

Александр Владиславович Дмитриев,  
к. филол. н.  
Санкт-Петербургский политехнический универ-  
ситет Петра Великого, Санкт-Петербург, Россия  
э-почта: avd84@list.ru

Alexander Vladislavovich Dmitriev,  
Cand. of Sc. (Philology)  
Peter the Great St. Petersburg Polytechnic  
University, St. Petersburg, Russia  
e-mail: avd84@list.ru

## **ОПЫТ ОНОМАСТИЧЕСКИХ РЕКОНСТРУКЦИЙ МЕТОДАМИ КОМПЬЮТЕРНОЙ, МАТЕМАТИЧЕСКОЙ И НЕЙРОСЕТЕВОЙ ОБРАБОТКИ В СИНХРОННО-ДИАХРОННОМ АСПЕКТЕ**

В статье рассматривается применение современных компьютерных и математических методов для систематизации региональной топонимии и проведения ономастических реконструкций. Представлен анализ основных направлений цифровой ономастики: топонимических баз данных, ГИС-технологий, ономотометрии и нейросетевых моделей для автоматического распознавания топонимов.

*Ключевые слова:* цифровая ономастика; топонимические базы данных; ГИС-технологии, ономотометрия; нейросетевые модели; BERT-ономастика; ономастические реконструкции.

## **EXPERIENCE IN ONOMASTIC RECONSTRUCTIONS USING COMPUTER, MATHEMATICAL AND NEURAL NETWORK PROCESSING IN SYNCHRONIC-DIACHRONIC ASPECT**

The article examines the application of modern computer and mathematical methods for systematizing regional toponymy and conducting onomastic reconstructions. The analysis of the main directions of digital onomastics is presented: toponymic databases, GIS technologies, onomatometry, and neural network models for automatic toponym recognition.

*Key words:* digital onomastics; toponymic databases; GIS technologies; onomatometry; neural network models; BERT-onomastics; onomastic reconstructions.

Развитие цифровых технологий за последние десятилетия коренным образом трансформировало методологию ономастических исследований как в международном, так и в российском научном пространстве. Цифровые методы открыли новые горизонты для систематизации, визуализации и анализа топонимических данных, что привело к появлению целого направления – цифровой ономастики.

В европейском научном сообществе значительный вклад внесли исследования П. Видесотта по ономотометрии (Videsott P. Onomatometrie: Numerische Taxonomie als Methode der Namenforschung. Tübingen: Max Niemeyer Verlag, 2004. 328 p.), а также ономастические проекты Польши, Германии, Великобритании, Испании, Франции, Китая, Индии, представленные на сайте блогера-ономаста Е. Шохенмайера (URL: <https://e-onomastics.blogspot.com/>).

В России также сформировались влиятельные региональные топонимические школы: Петербургская (А. С. Герд, А. В. Дмитриев), Московская (А. Л. Шилов), Вологодская (Ю. И. Чайкина), Карельская (И. И. Муллонен, Е. В. Захарова,

Н. Н. Мамонтова), Екатеринбургская / Уральская (А. К. Матвеев, Е. Л. Березович), Удмуртская (М. Г. Атаманов), Марийская (А. Н. Куклин), Мордовская (И. К. Инжеватов), Коми (А. Г. Мусанов) и др.

Синтез международного опыта и отечественных разработок создает плодотворную почву для инновационных подходов к ономастическим реконструкциям – довольно молодому направлению в рамках цифровой ономастики. Это направление зародилось на стыке исторической географии и картографии, прикладной ономастики, компьютерных методов обработки языковых данных, компьютерного зрения и автоматического распознавания топонимов, а также ГИС-технологий.

Развитие цифровой ономастики происходит в нескольких магистральных направлениях: информационные сайты и каталоги; топонимические базы данных и корпуса; ГИС-решения для систематизированного описания ономастикона; количественная ономастика (ономатометрия); нейросетевые решения (BERT-ономастика).

Среди крупнейших международных баз данных следует выделить: 1) GeoNames – база из около 25 млн географических названий, включая 2,8 млн населенных пунктов с координатами, данными о высотах, населении и административном делении (URL: <http://www.geonames.org/>); 2) PELAGIOS – специализированная база античных и средневековых топонимов (около 2 млн записей), связывающая исторические места с современными координатами (URL: <https://pelagios.org/>); 3) UNGEGN World Geographical Names Database – официальная база ООН, содержащая около 200 000 стандартизированных названий стран, городов и регионов на разных языках (URL: <https://unstats.un.org/unsd/geoinfo/geonames/>); 4) Getty Thesaurus of Geographic Names (TGN) – около 2,5 млн записей исторических и современных мест с акцентом на искусство и культуру (URL: <https://www.getty.edu/research/tools/vocabularies/tgn/>).

В России крупнейшими топонимическими картотеками располагают Уральский федеральный университет, кафедра русского языка и общего языкознания (Екатеринбург) – более 2 млн единиц; Институт языка, литературы и истории КарНЦ РАН (Петрозаводск) – более 300 000 единиц; Топонимическая картотека Якутии М.С. Иванова.

Первым автоматизированным ресурсом по топонимике Европейского Севера России стал TORIS (1997–2004 гг.) (URL: <http://toris.krc.karelia.ru/papers/source/spb2004/index.ru.phtml>).

Однако в нем отсутствовала возможность просмотра топонимов на карте и послойной навигации во времени.

Среди отечественных региональных проектов можно отметить следующие: 1) топонимическая база данных Костромской области (Мищенко О. В. Электронная база данных «Топонимия Костромской области»: обработка и систематизация полевого материала // Вопросы ономастики. 2011. №2(11). С. 89–104.); 2) обучающая база данных «Московская топонимика» (Юе Ц., Васильева Н. В. Параметризация топонимического материала как основа построения обучающей базы данных «Московская топонимика» // Вестник Пермского национального исследовательского политехнического университета. Проблемы языкознания и педагогики. 2019. № 4. С. 20–29); 3) топонимическая база данных на основе армянских и грузинских текстов (ВИЭ, 2020–2021) (URL: [https://hist.hse.ru/southcaucasus/project2020\\_21](https://hist.hse.ru/southcaucasus/project2020_21)); 4) база данных русскоязычных топонимов стран Балтии (Жураковская Н. В., Тесленок С. А. Создание базы данных русскоязычных топонимов стран Балтии // Огарев-Online. 2021. №5(158)); 5) база данных ойконимов Республики Саха (Якутия) (Старости-на А. С. и др. База данных ойконимов Республики Саха (Якутия): варианты написания населенных пунктов: свидетельство о регистрации базы данных, 2022).

В последние годы геоинформационные системы (ГИС) стали играть все большую роль в проведении топонимических исследований. Создание и реализация онлайн-ГИС является одной из тенденций развития исторической информатики. В Европе это такие масштабные картографические проекты, как "Endings of places in Poland", "Suffixes of German towns", "Mapping the toponymy of Great Britain", "Surname Regions in Asturian", а также исследования распространения фамилий в различных странах: "Kleiner deutscher Familiennamenatlas", "Frequency of family names with meaning smith in Austria", "Chinese surname distribution" и "Swedish Regional Distribution of Names". Все эти проекты представлены на сайте Е. Шохенмайера.

В России лидером в области разработки современных топонимических ресурсов на основе ГИС-технологий выступает Карельский научный центр (г. Петрозаводск). За последние 20 лет ими были разработаны локальная ИАС «Топонимия Карелии» (2003–2010 гг.); ГИС «Электронная картотека топонимов Восточного Обонежья»; открытая ГИС по топонимии Карелии на примере Арктической зоны республики (2021 г.). Все проекты были поддержаны грантами РГНФ, РФФИ (позже – РНФ).

Особую ценность представляет ГИС «Топонимия Карелии» – TopKar (Топонимия Карелии (URL: <http://topkar.krc.karelia.ru/ru>), разработка которой реализуется группой исследователей Карельского научного центра РАН в рамках гранта РНФ № 22-28-00362. Этот ресурс позволяет сохранять, систематизировать и исследовать названия географических объектов Карелии и сопредельных областей на карельском, финском и русском языках.

Среди других отечественных региональных ГИС-проектов можно отметить: 1) ГИС населенных пунктов казахстанского Приуралья (Учеваткина Н. В., Ивлиева Н. Г. О разработке топонимической базы данных ГИС населённых пунктов казахстанского Приуралья // XLIV Огаревские чтения: материалы науч. конф. (Саранск, 08–15 декабря 2015 г.). Саранск: Изд-во Мордов. ун-та, 2016. С. 481–487); 2) база топонимических данных Республики Башкортостан (Абсалямов Ю. М., Сиразитдинов З. А. Создание базы топонимических данных Республики Башкортостан для географических информационных систем // Tiltanym. 2022. № 185(1). С. 123–140); 3) электронные топонимические карты территории Мордовии (URL: <https://journal.mrsu.ru/arts/sozdanie-elektronnyx-toponimicheskix-kart-territorii-mordovii>); 4) историческая ГИС «Источники по исторической географии Бежецкого Верха» (Фролов А. А., Голубинский А. А. Историческая ГИС «Источники по исторической географии Бежецкого Верха» // Информационный бюллетень ассоциации История и компьютер. 2014. №42. С. 146–148); 5) исторический атлас Бежецкой пятины (современные Новгородская и Тверская области) (Степанова Ю. В., Гаврилов П. В., Кутаков С. С. Владычная волость Удомля в Бежецкой пятине Новгородской земли в конце XV–первой половине XVI в.: историко-географическая реконструкция в ГИС // Genesis: исторические исследования. 2021. №12. С. 332–345).

Примером комплексного подхода к топонимической реконструкции является проект «Топонимическая база данных Западной Ингерманландии с компонентом картографической визуализации», выполняемый на базе Лаборатории цифровой лингвистики Высшей школы лингвистики и педагогики СПбПУ Петра Великого под руководством к. т. н., доц. М. В. Болсуновской и к. ф. н., доц. А. В. Дмитриева. Новизна исследования заключается в применении современных методик и инструментальных средств геоинформационного моделирования, привлечении богатого архивного и научно-теоретического материала (русские, финские, эстонские и шведские источники), а также в программной возможности послойной навигации и передвижения по интерактивной карте во времени.

Отдельное направление исследований представляет ономотометрия, или количественная ономастика. Впервые оно было системно представлено в работе П. Видесотта. Ономотометрия определяется как метод количественного анализа больших массивов данных об именах (топонимах, антропонимах и т. д.) с целью выявления скрытых структур и закономерностей в их пространственном распределении. Метод позволяет: выявлять общие закономерности в распределении имен на определенной территории; находить области со схожей структурой имен («изонимические структуры»); проводить количественный анализ большого объема ономастических данных.

Одним из перспективных направлений современной цифровой ономастики является нейроономастика, занимающаяся разработкой ономастических нейромоделей. Здесь, пожалуй, можно отметить такие системы автоматического распознавания и разрешения топонимических референций, как TopoBERT (Zhou B., Zou L., Hu Y., Qiang Y., Goldberg D. TopoBERT: a plug and play toponym recognition module harnessing fine-tuned BERT // *International Journal of Digital Earth*. 2023. Vol. 16, №1. P. 3045–3064), NeuroTPR (Wang J., Hu Y., Joseph K. NeuroTPR: A Neuro-Net Toponym Recognition model for extracting locations from social media messages // *Transactions in GIS*. 2020. Vol. 24, №3. P. 719–735.) и DM-NLP (Ma C., Zheng H., Xie P., Li Ch., Li L., Si L. DM-NLP at semeval-2018 task 8: neural sequence labeling with linguistic features // *Proceedings of the 12th International Workshop on Semantic Evaluation*. New Orleans: Association for Computational Linguistics, 2018. P. 707–711).

Современные нейросетевые разработки предлагают специализированные решения для различных аспектов топонимического анализа: TopoBERT ( $F1=0.86-0.88$ ) эффективно обрабатывает стандартные тексты через архитектуру BERT с механизмом внимания, но требует значительных вычислительных ресурсов; NeuroTPR специализируется на неформальных текстах социальных медиа, включая обработку хэштегов и @mentions через многоуровневую архитектуру с BiLSTM; DM-NLP демонстрирует превосходные результаты при анализе научных и структурированных текстов ( $F1=0.89-0.91$ ), интегрируя синтаксический анализ с машинным обучением; дополняет эту экосистему HistBERT, ориентированный на диахронический анализ исторических топонимов, что в совокупности формирует мощный инструментарий для ономастических реконструкций.

Таким образом, развитие цифровых методов в ономастических исследованиях открывает новые возможности для систематизации, сохранения и анализа топонимического наследия. Интеграция традиционных лингвистических подходов с современными компьютерными, математическими и нейросетевыми методами позволяет выявлять ранее скрытые закономерности в топонимическом материале и реконструировать историко-культурный ландшафт в синхронно-диахронном аспекте.