

УДК 008**Возможности использования искусственного интеллекта в
ревитализации исчезающих языков****Петров Алексей Иванович**

Аспирант,
Арктический государственный институт культуры и искусств,
677000, Российская Федерация, Якутск, ул. Орджоникидзе, 4;
e-mail: petrovai05@mail.ru

Аннотация

Статья исследует возможности применения искусственного интеллекта (ИИ) для ревитализации исчезающих языков на примере юкагирского языка. Рассматриваются современные технологии ИИ в обработке естественного языка для передачи редких языков новым поколениям. Анализируется потенциал ИИ в разработке образовательных ресурсов, включая виртуальных языковых ассистентов, интерактивных платформ с обратной связью в режиме реального времени, самоучителей, разговорников, аудиоматериалов с натуральным звучанием, созданных ИИ. Оценивается возможность повышения доступности материалов на юкагирском языке путем применения ИИ для расширения контента, как переводы текстов, книг, сайтов, разработка мультимедийных материалов для театров, концертов, конференций. Отдельное внимание уделяется применению ИИ для подготовки юридических документов и машинного перевода в ознакомительных целях. Перечисленное может стать основой для дальнейших исследований, способствовать разработке лексики и расширению сфер применения языка за пределами традиционного образа жизни и хозяйственной деятельности, что открывает перспективы возрождения языка.

Для цитирования в научных исследованиях

Петров А. И. Возможности использования искусственного интеллекта в ревитализации исчезающих языков // Культура и цивилизация. 2024. Том 14. № 11А. С. 21-27.

Ключевые слова

Этнокультурное наследие, языковая ревитализация, юкагирский язык, искусственный интеллект, цифровая лингвистика, сохранение культурной идентичности, мультимедийные технологии.

Введение

В последние десятилетия проблема исчезающих языков становится все более актуальной в условиях глобализации и доминирования нескольких мировых языков, что угрожает культурному многообразию и приводит к утрате этнокультурного наследия. Сохранение и ревитализация таких языков является важной задачей, направленной на поддержание культурной идентичности малочисленных народов.

На 2012 год, в мире насчитывается около 6500–7000 языков. Различия в основном вызваны неоднозначностью определения «язык» и «диалект» [Pereltsvaig, 2012, 11].

Согласно данным ЮНЕСКО, в 2003 году 96% населения мира говорит лишь на 3% от количества языков в мире. ЮНЕСКО оценивает, что по крайней мере 50% из более чем 6000 языков мира теряют носителей. Таким образом, к концу 21-го века около 90% всех языков мира могут быть заменены доминирующими языками. Практически во всех странах мира существует острая необходимость в поддержании языков на различных уровнях исчезания [UNESCO, 2003, 4].

Исчезание каждого языка – это утрата уникального человеческого опыта, отражающего мышление и культурные ценности, являясь неотъемлемой частью культуры и основным средством ее передачи, и выражением национальной ментальности каждого народа. Утрата языка для носителей связана с потерей части этнической и культурной самобытности, памяти и традиций предков.

Актуальность исследования выдвинутой темы обусловлена тем, что применение ИИ в ревитализации исчезающих языков открывает новые возможности для их сохранения и возрождения. Возможности рассматриваются на примере юкагирского языка, который также относится к исчезающим. Что в свою очередь поддерживает культурное и языковое разнообразие, а также способствует более глубокому пониманию роли технологий в сохранении этнокультурного наследия.

Одной из основных проблем ревитализации исчезающих языков, можно выделить отсутствие устойчивой языковой среды и недостаток ресурсов для обучения и поддержки исчезающих языков, что требует применения инновационных цифровых решений.

Материалы и методы исследования

В настоящей статье материалом исследования выступили работы российских и зарубежных ученых, касающиеся исчезающих языков, ревитализации языков, цифровой лингвистики и ИИ, применительно к их использованию в ревитализации юкагирского языка. Исследование было выполнено аналитическим методом междисциплинарного изучения выявленной проблемы.

Результаты исследования и их обсуждение

Анализ проблемы ревитализации языков выявил, что основным индикатором для оценки витальности (жизнеспособности) языка является определение механизмов передачи того или иного языка из поколения в поколение его носителей.

В российской практике ревитализации языков выявлены следующие механизмы:

- образование (на всех уровнях государственной образовательной системы и вне государственной образовательной системы);
- контент (широта и доступность материалов на языке, формат, подходящий под различные

- целевые аудитории, юридическая чистота материалов);
- культура и социум (внутри семьи, семейный досуг, развлекательные мероприятия, фестивали, театры, вывески, таблички, меню, ценники);
 - любые документы, имеющие юридическую силу (законодательные, регламентные, внутрикорпоративные);
 - документация и кодификация (создание базы для научных исследований, базы для последующих учебных материалов).

Данная классификация основана на крупных объединениях схожих практик, с разделением на подкатегории и возможностью добавления новых практик [Винклер, 2021, 15].

Эксперты ЮНЕСКО выделяют 6 степеней угрозы исчезновения языка. Стоит отметить, что даже 5-й «безопасный» уровень не гарантирует жизнеспособность языка, поскольку в любой момент носители языка могут прекратить передавать свой язык следующему поколению [UNESCO, 2003, 9].

Таблица 1 - Уровень исчезновения языка на основе количества говорящих на этом языке [UNESCO, 2003, 9]

Уровень	Оценка	Количество говорящих на этом языке
Безопасный	5	Язык используется людьми всех возрастов, от детей и старше.
Небезопасный	4	Язык используется некоторыми детьми во всех сферах (контекстах) или используется всеми детьми только в некоторых сферах.
Под угрозой исчезновения	3	В основном язык используется поколением уровня родители и старше.
Серьезная угроза исчезновения	2	В основном язык используется поколением уровня дедушек и бабушек и старше.
Критическая угроза исчезновения	1	В основном язык используется немногими носителями поколения прабабушек и прадедушек.
Исчезнувший язык	0	Никто не говорит на этом языке.

На основе данной таблицы можно определить языки, которые находятся в серьезной и критической угрозе исчезновения. Например, юкагирский язык находится на уровне критической угрозы исчезновения. Юкагирский язык в основном распространен на северо-востоке Якутии, а также в районе Чукотского автономного округа и частично в Магаданской области. Юкагирский язык делится на тундренный и лесной [Шадрин, 2022, 67]. Согласно Всероссийской переписи населения Российской Федерации 2010 года, число людей, считающих юкагирский язык родным, составляет 370 чел., из них большая часть знают только отдельные фразы и слова. Есть сведения о том, что количество владеющих юкагирским языком на разных уровнях, не превышает 100 чел. [Шадрин, 2022, 67]. И они представляют собой преимущественно представителей старшего и лишь частично среднего поколений.

Юкагирский язык бытовал лишь в изустной среде, в неформальном общении между сородичами. Первая книга на юкагирском языке была опубликована лишь в 1969 г., это был сборник стихов Улуро Адо «Кодэ, мэру: нийуол» [Шадрин, 2022, 71].

Несмотря на наличие разработанной письменности, обучение юкагирскому языку в детских садах и школах сталкивается с серьезными проблемами, такими как нехватка учебных материалов, недостаток квалифицированных преподавателей-юкагиров, а также крайне низкий интерес к изучению языка у детей-юкагиров и отсутствие поддержки для его использования в семейной среде [Лукина, 2023, 151].

Как показывают публикации ЮНЕСКО, процесс ревитализации языка зависит от уровня грамотности этнического сообщества, владеющего данным языком.

**Таблица 2 - Уровень исчезновения языка на основе грамотности населения
[UNESCO, 2003, 14]**

Уровень исчезновения языка	Оценка	Доступность письменных материалов
Безопасный	5	Существует устоявшаяся орфография, традиция грамотности, в литературе и в повседневности, в СМИ, в административной сфере и образовании.
Небезопасный	4	Письменные материалы существуют, и в школе дети развивают грамотность на этом языке. Письменность на языке не используется в административной сфере.
Под угрозой исчезновения	3	Письменные материалы существуют, и дети могут сталкиваться с письменной формой языка в школе. Грамотность не продвигается через печатные СМИ.
Серьезная угроза исчезновения	2	Письменные материалы существуют, но они могут быть полезны лишь для некоторых членов сообщества. Обучение грамотности на этом языке не является частью школьной программы.
Критическая угроза исчезновения	1	Практическая орфография известна обществу, и некоторые материалы пишутся на этом языке.
Исчезнувший	0	Орфография недоступна для общества.

Исходя из данной таблицы, следует отметить, что динамику исчезновения юкагирского языка можно отнести к категории от серьезной угрозы исчезновения до критической угрозы исчезновения.

Согласно базе практик ревитализации, научного центра по сохранению, возрождению и документации языков России Института языкознания РАН, на 2024 год не зарегистрировано ни единой практики ревитализации юкагирского языка (ни южного, ни северного).

В настоящее время в качестве практик ревитализации не использованы технологии ИИ. Нами предпринят анализ имеющихся теоретических и практических разработок в данном контексте.

ИИ, имитирующий человеческий интеллект, имеет раздел машинного обучения. Обработка естественного языка – это специализированная область машинного обучения, которая фокусируется на возможности понимать, интерпретировать и генерировать человеческий язык.

Так, например, ИИ, созданный на модели «seq2seq» для обработки естественного языка, и использующий нейронную сеть глубокого обучения, способен обучаться на определенной выборке текстовых материалов. Несмотря на свои ограничения (только короткие диалоги и только на частых словах/фразах), по состоянию на 2018 год, широко применялся и достаточно хорошо отвечал на заданные вопросы, с подбором наиболее подходящих слов по вероятности (частоте) их связи в базовом материале, использованном для обучения ИИ [Goyal, 2018, 191–194]. В 2016 году «Google Translate» был улучшен с применением архитектуры «seq2seq» модели.

В 2022 году исследователи Массачусетского технологического института, Корнеллского университета и университета Макгилла, продемонстрировали систему ИИ, которая может самостоятельно выявлять правила и закономерности в человеческом языке. При наличии примеров с различными грамматическими формами (такими как, время, род, суффиксы), эта модель ИИ самостоятельно выявляет правила, по которым строятся грамматические формы. Стоит отметить, что описанная модель ИИ самостоятельно выявляет грамматические формы из небольшого объема вводных данных [Ellis, 2022, 1–4].

Подобная система ИИ, обученная исчезающему языку, как например, юкагирский язык, имеет широкие возможности в ревитализации юкагирского языка на разных уровнях во всех

категориях, принятых институтом языкознания РАН. На наш взгляд, эти возможности могут быть использованы в следующих направлениях:

- 1) В системе образования, ИИ способен оказать поддержку в изучении юкагирского языка как в рамках образовательной системы, так и вне ее, с использованием виртуальных языковых ассистентов, интерактивных платформ с обратной связью в режиме реального времени, а также мобильных приложений различного назначения, в том числе с применением систем виртуальной и дополненной реальности.

При этом преподавание юкагирского языка может быть как в виде отдельной дисциплины, так и по иным дисциплинам на юкагирском языке.

Стоит также отметить возможность ИИ в создании самоучителей, разговорников и аудиоматериалов (озвучивание текста с натуральным звучанием), что поможет сделать процесс обучения более доступным и увлекательным.

Примерами могут служить онлайн-переводчик «Яндекс Переводчик» и виртуальный ассистент «Яндекс Алиса», в которых применяется ИИ, но на данный момент эти приложения доступны только на распространенных языках мира.

- 2) Контент – это возможность получения информации разного рода на родном языке, как в виде текста, так и в виде аудио/речи.

ИИ со знанием юкагирского языка может переводить отдельные тексты, и целые книги, популярные интернет-блоги, известные сайты, российские или иностранные журналы и другие формы текстового формата как заранее, так и в режиме реального времени. Также возможно озвучивание и генерация видео с лицом и мимикой виртуального ассистента/учителя/друга.

Подобные ресурсы полезны для привлечения молодежи к изучению языка, что в свою очередь может способствовать повседневному использованию языка цифровой средой.

Примером может служить перевод видеотрансляций в режиме реального времени в Яндекс Браузере, с применением ИИ «YandexGPT», доступность языков ограничена. И чат-боты с ИИ, как например ChatGPT, YandexGPT (в составе «Яндекс Алиса»), могут генерировать тексты достаточно хорошего качества, особенно на те темы и сферы, в которых много исходного материала, даже если исходный материал на другом языке. Список доступных языков пока еще ограничен.

- 3) Культура и социум. Эта категория связана с функционированием языка в обществе и социуме. ИИ может быть использован для создания песен, стихов или прозы на юкагирском языке, создания развлекательных материалов, и проведении культурных мероприятий. Так, например, ИИ может помочь в создании мультимедийных материалов (изображений, аудио, видео) для фестивалей, театров, концертов, конкурсов, праздников, а также использоваться в спортивных играх и повседневных занятиях в семье.

Что может расширить контент в сферах, связанных с современными технологиями, наукой, искусством, креативными технологиями, туризмом и тд.

Примерами могут служить ИИ «DALL-E», «Midjourney», «Suno AI» и «Runway Gen3», которые могут генерировать изображения, тексты в виде изображений, музыку и видео.

- 4) Законодательство и документы. ИИ может способствовать переводу юридических документов на юкагирский язык. Полностью автоматизированный перевод может быть использован для публикации переводов документов в ознакомительных целях. Кроме того, ИИ может помочь в разработке систем документооборота для административных процессов, таких как судопроизводство и бюллетени на юкагирском языке.

- 5) Документация и кодификация. ИИ может способствовать сбору информации о языке, как в ходе полевых исследований и опросов, так и через оцифровку существующих

текстовых и аудио архивов, которые станут основой для дальнейших исследований и учебных программ.

ИИ также может содействовать разработке лексики и созданию словарей. Используя технологии обработки естественного языка, можно генерировать новые тексты и расширять сферы применения языка за пределами традиционного образа жизни и хозяйственной деятельности, таких как охота и рыбалка.

ИИ также может участвовать в процессе регламентации и стандартизации языка, предлагая решения для унификации правописания и создания альтернативных систем письменности, таких как латинизация или кириллизация. Это обеспечит дальнейшую кодификацию юкагирского языка и его диалектов, а также их интеграцию в цифровую среду.

Заключение

Юкагирский язык следует отнести к числу исчезающих языков из-за небольшого числа его носителей, их возраста и ограниченной письменности. В современных условиях использование систем ИИ предоставляет уникальные возможности для ревитализации таких исчезающих языков, как юкагирский.

Выявленные нами возможности ревитализации языков с применением технологий ИИ, обученного юкагирскому языку, нуждаются в практическом воплощении на разных уровнях (глубине) во всех категориях.

Библиография

1. Винклер М. А., Павлова О. М. Российские практики сохранения, возрождения и развития языков // Институт языкознания Российской академии наук URL: https://iling-ran.ru/languages_of_russia/2021_stage1/doc3.pdf (дата обращения: 30.10.2024).
2. Лукина М. П. «Исчезающие» юкагирские языки: к проблеме языкового сдвига. // Современная наука: актуальные проблемы теории и практики. Серия: Гуманитарные науки. - 2023. - №2. - С. 151.
3. Шадрин В. И. Юкагирский язык: социолингвистическая ситуация и общественные инициативы. Цифровизация языкового и культурного наследия коренных народов Арктики // Цифровизация языкового и культурного наследия коренных народов арктики: сборник статей по материалам Всероссийской научно-практической конференции. - Якутск: Северо-Восточный федеральный университет имени М. К. Аммосова, 2022. - С. 67–76.
4. Ellis K., Albright A., Solar-Lezama A., Synthesizing theories of human language with Bayesian program induction. // Nature communications. - 2022. - №13. - С. 5024.
5. Goyal P., Pandey S., Jain K. Deep Learning for Natural Language Processing: Creating Neural Networks with Python. - Berkeley: Apress, 2018. - 290 с.
6. Pereltsvaig A., Languages of the World. An introduction. - New York: Cambridge University Press, 2012. - 298 с.
7. UNESCO Ad hoc Expert Group on Endangered Languages. Language Vitality and Endangerment. - Paris: UNESCO, 2003. - 27 с.

The Potential of Artificial Intelligence in the Revitalization of Endangered Languages

Aleksei I. Petrov

Postgraduate Student,
Arctic State Institute of Culture and Arts,
677000, 4 Ordzhonikidze str., Yakutsk, Russian Federation;
e-mail: petrovai05@mail.ru

Aleksei I. Petrov

Abstract

The article explores the possibilities of using artificial intelligence (AI) for the revitalization of endangered languages, using the Yukaghir language as an example. It examines modern AI technologies in natural language processing for transmitting rare languages to new generations. The potential of AI in developing educational resources is analyzed, including virtual language assistants, interactive platforms with real-time feedback, self-study guides, phrasebooks, audio materials with natural sound created by AI. The possibility of increasing the accessibility of materials in the Yukaghir language through the use of AI for content expansion, such as translations of texts, books, websites, and the development of multimedia materials for theaters, concerts, and conferences, is assessed. Special attention is paid to the use of AI for preparing legal documents and machine translation for informational purposes. The listed applications can serve as a basis for further research, contribute to the development of vocabulary, and expand the scope of language use beyond traditional lifestyles and economic activities, opening up prospects for language revival.

For citation

Petrov A.I. (2024) Vozmozhnosti ispol'zovaniya iskusstvennogo intellekta v revitalizatsii ischezayushchikh yazykov [The potential of artificial intelligence in the revitalization of endangered languages]. *Kul'tura i tsivilizatsiya* [Culture and Civilization], 14 (11A), pp. 21-27.

Keywords

Ethnocultural heritage, language revitalization, Yukaghir language, artificial intelligence, digital linguistics, preservation of cultural identity, multimedia technologies.

References

1. Winkler M. E, Pavlova O. Russian practices of preservation, revival, and development of languages // Institute of Linguistics of the Russian Academy of Sciences URL: https://iling-ran.ru/languages_of_russia/2021_stage1/doc3.pdf (date of access: 10/30/2024).
2. Lukina M. "Endangered" yukaghir languages: to the problem of language shift. // *Modern Science: Current Problems of Theory and Practice*. Series: Humanities. - 2023. - №2. – C. 151.
3. Shadrin V. Yukaghir language: sociolinguistic situation and public initiatives. Digitalization of linguistic and cultural heritage of indigenous peoples of the Arctic // *Digitalization of the linguistic and cultural heritage of the indigenous peoples of the Arctic: a collection of articles based on the materials of the All-Russian scientific and practical conference*. - Yakutsk: North-Eastern Federal University named after M.K. Ammosov, 2022. – C. 67-76.
4. Ellis K., Albright A., Solar-Lezama A., Synthesizing theories of human language with Bayesian program induction. // *Nature communications*. - 2022. - №13. - C. 5024.
5. Goyal P., Pandey S., Jain K. *Deep Learning for Natural Language Processing: Creating Neural Networks with Python*. - Berkeley: Apress, 2018. - 290 c.
6. Pereltsvaig A., *Languages of the World. An introduction*. - New York: Cambridge University Press, 2012. - 298 c.
7. UNESCO Ad hoc Expert Group on Endangered Languages. *Language Vitality and Endangerment*. - Paris: UNESCO, 2003. – 27 c.