

Коканова Е.С., Шняков П.Е.

**СПЕЦИФИКА РАЗРАБОТКИ НЕНЕЦКО-РУССКОГО И
РУССКО-НЕНЕЦКОГО ОНЛАЙН-СЛОВАРЯ[©]**

*Северный (Арктический) федеральный университет
имени М.В. Ломоносова,*

Россия, Архангельск, e.s.kokanova@narfu.ru, p.shnyakov@narfu.ru

Аннотация. В статье рассматриваются ключевые аспекты разработки ненецко-русского и русско-ненецкого онлайн-словаря, направленного на поддержку ненецкого литературного языка, находящегося под угрозой исчезновения. Описывается влияние суффиксальной морфологии ненецкого языка на структуру и содержательную организацию словарной базы. Особое внимание уделено обзору лексикографического программного обеспечения, в частности для языков малых народов России, применению методов обработки естественного языка и тонкой настройке GPT-2 на корпусе текстов на ненецком языке. Это открывает новые возможности для генерации текстов на ненецком языке и способствует расширению многоязычных корпусов и интеграции ненецкого языка в системы нейронного машинного перевода. Сформированные методологические принципы позволили разработать функциональный, адаптивный и удобный интерфейс онлайн-словаря с поддержкой ввода специфичных символов ненецкого алфавита и интеграцией мультимедийного контента и определить перспективы использования искусственного интеллекта для сохранения языков малочисленных народов и интеграции этих языков в цифровое пространство.

Ключевые слова: онлайн-словарь; ненецкий литературный язык; методы обработки естественного языка; нейронный машинный перевод; тонкая настройка GPT-2.

Поступила: 15.07.2024

Принята к печати: 10.06.2025

Kokanova E.S., Shnyakov P.Y.
Specifics of designing the Nenets-Russian and
Russian-Nenets online dictionary[©]

*Northern (Arctic) Federal University,
Russia, Arkhangelsk, e.s.kokanova@narfu.ru, p.shnyakov@narfu.ru*

Abstract. The paper deals with specifics of designing the Nenets-Russian and Russian-Nenets online dictionary aimed at supporting the endangered Nenets literary language. The impact of the Nenets suffixal morphology on the structure and content of the lexical database is analyzed, with a particular focus on Russian-to-Nenets translation. The study reviews specialized lexicographical software for the languages of Russia's minority groups, the application of natural language processing methods, and fine-tuning GPT-2, for the Nenets language corpus. This approach opens up new possibilities for text generation in Nenets and contributes to expanding multilingual corpora and integrating the Nenets language into neural machine translation. The paper also discusses the methodological principles for creating a functional and user-friendly interface that supports specific Nenets alphabet symbols and includes multimedia content. The potential of AI technologies are considered for the preservation of minority languages and their inclusion in the digital space are considered.

Keywords: online dictionary; Nenets literary language; NLP methods; neural machine translation; fine-tuning GPT-2.

Received: 15.07.2024

Accepted: 10.06.2025

Введение

Ненецкий язык является малоресурсным, находится под угрозой исчезновения и включен в категорию «критически уязвимых»¹ языков. Несмотря на кризисное положение, язык продолжает развиваться, пополняясь новой лексикой, отражающей современные социальные и политические реалии [Люблинская, 2019, с. 9]. Его сохранение становится важной задачей для носителей региона и страны², поскольку утрата языка означает потерю уникальной культуры, традиций и межпоколенческих знаний [Новожилов, 2022, с. 370].

Согласно статистическим данным, только 3% от общей численности населения Ненецкого автономного округа говорит на не-

© Kokanova E.S., Shnyakov P.E., 2025

¹ Лесные ненцы. – URL: <https://en.wal.unesco.org/languages/forest-nenets>

² О проекте Сибирского федерального университета. – URL: <https://snoword.ru/>

нечком языке [Бобрикова, 2024, с. 30]. Это вызвано политическими изменениями, утратой традиции передачи ненецкого языка и сокращением времени на его изучение в образовательных учреждениях – основополагающей сфере регулирования языковой ситуации [Белолюбская, 2014, с. 112]. Предпочтение отдается русскому языку как межнациональному средству общения из-за его коммуникативных возможностей [Платова, 2020, с. 555]. Это положение актуально для всех языков малых народов России, существующих с русским и другими влиятельными языками этнических групп [Арефьев, 2015, с. 76].

Правильно сформированные двуязычные образовательные программы, документирование языков и политические преобразования помогают сохранить языки малых народов [Бромхам, 2022, с. 170]. Отмечается, что изменения в системе образования, особенно в методах преподавания, стимулируют интерес носителей к изучению родного языка [Люблинская, 2024, 95]. Одной из задач по сохранению ненецкого языка в области образования является «создание портала ненецкого языка, включающего в себя онлайн-словари...» [Сабуров, 2023, с. 205].

В ответ на острую необходимость в специализированном языковом ресурсе настоящее исследование описывает специфику разработки первого онлайн-словаря для языковых пар русский – ненецкий и ненецкий – русский, основанного на большеземельском говоре тундрового диалекта ненецкого языка. Проект онлайн-словаря создан в связи с отсутствием подобных цифровых ресурсов и потребностью в создании многоязычного параллельного корпуса, который необходим для нейронного машинного перевода (НМП) [Шняков, 2024].

Ненецко-русский и русско-ненецкий онлайн-словарь¹ включает 4860 параллельных сегментов и 285 новостных статей, содержащих в среднем 345 ненецких и 387 русских слов в переводе; для точности и удобства перевода реализована функция вывода схожих слов на основе алгоритма расстояния Левенштейна.

Целью данного исследования является разработка методологических принципов создания онлайн-словаря для ненецкого литературного языка, предполагающая решение следующих основных задач:

¹ Ненецко-русский и русско-ненецкий онлайн-словарь. – URL: <https://nenrusdict.narf.ru>

- изучить существующие лексикографические ресурсы по разработке онлайн-словаря;
- выявить актуальные подходы в лексикографическом программном обеспечении (ПО), адаптированном под ненецкий язык;
- определить методологические принципы, необходимые для разработки первого подобного цифрового инструмента, включая методы наполнения и аугментации, специфику интерфейса и функцию поиска на русском и ненецком языках;
- оценить перспективы и ограничения дальнейшего расширения инструмента, включая добавление мультимедийных и вспомогательных материалов для поддержания интереса к ненецкому языку.

К ключевым этапам проекта относятся: лингвистический анализ, формирование базы словарных материалов, а также разработка и тестирование веб-приложения. Каждый из этапов сопровождается определенными сложностями, обусловленными агглютинативно-суффиксальной природой¹ малоресурсного языка, оказывающей существенное влияние на процесс сбора данных, обработку текста на ненецком языке (Natural Language Processing, NLP) и создание словарной статьи. Сравним образование форм глагола *плакать* по лицам в русском и ненецком языках:

я заплáкал – мань ярумад’;
ты заплáкал – пыдар яруман;
он заплáкал – пыда ярума;
она запла́кала – пыда ярума.

Типичная структура ненецкого слова включает корень, за которым следуют словообразовательные и словоизменительные аффиксы; в современном ненецком языке нет грамматической категории рода, что показывают вышеупомянутые примеры.

Лесной диалект ненецкого языка более распространен в ненецкоговорящей онлайн-среде по сравнению с тундровым диалектом [Шняков, 2024]; данные для тундрового диалекта крайне необходимы и для онлайн-словаря, и для модели НМП. Сбор данных осуществляется с помощью возможностей Python и VK API, оптического распознавания символов (OCR) и ручного редактирования.

Представить процесс создания такого масштабного проекта без технологий искусственного интеллекта (ИИ) в настоящее время невозможно, поэтому применены методы NLP: морфологический

¹ Ненецкий язык. – URL: <https://minlang.iling-ran.ru/lang/neneckiy-yazyk>

анализ текста на ненецком языке с использованием ruTmorph² для удаления русских слов и векторизация слов для семантического поиска по корпусу с Word2Vec. Используется также глубокое обучение (Deep Learning, DL) для разработки модели НМП для языковой пары русский – ненецкий, что позволит начать работу в направлении онлайн-переводчика¹.

Тонкая настройка предобученной модели GPT-2 на корпусе текстов на ненецком языке предлагается в качестве дополнительной технологии ИИ в процессе разработки онлайн-словаря. Она открывает новые возможности для генерации текста на ненецком языке, который является недостаточно исследованным, и позволяет оценить применение заранее обученных больших языковых моделей (Large Language Models, LLMs) для выполнения задачи НМП. Дообученную модель возможно использовать для аугментации текстовых данных: увеличения качественного многоязычного параллельного корпуса и ускорения внедрения ненецкого языка в системы НМП.

Обзор лексикографических инструментов

В настоящее время существуют онлайн-словари и онлайн-переводчики² языков малых народов России, предоставляющие функции поиска слов и вывода эквивалентных переводов на другой язык. Цифровые решения имеют преимущества по сравнению с печатными словарями, обеспечивая более удобный доступ к информации и расширенные возможности для обучения [Ольшвант, 2020]. Их возможности поиска, визуализации, гиперссылок и обновления данных играют ключевую роль в поддержании их полезности и актуальности в образовательной среде [Агапова, 2014]. Функционал подобных словарей включает не только ввод с клавиатуры, но и поиск по отдельным буквам или наборам символов для упрощения нахождения нужного слова. Например, подобная архитектура реализована в русско-коми словаре³. В онлайн-словарях часто возникает проблема ввода символов, специфичных

¹ Как цифровизация помогает сохранить языки коренных народов России. – URL: <https://gramota.ru/journal/stati/tekhnologii/kak-tsifrovizatsiya-pomogaet-sokhranit-yazyki-korennykh-narodov-rossii>

² Нейросети Яндекса помогут сохранить языки народов России и ближе познакомиться с их культурой. – URL: <https://yandex.ru/company/news/01-02-07-2024>

³ Учим коми язык. – URL: <http://komikyv.ru>

для каждого языка, ввиду различий в алфавитах. Для решения этой проблемы рядом с полем ввода добавляются кнопки для ввода специфичных символов, как, например, в русско-бурятском словаре¹. В некоторых случаях возможно объединение букв для упрощения поиска слов. В осетинско-русском словаре² вместо символа «æ» можно использовать «ae». Разработчики также предлагают ссылки на скачивание клавиатуры соответствующего языка, как на сайте якутского словаря³.

Некоторые онлайн-словари интегрированы с медиаконтентом на национальном языке, что способствует более глубокому погружению в изучаемый язык. Например, русско-чувашский онлайн-словарь⁴ предлагает пользователям доступ к песням на чувашском языке. В связи с этим важным элементом лексикографического ПО является возможность прослушивания произношения слов, позволяющая улучшить навыки аудирования у пользователей.

Интеграция современных технологий для мобильных устройств в разработку словарей повышает их доступность и удобство, обеспечивая использование в любых условиях [Сусилана, 2022]. Мобильные и веб-приложения интегрируют словари в образовательный процесс, предлагая пользователям возможность изучения и практики языка в любое время и в любом месте [Докторов, 2023]. Это значительно расширяет функциональные возможности и практическую ценность словарей.

Вместе с тем дружественный интерфейс онлайн-словаря во многом зависит от его архитектуры и дизайна. Перегрузка веб-страницы излишней информацией может затруднить пользование словарем. Интуитивно понятный интерфейс облегчает работу со словарем, особенно для пользователей, малознакомых с цифровыми технологиями. Следовательно, создание современного внешнего вида веб-приложения и использование элементов народного орнамента в дизайне позволяют проследить связь продукта с культурой наенцев и улучшать пользовательскую культуру.

¹ Русско-бурятский, бурятско-русский электронный словарь. – URL: <https://burlang.ru>

² Осетинский язык online. – URL: <https://ironau.ru>

³ Якутский словарь. – URL: <https://sakhatyla.ru>

⁴ Русско-чувашский онлайн-словарь и переводчик. – URL: <https://chuvashtrepublic.ru>

Обсуждение

За последнее столетие ненецкий язык претерпел значительные изменения. Кириллический алфавит пришел на смену латинскому и включил уникальные символы *у*, ' (звонкая тасер'') и '' (глухая тасер''). Для удобства пользователей поисковая строка должна поддерживать ввод этих символов и обрабатывать запросы с опечатками, обеспечивая динамическое автодополнение и фильтрацию слов с метками и URL-адресами для перехода к страницам со словарными статьями (рис. 1). В каждой словарной статье должно быть указано слово, его эквиваленты на другом языке, информация о говоре, диалекте и источнике. Важным компонентом словарной статьи является толкование или определение многозначного слова, например, глагол *бегать* в ненецком языке может иметь множество значений, связанных с тем, кто и как совершает это действие [Коканова, 2023, с. 89].

термин¹ часть речи **Н** **И** **М**
тематика или принадлежность толкование или определение (значение)
диалект и говор
морфологические особенности
информация о склонении или спряжении
синтаксические особенности
эквивалент
толкование или определение (значение) эквивалента

Примеры использования

Первый пример, содержащий термин. — Перевод примера.

Второй пример, содержащий термин. — Перевод примера.

Иные формы слова

термин во мн. в. множественный вид
термин в д-мн. в. длительно-множественный вид

Источник

Источник

Словарь, откуда заимствована информация о слове и его особенностях
Автор или авторы
Издательство и информация о нем
Год публикации словаря

Рис. 1. Пример организации словарной статьи в онлайн-словаре

Интеграция предобученной нейронной сети для генерации изображений с использованием модели Stable Diffusion v1.5 повысит интерактивность словаря (рис. 2).



Рис. 2. Пошаговая генерация изображения для «кота» хора (кот)

Модификация проекта позволит внедрить собранные языковые параллельные корпуса в образовательные игровые приложения для детей, сходные с подобными реализациями: Studycat, Little Pim, Dinolingo и Duolingo (рис. 3).

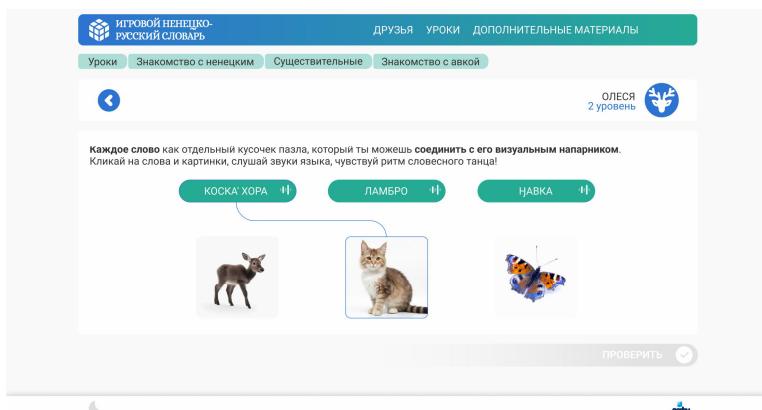


Рис. 3. Пример лексического интерактивного задания в игровом режиме

Для расширения доступности ненецкого языка в сети Интернет важно интегрировать новостные и информационные статьи, фольклорные сборники и другие важные материалы в онлайн-словарь [Поликарпов, 2023, с. 120]. Это обеспечит автоматический и динамический вывод современных текстов, упрощая изучение лексики. Благодаря современным моделям DL с архитектурой трансформер возможно генерировать текст на ненецком языке и в перспективе расширять собранный многоязычный параллельный корпус.

Результаты теоретического обзора лексикографического ПО позволяют сформировать методологические принципы и актуали-

зировать технологические операции, на которых базируется разработка онлайн-словаря для ненецкого языка:

- поддержка уникальных символов ненецкого языка, таких как ю и ѿ;
- внедренные алгоритмы динамического автодополнения и обработки запросов с учетом возможных ошибок пользователя;
- представление многоуровневой лексической информации (эквиваленты на русском языке, говор, диалект и происхождение), позволяющее учитывать специфику значения в зависимости от контекста;
 - интеграция современных нейросетевых моделей (Stable Diffusion) для создания визуальных элементов, повышающая интерактивность ресурса;
 - адаптация ресурса под различные устройства для повышения доступности;
 - потенциал использования системы как образовательного игрового приложения для детей, мотивирующего подрастающее поколение к изучению родного языка.

Это помогает разработать лексикографический инструмент, необходимый для сохранения и изучения ненецкого языка и многих элементов культуры, которые исчезают в связи с изменениями традиционного образа жизни [Раздымаха, 2021]. Интересным представляется применение генеративных нейросетей для расширения текстового корпуса.

GPT (Generative Pre-trained Transformer) – это архитектура модели DL, предназначенная для генерации связного текста на основе заданного контекста. Выбор GPT-2¹ обусловлен ключевыми факторами: потенциалом для увеличения объема текстовых данных на ненецком языке, открытыми весами и возможностями ее обучения на малоресурсном языке, знанием о языковых закономерностях, возможностью обучения на больших объемах текстовых данных.

Для реализации и настройки GPT-2 используется библиотека transformers и платформа Hugging Face², которая предоставляет инструменты для загрузки нейросетевых моделей. Сначала загружается предварительно обученная модель и токенизатор. Конвейер для генерации текста настраивается с помощью функции pipeline, где указываются модель, токенизатор и максимальная длина генерируемого текста.

¹ OpenAI GPT2. – URL: https://huggingface.co/docs/transformers/en/model_doc/gpt2

² Конвейеры для инференса. – URL: https://huggingface.co/docs/transformers/en/pipeline_tutorial#pipelines-for-inference

Для настройки модели на ненецком языке необходимо загрузить корпус текстов и подготовить данные. Это включает создание объекта TextDataset с использованием токенизатора. Далее формируются пакеты данных для обучения модели с помощью DataCollator ForLanguageModeling.

Параметры обучения включают:

- директорию для сохранения результатов (output_dir) и флаг для перезаписи содержимого;
- количество эпох обучения, размер пакета, частоту сохранения и ограничение на количество сохраненных моделей;
- установку флага prediction_loss_only для экономии пространства на диске.

Процесс обучения начинается с создания объекта Trainer, передачи модели, корпуса и параметров обучения, а также вызова метода train() (рис. 4).



Рис. 4. Процесс обучения модели

Проверка обученной модели включает генерацию текста на ненецком языке с использованием токенизатора и метода generate. Сгенерированный текст может не всегда быть грамматически и стилистически корректным, но он значительно лучше, чем у модели без тонкой настройки на ненецком языке (табл. 1).

Как видно из таблицы, модель, обученная на корпусе ненецких текстов, значительно лучше справляется с задачей генерации текста на ненецком языке по сравнению с моделью без тонкой настройки. В столбце «Содержание ответа модели GPT-2» не представляется возможным показать перевод на русский язык, так как модель «галлюцинирует», то есть выдает решения, не подтвержденные данными ее обучения, или вымышленные ответы, не подкрепляемые действительностью. Для достижения еще более высоких результатов рекомендуется продолжать обучение модели на качественных и разнообразных данных.

Таблица 1

Сгенерированные ответы на некоторые запросы

№	Запрос с переводом	Содержание ответа модели GPT-2	Содержание ответа обученной на корпусе модели GPT-2 с переводом
1	Ань торова (Здравствуйте)	Ань торова позразния эидеция в призудает в ис	Ань торовань тара. Тарем' нянна дани' тара-на ненэц" нянна дани' тара (Здравствуйте. Этот ненецкий участник мой)
2	Нани' торова! (Здравствуй!)	Нани' торова! Еда и об-рэвно ужду IPS, Ражт, Со	Нани' торова! Тарем' нянна манзараванда тара-на дани' (Здравствуй! Этот участник лодыры)
3	Лакамбой (До свидания)	Лакамбой Такведно и офял в разь оть Va ма, ф	Лакамбой' нянда нани' тарем' манэ"нада (До свидания, так он увидел)
4	Серон сававна мина" (Хорошо идут дела)	Серон сававна мина"оргияния\n\nСерор он сававна	Серон сававна мина". Тарця сававна ми" са-вавна ми" (Хорошо идут дела. Такой вести)
5	Мань манзаян' мицам' (Я иду из дома на работу)	Мань манзаян' мицам'. набетовденномномость к	Мань манзаян' мицам'. Тарця нани' тэванда нани' (Я иду из дома на работу. Такой около)

Заключение

Разработанный онлайн-словарь интегрирует 4860 сегментов и 285 новостных статей на ненецком и русском языках с автоматическим дополнением лексических единиц примерами употребления и реализацией алгоритма поиска похожих слов с применением расстояния Левенштейна. Лексикографическое ПО является не просто справочным ресурсом, но и учебным инструментом для изучающих ненецкий язык.

На основе обзора современных лексикографических инструментов выявлены ключевые подходы и технологии для разработки ненецко-русского и русско-ненецкого онлайн-словаря: динамический поиск, поддержка ввода специфичных символов, интеграция медиаконтента и интуитивно понятный интерфейс. Например, в разработанном продукте:

- поисковая строка обрабатывает опечатки и предлагает альтернативные варианты при некорректном вводе, включая неверную раскладку клавиатуры;
- предусмотрена функция поиска слов по начальной букве или набору от двух до трех символов;
- словарная статья базируется на учете лингвистических особенностей ненецкого языка, включая грамматические формы лексем, такие как (1) множественное и длительно-множественное число (например, для глагола *сюрберц* ‘бегать (о человеке)’ длительно-множественная форма представлена как *сюрмберц*); (2) данные о принадлежности слова к конкретному диалекту и говору (например, для существительного *увака* (*авка*) указано: «тундровый диалект, большеземельский говор») и информация об источнике, из которого взята лексема (например, для слова *Нэрм* ‘Арктика’ указан источник – «Русско-ненецкий словарь» А.П. Пырерка и Н.М. Терещенко); (3) информация о лексико-грамматических особенностях ненецкого языка уточнялась у носителей языка (например, в Ассоциации ненецкого народа «Ясавэй»);
- продумана логика вывода омонимов из базы данных, включающей 7413 ненецких и 3542 русских слов, для обеспечения четкой структуры информации и упрощения восприятия пользователем, исключения путаницы: при выборе слова с омонимами на веб-странице автоматически выводится список его значений с нумерацией; например, для лексемы *банка* представлены два значения: первое – *банка* как емкость для жидкостей или сыпучих продуктов (посуды), второе – *банка* как подводная отмель;
- веб-приложение оптимизировано для адаптивного отображения на экранах различных типов устройств.

Дальнейшее совершенствование онлайн-словаря включает применение DL и методов обучения LLM с целью расширения текстовых ресурсов на ненецком языке и улучшения интеграции языка в системы НМП.

Список литературы

- Агапова Н.А., Картофелева Н.Ф. О принципах создания электронного словаря лингвокультурологического типа: к постановке проблемы // Вестник Томского государственного университета. – 2014. – № 382. – С. 6–10.
- Арефьев А.Л. О языках коренных малочисленных народов России // Социологические исследования. – 2015. – № 8. – С. 75–84.
- Белолюбская В.Г. Языки коренных народов Севера: история и современность // Высшее образование в России. – 2014. – № 5. – С. 110–115.
- Бобрикова М.И. Современное состояние преподавания ненецкого языка на примере Ненецкого автономного округа // Русский Север и Арктика: фундаментальные проблемы истории и современности. – Архангельск : Сев. (Арктич.) федер. ун-т им. М.В. Ломоносова, 2024. – С. 29–34.

- Глобальные предикторы исчезновения языков и будущее лингвистического разнообразия / Бромхам Л., Диннейдж Р., Скиргард Х., Ричи А., Кардилло М., Меакинс Ф., Гринхилл С., Хуа Ся // Природная экология и эволюция. – 2022. – Т. 6. – С. 163–173. – DOI: 10.1038/s41559-021-01604-y
- Докторов А.А., Сидорова А.А., Дмитриева О.Н. Разработка мобильного приложения для изучения английского языка на основе сказки о Кольмане // МНКО. – 2023. – № 2(99). – С. 196–199. – DOI: 10.24412/1991-5497-2023-299-196-199.
- Коканова Е.С. Когнитивно-семантический подход к лексикографическому портфолио глаголов движения при создании ненецкого онлайн-словаря // Когнитивные исследования языка. – 2023. – № 3-2(54). – С. 88–91.
- Люблинская М.Д., Валенкова В.В. Преподавание языка как способ укрепления национального самосознания // Социолингвистика. – 2024. – № 1(17). – С. 91–103. – DOI: 10.37892/2713-2951-1-17-91-103
- Люблинская М.Д., Пушкирева М.В. Создание новой терминологии на тундровом ненецком языке // Родной язык. – 2019. – № 2. – С. 5–32.
- Новожилов А.Г. Сохранение языков коренных малочисленных народов в условиях глобализации (опыт северо-запада России) // Русин. – 2022. – № 67. – С. 361–374. – DOI: 10.17223/18572685/67/20
- Ольшванг О.Ю., Архипова И.С., Олехнович О.Г. Разработка электронного латинско-русского словаря для студентов медицинских специальностей // МНКО. – 2020. – № 1(80). – С. 15–17. – DOI: 10.24411/1991-5497-2020-00006
- Платова Г.В., Кильдяшова Т.А. Современный статус и перспективы ревитализации ненецкого языка в ненецком автономном округе // Ломоносовские научные чтения студентов, аспирантов и молодых ученых – 2020 / Сев. (Арктич.) федер. ун-т им. М.В. Ломоносова ; сост. Ю.С. Кузнецова. – Архангельск : Сев. (Арктич.) федер. ун-т им. М.В. Ломоносова, 2020. – Т. 1. – С. 552–557.
- Поликарпов А.М., Латышева О.Е. Перевод фольклорных текстов как средство сохранения ненецкого языка: интегративный подход // Вестник ВолГУ. Серия 2: Языкознание. – 2023. – № 3. – С. 118–128. – DOI: 10.15688/jvolsu2.2023.3.11
- Раздымаха С.Т. Праздник встречи солнца: обрядовая культура как элемент устного народного творчества ненцев // Культура и цивилизация. – 2021. – Т. 11, № 1-1. – С. 161–165. – DOI: 10.34670/AR.2020.45.59.019
- Сабуров А.А., Никифоров А.С., Минчук О.В. Состояние сохранности ненецкого языка в Ненецком автономном округе: по материалам социолингвистического исследования // Арктика и Север. – 2023. – № 50. – С. 189–210. – DOI: 10.37482/issn2221-2698.2023.50.189
- Сусилана Р., Рулляна Г., Ардианаша, Вуланари Я. Словарь педагогики: разработка веб-приложения // Международный журнал по вопросам образования. – 2022. – Т. 15, № 1. – С. 197–218. – DOI: 10.29333/iji.2022.15112a
- Шняков П.Е. Разработка модели нейронного машинного перевода в языковой паре русский – ненецкий // Ломоносовские научные чтения студентов, аспирантов и молодых ученых – 2024 / Сев. (Арктич.) федер. ун-т им. М.В. Ломоносова. – Архангельск : ИД САФУ, 2024. – Т. 2. – С. 745–750.
- Шняков П.Е. Сбор данных на лесном диалекте ненецкого языка и построение векторной модели с определением контекстной близости слов // От машинного перевода к машинному обучению : сб. науч. статей (27–28 октября 2023 г.) /

отв. ред. Е.С. Коканова ; Сев. (Арктич.) федер. ун-т им. М.В. Ломоносова. – Архангельск : Сев. (Арктич.) федер. ун-т им. М.В. Ломоносова, 2024. – С. 80–86.

References

- Agapova, N.A., Kartofeleva, N.F. (2014). On the principles of creating an electronic dictionary of the linguoculturological type: raising the problem. *Tomsk State University Journal*, 382, 6–10.
- Arefiev, A.L. (2015). O yazykakh korennnykh malochislennykh narodov Rossii. *Sociological Studies*, 8, 75–84.
- Belolyubskaya, V.G. (2014). The system of education in the languages of the indigenous minorities in the Republic of Sakha (Yakutia): history and modernity. *Higher Education in Russia*, 5, 110–115.
- Bobrikova, M.I. (2024). Sovremennoe sostoyanie prepodavaniya nenetskogo yazyka na primere Nenetskogo avtonomnogo okruga. In *Russian North and Arctic: Fundamental Issues of History and Modernity* (pp. 29–34). Arkhangel'sk: Lomonosov Northern (Arctic) Federal University.
- Bromham, L., Dinnage, R., Skirgård, H., Ritchie, A., Cardillo, M., Meakins, F., Greenhill, S., Hua, Xia. (2022). Global predictors of language endangerment and the future of linguistic diversity. *Nature Ecology & Evolution*, 6, 163–173. DOI: 10.1038/s41559-021-01604-y
- Doktorov, A.A., Sidorova, A.A., & Dmitrieva, O.N. (2023). Development of a Mobile Application for Learning English Based on the Tale of Kolyman. *MNKO*, 2(99), 196–199. DOI: 10.24412/1991-5497-2023-299-196-199
- Kokanova, E.S. (2023). Cognitive semantic approach to the lexicographic portrait of motion verbs in creating a Nenets Online dictionary. *Cognitive Studies of Language*, 3(54), 88–91.
- Lyublinskaya, M.D., Pushkareva, M.V. (2019). Vocabulary development in Tundra Nenets. *Rodnoy yazyk*, 2, 5–32.
- Lyublinskaya, M.D., Valenkova, V.V. (2024). Language teaching as a way to strengthen national identity. *Sociolinguistics*, 1(17), 91–103. DOI: 10.37892/2713-2951-1-17-91-103
- Novozhilov, A.G. (2022). Preserving the languages of indigenous minorities under globalization (a case study of the Russian North-West). *Rusin*, 67, 361–374. DOI: 10.17223/18572685/67/20
- Ol'shvang, O.Y., Arkhipova, I.S., Olekhovich, O.G. (2020). Development of an electronic Latin-Russian dictionary for students of medical specialties. *MNKO*, 1(80), 15–17. DOI: 10.24411/1991-5497-2020-00006
- Platova, G.V., Kildyashova, T.A. (2020). Sovremennyy status i perspektivy revitalizatsii nenetskogo yazyka v Nenetskem avtonomnom okruse. *Lomonosovskie nauchnye chteniya studentov, aspirantov i molodykh uchenykh – 2020*, 1, 552–557.
- Polykarpov, A.M., Latysheva, O.E. (2023). Translation of folklore texts as a means of preserving the Nenets language: an integrative approach. *Science Journal of VolsU. Linguistics*, 3, 118–128. DOI: 10.15688/jvolsu2.2023.3.11
- Razdymakha, S.T. (2021). Celebration of the sun: ritual culture as an element of oral folk art of the Nenets. *Culture and Civilization*, 11(1-1), 161–165. DOI: 10.34670/AR.2020.45.59.019

- Saburov, A.A., Nikiforov, A.S., Minchuk, O.V. (2023). Preservation of the Nenets language in the Nenets Autonomous Okrug: based on sociological survey. *Arctic and North*, 50, 155–174. DOI: 10.37482/issn2221-2698.2023.50.189
- Shnyakov, P.E. (2024). Data collection on the Forest dialect of the Nenets language and building a Vector Model to determine contextual word similarity. In *Ot mashinnogo perevoda k mashinnomu obucheniyu* (pp. 80–86). Arkhangelsk.
- Shnyakov, P.E. (2024). Razrabotka modeli neyronnogo mashinnogo perevoda v yazykovoy pare russkiy – nenetskiy. *Lomonosovskie nauchnye chteniya studentov, aspirantov i molodykh uchenykh – 2024*, 2, 745–750.
- Susilana, R., Rullyana, G., Ardiansah, & Wulandari, Y. (2022). Pedagogia Dictionary: Web Application Development. *International Journal of Instruction*, 15(1), 197–218. DOI: 10.29333/iji.2022.15112a.

Сведения об авторах

Коканова Елена Сергеевна – кандидат филологических наук, доцент, заведующая базовой кафедрой технологий и автоматизации перевода в бюро переводов «АКМ-Вест», Северный (Арктический) федеральный университет имени М.В. Ломоносова, Россия, Архангельск, e.s.kokanova@narfu.ru

Шняков Павел Евгеньевич – ассистент кафедры информационных систем и информационной безопасности, Северный (Арктический) федеральный университет имени М.В. Ломоносова, Россия, Архангельск, p.shnyakov@narfu.ru

About the authors

Kokanova Elena Sergeevna – Ph. D. of Philology, Associate Professor, Head of the Department of Translation Technology and Practice at AKM-WEST, Northern (Arctic) Federal University, Russia, Arkhangelsk, e.s.kokanova@narfu.ru

Shnyakov Pavel Yevgenyevich – Assistant Professor, Department of Information Systems and Information Security, Northern (Arctic) Federal University, Russia, Arkhangelsk, p.shnyakov@narfu.ru