

DOI: 10.15514/ISPRAS-2025-37(6)-27



Язык памятников прибалтийско-финской письменности XVII-XIX вв.: комплексный анализ на базе лингвистической платформы LingvoDoc (введение в проект)

С.В. Нагурная, ORCID: 0000-0002-6233-8045 <kov@krc.karelia.ru>

*Институт языка, литературы и истории Карельского научного центра РАН,
Россия, 185910, Республика Карелия, г. Петрозаводск, ул. Пушкинская, д. 11,*

Аннотация. В статье дана характеристика проекта, реализация которого начата в текущем году в Институте языка, литературы и истории Карельского научного центра РАН, – «Язык памятников прибалтийско-финской письменности XVII-XIX вв.: комплексный анализ на базе лингвистической платформы LingvoDoc». Платформа LingvoDoc – цифровое хранилище, предназначенное для резервирования языковых данных, инструменты которого предоставляют возможность одновременно осуществлять обработку языкового материала и проводить в онлайн-режиме анализ фонетических, морфологических, лексических особенностей языка. Размещение текстов памятников карельской и вепсской письменностей на платформе LingvoDoc позволит не только решать исследовательские задачи (текстологический анализ, выявление диалектной специфики, создание конкордансов и т.д.), но и выйти на решение вопросов языковой документации. Метод обработки больших данных обеспечит релевантность результатов.

Ключевые слова: карельский язык; вепсский язык; платформа LingvoDoc; памятники письменности; анализ текста; документация языков.

Для цитирования: Нагурная С.В. Язык памятников прибалтийско-финской письменности XVII-XIX вв.: комплексный анализ на базе лингвистической платформы LingvoDoc (введение в проект). Труды ИСП РАН, том 37, вып. 6, часть 2, 2025 г., стр. 169–176. DOI: 10.15514/ISPRAS-2025-37(6)-27.

Благодарности: Статья подготовлена за счет гранта Российского научного фонда No 25-78-20006. Результаты получены с использованием услуг Центра коллективного пользования Института системного программирования им. В.П. Иванникова РАН – ЦКП ИСП РАН.

The Language of Baltic-Finnic Literary Monuments of the 17th–19th Centuries: A Comprehensive Analysis Based on the LingvoDoc Linguistic Platform (Introduction to the Project)

S.V. Nagurnaja, ORCID: 0000-0002-6233-8045 <kov@krc.karelia.ru>

*Institute of Linguistics, Literature and History of KarRC RAS,
11, Pushkinskaya st., Petrozavodsk, 185910, Russia.*

Abstract. The article describes a project, the implementation of which began this year at the Institute of Language, Literature and History of the Karelian Scientific Centre of the Russian Academy of Sciences: “The Language of the Monuments of the Baltic-Finnic Literature of the 17th-19th Centuries: A Comprehensive Analysis Based on the LingvoDoc Linguistic Platform.” The LingvoDoc platform is a digital repository designed to back up language data. Its tools allow for the simultaneous processing of language material and the online analysis of phonetic, morphological, and lexical features of the language. Placing texts from Karelian and Vepsian scripts on the LingvoDoc platform will not only enable research tasks (textual analysis, identifying dialectal specifics, creating concordances, etc.) but also address issues of language documentation. Big data processing will ensure the relevance of the results.

Keywords: Karelian language; Vepsian language; LingvoDoc platform; written monuments; text analysis; language documentation.

For citation: Nagurnaya S.V. Language of Baltic-Finnic written monuments of the 17th-19th centuries: a comprehensive analysis based on the LingvoDoc linguistic platform (introduction to the project). *Trudy ISP RAN/Proc. ISP RAS*, vol. 37, issue 6, part 2, 2025, pp. 169-176 (in Russian). DOI: 10.15514/ISPRAS-2025-37(6)-27.

Acknowledgements. The reported study was granted by Russian Science Foundation, project no. 25-78-20006. The results were obtained using the services of the Ivannikov Institute for System Programming (ISP RAS) Data Center.

1. Введение

В последние годы значительно возрос интерес исследователей к вопросам истории языка. Интерпретация лингвистического материала памятников позволяет реконструировать исторические варианты языков. Кроме того, для ряда языков и диалектов в ходе исследований уточняется «...релевантность диалектно-дифференцирующих изоглосс, на основе которых строятся их современные классификации» [1]. Не являются исключением и прибалтийско-финские языки Карелии, карельский и вепсский, письменные стандарты которых активно развиваются с начала 1990-х гг., но формирование их письменных традиций уходит в глубь веков.

Письменное наследие карельского и вепсского языков не отличается многообразием, поэтому каждый из имеющихся памятников письменности – это уникальный ресурс для изучения истории языков и народов. Памятники письменности – важнейшие документы не только исторической фонетики, грамматики, лексикологии и диалектологии языков, но и этногенеза говорящих на нем этнических групп. Каждый из них, независимо от жанра, является незаменимой деталью, запечатлевшей язык в определенный момент его развития.

В текущем году Российским научным фондом поддержан реализуемый в Институте языка, литературы и истории Карельского научного центра РАН проект «Язык памятников прибалтийско-финской письменности XVII-XIX вв.: комплексный анализ на базе лингвистической платформы LingvoDoc». В соответствии с конкурсной документацией, исследование осуществляется на базе существующей научной инфраструктуры мирового уровня, которую представляет собой система LingvoDoc [2], предназначенная «для составления, анализа и хранения словарей, корпусов и конкордансов языков и диалектов».

В числе целей предпринимаемого исследования:

- 1) текстологическое изучение материалов известных на сегодняшний день старокарельских и старовепских рукописных и печатных памятников письменности, уточнение связей между ними;
- 2) целостное описание графо-фонетических, морфологических, лексических систем памятников;
- 3) определение (уточнение) диалектной специфики памятников;
- 4) введение в научный оборот нового лингвистического материала, что позволит представить историю карельского и вепского языков как непрерывный процесс.

В соответствии с перечисленными целями одной из важнейших задач при реализации проекта представляется целенаправленная работа по выявлению, оцифровке, расшифровке и созданию на базе платформы LingvoDoc коллекции текстов разножанровых памятников карельской и вепской письменностей. Инструментарий лингвистической платформы позволяет осуществлять расширенный фонологический и морфологический анализ текстов памятников, а также произвести диалектную локализацию языка ранних письменных источников. В результате выполнения указанных задач одновременно будет решаться одна из основных проблем современного филологического знания – цифровое сохранение миноритарных языков.

2. Обзор проекта

2.1 Предметная область исследования

Официально карельская и вепская письменности получили развитие в 1930-е гг., однако ранние рукописные памятники фиксировались значительно раньше [3-5]. В силу своей уникальности и единичности каждый из них представляет собой особую ценность для исследователей.

К самым ранним прибалтийско-финским записям относятся берестяные грамоты, обнаруженные в Новгороде – грамота № 292 с записью заклинания от молнии (1240-1260 гг.), грамота № 403, содержащая русско-карельский словарь (1360-1380 гг.) и др. Используемый в берестяных грамотах язык мог быть отражением некоего северо-восточного прибалтийско-финского койне – языковой формы, объединившей черты нескольких прибалтийско-финских диалектов. Языковая форма зародилась на полиэтнической древненовгородской территории [6]. Грамоты содержат собственные личные имена и топонимы с карельскими элементами, а также лексику прибалтийско-финского, в том числе карельского, происхождения.

Фиксации карельских топонимов и антропонимов в изобилии встречаются в новгородских и шведских документах позднего Средневековья.

Основную часть ранних рукописных и печатных памятников XVII-XIX вв. представляют словарные записи, грамматические очерки в этнографических описаниях, переводы текстов духовного содержания и учебные пособия.

В XVII в. кириллицей были записаны уникальные тексты десяти карельских заговоров, в которых можно найти черты всех наречий карельского (людиковского, ливвиковского и собственно карельского) и вепского языков, неоднородные по своим фонетико-морфологическим признакам. Язык этих текстов является предшественником карельского и вепского языков, поскольку процесс формирования наречий карельского языка в XVII в. еще продолжался. Этот источник ждет дальнейшего исследования.

К XVII в. относится русско-карельская словарная запись из списков Азбучного патерика библиотеки Соловецкого монастыря. К настоящему времени установлено, что время создания записи – 1666-1668 годы [7]. Также определены языковые особенности, позволяющие верифицировать карельскую диалектную основу словарной записи и описать характерные признаки карельского языка XVII в.

Не так давно обнаружена ранее неизвестная карело-русская запись 1668 года в авторском сборнике древнерусского книжника Прохора Коломнятина. Запись содержит 600 лексем и отражает язык тверских карелов [8].

Ранние опубликованные лексические материалы на карельском языке в некоторой степени известны благодаря историко-филологическим исследованиям ученых-этнографов XVIII-XIX вв. В поисках древнего мирового языкового универсума ученые сопоставляли лексические формы разных языков. Результатом такой деятельности стал, в частности, словарь тезаурусного типа П. Палласа «Сравнительные словари всех языков и наречий», в который вошли карельские лексические материалы. В настоящее время определена диалектная специфика карельских материалов словаря [9-10]. Исследование осуществлялось с применением инструмента «Анализ когнатов» на лингвистической платформе LingvoDoc.

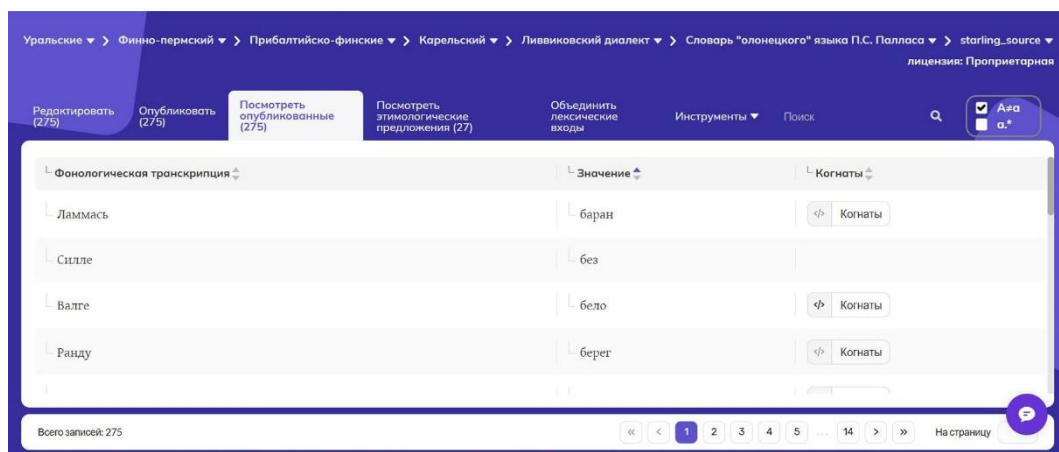


Рис. 1. Словарь «олонецкого» языка П.Палласа на платформе LingvoDoc.

Fig. 1. Dictionary of the "Olonets" language by P. Pallas on the LingvoDoc platform.

Практически все известные памятники карельской и вепсской письменностей были созданы на кириллической графике. Программы LingvoDoc позволяют обрабатывать тексты с сохранением графики, что имеет важное значение для описания памятников письменности.

Редким исключением являются, к примеру, записи члена-корреспондента Петербургской академии наук Федора Кеппена, который, доказывая родство индоевропейского и финно-угорского племен, в одной из своих работ в составе общей финно-угорской лексики привел ряд карельских и вепсских слов, зафиксированных латиницей.

Сведения о карелах и карельском языке содержались в этнографических описаниях Олонецкой и Архангельской губерний, некоторые из которых включали языковые материалы. Примером может служить произведение известного российского ученого-этнографа В.А. Дашкова «Описание Олонецкой губернии в историческом, статистическом и этнографическом отношениях» 1841 года, в заключительной части которого содержатся грамматический очерк карельского языка и небольшой карельско-русский словарь (чуть более 70 слов, записанных русскими буквами).

Значительная по объему и разнообразию карельская лексика представлена в труде А.В. Старчевского «Проводник и переводчик по отдаленнейшим окраинам России», опубликованном в 1889 г. в Санкт-Петербурге. Современные исследования карельского материала показали, что в работе использована разноречная лексика [11].

Памятников вепсской письменности на сегодняшний день известно значительно меньше, чем карельской. Прежде всего, это указанные выше заговоры XVII в. Рукопись с записью заговоров обнаружил в Заонежье российский филолог В.И. Срезневский. Известный финский языковед Э. Сетяля, консультировавший В.И. Срезневского, предположил, что нерусские

записи сделаны на вепском языке. При последующей работе с документом ученые пришли к заключению, что записи являются сплавом нескольких близкородственных языков южной Карелии, и часть из них – это вепско-людиковские заговоры, записанные кириллицей.

Второй, более поздний, памятник относится к середине XIX в. Он был обнаружен в делах канцелярии Олонецкого губернатора за 1848 г. и представляет собой рукописный русско-вепский словарь, составленный вытегорским исправником, который назвал свою работу «Лексиконом карельского языка». Лингвистическое исследование памятника показало, что в нем представлен куштозерский говор вепского языка, входящий в группу восточных, вологодских средневепских говоров [12]. Поскольку куштозерские вепсы давно ассимилировались, данный «Лексикон» представляет несомненный интерес для вепской исторической диалектологии.

Записи карельской и вепской лексики встречались в неопубликованных трудах по финно-угроведению. В частности, в Архиве Российской академии наук в С.-Петербурге сохранилась рукопись многоязычного глоссария А.М. Шегрена, включающего 360 лексем с переводами на саамский, ненецкий, коми языки, а также на ливвиковское наречие карельского языка и вепский язык. Рукопись отражает полевую работу Шегрена 1824-1829 гг., когда ученый находился в длительной экспедиции, побывав, в том числе, в Олонецкой и Северной Карелии.

В ходе реализации анонсируемого проекта нам удалось познакомиться с фондом А.М. Шегрена. В результате начатой обработки карельских и вепских материалов часть из них уже помещена на платформу LingvoDoc (в частности, список русских слов с переводом на «чудский» язык из фонда № 94 и др.).

2.2 Значение исследования

Сравнительно-сопоставительное исследование вводимых текстов памятников письменностей карельского и вепского языков на лингвистическую платформу LingvoDoc будет способствовать не только решению задач, стоящих перед сравнительно-историческим языкознанием (реконструкция исторических форм языков, исследование диахронических процессов и т.д.), но и решению одной из фундаментальных проблем современного гуманитарного знания – цифровой документации языков. В настоящее время документальная лингвистика является активно развивающейся областью прикладной лингвистики. Современные исследования в сфере документации языков направлены на обеспечение долгосрочных, всеобъемлющих и многоцелевых записей лингвистических практик, характерных для речевых сообществ. В числе таких лингвистических практик – памятники письменности.

Одна из важнейших целей документирования языка – предоставление данных для дальнейших теоретических и прикладных исследований языков, находящихся под угрозой исчезновения. Для языковой документации используются различные инструменты лингвистических технологий. Современный лингвистический исследовательский инструментарий, сосредоточенный на базе платформы LingvoDoc, позволяет посмотреть на уникальный материал памятников карельской и вепской письменности под новым углом, найти ответы на вопросы исторической фонетики, грамматики, лексикологии и диалектологии, которые до настоящего времени оставались нерешенными.

Теоретическая значимость исследования обусловлена возможностью реконструкции грамматических систем карельского и вепского языков анализируемого периода и определения хронологии основных языковых явлений. Важным представляется также систематизация принципов кириллической транскрипции, характерных для памятников письменности прибалтийско-финских языков данного периода.

Изучение памятников письменности имеет также важное прикладное значение для современной ситуации, поскольку формирование современной письменной традиции невозможно без учета истории карельского и вепского языков. Анализ материалов

памятников разных жанров, разного временного и территориального охвата с применением современных методик позволяет отследить историю языковых явлений, определить основные направления их контактирования, выявить архаизмы. Изучение истории языка, истории становления письменности является отправной точкой в процессе создания нормированных языков. Выявление и исследование исторических форм языка может способствовать развитию и обогащению его определенных фонетических, морфологических и синтаксических свойств. Проект имеет выход в современную практику развития новописьменных вариантов карельского и вепсского языков, поскольку корпусом карельских и вепсских старописьменных текстов можно воспользоваться при составлении исторической грамматики, толкового, этимологического и диалектных словарей, в преподавании вузовских курсов (история языка, диалектология, лексикология, финно-угроведение, феннистика и т.д.). Основными мероприятиями, осуществляемыми в рамках реализации проекта, таким образом, являются:

- инвентаризация и систематизация имеющихся в распоряжении лингвистов рукописных и печатных памятников карельского и вепсского языков, их текстологический анализ;
- выявление новых памятников письменности;
- оцифровка, расшифровка и аннотирование текстов;
- загрузка материалов на лингвистическую платформу LingvoDoc, создание конкордансов;
- анализ когнатов, определение диалектной принадлежности текстов памятников, описание фонологических особенностей;
- создание морфологических словарей памятников, выявление и описание морфологических особенностей текстов, их сопоставительный анализ;
- анализ лексического состава памятников письменности;
- разработка базы данных «Памятники письменности карельского и вепсского языков» и создание на ее основе электронного путеводителя.

Таким образом, системный анализ памятников карельской и вепсской письменностей позволит существенно расширить имеющиеся на сегодняшний день представления о развитии и становлении карельского и вепсского языков, осуществить реконструкцию языковых форм определенного исторического периода. Размещение текстов памятников на платформе LingvoDoc, в свою очередь, предоставит возможность использования современных методов анализа этих текстов и их цифрового документирования.

Список литературы / References

- [1]. Кошелюк Н. А. Чусовой словарь из архива Г. Ф. Миллера: особенности и его значение для классификации мансийских диалектов // *Oriental Studies*, № 5, 2023 г., стр. 1309–1324. DOI: 10.22162/2619-0990-2023-69-5-1309-1324. / Koshelyuk N. A. Dictionary of Chusovaya Mansi from G. F. Müller's Archives: Its Features and Significance for Mansi Dialect Classification. *Oriental Studies*, 2023, No 5, pp. 1309–1324 (in Russian). DOI: 10.22162/2619-0990-2023-69-5-1309-1324.
- [2]. Платформа LingvoDoc, Институт языкознания Российской академии наук, Институт системного программирования им. В.П. Иванникова Российской академии наук. Доступно по ссылке: <https://lingvodoc.ispras.ru/>, обращение 07.10.2025.
- [3]. Нагурная С.В. Карельская письменность. Народы Карелии: историко-этнографические очерки. Петрозаводск, Periodika, 2019, стр. 65–77. / Nagurnaya S.V. Karelian script. Peoples of Karelia: historical and ethnographic essays. Petrozavodsk, Periodika, 2019, pp. 65-77 (in Russian).
- [4]. Памятники лексикографии в «Цветнике» Прохора Коломнятина 1668 года. Исследования и тексты. М.; СПб., Альянс-Архео, 2023, 634 с. / Monuments of lexicography in Prokhor Kolomnyatin's

- "Flowerbed" of 1668. Research and texts. Moscow, Saint Petersburg, Alliance-Archeo, 2023, 634 p. (in Russian).
- [5]. Муллонен И. И., Новак И. П. Древнекарельские языковые особенности в памятниках письменности второй половины XVII — начала XVIII в. Вестник Санкт-Петербургского университета. Язык и литература, № 21 (3), 2024 г., стр. 720–740. DOI: 10.21638/spbu09.2024.312. / Mullonen I. I., Novak I. P. Old-Karelian language features in written language heritage of the second half of the 17th — early 18th centuries. Vestnik of Saint Petersburg University. Language and Literature, 2024, No 21 (3), pp. 720–740 (in Russian). DOI: 10.21638/spbu09.2024.312.
- [6]. Laakso J. Vielä kerran itämerensuomen vanhimmista muistomerkeistä. Virittäjä, 4, 1999, pp. 531–555.
- [7]. Муллонен И. И., Панченко О. В. Первый карельско-русский словарь и его автор афонский архимандрит Феофан. Петрозаводск, ПетрГУ, 2013, 116 с. / Mullonen I. I., Panchenko O. V. The first Karelian-Russian dictionary and its author, the Athonite archimandrite Feofan. Petrozavodsk, PetrSU, 2013, 116 p. (in Russian).
- [8]. Савельева Н. В. Неизвестные памятники лексикографии в «Цветнике» Прохора Коломнятина. Русская литература, № 3. 2019 г., стр. 54–63. / Savelyeva N. V. Unknown monuments of lexicography in Prokhor Kolomnyatin's "Flowerbed". Russian literature, 2019, No 3, pp. 54–63 (in Russian).
- [9]. Новак И. П. Определение диалектной специфики карельских материалов словаря П. С. Палласа. Финно-угорский мир, №1, 2024 г., стр. 33–49. DOI: 10.15507/2076-2577.016.2024.01.33-49. / Novak I. P. Definition of dialect specificity of Karelian materials of P. S. Pallas' dictionary. Finno-Ugric world, 2024, No 1, pp. 33–49 (in Russian). DOI: 10.15507/2076-2577.016.2024.01.33-49.
- [10]. Новак И. П., Нагурная С. В. Лингвистический анализ карельского языкового материала словаря П. С. Палласа. Финно-угорский мир, №3, 2023 г., стр. 287–300. DOI: 10.15507/2076-2577.015.2023.03.287-300. / Novak I. P., Nagurnaya SV. Linguistic analysis of Karelian language material in P. S. Pallas's dictionary. Finno-Ugric World, 2023, No 3, pp. 287–300 (in Russian). DOI: 10.15507/2076-2577.015.2023.03.287-300.
- [11]. Новак И. П., Нагурная С. В. Определение диалектной принадлежности карельского языкового материала «Проводника и переводчика по отдаленнейшим окраинам России» А. В. Старчевского. Финно-угорский мир, №1, 2022 г., стр. 20–32. DOI: 10.15507/2076-2577.014.2022.01.20-32. / Novak I. P., Nagurnaya S. V. Dialectal attribution of the Karelian language in "Guide and translator to remote outskirts of Russia" by A. V. Starchevsky. Finno-Ugric World, 2022, No 1, pp. 20–32. (In Russian). DOI: 10.15507/2076-2577.014.2022.01.20-32.
- [12]. Баранцев А. П. О рукописном русско-вепском словаре середины прошлого века. Ученые записки Тартуского государственного университета, № 344, 1975 г., стр. 43–55. / Barantsev A. P. On the handwritten Russian-Veps dictionary of the middle of the last century. Scientific Notes of Tartu State University, 1975, No 344, pp. 43–55 (In Russian).

Информация об авторах / Information about authors

Светлана Викторовна НАГУРНАЯ – кандидат филологических наук, старший научный сотрудник Института языка, литературы и истории Карельского научного центра РАН. Сфера научных интересов: история языка, документальная лингвистика, социолингвистика, лексикология.

Svetlana Viktorovna NAGURNAYA – Cand. Sci. (Philology), Senior Researcher of the Institute of Language, Literature and History of the Karelian Research Center of the Russian Academy of Sciences. Research interests: history of language, documentary linguistics, sociolinguistics, lexicology.

