯ], «Материалы диалектологического атласа тюркских языков СССР» [МДАТЯ] — оригинальные версии карт, опубликованных в этом издании, размещены на специальном электронном ресурсе³.

Безусловно, эпоха автоматической обработки больших массивов данных и в особенности развитие возможностей электронного картографирования способствовали интенсификации исследований в области ареальной лингвистики, как все чаще стали называть эту область науки. Можно отметить «Лингвистический атлас вепского языка» [ЛАВЯ], включающий около 200 карт и по своим принципам во многом ориентирующийся на своего «старшего брата» — международный проект «Atlas Linguarum Fennicarum» [ALFE]. На основании исследований, предпринятых при подготовке «Лингвистического атласа вепского языка», оказывается возможным реконструировать вепскую историческую территорию, уточнить ее членение и особенности контактов с карельским и русским языками [Зайцева, Муллонен 2018].

¹URL: <https://da.ruslang.ru/> (дата обращения: 28.08.2026).

²URL: <https://slavatlas.org/database.html> (дата обращения: 28.08.2026).

³URL: <https://philology.nsc.ru/resources/datya.php> (дата обращения: 28.08.2026).

Появляются и локальные проекты вроде электронного ресурса «Диалектологический атлас обско-угорских и самодийских языков, распространенных на территории Ханты-Мансийского автономного округа — Югры», размещенного на сайте Обско-угорского института прикладных исследований и разработок⁴, или серии электронных диалектологических атласов, подготовленных сотрудниками Института русского языка им. В. В. Виноградова РАН и НИУ ВШЭ: «Атлас русских говоров Башкирии», «Диалектологический атлас русских народных говоров Волгоградской области», «Русские говоры Среднего Поволжья: Чувашская Республика, Республика Марий Эл»⁵.

Особую значимость картографирование имеет для топонимии — в силу того, что по своей сути она теснее всего связана с географическими реалиями и их расположением на местности. Поскольку топонимический материал является более массовым и при этом для многих территорий он гораздо хуже собран, данная задача долгое время оставалась нерешенной. Однако в последние годы намечается постепенный переход от топонимических баз данных к созданию топонимических атласов [Гордова 2015; Гордова и др. 2022] и топонимических геоинформационных систем.

История обработки топонимических данных в Уральском университете

С 1961 г. (первые одиночные выезды совершались и ранее, в конце 1950-х гг.) в Уральском государственном университете начала свою непрерывную работу Топонимическая экспедиция. За долгие годы экспедицией были полностью обследованы территории Архангельской и Вологодской областей, большая часть Костромской и Свердловской областей, а также частично охвачены и многие другие регионы (Республика Карелия, Мурманская, Ярославская, Челябинская, Омская области и др.). Со временем все громче заявляла о себе проблема хранения собранного массива данных. Сначала материалы хранились только в полевых блокнотах собирателей, но очень скоро стало понятно, что нужна систематизированная форма хранения, которая облегчала бы поиск. Так возникли бумажные картотеки, которые сначала имели относительно свободную форму. Уже в 1980-е гг. предпринимаются попытки составить формализованную картотеку (имея в виду возможную машинную обработку данных в будущем), растет количество вспомогательных картотек: субстратная — где отдельно были вынесены карточки

⁴URL: <https://atlas.oupiir.ru/> (дата обращения: 28.08.2026).

⁵URL: <https://atlases.ruslang.ru/> (дата обращения: 28.08.2026).

с топонимией субстратного происхождения; основная — где топонимы были сгруппированы по субстратным основам (т. е. не учитывались дополнительные определения, приставки, предлоги и т. п., которые могли нарушать алфавит); обратная — где карточки были продублированы таким образом, что каждый топоним был написан в обратной последовательности, что позволяло вести поиск потенциальных формантов; и т. д.

Следующим этапом в истории обработки топонимических данных стало создание прототипа базы данных топонимической картотеки УрГУ в 1998 г. Прототип был реализован на базе СУБД MS Access (с поддержкой запросов на SQL). Был разработан интерфейс, позволяющий осуществлять операции ввода, корректировки и поиска данных [подробнее см.: Николаева 2000]. В своем нынешнем виде база данных Топонимической экспедиции УрГУ (ныне УрФУ) функционирует с 2000 г. Сначала база данных служила в первую очередь для формализации топонимических данных, позднее появились разделы для обработки лексических данных, онимов (антропонимов, катойконимов, этнонимов и пр.), этнографии и данных письменных источников. В настоящее время в базе данных выделяется несколько территориальных секторов. Перечислим их по хронологии появления в электронной форме: Русский Север (Архангельская и Вологодская области), Костромская область, Урал (прежде всего Свердловская область, но также введены данные полевых сборов по Пермской и Челябинской областям), Беломорье (прежде всего данные полевых сборов по Республике Карелия и Мурманской области).

Структура базы данных

Структура базы данных включает в себя в формализованном виде всю информацию, отраженную на бумажной карточке. В современных условиях (начиная с 2019 г.; для полевых материалов Костромской области — с 2011 г.) ввод собранного материала осуществляется из полевых блокнотов напрямую в базу данных, минуя этап бумажной картотеки. Подробно структура базы данных и ее особенности рассматриваются в публикациях [Николаева 2000; Мищенко 2011].

В текущей версии базы данных дополнительно включена географическая привязка данных в виде указания географических координат (подробнее см. далее). Таким образом, в современном виде топонимическая карточка базы данных с ГИС-компонентом имеет следующую структуру: 1) индекс («русское» или «субстратное»); 2) область; 3) район; 4) сельсовет; 5) населенный пункт; 6) широта; 7) долгота; 8) топоним; 9) вид географического объекта; 10) вид географического объекта дополнительный; 11) топографическая

привязка (с указанием расстояния в километрах, берега реки и т. п.); 12) варианты топонима; 13) помета («официальное» / «неофициальное», «старое», «местное» и т. п.); 14) мотив (мотивировочный контекст, позволяющий раскрыть мотив номинации); 15) примечания; 16) дата сбора (год и месяц); 17) ФИО информанта; 18) ФИО собирателя.

Паспортизация данных

Отдельную проблему представляет собой паспортизация данных. Когда собиратели работали исключительно с бумажными карточками, в паспорте карточки указывалось место сбора конкретного топонима — тот населенный пункт, в котором работали участники экспедиции. Постепенно у этого подхода обнаружилось несколько проблемных зон: 1) изменения в административно-территориальном делении, которые накапливаются с течением времени (один и тот же населенный пункт мог менять свое подчинение и в следующий период сбора относиться к другому сельсовету, а иногда и к другому району); 2) указание в паспорте карточки названия населенного пункта, отсутствующего в справочниках административно-территориального деления ввиду следующих причин: а) населенный пункт прекратил существование; б) указано старое название населенного пункта, изменившееся вследствие переименования; в) указано неофициальное название населенного пункта.

По мере увеличения общего стажа Топонимической экспедиции УрФУ (в этом году будет отмечаться 65-летний юбилей с даты первого официального выезда) возникла потребность в сопоставлении разновременных данных, собранных на одной и той же территории: в некоторые районы выезды происходили несколько раз с интервалом в 20, а иногда и в 40 лет. С начала создания базы данных была очевидна и последовательно проводилась ориентация на строгую административную привязку каждого языкового факта: для каждого топонима в паспорте карточки указывались не только область и район, но и название деревни и сельсовета. Однако эта практика имеет один существенный недостаток: административно-территориальное деление на обследованных территориях за прошедшее с момента первых экспедиций время неоднократно менялось, вследствие чего некоторые населенные пункты на следующем витке сборов оказывались в другом сельсовете или даже в соседнем районе, а некоторые просто прекратили свое существование и идентифицировать их местонахождение в настоящий момент не всегда возможно. Кроме того, при наборе в базу данных материалов экспедиций 1960-х гг. выяснилось, что собиратели нередко указывали в паспорте не официальное название деревни, а то, которое сообщил им информант. Идентификация таких

местных названий и соотнесение их с официальными наименованиями также представляет собой дополнительную сложность. Иногда такая идентификация возможна средствами самой картотеки, если эти варианты были тщательно зафиксированы собирателями, но нередко приходится обращаться к справочникам административно-территориального деления прошлых лет, а в ряде случаев и к дореволюционным спискам населенных мест (особенно ярко эта проблема проявилась при занесении в базу данных полевых материалов по топонимии Свердловской области, собранных в 1960-е гг.).

Кроме того, стремление к строгой паспортизации породило еще одну проблему — на какое состояние административно-территориального деления ориентироваться, если оно постоянно меняется? Эта проблема была осознана практически сразу, поскольку уже к моменту создания базы в начале 2000-х гг. накопились значительные расхождения между текущей ситуацией и тем, что фиксировали экспедиции 1960–1980-х гг. Было принято решение в каждом регионе ориентироваться на определенный справочник административно-территориального деления — обычно справочник 1970–1980-х гг. (лучше отражающий экспедиционную реальность того времени и более полный по сравнению с современной ситуацией). Так, при внесении данных по топонимии Свердловской области решено было опираться на справочники 1956 и 1978 гг. Однако в ходе набора карточек выяснилось, что этого недостаточно: исследователям пришлось обращаться к справочнику [СНПУО], содержащему ныне отсутствующие в справочниках и на картах населенные пункты.

Возможен и другой подход. Он реализован в еще одной давно и активно развивающейся топонимической базе данных, создаваемой в Институте языка, литературы и истории Карельского научного центра РАН [см.: Муллонен, Лялля 2005] и известной как ГИС «Топонимия Карелии». В этой базе данных при паспортизации каждого топонима указывается два населенных пункта: по состоянию на 1926 г. (этот момент отражает наиболее разветвленную ойконимическую систему, далее количество населенных пунктов начинает постепенно сокращаться) и современный (по состоянию на 2013 г., когда это решение было окончательно зафиксировано).

Геопространственная идентификация данных

На следующем этапе развития топонимической базы данных возникла задача картографирования топонимических данных и разработки геоинформационной системы, позволяющей исследователям более эффективно применять методы ареальной лингвистики. Для решения этой задачи к работе были

привлечены научные сотрудники Института математики и механики УрО РАН А. С. Игумнов и П. А. Васёв, которыми были предложены различные методы решения такой задачи.

Общую логику решения можно разделить на четыре этапа.

1. Создание алгоритма (разработчики — А. С. Игумнов, П. А. Васёв, Институт математики и механики УрО РАН), позволяющего сопоставить исходную топонимическую базу данных в MS Access с базой данных географических координат населенных пунктов «Яндекса» (RussianOikonymsDataset). Сопоставление по алгоритму считалось надежным при совпадении записи в топонимической базе данных с данными «Яндекса» на 90 % и более.

2. Уточнение географических координат и определение местоположения населенных пунктов, для которых алгоритм не смог установить географические координаты. Записи в базе данных, для которых совпадение было менее 90 %, проверялись вручную. Причины возникновения расхождений между топонимической базой данных и базой данных «Яндекса», которые приводили к недостаточно корректной работе алгоритма, делятся на три группы:

- не найден населенный пункт (решение: поиск по картам «Яндекса», Google, OpenStreetMaps, familio.org);
- несколько «омонимичных» записей в базе «Яндекса» (решение: разграничение средствами исходной топонимической базы данных либо при помощи информации об административно-территориальном делении);
- несколько «омонимичных» записей в исходной топонимической базе данных (решение: поиск по картам «Яндекса», Google, OpenStreetMaps, familio.org).

2.1. Для территории Костромской области, где бóльшая часть экспедиций проходила 2000–2020-е гг. (и продолжается в настоящее время), а самые ранние выезды относятся к 1970-м гг., оказалось возможно уточнение в ручном режиме местоположения каждого населенного пункта, не установленного алгоритмом: либо средствами самой базы данных, либо с помощью современных картографических сервисов и с опорой на знание местности участников современных экспедиций.

2.2. Для территории Белозерья ручная обработка нераспознанных населенных пунктов оказалась практически невозможной и в силу их массовости, и в силу большей давности записей (часть записей относится к 1960-м гг.), поэтому для этой территории на данном этапе реализовано другое решение: все топонимические записи, содержащиеся в паспорте нераспознанные населенные пункты, привязывались к центру сельсовета (к счастью, его во всех случаях удалось установить без каких-либо затруднений).

3. Разработка методов экспорта данных из топонимической базы данных MS Access в форму, пригодную для визуализации.

4. Выявление необходимых параметров лингвогеографического представления топонимии и визуализация полученных данных средствами QGIS (разработка графического интерфейса базы данных и вариантов отображения топонимов на карте).

4.1. Отображение всех топонимов на карте.

4.2. Отображение на карте топонимов, соответствующих заданным критериям поиска.

4.3. Возможность отображения комбинированного поиска: по топониму и по виду объекта.

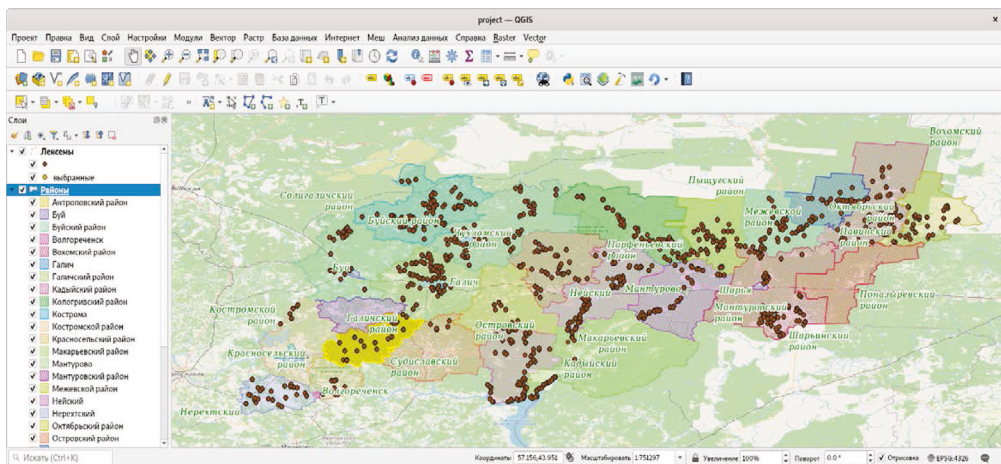
4.4. Возможность создания отчета: сохранение фрагмента карты с полным списком топонимов, условными обозначениями; указанием границ районов.

Что привязывать к карте: результаты и некоторые следствия

1. Привязка места сбора

Выше было предложено решение для задачи привязки к карте населенного пункта, т. е. в нашем случае места сбора. Опишем результаты, которые может дать такой подход, а также его сильные и слабые стороны.

Важное достоинство указанного метода заключается в том, что он позволяет за короткое время привязать к карте значительное количество записей (несколько тысяч), используя географические координаты, соответствующие полю «Населенный пункт» в топологической базе данных. Это решение подходит для картографирования лексических, этнографических и фольклорных данных, когда исследователю важно видеть место фиксации и распространение языковых фактов. Кроме того, такой подход обеспечивает сопоставленность с другими разделами базы данных, т. е. дает возможность сопоставить распространенность какого-то диалектного слова в апеллятивной лексике и, например, в фольклоре; сопоставить распространенность какого-то личного имени в топонимии и в прозвищах; и т. п. На данный момент привязка места сбора реализована в двух разделах базы данных — в базе данных «Диалектная лексика Костромской области» (точная привязка всех населенных пунктов, упоминаемых в базе, — фактически все эти населенные пункты привязаны также для топонимии, антропонимии и этнографии) и в базе данных «Топонимия Белозерья» (точная привязка большинства населенных пунктов, в «темных» случаях — привязка центра сельсовета), при этом есть возможность визуализации границ районов (см. ил. 1), что важно для сопоставления с данными диалектных словарей и некоторых других задач. Всего удалось привязать 46 тыс. записей.

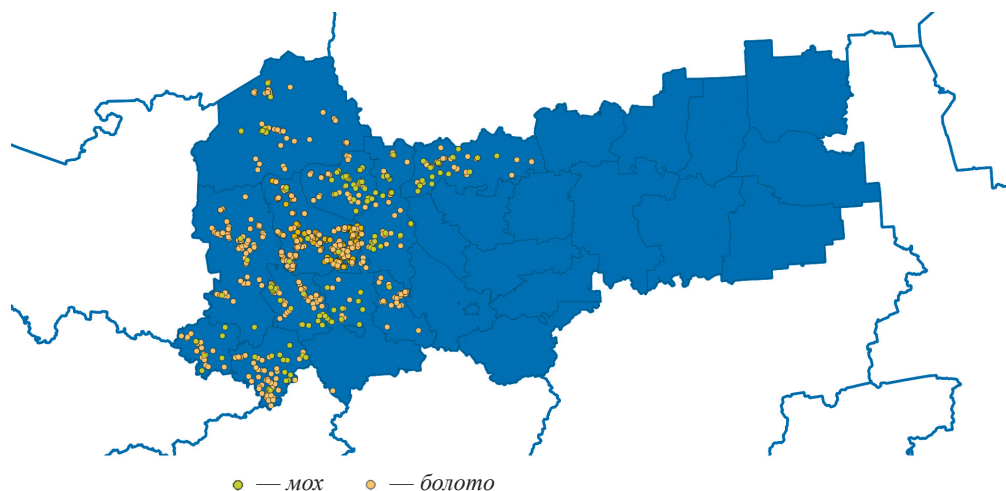


Ил. 1. Визуализация границ районов в базе данных
«Диалектная лексика Костромской области».

Карта отражает все лексемы,
собранные на территории Костромской области

Недостаток метода заключается в том, что он не отражает реальное географическое положение объектов — он показывает только те населенные пункты, где этот топоним известен. Кроме того, при автоматической обработке больших объемов данных оказывается алгоритмически невозможно различить две ситуации: 1) когда на одной территории есть несколько одноименных объектов, что периодически встречается, особенно для топонимов, образованных от географических терминов либо от распространенных антропонимов; 2) когда имеет место дублирование записей (т. е. один и тот же топоним многократно записан в разных населенных пунктах — это особенно характерно для названий относительно крупных рек и крупных населенных пунктов). Эта проблема может быть снята, если предусмотреть процедуру очистки исходной базы данных от дублирующих записей и подготовку промежуточной базы данных, которая будет связываться с картой. Такая очистка данных может быть произведена вручную, но для больших объемов данных это потребует значительных трудозатрат. Представляется возможной и разработка специального алгоритма, который по некоторым аккуратно подобранным критериям сможет автоматически удалить часть дублирующих записей.

Тем не менее для больших массивов данных результаты картографирования нередко оказываются достаточно информативными даже в условиях неочищенных (полевых) данных — безотносительно реального количества объектов (см. ил. 2).



Ил. 2. Ареалы топонимических формантов *-мох* и *-болото* на территории Белозерья

Для топонимии рассмотренный метод геопространственной идентификации не является единственным — теоретически возможно привязать к карте не место фиксации, а сам топоним.

2. Привязка топонима

Этот метод геопространственной идентификации топонимических данных обладает двумя основными преимуществами: он достоверно показывает расположение на местности объектов, обозначенных соответствующими топонимами, и одновременно снимает проблему дублирования записей. Очевидным недостатком этого метода является его ресурсоемкость. Для того чтобы привязать к карте сам топоним, потребуется либо ручная обработка (т. е. привязка каждого топонима вручную, что требует больших трудозатрат в условиях неразмеченных данных и нередко невозможно при работе с письменными источниками — лоциями, актами монастырей, списками топонимов и пр.), либо применение ресурсоемких инструментов обработки естественного языка (Natural Language Processing), таких как большие языковые модели (LLM), для разбора и формализации привязок, написанных в материалах Топонимической экспедиции на естественном языке, часто в довольно свободной форме. При этом оба этих механизма никак не решают проблему корректной привязки линейных объектов (прежде всего рек и отчасти дорог), которые в силу своей протяженности фактически должны быть привязаны к нескольким точкам, и значительных площадных объектов (крупных болот и в некоторых случаях

озер). Кроме того, при данном подходе осложняется возможность адекватного соотнесения разных разделов базы данных — топонимии и лексики, топонимии и антропонимии, и пр. Текущая структура базы данных не позволяет сделать одновременно привязки по месту сбора и по самому топониму (по местоположению называемого топонимом объекта), однако такая возможность может быть реализована в будущем.

Тем не менее существуют некоторые примеры удачной реализации такого подхода к привязке топонима, когда он оказывается единственно возможным в силу стоящих перед исследователями задач.

Хорошим примером реализации подхода привязки к карте самих топонимов является ГИС «Топонимия Карелии», где каждый топоним привязан к квадрату топографической карты (для линейных объектов — к нескольким квадратам). Это оказывается возможным в первую очередь потому, что картографирование топонимов происходит еще в полевых условиях, в ходе работы с информантами. Аналогичный подход реализован в родственном онлайн-проекте «ТопКар — Топонимия Карелии»⁶ [подробнее о проекте см.: Захарова 2022], включающем свыше 60 тыс. названий.

Привязка к карте самих топонимов также оказывается единственно возможным методом при исследовании распространения конкретных топонимических основ, когда речь идет об относительно небольших объемах данных, которые реально перепроверить вручную. Показательные примеры таких карт представлены в трудах Уральской ономастической школы, прежде всего ее основателя А. К. Матвеева [2001–2015; и др.], в исследованиях Петрозаводской ономастической школы [см., например: Муллонен, 2012; 2025; Захарова и др. 2018; и мн. др.], в работе, посвященной этнолокальным группам в Белозерье по данным топонимии, коллективных прозвищ и топонимических легенд [Макарова 2017], а также в статье [Березович и др. 2015], где рассматриваются различные варианты реализации прибалтийско-финской основы *hiisi*- в топонимии Русского Севера.

Заключение

Сделанные в статье наблюдения показывают, что оба рассмотренных метода геопространственной идентификации топонимии (привязка места фиксации и привязка объекта, поименованного топонимом) важны в исследовательской практике. Привязка места сбора позволяет быстро картографировать большие массивы данных и соотносить друг с другом разные типы

⁶URL: <http://topkar.krc.karelia.ru/ru> (дата обращения: 28.08.2026).

языковых фактов (топонимию, антропонию, лексику и т. п.), что важно для комплексных междисциплинарных исследований. Привязка самих топонимов (объектов, поименованных топонимом) нередко имеет принципиальное значение для микрорегиональных исследований, для изучения локальных языковых процессов и при ограниченном наборе данных. При этом она предполагает большее количество ручной обработки, а также наличие более чистых исходных данных, что не всегда возможно, особенно при работе с материалами письменных источников.

Представляется, что перспективы языковых ГИС Уральского университета выглядят следующим образом. На первом этапе необходимо произвести автоматическую привязку всех собранных полевых данных (ономастических, лексических, этнографических) по месту фиксации. На следующем этапе для топонимических материалов следует осуществить очистку данных (удаление дублирующих записей, снятие «омонимии» для протяженных и/или крупных географических объектов). Далее с применением инструментов больших языковых моделей (LLM) нужно реализовать процедуру автоматического разбора и формализации привязок и последующей автоматической пространственной идентификации топонимов по местоположению конкретного объекта в тех случаях, где это в принципе возможно. Таким образом, вероятно, решением может стать создание базы данных одновременно с двумя поисковыми «входами» для привязок — по месту фиксации и по самому топониму.

Источники

- Гордова Ю. Ю. Топонимический атлас Рязанской области. М. : Наука, 2015.
- Гордова Ю. Ю., Мельник В. И., Мудрак О. А., Коряков Ю. Б., Девяткина Е. М. Топонимический атлас Тамбовской области. М. : Ленанд, 2022.
- ДАРЯ — Диалектологический атлас русского языка. Центр Европейской части СССР. Вып. 1 : Фонетика / под ред. Р. И. Аванесова, С. В. Бромлей. М. : Наука, 1986 ; Вып. 2 : Морфология / под ред. С. В. Бромлей. М. : Наука, 1989 ; Вып. 3 : Синтаксис. Лексика / под ред. О. Н. Мораховской. М. : Наука, 1996.
- ДАБЯ — Диалектологический атлас башкирского языка / гл. ред. Ф. Г. Хисамитдинова. Уфа : Гилем, 2005.
- ЛАВЯ — Зайцева Н. Г., Муллонен И. И., Мызников С. А., Жукова О. Ю., Бродский И. В. Лингвистический атлас вепсского языка / под общ. ред. Н. Г. Зайцевой. СПб. : Нестор-История, 2019.
- МДАТЯ — Материалы диалектологического атласа тюркских языков СССР / сост. Е. И. Убрятова, Н. Н. Широкова ; отв. ред. Н. Б. Кошкарева, Н. Н. Федина, А. В. Байыр-оол. Новосибирск : Академиздат, 2025.
- СНПУО — Список населенных пунктов Уральской области : в 16 т. / под ред. И. Н. Гридина и др. Свердловск : Орготд. Уралоблисполкома, Уралстатупр-ния и окружн. исполкомов, 1928.

ALFE — Atlas Linguarum Fennicarum. O. 1–3 / päätoim. T. Tuomi. Helsinki : Suomalaisen Kirjallisuuden Seura : Kotimaisten kielten tutkimuskeskus, 2004–2010.

Исследования

- Березович Е. Л., Макарова А. А., Муллонен И. И. «Демонологические» названия в русской и субстратной топонимии Русского Севера и Карелии (на материале основ *čert-* и *hiisi-*) // *Linguistica*. 2015. Vol. 55, No. 1 : Slavic / Non-Slavic Language Contact in the Area of Diachronic Onomastics. S. 187–205. <https://doi.org/10.4312/linguistica.55.1.187-205>
- Зайцева Н. Г., Муллонен И. И. Формирование диалектной карты вепсского языка (на материале «Лингвистического атласа вепсского языка») // *Вопросы языкознания*. 2018. № 6. С. 85–103. <https://doi.org/10.31857/S0373658X0002021-2>
- Захарова Е. В. Разработка геоинформационного ресурса по топонимии Карелии: историко-культурный аспект // *Финно-угорский мир*. 2022. Т. 14, № 4. С. 463–473. <https://doi.org/10.15507/2076-2577.014.2022.04.463-473>
- Захарова Е. В., Кузьмин Д. В., Муллонен И. И., Шибанова Н. Л. Топонимные модели Карелии в пространственно-временном контексте. М. : Изд. дом ЯСК, 2018.
- Макарова А. А. К вопросу о картографировании этнолокальных групп Белозерья по ономастическим данным // *Известия Уральского федерального университета. Сер. 2 : Гуманитарные науки*. 2017. Т. 19, № 4 (169). С. 63–83. <https://doi.org/10.15826/izv2.2017.19.4.064>
- Матвеев А. К. Субстратная топонимия Русского Севера : в 4 ч. Екатеринбург : Изд-во Урал. ун-та, 2001–2015.
- Мищенко О. В. Электронная база данных «Топонимия Костромской области»: обработка и систематизация полевого материала // *Вопросы ономастики*. 2011. № 2 (11). С. 89–104.
- Муллонен И. И. Природные и культурные факторы формирования вепсской этнической территории // *Труды КарНЦ РАН. Сер. Гуманитарные исследования*. 2012. Вып. 3, № 4. С. 13–24.
- Муллонен И. И. Картографирование как инструмент исследования локальной топонимической системы // *Финно-угорский мир*. 2025. Т. 17, № 4. С. 430–441. <https://doi.org/10.15507/2076-2577.017.2025.04.430-441>
- Муллонен И. И., Ляля Е. В. Геоинформационная аналитическая система «Топонимия Заонежья» // *Вопросы ономастики*. 2005. № 2. С. 86–96.
- Николаева Е. С. Электронная картотека топонимии Русского Севера: проблемы и перспективы // *Финно-угорское наследие в русском языке / под ред А. К. Матвеева. Вып. 1. Екатеринбург : Изд-во Урал. ун-та, 2000. С. 143–154.*

References

- Berezovich, E. L., Makarova, A. A., & Mullonen, I. I. (2015). “Demonologicheskie” nazvaniia v russkoi i substratnoi toponimii Russkogo Severa i Karelii (na materiale osnov *čert-* i *hiisi-*) [“Demonological” Root *čert-* and *hiisi-* in Russian Toponymy of the Russian North and the Republic of Karelia]. *Linguistica*, 55(1), 187–205. <https://doi.org/10.4312/linguistica.55.1.187-205>
- Makarova, A. A. (2017). K voprosu o kartografirovanii etnolokalnykh grupp Belozeria po onomasticheskim dannym [On the Mapping of Local Ethnic Groups in Belozerye with Reference to Onomastic Data]. *Izvestiya Uralskogo federalnogo universiteta. Ser. 2: Gumanitarnye nauki*, 19(4), 63–83. <https://doi.org/10.15826/izv2.2017.19.4.064>
- Matveyev, A. K. (2001–2015). *Substratnaia toponimiia Russkogo Severa* [The Substrate Toponymy of the Russian North] (Pts. 1–4). Ekaterinburg: Ural University Press.

- Mishchenko, O. V. (2011). Elektronnaia baza dannykh “Toponimiiia Kostromskoi oblasti”: obrabotka i sistematizatsiia polevogo materiala [Kostroma Region Toponymy, an Electronic Database: Field Material Processing and Systematization]. *Voprosy onomastiki*, 2(11), 89–104.
- Mullonen, I. I. (2012). Prirodnye i kul’turnye faktory formirovaniia vepsskoi etnicheskoi territorii [Natural and Cultural Factors behind the Formation of the Vepsian Ethnic Area]. *Trudy KarNTs RAN. Seriiia Gumanitarnye issledovaniia*, Iss. 3, No. 4, 13–24.
- Mullonen, I. I. (2025). Kartografirovanie kak instrument issledovaniia lokalnoi toponimicheskoi sistemy [Mapping as a Tool for Investigating a Local Toponymic System]. *Finno-Ugric World*, 17(4), 430–441. <https://doi.org/10.15507/2076-2577.017.2025.04.430-441>
- Mullonen, I. I., & Lyallya, E. V. (2005). Geoinformatsionnaia analiticheskaiia sistema “Toponimiiia Zaonezhia” [Geographical Information and Analytical System “Toponymy of Zaonezhje”]. *Voprosy onomastiki*, 2, 86–96.
- Nikolaeva, E. S. (2000). Elektronnaia kartoteka toponimii Russkogo Severa: problemy i perspektivy [An Electronic Card Index of the Toponymy of the Russian North: Problems and Prospects]. In A. K. Matveyev (Ed.), *Finno-ugorskoe nasledie v russkom iazyke* [Finno-Ugric Heritage in the Russian Language] (Iss. 1, pp. 143–154). Ekaterinburg: Ural University Press.
- Zaitseva, N. G., & Mullonen, I. I. (2018). Formirovanie dialektnoi karty vepsskogo iazyka (na materiale “Lingvisticheskogo atlasa vepsskogo iazyka”) [Formation of a Dialect Map of the Veps Language (Based on the “Linguistic Atlas of the Veps Language”)]. *Voprosy jazykoznaniia*, 6, 85–103. <https://doi.org/10.31857/S0373658X0002021-2>
- Zakharova, E. V. (2022). Razrabotka geoinformatsionnogo resursa po toponimii Karelii: istoriko-kulturnyi aspekt [Development of a Geographic Information System for Toponyms of Karelia: Historical and Cultural Aspect]. *Finno-Ugric World*, 14(4), 463–473. <https://doi.org/10.15507/2076-2577.014.2022.04.463-473>
- Zakharova, E. V., Kuzmin, D. V., Mullonen, I. I., & Shibanova, N. L. (2018). *Toponimnye modeli Karelii v prostranstvenno-vremennom kontekste* [Toponymic Models of Karelia in the Spatio-Temporal Context]. Moscow: Izd. dom YaSK.