

**Медяникова Ольга Николаевна**

студент

руководитель Кривошеева Елена Игоревна

кандидат филологических наук, доцент

Тихоокеанский государственный

университет

г. Хабаровск, Россия

**Olga Medyannikova**

Student

under the guidance of

Krivosheeva Elena Igorevna

PhD in Philology, Associate Professor

Pacific National University

Khabarovsk, Russia

2021102242@pnu.edu.ru

## ИСПОЛЬЗОВАНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ РАБОТЫ С ЯПОНСКИМ ЯЗЫКОМ: ВЫЯВЛЕНИЕ ОШИБОК И ОЦЕНКА КАЧЕСТВА ПЕРЕВОДА

В данной статье тестируются и анализируются результаты способностей сервисов-переводчиков, таких как Яндекс Переводчик, Google.Translate и ChatGPT, использующих нейросетевой подход для перевода текста с японского на русский язык. В качестве материала для исследования привлечена статья из японской прессы с лингвокультурологическим наполнением, что призвано создать дополнительную сложность для тестируемых систем. Авторами оценивается качество перевода относительно выработанной шкалы критериев, выявляются достоинства и недостатки каждой из систем, а также прогнозируются будущие тенденции переводческой деятельности в паре человек – искусственный интеллект.

*Ключевые слова: машинный перевод; нейросети; нейронный машинный перевод; искусственный интеллект; ИИ; японский язык.*

## USING NEURAL NETWORKS TO WORK WITH THE JAPANESE LANGUAGE: IDENTIFYING MISTAKES AND EVALUATING THE QUALITY OF TRANSLATIONS

This article tests and analyzes the results of the abilities of translation services such as Yandex Translator, Google.Translate and ChatGPT, using a neural network approach to translate text from Japanese into Russian. An article from the Japanese press with linguistic and cultural content was used as material for the study, which is designed to create additional complexity for

the tested systems. The authors evaluate the quality of translation relative to the developed scale of criteria, identify the advantages and disadvantages of each of the systems, and predict future trends in translation activities in a human – artificial intelligence pair.

*Key words: Machine translation; neural networks; neural machine translation; artificial intelligence; AI, Japanese language.*

В современном мире информационные технологии оказывают значительное влияние на разные сферы жизни человека. Не является исключением и переводческая деятельность, которая, в попытке адаптироваться к новому технологическому ритму, внедряет передовые технологии и активно выстраивает новые алгоритмы работы с привлечением искусственного интеллекта (ИИ). Сегодня ИИ является лидирующим инструментом, широко применяющимся для анализа, понимания, извлечения информации и генерации текста на естественных языках. Это включает в себя задачи, такие как распознавание и синтез речи, семантический анализ, разрешение неоднозначности, классификация текстов и многое другое.

Обработка естественного языка достигла значительного прогресса, позволяя создавать более точные и эффективные модели для перевода и обработки текста. Тем не менее многие задачи в обработке естественного языка все еще остаются вызовом из-за сложности естественного языка и его контекстуальной природы [1, с. 40].

С целью анализа качества выполненного перевода с японского на русский язык, а также для определения достоинств и недостатков в работе ряда систем ИИ было принято решение осуществить перевод японской статьи о русской бане с интернет-ресурса GetNaviWeb, созданной при поддержке газеты Mainichi [2]. Данный материал представляется интересным для анализа в аспекте передачи русских реалий, а также для исследования возможностей ИИ при работе с текстовым материалом.

В качестве инструментов перевода выступают такие сервисы, как Яндекс Переводчик [3], Google.Translate [4] и ChatGPT [5].

Для начала обратимся к заголовку статьи, где для обозначения русской реалии «баня» использовано заимствованное из немецкого языка слово *サウナ* (сауна), не имеющее отношения к русской культуре и не совпадающее по коннотативному значению. Так, отличия сауны от бани заключаются в следующем [6]:

1. температура (сауна: 100 градусов и выше; баня: около 60 градусов на нижних полках и около 100 градусов на верхних);

2. влажность (сауна: в пределах 20 %, сухой воздух; баня: воздух менее сухой);

3. камни для топки помещения (сауна: камни находятся в открытом состоянии; баня: в печи с закрытой дверцей);

4. атрибутика (сауна: березовый или иной веник пересыхает и становится бесполезным; баня: из-за влажности и температуры веник остается в рабочем состоянии).

В японском языке есть заимствованное из русского языка слово – バニヤ (баня), но ИИ при переводе «уравнил» サウナ (сауна) и バニヤ (баня), что является недостатком с лингвокультурологической точки зрения.

Далее определим критерии соответствия, относительно которых будет производиться оценка качества выполненного перевода с привлечением указанных выше программ ИИ:

- 1) смысловое: ошибки, искажающие суть предложения;
- 2) лингвистическое: лексические, орфографические, грамматические, пунктуационные ошибки; опечатки;
- 3) терминологическое: несоответствие терминологии;
- 4) стилистическое: использование разных функциональных стилей; непоследовательность в тексте; искажение экспрессивного фона оригинала и авторской оценки;
- 5) локализационное: ошибки, связанные с переводом реалий.

Текст статьи с помощью указанных сервисов был переведен полностью, не фрагментарно. Обратимся к полученным результатам.

Смысловые ошибки. Пример 1: ロシアのサウナは湿度が高く「蒸し風呂」とも呼ばれ、都市部ではアパートに「一人用サウナ」がついているほど。(росиа но сауна ва ондо га такаку «мусибуро» томо ёбарэ, тосибу дэва апа:то ни «хитори ё: сауна» га цуитэ иру ходо.)

Google.Translate: ‘Русские сауны еще называют «паровыми банями» из-за их высокой влажности, а в городских условиях – нет’.

Указанный онлайн-переводчик искажает смысл предложения до такой степени, что основная мысль статьи теряется. Мы предлагаем следующий перевод: ‘Русские бани еще именуется “парными” из-за высокой влажности, кроме этого в городских условиях размещают так называемые «индивидуальные сауны», то есть специальные бани-кабинки’. При сравнении вариантов заметна существенная потеря смысла, потому что сервис проигнорировал вторую часть предложения. Помимо этого, видится необходимость объяснить, что конкретно имел в виду автор под термином «индивидуальные сауны».

Лингвистические ошибки. Пример 1: 今回はこのロシア式サウナの特徴と、モスクワっ子が憧れるサウナの聖地「サンドゥニ」をご紹介します。(конкай ва коно росиасики сауна но токутё: то, Мосукувакко га акогарэру сауна но сэйти «сандуни» о госё:кай симасу.)

Яндекс Переводчик: ‘На этот раз мы расскажем об особенностях настоящей русской бани и о святом месте –7 сауне “Сандуни”, которой восхищаются московские дети’.

モスクワっ子 (мосукувакко) – «(коренной) москвич» [7]. Слово состоит из названия города Москва – モスクワ (мосукува) и иероглифа ‘ребенок’ 子 (ко) с удвоением, что в комбинации означает ‘человек, родившийся и выросший в Москве’. ‘Московский ребенок’ на японском записывается иначе – モスクワの子 (мосукува но ко). Таким образом, сервис неправильно определил значение слова, основываясь лишь на его составляющих.

Пример 2: *その理由にはロシアの独特な入浴スタイルがあり、休憩室のようなスペースで、食事やお茶をつまみながら、休憩スペースとサウナを行き来するのです。*

(соно рию: ни ва росиа но докутоку на ню:ёку сутаиру га ари, кю:кэйсицу но ё:на супэ:су дэ, сё:кудзи я отя о цумаминагара, кю:кэй супэ:су то сауна о юкики суру но дэсу.)

Google.Translate: ‘Причиной этого является уникальный в России стиль купания, когда вы перемещаетесь между зоной отдыха и сауной, одновременно едя или пью чай в пространстве, похожем на комнату отдыха’.

В русском языке считается ненормативным употребление деепричастия несовершенного вида с такими глаголами, как, как «есть» и «пить». Для перевода предложения следует заменить выбранную грамматическую конструкцию на другую или подобрать подходящие по смыслу синонимы. Например, Яндекс Переводчик выдал такой результат: ‘наслаждаясь едой или чашечкой чая’.

Терминологические ошибки. Пример: *後述する高級サウナの「サンドウニ」では、「叩き師」なる方たちが白樺で利用者の身体を万遍なく叩くそうですが、その加減がとても上手で、まるでマッサージのようだと評判。* (ко:дзюцу суру ко:кю: сауна но «сандуни» дэ ва, «татакиси» нару кататати га сиракаба га риё:ся но карада о манбэн наку татаку со: дэсуга, соно кагэн га тотэмэ дзё:дзу дэ, марудэ массадзи но ё: да то хё:бон.)

Яндекс Переводчик: ‘В «Сандуни», сауне высокого класса, о которой мы расскажем ниже, кажется, что люди, которые становятся «хлопушками», ударяют по телу пациента березовым веником миллион раз, но регулировка очень хорошая, и считается, что это похоже на массаж’.

Яндекс Переводчик, Google.Translate и ChatGPT перевели слово 叩き師 (татаки) как ‘люди-хлопушки’, ‘колотушки’, ‘ударники’ соответственно, однако эта лексическая единица означает ‘банщик’, то есть человек, профессионально отбивающий части тела людей веником из разных пород дерева во время банных процедур.

Стилистические ошибки. Пример: *その理由にはロシアの独特な入浴スタイルがあり、休憩室のようなスペースで、食事やお茶をつまみながら、休憩スペースとサウナを行き来するのです。*

(соно рию: ни ва росиа но докутоку на ню:ёку сутаиру га ари, кю:кэйсицу но ё:на супэ:су дэ, сё:кудзи я отя о цумаминагара, кю:кэй супэ:су то сауна о юкики суру но дэсу.)

ChatGPT: ‘Русская особенность – специфика принятия ванны, когда люди ходят между отдыхальной, где перекусывают и пьют чай, и сауной’.

Ранее оригинал этого предложения был рассмотрен в группе лингвистических ошибок. Тем не менее уже другой сервис допустил неточности при переводе, так как термина «отдыхальня» не зафиксировано ни в одном

русском толковом словаре. Предположительно, это слово является сочетанием высокопарного существительного «опочивальня» и глагола «отдыхать» и является некорректным.

Локализация. Пример 1: 憧れの聖地「サンドゥニ」とは?

(акогарэ но сэйти «сандуни» то ва?)

Google.Translate: ‘Что представляет собой заветное священное место «Санденис»?’

サンドゥニ (сандуни), то есть наименование места, сервисы по большей мере переводили с помощью транслитерации с японского варианта, однако в некоторых ситуациях название бани меняло форму перевода по непонятным причинам, как, например, в этом подзаголовке от Google.Translate. Тем не менее даже вариант «сандуни» не является правильным. サンドゥニ (сандуни) – это сандуновская баня или баня «Сандуны», московская публичная баня.

Яндекс Переводчик, Google.Translate и ChatGPT не справились с переводом целой статьи, допуская ошибки по всем критериям оценивания. При точечном выборе предложений для перевода результаты в несколько раз лучше, но не идеальны. Таким образом, перевод, предложенный искусственным интеллектом, требует постредактирования со стороны человека по всем аспектам лингвистики (синтаксическому, лексическому, стилевому и т.д.).

Важно осознавать, что на данном этапе нейросети – это инструмент переводчика, его помощник, но не его конкурент. Преимущества применения искусственного интеллекта при переводе включают высокую скорость и создание общего понимания текста. Тем не менее при работе с текстовым материалом допускаются грубые ошибки или неточности. В таком случае специалист проводит критический анализ результата и выполняет постредактирование текста. Кроме того, выявлена прямая взаимосвязь между объемом информации и качеством перевода: чем более узконаправлен и объемнее текст, тем больше ошибок.

## ЛИТЕРАТУРА

1. Бурнашев Р. Ф., Анварова Л. А. Применение нейронных сетей в автоматическом переводе и обработке естественного языка // Universum: технические науки. 2024. № 4 (121). С. 39–43.

2. 裸の付き合いは国境を越える! 「サウナ」から見るロシア人と日本人の共通点 (GetNavi web) [Electronic resource] // Mainichi. URL: <https://mainichi.jp/articles/20180526/gnw/00m/040/000000c> (accessed: 16.06.2024).

3. Яндекс Переводчик [Электронный ресурс]. URL: <https://translate.yandex.ru/> (дата обращения: 17.06.2024).

4. Google Translate [Electronic resource]. URL: <https://translate.google.ru/> (accessed: 17.06.2024).

5. ChatGPT [Electronic resource]. URL: <https://chat-gpt.org/chat> (accessed: 17.06.2024).

6. Разница между баней и сауной [Электронный ресурс] // TheDifference.ru. Об отличиях. URL: <https://thedifference.ru/otliche-bani-ot-sauny/#:~:text=Воздух%20в%20саунах%2C%20в%20отличие,в%20парной%20выше%20ста%20градусов> (дата обращения: 18.06.2024).

7. goo辞書 [Electronic resource]. URL: <https://dictionary.goo.ne.jp/> (accessed: 11.06.2024).