

СОВРЕМЕННЫЕ ПРОБЛЕМЫ И РЕШЕНИЯ В ОБЛАСТИ МАШИННОГО ПЕРЕВОДА

Я. Д. Бакуменко^а, А. Н. Таджибова^б

Сургутский государственный университет, г. Сургут, Российская Федерация

^а ORCID: <http://orcid.org/0009-0003-2372-5397>, ✉ bakumenko_yad@edu.surgu.ru

^б ORCID: <http://orcid.org/0009-0006-3239-6902>, tadzhibova_an@surgu.ru

Аннотация: машинный перевод играет ключевую роль в современном обществе, обеспечивая доступ к информации на различных языках в условиях глобализации. В данной статье проводится комплексный анализ проблем и достижений в области машинного перевода, с акцентом на использование передовых технологий, таких как нейронные сети и трансформеры, для повышения точности перевода. Рассматриваются основные метрики качества перевода, включая BLEU, METEOR и POS-BLEU, и их применение к различным системам перевода, таким как DeepL, Яндекс и Google Translate. Основное внимание уделяется сравнению эффективности этих систем с точки зрения лексического соответствия, семантической точности и синтаксической структуры. Для более глубокой оценки перевода была разработана специализированная программа на Python, которая проводит автоматический анализ перевода на основе вышеуказанных метрик. Программа выявляет слабые места в переводах и генерирует отчеты для дальнейшей оптимизации. По данным исследования, DeepL и Яндекс демонстрируют высокие результаты по метрике BLEU, что указывает на хорошее лексическое соответствие и точность синтаксической структуры. Однако все три системы сталкиваются с трудностями в области синтаксиса, лексики и использования артиклей, что требует дальнейших улучшений. В работе подчеркивается важность многогранной оценки качества перевода и необходимости интеграции машинного анализа с экспертным контролем. Также рассматривается роль этих технологий в специализированных областях, где точность перевода имеет критическое значение. Авторы отмечают необходимость дальнейших исследований в области машинного перевода для улучшения качества перевода и его адаптации к различным языковым и культурным контекстам.

Ключевые слова: машинный перевод, нейронные сети, трансформеры, многоязычные модели, оценка качества, BLEU, METEOR, POS-BLEU.

Для цитирования: Бакуменко Я. Д., Таджибова А. Н. Современные проблемы и решения в области машинного перевода. *Успехи кибернетики*. 2025;6(2):47–59.

Поступила в редакцию: 14.04.2025.

В окончательном варианте: 09.06.2025.

CURRENT CHALLENGES AND APPROACHES IN MACHINE TRANSLATION

Ya. D. Bakumenko^a, A. N. Tadzhibova^b

Surgut State University, Surgut, Russian Federation

^a ORCID: <http://orcid.org/0009-0003-2372-5397>, ✉ bakumenko_yad@edu.surgu.ru

^b ORCID: <http://orcid.org/0009-0006-3239-6902>, tadzhibova_an@surgu.ru

Abstract: we conducted a comprehensive analysis of the current challenges and recent advancements in machine translation (MT), which plays a vital role in enabling multilingual communication in today's globalized world. We focused on the application of advanced technologies – neural networks and transformer architectures – to improve translation accuracy. We evaluated leading MT systems (DeepL, Yandex Translate, and Google Translate) using standard quality metrics: BLEU, METEOR, and POS-BLEU. Our evaluation emphasized lexical matching, semantic fidelity, and syntactic structure. We developed a Python-based tool that automatically assessed translation quality using these metrics. The program identified weaknesses in translations and generated detailed reports to guide optimization. Our results showed that DeepL and Yandex achieved higher BLEU scores, indicating better performance in lexical and syntactic accuracy. However, all three systems exhibited issues with syntax, vocabulary choices, and article usage, highlighting areas that require further development.

We emphasize the importance of multi-dimensional evaluation – combining automated metrics with expert assessment. We also examined MT performance in specialized domains, where high translation precision is critical. Our findings underscore the need for continued research to further improve MT systems and adapt them to diverse linguistic and cultural contexts.

Keywords: machine translation, neural networks, transformers, multilingual models, quality evaluation, BLEU, METEOR, POS-BLEU.

Cite this article: Bakumenko Ya. D., Tadzhibova A. N. Current Challenges and Approaches in Machine Translation. *Russian Journal of Cybernetics*. 2025;6(2):47–59.

Original article submitted: 14.04.2025.

Revision submitted: 09.06.2025.

Введение

Машинный перевод (МП) является необходимым в условиях современной глобализации и цифровизации, когда потребность в доступе к информации на различных языках стала неотъемлемой частью нашей жизни.

Термин *машинный перевод* понимается, по крайней мере, в двух смыслах. В узком смысле — это процесс перевода текста с одного языка на другой с использованием компьютера, реализуемый полностью или почти полностью. Машинный перевод в широком смысле — это область научных исследований, находящаяся на стыке лингвистики, математики, кибернетики, и имеющая целью построение систем, реализующих машинный перевод в узком смысле [1, с. 3].

Машинный перевод является автоматизированным процессом перевода текста с одного языка на другой с использованием алгоритмов. Идея его создания появилась в 1947 году, а первая демонстрация системы состоялась в 1954 году в штаб-квартире IBM. Основная задача машинного перевода — воспроизвести содержание текста с учетом лингвистических и культурных особенностей. Однако, в отличие от человека, компьютерные системы не способны полноценно понимать смысл и контекст, а лишь следуют заложенным алгоритмам и словарям. Это создает ряд трудностей, особенно при передаче оттенков значений, интонаций и сложных грамматических конструкций. Для эффективной работы машинного перевода требуются точные языковые правила и универсальные модели, что делает его разработку междисциплинарной задачей. Немаловажным аспектом является адаптация лингвистической терминологии и формализация языковых единиц, особенно для языков с богатой морфологией и сложным синтаксисом.

Цель исследования

Целью данного исследования является анализ проблем и решений в области машинного перевода, включая использование нейронных сетей для повышения точности перевода, а также учет культурных реалий в процессе перевода. Особое внимание уделено сравнению популярных систем машинного перевода (DeepL, Яндекс, Google Translate) и нейронной сети GPT, с акцентом на использование метрик качества перевода, таких как BLEU, METEOR и POS-BLEU.

Материалы и методы исследования

В процессе исследования использовались эмпирические данные и теоретические исследования в области машинного перевода, включая работы, посвященные нейронным сетям, трансформерам, многоязычным моделям и методам оценки качества перевода, BLEU, METEOR и другим популярным метрикам. Также были проанализированы современные системы машинного перевода, такие как Google Translate, DeepL и Яндекс.

Для более глубокой оценки качества перевода нами была разработана программа на языке Python, которая проводит автоматический анализ с использованием вышеуказанных метрик. Программа оценивает каждое предложение по метрикам BLEU, METEOR и POS-BLEU, позволяет выявить проблемные зоны в переводах и генерирует отчет, который используется для дальнейшей оптимизации и улучшения качества перевода.

Перевод — это не просто преобразование слов и предложений на другом языке, но также и учет контекста, многозначности отдельных слов. Многозначные слова особенно затрудняют задачу: одно и то же слово может означать разные вещи в зависимости от ситуации. Например, слово «ключ» в русском языке может обозначать как предмет для открытия дверей, так и источник воды. Если система не распознает правильное значение с учетом контекста, перевод может оказаться ошибочным. Аналогичные сложности возникают и с анафорическими связями, такими, как местоимения. В предложении «Иван пришел в дом, и он был пуст» местоимение «он» должно относиться к «дому», а не к «Ивану», что требует от системы способности анализировать контекст и структуру текста. Для решения этих проблем разрабатываются правила, касающиеся фразеологического анализа и распознавания контекста.

Проблемы, связанные с контекстуальной неоднозначностью, требуют внедрения и более сложных методов. Одним из таких методов является обучение на обширных объемах данных, что позволяет системе «видеть» множество примеров использования слов в различных контекстах и правильно определять их значения в зависимости от ситуации. Для этого создаются и анализируются большие текстовые корпуса, охватывающие разные стили и темы, чтобы помочь системе точнее передавать смысл фраз. Трансформеры, такие как модели BERT и GPT, значительно улучшают понимание контекста и помогают сохранять точность смысловой передачи. Данные модели анализируют текст комплексно, одновременно обрабатывая все компоненты предложения, следовательно, это дает возможность более глубоко «понимать» значения слов в контексте всего текста. Модели обучаются на параллельных текстах и затем дообучаются для более точного перевода.

Недавние эмпирические улучшения, достигнутые благодаря трансферному обучению с использованием языковых моделей, продемонстрировали, что предварительное обучение языковой модели эффективно для улучшения многих задач обработки естественного языка. К ним относятся задачи на уровне предложений, такие как логический вывод на естественном языке и перефразирование, которые направлены на предсказание взаимосвязей между предложениями путем их целостного анализа, а также задачи на уровне токенов, например, распознавание именованных объектов и ответы на вопросы, где требуются модели для получения детализированной информации [2].

Для того чтобы адаптировать перевод для самых разнообразных текстов, создаются специализированные подкорпусы данных, предназначенные для конкретных областей, что позволяет более точно учитывать специфическую терминологию и стиль. Различия в синтаксисе и морфологии между языками также требуют крайне точной настройки систем; система должна учитывать порядок слов и правила изменения окончаний, которые особенно важны для языков с развитой грамматикой.

Значительного прогресса в областях обработки языка удалось добиться на более низких уровнях обработки речи: для письменной формы языка к ним относятся сегментация текста, лексический, морфосинтаксический и синтаксический анализ; для обработки устной речи — распознавание речи, преобразование текста в речь и распознавание говорящего. В результате многие приложения теперь вошли в наше повседневное использование. В том, что касается обработки письменной речи — это проверка орфографии и грамматики, одноязычные межъязыковые поисковые системы, машинные онлайн-переводчики, в области обработки устной речи стоит отметить голосовые системы GPS, системы диктовки, транскрипцию и автоматическую индексацию аудиовизуального контента. Данный список показывает, что многие из этих приложений объединяют устную и письменную формы языка (транскрипция речевого сообщения в текст, перевод текста в речь) [2, с. 155].

Помимо многозначности, одной из сложных проблем в области машинного перевода является адекватный перевод фразеологизмов. Фразеологизмы представляют собой устойчивые лексические сочетания, значение которых определяется целым выражением, а не отдельными словами. Так, английские выражения, такие как «it's high time» (давно пора), «take your time» (не торопись) или «help yourself» (угощайся), не имеют буквального эквивалента в других языках, но их смысл можно передать соответствующими выражениями. С точки зрения смысловой связанности компоненты фразеологизмов можно разделить на образные и необразные. Необразные фразеологизмы сохраняют значение составляющих слов, но их элементы нельзя менять, как, например, в словосочетаниях «играть роль» и «иметь значение». Такие фразеологизмы чаще встречаются в научной и деловой речи.

Образные фразеологизмы, напротив, имеют значение, которое не является суммой значений отдельных слов. В языке встречаются такие примеры, как «погнаться за двумя зайцами», «попасть как кур в ошип» или «лиха беда начало» на русском языке и «through thick and thin» (во что бы то ни стало), «tooth and nail» (не жалея сил, засучив рукава), «hit the nail on the head» (попасть в точку) на английском.

Сложность машинного перевода фразеологизмов заключается в том, что их компоненты теряют свой буквальный смысл и их необходимо переводить не дословно, а с учетом культурных и языковых особенностей. В зависимости от ситуации для перевода фразеологизмов переводчик может использовать различные методы: фразеологический эквивалент, фразеологический аналог, калькирование или описательный перевод.

Машинные системы часто переводят такие фразы неверно, что приводит к искажению смысла. Чтобы решить эту проблему, современные технологии, основанные на нейронных сетях, обучаются

на больших объемах данных. Модели анализируют примеры перевода идиом в различных контекстах, чтобы подбирать подходящие аналоги на другом языке. Алгоритмы BERT и GPT способны учитывать контекст и заменять, например, «a piece of cake» на «проще простого» в русском языке. Это помогает сохранить точное значение выражений.

Следует отметить, что на 95% языков говорит только 6% населения мира. Некоторые лингвисты считают, что 90% языков исчезнут в течение столетия. Таким образом, по наличию данных о языке и автоматических систем его обработки можно выделить языки с большим или меньшим объемом ресурсов или полным их отсутствием, а также устные языки без системы письма. Наличие данных имеет решающее значение для развития используемых систем, часто основанных на статистическом подходе. Системы перевода не способны поддерживать такие языки на уровне более популярных, таких как английский или китайский. Например, языки с развитой морфологией, такие как финский или венгерский, имеют сложную структуру слова, включающую многочисленные суффиксы, падежи и склонения. Похожая проблема возникает и с диалектами, где словарный состав и синтаксические особенности могут существенно отличаться даже в пределах одного языка.

В этой связи машинный перевод требует существования параллельных корпусов, однако их число ограничено. Поэтому мы стараемся восполнить данный пробел путем разработки методов, использующих «шумные» параллельные корпуса текстов, сопоставимые корпуса (тексты одной тематики на разных языках) или квазисопоставимые корпуса, которые являются более доступными, особенно благодаря распространению Интернета [2, с. 161].

Facebook AI создала модель M2M-100, которая обучается сразу на параллельных текстах многих языков. Благодаря этому подходу она умеет работать с редкими языками, используя общие лингвистические закономерности. Например, опыт перевода между испанским и португальским помогает точнее переводить на каталонский, поскольку эти языки похожи по структуре. В отличие от традиционных систем, ориентированных на переводы через английский, M2M-100 сразу переводит между 100 языками и поддерживает 9 900 направлений. Для этого она группирует языки по близости и применяет «случайную перенаправку» параметров внутри каждой группы. Модель обучалась на большом корпусе CommonCrawl, что позволило сохранить точность даже там, где не было пар с английским языком. В итоге прямой перевод между неанглоязычными парами показал более высокие оценки BLEU по сравнению с двуязычными моделями, использующими данные WMT систем и работающими через английский.

Помимо этого, в M2M-100 внедрен метод переноса обучения: сначала модель учится на одном, «большом» языке, а затем дообучается на небольшом наборе данных редкого языка. После обучения на немецком ей проще адаптироваться к люксембургскому. Также применяется активное обучение: модель выбирает для дообучения сложные фразы, которые затем переводят носители языка или эксперты. Эти проверки постепенно наращивают корпус, что повышает качество перевода для малораспространенных языков и диалектов, поддерживая языковое разнообразие.

Следует также подчеркнуть, что в секторе информационных технологий значительные усилия в сфере развития многоязычия предпринимают и другие крупные американские компании. Одной из ведущих компаний в этом направлении является Google, которая значительно расширила количество поддерживаемых языков и улучшила качество своих переводческих сервисов. По состоянию на июнь 2024 года Google Translate поддерживает 243 языка, что на 98 языков больше, чем было доступно ранее. Этот сервис продолжает активно развиваться, добавляя новые языки и улучшая качество перевода. В 2024 году было добавлено 110 новых языков, включая редкие и исчезающие языки — кантонский, манкс, н'ко, панджаби (шахмуки), тамазигт (амазиг) и ток-писин. Благодаря этим обновлениям Google расширяет охват более чем на 614 миллионов человек по всему миру. Google активно работает и над сохранением исчезающих языков, таких как диула, волоф, бауле и других. Кроме того, Google использует новейшие модели PaLM 2, которые позволяют эффективно обучать нейросети для перевода между тесно связанными языками. Модели PaLM 2 также улучшили работу с региональными языками — башкирским, чеченским, осетинским и крымскотатарским.

Компания Microsoft также активно развивает свои языковые технологии. На июнь 2025 года Microsoft Translator поддерживает 179 языков и диалектов, включая 21 вариант испанского языка и 16 вариантов арабского. Важно отметить, что компания продолжает улучшать свои инструменты, например, с Microsoft Translator Pro, который был выпущен в ноябре 2024 года. Этот сервис теперь включает

точный перевод речи в реальном времени с поддержкой транскрипции на одном языке, улучшенной многозадачности и шумоподавления, что значительно расширяет возможности перевода в условиях динамичных ситуаций. Microsoft также предлагает поддержку офлайн-переводов в своих приложениях для Android и iOS, улучшая доступность для пользователей с ограниченным Интернет-доступом. Помимо этого, Microsoft интегрировала свои языковые инструменты в популярные офисные программы Microsoft 365, где доступна проверка орфографии и грамматики на 126 языках, включая учет региональных вариантов на 233 и 61 языках для орфографии и грамматики соответственно. Это демонстрирует стремление компании сделать свои инструменты доступными для широкой аудитории пользователей по всему миру.

Технические ограничения также играют значительную роль в определении качества перевода. Сегодня существуют различные типы систем машинного перевода, включая полностью автоматизированные, машинно-ассистируемые переводы (МАНТ и НАМТ) и системы переводческой памяти (Translation Memory, TM). В зависимости от роли человека в процессе перевода системы могут по-разному справляться с задачей, однако в каждом случае на точность влияет ряд факторов. Например, системы TM, которые широко используются профессиональными переводчиками (DL Trados Studio, MemoQ, Wordfast, OmegaT, Across и др.), позволяют сохранять уже переведенные фразы или части текста, чтобы предложить их для повторного использования. Это снижает трудоемкость перевода и улучшает качество за счет использования проверенных переводов.

С переходом к нейронному машинному переводу (NMT) технологии стали учитывать более широкий контекст. Первая волна NMT базировалась на архитектуре «энкодер–декодер» с рекуррентными нейронными сетями, однако с развитием Transformer-модели в 2017 году качество перевода значительно возросло благодаря механизму внимания. В отличие от более старых систем машинного перевода, NMT устранил многие недостатки предыдущих подходов, а именно проблемы со связанностью текста. Однако, несмотря на значительный прогресс, полностью автоматизированные системы (например, Google Translate) и другие сервисы все еще сталкиваются с рядом ограничений: трудности при переводе специализированных текстов, дефицит параллельных данных и проблемы с доменным рассогласованием (ситуацией, при которой тематическая, жанровая или стилевая область “domain” текстов, на которых обучалась модель машинного перевода, не совпадает с областью входных или целевых текстов, которые системе приходится переводить на практике).

В процессе совершенствования системы машинного перевода компания Яндекс применяет несколько ключевых методов, направленных на повышение точности перевода, особенно с русского языка. Одним из таких методов является использование морфологических трансформаций. Основное преимущество этого подхода заключается в том, что он помогает преобразовать русские предложения в более удобные для перевода формы. Как известно, русский язык обладает богатой морфологией, и это часто является причиной трудностей в переводе. Применение морфологических трансформаций решает проблему изменяющихся окончаний и форм слов, что, в свою очередь, существенно улучшает точность перевода.

Кроме того, следует отметить, что Яндекс использует модели последовательностей операций (Operation Sequence Model). Этот подход интегрирует операции переноса и лексические переводы в последовательности минимальных единиц перевода. Такой метод упрощает процесс перевода на уровне синтаксиса и порядка слов, что особенно важно при переводе между языками с различной грамматической структурой. Обратим внимание на то, что этот подход позволяет решить одну из основных проблем машинного перевода, связанную с различиями в грамматических системах. В дополнение к вышеупомянутым методам Яндекс использует подход трансфем (transfemes), который преобразует слова в более структурированные единицы перевода. Это, безусловно, помогает снизить сложность перевода между языками с разными грамматическими и синтаксическими особенностями. Применение этих методов значительно улучшает точность перевода, что подтверждается результатами экспериментов, проведенных на тестах WMT13 и WMT14. Результаты этих тестов показали повышение показателей BLEU.

Не менее важным инструментом в области машинного перевода является переводчик DeepL от одноименной немецкой компании, специализирующейся на разработке технологий машинного перевода с использованием искусственного интеллекта. В исследовании Линлина Ли 2024 г. был проведен сравнительный анализ качества перевода с использованием DeepL, Google Translate и Microsoft

Translator. Оценка проводилась по таким параметрам, как точность, плавность и естественность, с применением стратегии «расчета семантического сходства».

Одним из важных выводов исследования стало то, что DeepL продемонстрировал 94% точности перевода, что значительно превосходит результаты Google Translate (86%). В исследовании использовались различные текстовые корпуса и точность перевода была оценена в общем контексте, без уточнения конкретной языковой пары. В части плавности перевода DeepL показал более высокий результат, достигнув 146% в изменении длины предложений. DeepL также оказался лидером среди трех сервисов, особенно в отношении культурной адаптации (92%) и использования идиоматических выражений (94%) [18, с. 715–716].

Процесс работы DeepL включает несколько стадий. На первом этапе текст разделяется на категории, такие как профессиональная терминология, общий словарный запас и простые предложения. Это позволяет точнее выбирать и обрабатывать слова, что, в свою очередь, улучшает качество перевода. На следующем этапе DeepL использует технологии тегирования частей речи и синтаксического анализа для грамматической корректности перевода. Как правило, ошибки на этом этапе могут привести к некорректному переводу, поэтому важно обеспечить правильную обработку синтаксиса и контекста. Для оценки качества перевода DeepL использует методику, основанную на расчете семантического сходства. Это позволяет объективно сравнить перевод с оригиналом и другими переводами. В результате система демонстрирует превосходство в точности перевода, особенно для более стилистически сложных текстов. Однако исследование также показало, что Google Translate имеет небольшое преимущество в адаптации к контексту и культурным особенностям. Это открывает возможности для дальнейших улучшений в области машинного перевода.

Перспективы развития машинного перевода тесно связаны с нейронными сетями, а именно с трансформерами seq2seq, BERT, T5 и GPT. Эти модели позволяют значительно точнее моделировать контекст, справляясь с трудными синтаксическими структурами. Языковые модели, такие как GPT, показывают высокий уровень генерации текста, но, несмотря на это, они все еще сталкиваются с проблемами точности при интерпретации сложных структур. Это, в свою очередь, требует дальнейших улучшений, таких как fine-tuning для специфических задач и Zero-Shot Learning, который позволит моделям обучаться на новых языках с минимальными данными.

Будущее машинного перевода заключается в развитии гибридных и интерактивных подходов, которые объединяют нейронные и статистические методы для повышения точности. Интерактивные системы, где пользователи могут вносить коррективы в перевод, позволяют адаптировать его под специфические задачи, особенно в таких областях, как юриспруденция или медицина, где требуется высокая точность. В дополнение к этому ожидается развитие мультиформатных моделей, которые объединяют текст и изображения, а также адаптируются под контекст пользователя. Это откроет новые возможности для профессиональных приложений, что, безусловно, повысит качество и доступность машинного перевода. Более того, для отслеживания прогресса в сфере машинного перевода используются различные системы оценки, такие как BLEU (Bilingual Evaluation Understudy) и METEOR (Metric for Evaluation of Translation with Explicit ORdering). Метрика BLEU, одна из первых автоматизированных метрик, измеряет совпадение между машинным переводом и эталонными переводами на основе вычисления n-грамм (последовательностей слов). Чем больше таких совпадений, тем выше оценка, что указывает на схожесть с человеческим переводом. METEOR предлагает более гибкий подход, принимая во внимание лексические варианты — синонимы и морфологические изменения слов. Она включает этапы выравнивания слов, подсчет совпадений по смыслу, а также штрафы за отклонение от порядка слов, что делает ее более чувствительной к качеству перевода.

Современные исследователи работают над усовершенствованием этих метрик, добавляя учет синтаксических и семантических аспектов, чтобы они могли оценивать точность и естественность перевода более комплексно. В дополнение к традиционным метрикам оценки качества машинного перевода, которые фокусируются на лексическом и частично семантическом соответствии между эталонным и машинным переводами, все большее внимание уделяется метрикам, основанным на синтаксическом анализе. Такие метрики позволяют оценить, насколько хорошо машинный перевод сохраняет грамматическую структуру исходного текста. Они анализируют соответствие синтаксических деревьев перевода и референтного текста, выявляя отклонения и ошибки. Метрика POS tagging анализирует последовательность частей речи (POS), используемых в машинном переводе, и сравнивает ее с референт-

ным текстом. Она оценивает, насколько корректно соблюдены грамматические правила языка перевода. Эти улучшения позволяют моделям машинного перевода ориентироваться на более высокий уровень адекватности и плавности перевода.

Результаты исследования и их обсуждение

В рамках данного исследования была разработана программа на языке Python. Программа предназначена для автоматической оценки качества машинного перевода с использованием трех популярных метрик — BLEU, METEOR и POS-BLEU. Она использует несколько библиотек Python, включая nltk, Stanza и docx, чтобы провести анализ перевода. Эта программа анализирует перевод с точки зрения лексической точности, семантического соответствия и синтаксической структуры, что делает ее полезной для оценки и улучшения качества машинного перевода.

Программа начинает с чтения текстов перевода, которые необходимо оценить. Эталонный перевод и машинный перевод загружаются из текстовых файлов с помощью функции `read_translation_file`. Эта функция открывает файл и читает его содержимое, затем с помощью токенизатора из библиотеки nltk разбивает текст на предложения:

```
def read_translation_file(file_path):
    with open(file_path, 'r', encoding='utf-8') as file:
        text = file.read().strip()
    tokenizer = nltk.data.load('tokenizers/punkt/russian.pickle')
    sentences = tokenizer.tokenize(text)
    return sentences
```

Библиотека nltk применяется для токенизации текста, что позволяет делить текст на отдельные предложения. Токенизатор используется для обработки русского языка.

После токенизации текста каждое предложение разбивается на слова с использованием стандартного метода `.split()` в Python. Это необходимо для последующего анализа на уровне слов и фраз, что позволяет метрикам BLEU и METEOR корректно вычислять оценки:

```
def tokenize_and_split(sentences):
    return [sentence.split() for sentence in sentences]
```

POS tagging (частеречная разметка)

Для анализа синтаксической структуры перевода используется библиотека Stanza, которая выполняет POS-тегирование. В процессе этого анализа каждому слову в тексте присваивается соответствующая часть речи (например, существительное, глагол, прилагательное):

```
import stanza
stanza.download('ru') # Загрузка модели для русского языка
nlp = stanza.Pipeline('ru', processors='tokenize,pos', tokenize_pretokenized=False)
def pos_tag_sentences(sentences, nlp):
    tagged_sentences = []
    for sentence in sentences:
        doc = nlp(sentence)
        tags = [word.upos for sent in doc.sentences for word in sent.words]
        tagged_sentences.append(tags)
    return tagged_sentences
```

Модель `stanza.Pipeline` используется для обработки текста на русском языке и возвращает последовательности тегов частей речи, что необходимо для метрики POS-BLEU. После подготовки данных программа оценивает каждый перевод с использованием трех метрик.

BLEU оценивает лексическое совпадение между эталонным переводом и машинным переводом на основе n-грамм. Метрика BLEU используется для оценки точности перевода:

```
bleu_score = corpus_bleu(reference_corpus, machine_translation)
```

Для каждого предложения также вычисляется BLEU-оценка с помощью функции `sentence_bleu` из библиотеки nltk:

```
sent_bleu = sentence_bleu([ref], mt)
```

METEOR — более гибкая метрика, которая учитывает синонимы и морфологические изменения слов, а также порядок слов. METEOR лучше подходит для оценки семантического соответствия перевода:

```
score = single_meteor_score(ref_tokens, mt_tokens)
```

POS-BLEU оценивает синтаксическую правильность перевода, сравнивая POS-теги эталонного перевода и машинного перевода:

```
pos_bleu = sentence_bleu([ref_tags], mt_tags)
```

Программа проверяет, соответствует ли BLEU, METEOR и POS-BLEU пороговым значениям:

```
if sent_bleu < low_bleu_threshold:
    print(f'Низкий BLEU для предложения {i+1}: {sent_bleu:.4f}')
    print('Эталон:', ' '.join(ref))
    print('Машинный перевод:', ' '.join(mt))
    print('---')
```

Если оценка для предложения ниже заданного порога (например, 0.5), программа добавляет это предложение в отчет и выделяет его красным в документе, используя библиотеку docx.

После вычисления всех метрик и выделения низких оценок программа создает отчет в формате .docx. В отчете содержится общая информация о результатах анализа и подробности для каждого предложения:

```
document = Document()
document.add_heading('Отчет по оценке машинного перевода', 0)
document.add_heading('Общая оценка', level=1)
document.add_paragraph(f'Корпусный BLEU: {bleu_score:.4f}')
document.add_paragraph(f'Порог BLEU: {low_bleu_threshold:.4f}')
document.add_paragraph(f'Средний METEOR: {avg_meteor_score:.4f}')
document.add_paragraph(f'Порог METEOR: {low_meteor_threshold:.4f}')
document.add_paragraph(f'Корпусный POS-BLEU: {pos_bleu_corpus_score:.4f}')
document.add_paragraph(f'Порог POS-BLEU: {low_pos_bleu_threshold:.4f}')
```

В отчете отображаются все вычисленные значения для каждой метрики, а также указываются предложения с низкими оценками, выделенные красным цветом.

Программа позволяет проводить комплексный анализ качества перевода, выявляя слабые места перевода по каждой из метрик. Примером работы программы может служить анализ перевода научной статьи, выполненной с использованием различных машинных переводческих сервисов. Результаты показывают, что, несмотря на высокие оценки по метрике BLEU, некоторые предложения могут иметь низкую оценку по METEOR или POS-BLEU, что позволяет глубже понять, где именно требуется улучшение.

Программа обеспечивает автоматизацию процесса анализа качества перевода, позволяя пользователю получить полный отчет, который может быть использован для улучшения перевода и оптимизации работы машинных переводческих сервисов. Благодаря тому, что программа выделяет низкие оценки, она способствует более эффективному выявлению ошибок и упрощает анализ качества перевода, предоставляя ясную и структурированную информацию для дальнейшего использования.

Для оценки качества перевода была выбрана научная статья «Использование машинного обучения и нейронных сетей для автоматической верификации заданий в текстовом и графическом представлении и помощи преподавателю» авторов Н. О. Бесшапошников, М. С. Дьяченко и др., переведенная с русского языка на английский. Оценка качества перевода проводилась с использованием Google Translate, Яндекс Переводчика, DeepL и ChatGPT, а именно, модели o4-mini.

Таблица

Результаты анализа

	Яндекс переводчик	Google Translate	DeepL	ChatGPT
BLEU	0.5209	0.5095	0.3248	0.3074
METEOR	0.7395	0.7197	0.4806	0.5901
POS tagging	0.8967	0.8971	0.7942	0.8677
Среднее значение	0.7190	0.7088	0.5332	0.5884

В ходе проведенного исследования было установлено, что различия между сервисами могут оказывать значительное влияние на качество перевода. Использование трех метрик позволило получить всестороннюю картину эффективности этих сервисов при переводах с русского языка на английский язык.

Исследование показало, что Яндекс и Google демонстрируют лучшие результаты по метрике BLEU, что свидетельствует о высоком уровне лексического соответствия и синтаксической структуры перевода. Однако разница между ними минимальна. GPT показал более низкие значения по BLEU, что говорит о большем количестве перефразирований и изменении структуры предложения. Важно отметить, что Яндекс и Google также лидируют и по METEOR, что подтверждает более высокое сохранение смысла перевода в сравнении с DeepL.

Яндекс продемонстрировал отличные результаты по переводу научной статьи, обеспечивая высокий уровень точности, особенно при передаче контекста и синтаксической структуры на уровне предложения. В то же время Google показал более высокую точность в переводе специфических эквивалентов, где точность транслитерации и культурная адаптация имеют ключевое значение. Это указывает на то, что Яндекс лучше справляется с передачей общего контекста и структуры, тогда как Google показывает более высокую эффективность в переводе узкоспециализированных терминов.

Различия в производительности этих сервисов могут быть связаны с различиями в их архитектурах. DeepL использует исключительно нейронные сети, что позволяет модели анализировать весь контекст предложения комплексно. В свою очередь, Яндекс применяет гибридный подход, комбинируя нейронные сети и статистические методы, что делает его особенно эффективным при переводах, где важно правильно обработать редкие слова и специфические выражения. Google, как и Яндекс, использует гибридный подход, но его результаты по метрике BLEU и другим метрикам слегка уступают, что также указывает на возможные различия в подходах к обучению и архитектуре. Более того, мы можем предположить, что различия в результатах этих систем могут также быть обусловлены различиями в объемах и характере обучающих данных. Например, Яндекс, ориентированный на русскоязычные тексты, имеет более глубокое понимание структуры и специфики русского языка, что способствует большей точности перевода в данной языковой паре. В то время как DeepL, несмотря на высокие результаты в переводах между западноевропейскими языками, имеет меньший фокус на русском языке, что ограничивает его эффективность в переводах с и на русский. Что касается Google, его результаты несколько уступают Яндексу, вероятно, из-за меньшего объема выборки для русского языка в обучении системы, а также особенностей алгоритмов обучения, ориентированных на более широкий спектр языков. Это может объяснять, почему Google менее эффективен при переводах с русского языка.

Результаты ChatGPT по всем метрикам заметно хуже, чем у специализированных сервисов перевода. Это, вероятно, связано с тем, что ChatGPT не является специализированным переводчиком и не был обучен с учетом fine-tuning для задач машинного перевода. Хотя ChatGPT может генерировать текст на различных языках, его основная цель — это диалоговое взаимодействие и генерация осмысленных текстов, а не перевод. Отсутствие адаптации под конкретные задачи перевода, такие как работа с техническими текстами или специализированной лексикой, ограничивает точность его переводов.

Несмотря на высокие результаты по BLEU и METEOR, все системы сталкиваются с определенными проблемами в области синтаксиса, лексики и использования артиклей. Например, машинный перевод продолжает испытывать трудности с точной передачей сложных грамматических структур, что иногда приводит к потерям в точности перевода. Ошибки на уровне синтаксиса и лексики делают переводы менее естественными. В то же время стоит отметить, что системы машинного перевода

не всегда соблюдают тема-рематическое членение, что также может привести к искажению смысла текста. Нарушение правильного распределения информации по теме и реме может менять акценты в предложении, создавая путаницу или теряя важные детали контекста. Для исправления таких ошибок необходима дополнительная экспертная доработка, поскольку они требуют глубокого понимания текста и его структуры, что машины пока не могут обеспечить на должном уровне.

В ходе проведенного исследования также следует обратить внимание, что метрики оценки качества машинного перевода могут сильно зависеть от эталонного (экспертного) перевода и часто занижают оценки, если допущен сильный стилистический или лексический отход от оригинала. Это особенно актуально для метрики BLEU, которая ориентирована на точное совпадение слов и фраз между машинным переводом и эталонным переводом, не учитывая синонимы и перефразирование. Такая зависимость от эталонного перевода снижает объективность данной метрики, особенно при ее использовании в одиночку.

Примером может служить следующий случай:

Оригинал: «Существенной помощью может стать использование интеллектуальных помощников преподавателя и автоматизированных систем проверки, построенных методами машинного обучения и технологии нейронных сетей». *Экспертный перевод:* «Smart educator assistants and automated assessment tools based on machine learning and neural networks can significantly alleviate the problem». *Перевод Google:* «The use of intelligent teacher assistants and automated checking systems built using machine learning methods and neural network technology can be of significant help». Итоговая оценка BLEU: 0.0000 (но METEOR = 0.4808, POS-BLEU = 0.7440).

Метрика BLEU присудила переводу нулевую оценку, но она не отражает глубины ошибок, связанных с неправильным выбором лексических эквивалентов. Грамматическая метрика POS-BLEU показала, что грамматическая составляющая перевода сохранена, однако это не свидетельствует о полном соответствии перевода. В данном случае именно METEOR, учитывающий синонимы и перефразирование, оказался более информативным. Несмотря на то, что перевод Google в целом сохраняет адекватность передачи общего смысла текста, в нем не была обеспечена эквивалентность лексических единиц. Например, словосочетание «системы проверки» была переведено как «checking systems», что является ошибкой. Вместо этого следовало использовать «assessment tools», что точно соответствует контексту и смыслу оригинала.

В рамках нашего исследования дополнительно был взят текст рецензируемой статьи из области медицины, а также статья с экспертным переводом, который ранее не был опубликован в сети Интернет. Эти тексты были использованы для того, чтобы оценить эффективность машинного перевода в различных контекстах и с различной сложностью лексики. Все оценки, полученные по трем метрикам, показали высокие результаты, со средним значением (0.6896), что свидетельствует о хорошем качестве перевода. При этом важно отметить, что корреляции между переводами статей, уже имеющих экспертные переводы в сети, и текстами, ранее не опубликованными, не было обнаружено. Гипотеза о том, что сервисы машинного перевода используют уже готовые переводы из сети, не была подтверждена результатами нашего анализа. Напротив, оказалось, что системы машинного перевода способны генерировать уникальные переводы даже для статей, не имеющих предварительных переводов в сети.

Системы машинного перевода показали достаточно высокую эффективность, что, вероятно, связано с функциональным стилем выбранных нами текстов. Научные статьи имеют строгую структуру, фиксированную грамматику и часто используют интернациональные термины, что делает их легче для перевода. Кроме того, в научных текстах отсутствует имплицитность и слова не имеют двойственного значения. Однако перевод медицинских диагнозов, заключений и специфической терминологии остается более сложной задачей. Несмотря на это, современные сервисы машинного перевода достаточно адекватно справляются с научными текстами. Важно отметить, что для языковой пары русский-английский лучшим сервисом по результатам нашего исследования оказался Яндекс, что подтверждается его высокими показателями по метрикам качества перевода. Развитие технологий машинного перевода открыло новые возможности для глобальных взаимодействий, обмена опытом и научными открытиями, но тем не менее необходимо проводить постредактуру переводов экспертами, способными выявить и исправить ошибки, которые машина не в силах корректно интерпретировать.

Автоматическая оценка в сочетании с экспертным анализом позволяет повысить точность проверки переводов. Автоматизированные инструменты помогают быстро выявить ключевые ошибки и

несоответствия, освобождая переводчиков и редакторов от рутинных задач и позволяя им сосредоточиться на исправлении значимых недостатков. Тем самым сочетание машинного анализа и экспертного контроля делает процесс перевода более эффективным и надежным.

Заключение

Машинный перевод, бесспорно, занимает важное место в эпоху глобализации, облегчая доступ к информации. Современные системы МП, основанные на нейронных сетях и трансформерах, демонстрируют значительные успехи в воспроизведении сложных языковых структур, но также сталкиваются с множеством вызовов, таких как передача культурных нюансов, идиоматических выражений и перевод для малораспространенных языков. Разработки в области многоязычных моделей, трансферного обучения и активного обучения открывают новые возможности для улучшения качества перевода редких языков и диалектов. Будущее машинного перевода связано с развитием гибридных и интерактивных систем, которые учитывают пользовательский контекст и адаптируются к специфике текста. Метрики автоматической оценки продолжают совершенствоваться, позволяя более точно оценивать качество перевода. Включение синтаксических и семантических аспектов в эти метрики помогает лучше отслеживать прогресс, делая машинный перевод адекватным.

Для повышения качества машинного перевода необходимы дальнейшие исследования и усовершенствования в области лексического анализа, морфологии и синтаксиса, а также разработка методов, способных лучше справляться с переводом редких и сложных слов, особенно в специализированных областях. Оценка машинного перевода должна быть многогранной, и использование нескольких метрик, таких как BLEU, METEOR и POS-BLEU, является важным инструментом для повышения точности и адекватности перевода. Постоянное совершенствование этих технологий требует интеграции автоматической оценки с экспертным анализом, что поможет улучшить качество переводов и снизить количество ошибок.

Машинный перевод продолжает развиваться и демонстрирует значительные успехи в области обработки сложных языковых структур. Однако важно продолжать исследования для улучшения адаптации к редким языкам и сохранения точности перевода, особенно в контексте культурных различий.

ЛИТЕРАТУРА

1. Воронович В. В. *Машинный перевод: конспект лекций по специальности «Компьютерная лингвистика, Компьютерное обучение языкам»*. Минск: Белорусский государственный университет; 2013. 39 с.
2. *Net.lang: на пути к многоязычному киберпространству*: пер. с англ. / ред. перевода Е. И. Кузьмин, А. В. Паршакова. М.: Межрегиональный центр библиотечного сотрудничества; 2014. 464 с.
3. Jawahar G., Sagot B., Seddah D. What Does BERT Learn about the Structure of Language? *ACL 2019-57th Annual Meeting of the Association for Computational Linguistics*. 2019.
4. Devlin J. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805. 2018.
5. Rogers A., Kovaleva O., Rumshisky A. A Primer in BERTology: What We Know about How BERT Works. *Transactions of the Association for Computational Linguistics*. 2021;8:842–866.
6. Shukla A. et al. An Evaluation of Google Translate for Sanskrit to English Translation via Sentiment and Semantic Analysis. *Natural Language Processing Journal*. 2023;4:100025.
7. Caswell I. 110 New Languages are Coming to Google Translate. *Google Translate Blog*. 2024.
8. Fan A. et al. Beyond English-Centric Multilingual Machine Translation. *Journal of Machine Learning Research*. 2021;22:1–48.
9. Koehn P., Knowles R. *Six Challenges for Neural Machine Translation*. arXiv:1706.03872. 2017.
10. Maruf S., Saleh F., Haffari G. A Survey on Document-Level Neural Machine Translation: Methods and Evaluation. *ACM Computing Surveys (CSUR)*. 2021;54(2):1–36.
11. Stahlberg F. *Neural Machine Translation: A Review and Survey*. arXiv:1912.02047. 2019.
12. Fitria T. N. Performance of Google Translate, Microsoft Translator, and DeepL Translator: Error Analysis of Translation Result. *Al-Lisan: Jurnal Bahasa*. 2023;8(2):115–138.
13. Varela-Salinas M. J. et al. Google Translate and DeepL: Breaking Taboos in Translator Training. Observational Study and Analysis. *Ibérica*. 2023;45:243–266.

14. Borisov A., Galinskaya I. Y. Yandex School of Data Analysis Russian-English Machine Translation System for WMT14. *Proceedings of the Ninth Workshop on Statistical Machine Translation*. 2014:66–70.
15. Cambetta G., Di Nunzio G. M., Nosilia V. A Study on Automatic Machine Translation Tools: A Comparative Error Analysis Between DeepL and Yandex for Russian-Italian Medical Translation. *Umanistica Digitale*. 2021;10:139–163.
16. Kulikov V., Kulikova V., Yerkebulan G. Google/Yandex Translation Detection in the Patterns Identifying System of Multilingual Texts. *International Journal of Computers*. 2021;20(1):72–77.
17. Kamaluddin M. I. et al. Accuracy Analysis of DeepL: Breakthroughs in Machine Translation Technology. *Journal of English Education Forum (JEEF)*. 2024;4(2):122–126.
18. Linlin L. Artificial Intelligence Translator DeepL Translation Quality Control. *Procedia Computer Science*. 2024;247:710–717.
19. Sriram A. et al. *Cold Fusion: Training seq2seq Models Together with Language Models*. arXiv:1708.06426. 2017.
20. Egonmwan E., Chali Y. Transformer and seq2seq Model for Paraphrase Generation. *Proceedings of the 3rd Workshop on Neural Generation and Translation*. 2019:249–255.
21. Li Z. et al. Seq2seq Dependency Parsing. *Proceedings of the 27th International Conference on Computational Linguistics*. 2018:3203–3214.
22. Sutskever I. *Sequence to Sequence Learning with Neural Networks*. arXiv:1409.3215. 2014.
23. Ni J. et al. *Sentence-T5: Scalable Sentence Encoders from Pre-trained Text-to-Text Models*. arXiv:2108.08877. 2021.
24. Chen G. et al. Towards Making the Most of Cross-Lingual Transfer for Zero-Shot Neural Machine Translation. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*. 2022;1:142–157.
25. Rothman D. *Transformers for Natural Language Processing: Build, Train, and Fine-Tune Deep Neural Network Architectures for NLP with Python, Hugging Face, and OpenAI's GPT-3, ChatGPT, and GPT-4*. Birmingham: Packt Publishing Ltd; 2022.
26. Maltas S. I. et al. *Efficient Finetuning Strategies for Multilingual Neural Machine Translation*. Universitat Politècnica de Catalunya; 2024.
27. Bengesi S. et al. Advancements in Generative AI: A Comprehensive Review of GANs, GPT, Autoencoders, Diffusion Model, and Transformers. *IEEE Access*. 2024;PP(99):1–1.
28. Achiam J. et al. *GPT-4 Technical Report*. arXiv:2303.08774. 2023.
29. Stap D., Araabi A. ChatGPT is Not a Good Indigenous Translator. *Proceedings of the Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP)*. 2023:163–167.
30. Agarwal A., Lavie A. Meteor, M-BLEU and M-TER: Evaluation Metrics for High-Correlation with Human Rankings of Machine Translation Output. *Proceedings of the Third Workshop on Statistical Machine Translation*. 2008:115–118.
31. Babych B., Hartley T. Extending the BLEU MT Evaluation Method with Frequency Weightings. *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04)*. 2004:621–628.
32. Chen B., Cherry C. A Systematic Comparison of Smoothing Techniques for Sentence-Level BLEU. *Proceedings of the Ninth Workshop on Statistical Machine Translation*. 2014:362–367.
33. Papineni K. et al. BLEU: a Method for Automatic Evaluation of Machine Translation. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*. 2002:311–318.
34. Pérez D., Alfonseca E. Application of the BLEU Algorithm for Recognising Textual Entailments. *Proceedings of the First Challenge Workshop Recognising Textual Entailment*. 2005:9–12.
35. Banerjee S., Lavie A. METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments. *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*. 2005:65–72.
36. Denkowski M., Lavie A. Extending the METEOR Machine Translation Evaluation Metric to the Phrase Level. *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. 2010:250–253.
37. Denkowski M., Lavie A. Meteor Universal: Language-Specific Translation Evaluation for Any Target

- Language. *Proceedings of the Ninth Workshop on Statistical Machine Translation*. 2014:376–380.
38. Бешапошников Н. О., Дьяченко М. С., Леонов А. Г., Матюшин М. А., Орловский А. Е. Использование машинного обучения и нейронных сетей для автоматической верификации заданий в текстовом и графическом представлении и помощи преподавателю. *Успехи кибернетики*. 2020;1(2):35–41. DOI: 10.51790/2712-9942-2020-1-2-4.
 39. Wang F. Methods for Evaluating the Translation Quality of Artificial Intelligence Translator DEEPL Based on Multi-Translation Parallel Corpus. *2024 First International Conference on Software, Systems and Information Technology (SSITCON)*. 2024:1–5.
 40. Мыльников А. Г., Климов А. Е., Курбанниёзов Т. Ш., Буйместру Н. В., Черняева А. А., Гусарова Т. А. Экстренная тотальная дуоденопанкреатэктомия по поводу профузного желудочно-кишечного кровотечения, вызванного метастазами рака почки в поджелудочную железу. *Вестник Российского университета дружбы народов. Серия: Медицина*. 2023;27(2):246–253. DOI: 10.22363/2313-0245-2023-27-2-246-253.