



НАУЧНАЯ СТАТЬЯ

УДК 81'25

Дата поступления: 19.06.2024
рецензирования: 20.07.2024
принятия: 15.11.2024

NMT-перевод – основные модели, оценка качества

И.А. Борунов

Государственный университет просвещения, г. Мытищи, Российская Федерация
E-mail: ivanborun@gmail.com. ORCID: <http://orcid.org/0009-0001-3108-8004>

Аннотация: Настоящее исследование посвящено теме оценки качества моделей NMT-перевода, а также методам оценки качества полученных переводов. Оно имеет высокую актуальность в современных условиях работы переводчика, где используются инновационные технологии для облегчения рабочих процессов и повышения качества услуг. Целью исследования является всестороннее изучение современных моделей нейронного машинного перевода (NMT) и методов оценки качества переведенных текстов, выполненных с их помощью с акцентом на применение данных моделей для перевода юридической и законодательской литературы. Исследование направлено на выявление преимуществ и ограничений существующих NMT-технологий, а также на определение направлений для их дальнейшего совершенствования. Для достижения обозначенной цели были сформулированы следующие задачи: провести обзор и анализ основных моделей NMT, включая Seq2Seq, Transformer, BERT и их вариации с целью выявления их особенностей и применимости для перевода различных типов текстов; изучить и оценить качество перевода, выполненного с помощью современных NMT-систем; осуществить сравнительный анализ перевода юридических и законодательских текстов с использованием методов контекстуального, структурно-семантического и сравнительно-сопоставительного анализа; определить ограничения существующих NMT-моделей и разработать рекомендации для их дальнейшего улучшения и расширения их области применения; сформулировать выводы о перспективах применения NMT-технологий в профессиональной практике перевода и предложить направления для будущих исследований в данной области. В работе проведен анализ современных моделей NMT, включая Seq2Seq, Transformer, BERT и их вариации, исследованы особенности каждой модели и их применимость для различных языковых пар. Особое внимание уделено оценке качества NMT-переводов с использованием метрик BLEU, METEOR, TER, а также человеческой оценки качества перевода. Тексты выполненных с помощью NMT-систем переводов юридической литературы изучались при помощи методов контекстуального анализа, а также структурно-семантического и сравнительно-сопоставительного методов. Проведенное исследование позволяет сделать вывод о высокой эффективности современных моделей NMT и их способности генерировать качественные переводы на различных языках. Однако выявлены некоторые ограничения, требующие дальнейших исследований и улучшений. Результаты и выводы исследования могут быть полезны как для научного сообщества, так и для практиков в области (машинного) перевода и лингвистики. Кроме того, оно подчеркивает важность дальнейших исследований для улучшения качества автоматического перевода и расширения его области применения.

Ключевые слова: NMT-перевод; нейронные сети; Seq2Seq; Transformer; оценка качества; метрики, современные модели перевода.

Цитирование. Борунов И.А. NMT-перевод – основные модели, оценка качества // Вестник Самарского университета. История, педагогика, филология Vestnik of Samara University. History, pedagogics, philology. 2024. Т. 30, № 4. С. 220–227. DOI: <http://doi.org/10.18287/2542-0445-2024-30-4-220-227>.

© Борунов И.А., 2024

Иван Александрович Борунов – аспирант кафедры переводоведения и когнитивной лингвистики, Государственный университет просвещения, 141014, Российская Федерация, Московская область, г. Мытищи, ул. Веры Волошиной, 24.

SCIENTIFIC ARTICLE

Submitted: 19.06.2024
Revised: 20.07.2024
Accepted: 15.11.2024

NMT-translation – basic models, quality assessment

I.A. Borunov

Federal State University of Education, Mytishchi, Russian Federation
E-mail: ivanborun@gmail.com. ORCID: <http://orcid.org/0009-0001-3108-8004>

Abstract: This article is devoted to the topic of evaluating the quality of NMT-translation models, as well as methods of evaluating the quality of the obtained translations. The research we have conducted in the field of comparison of NMT-based translations of special texts is of high relevance in today's translator work environment, where modern technologies are used to facilitate work processes and improve the quality of services. The purpose of this study is to provide a comprehensive review of current neural machine translation (NMT) models and methods for assessing the quality of translated texts produced using them, with a focus on the application of these models to the translation of legal and legislative literature. The study aims to identify the advantages and limitations of existing NMT technologies, as

well as to identify areas for their further improvement. In order to achieve this goal, the following tasks were formulated: to review and analyze the main NMT models, including Seq2Seq, Transformer, BERT and their variations, in order to identify their peculiarities and applicability to the translation of different types of texts; to study and evaluate the quality of translation made with the help of modern NMT systems; to carry out a comparative analysis of the translation of legal texts using the methods of contextual analysis, structural-semantic and comparative-comparative analysis; to determine the limitations of the existing NMT models; to identify the advantages and limitations of the existing NMT models for the translation of legal texts. The paper analyses modern NMT models, including Seq2Seq, Transformer, BERT and their variations, investigating the features of each model and their applicability to different language pairs. Special attention is paid to the evaluation of NMT translation quality using BLEU, METEOR, TER metrics as well as human evaluation of translation quality. The texts of translations of specialized literature in various fields and publicistic texts on relevant topics made with the help of NMT systems were studied using contextual analysis methods, as well as structural-semantic and comparative-semantic methods. The research conducted allows us to conclude that modern NMT models are highly effective and capable of generating high-quality translations in different languages. However, despite the success achieved, some limitations have been identified that require further research and improvements. The study provides a better understanding of current NMT translation models, their applicability and translation quality. The results and conclusions of the study can be useful for both the academic community and practitioners in the field of (machine) translation and linguistics. Furthermore, the results and conclusions of this paper emphasize the importance of further research to improve the quality of automatic translation and to extend its field of application.

Key words: NMT translation; neural networks; Seq2Seq; Transformer; quality assessment; metrics.

Citation. Borunov I.A. NMT-translation – basic models, quality assessment. *Vestnik Samarskogo universiteta. Istorii, pedagogika, filologiya Vestnik of Samara University. History, pedagogics, philology*, 2024, vol. 30, no. 4, pp. 220–227. DOI: <http://doi.org/10.18287/2542-0445-2024-30-4-220-227>. (In Russ.)

© Borunov I.A., 2024

Ivan A. Borunov – postgraduate student of the Department of Translation Studies and Cognitive Linguistics, Federal State University of Education, 24, Vera Voloshina Street, Mytishchi, 141014, Russian Federation.

Введение

В современном мире рост объемов переводческой деятельности и повышение требований к качеству перевода ставят перед исследователями и практиками задачу поиска эффективных и инновационных подходов к переводу. Одним из ключевых направлений развития в области машинного перевода являются нейросетевые модели, а именно модели перевода с использованием нейронных сетей (NMT). NMT-перевод привлекает внимание исследователей и специалистов в области лингвистики, компьютерных наук и машинного обучения благодаря своей способности создавать переводы более естественного и качественного характера [Bahdanau 2014, p. 13].

В контексте данной области исследования, важными являются определения ключевых понятий. Нейросетевые модели перевода, или NMT, представляют собой подход к машинному переводу, основанный на использовании глубоких нейронных сетей для перевода текстовых последовательностей с одного языка на другой. Эти модели принесли существенные изменения в области машинного перевода и стали широко применяться в различных сферах [Карасев 2015, с. 117].

Одной из основных проблем, стоящих перед исследователями в области NMT-перевода, является оценка качества полученных переводов. Понимание того, насколько точными и естественными являются сгенерированные переводы, играет ключевую роль при выборе оптимальной модели и настройке ее параметров. В этом контексте актуальными становятся различные методы оценки качества перевода, такие как BLEU (Bilingual Evaluation Understudy), METEOR (Metric for Evaluation of Translation with Explicit ORdering), TER (Translation Error Rate), которые позволяют оценить соответствие перевода исходному тексту и качество перевода в целом.

В научной литературе по NMT-переводу существует множество работ, посвященных различным аспектам этой темы. Работы таких исследователей, как Улиткин И.А., Максименко О.И., Фаустова К.И., Шаляпина З.И. и многих других, оказывают значительное влияние на развитие современных методов и моделей перевода. Их труды способствуют углубленному пониманию процессов, лежащих в основе NMT-перевода, и разработке новых подходов к повышению качества и эффективности данной технологии [Улиткин 2016, с. 117; Максименко 2014, с. 6; Фаустова 2017, с. 83; Шаляпина 1996, с. 105].

Одним из ключевых преимуществ NMT-подхода является его способность учитывать контекст и зависимости между словами в предложении, что позволяет генерировать более связные и естественные переводы. Такие модели способны автоматически изучать языковые особенности и осуществлять перевод на основе обширных данных об оригинальных и целевых текстах.

Важным аспектом развития NMT является борьба с проблемой недостатка параллельных данных для обучения моделей. Для эффективного обучения нейросетевых моделей требуется большой объем параллельных корпусов текстов на разных языках. Специалисты занимаются разработкой методов, позволяющих обучать модели на небольших наборах данных или с применением методов активного обучения для повышения качества перевода.

Одной из ключевых задач при работе с нейросетевыми моделями перевода является оптимизация процесса обучения и настройки гиперпараметров модели. Это включает в себя выбор оптимальной архитектуры нейронной сети, определение методов регуляризации, улучшение процесса обучения с использованием техник, таких как обучение с учи-

телем и обучение без учителя, а также улучшение механизмов внимания для более точного перевода.

Кроме того, активно исследуются вопросы расширения функциональности NMT-моделей, например, включение механизмов генерации многоязычных переводов, адаптации к специфическим предметным областям или учета диалектов и разнообразия в языке. Все эти пути развития направлены на улучшение качества автоматического перевода и повышение эффективности работы с текстами на разных языках.

Таким образом, исследование текущей ситуации и тенденций в области NMT-перевода, включая модели, оценку качества и научные труды, представляет собой актуальную и перспективную задачу, способную привнести новые знания и подходы в развитие машинного перевода.

Цель и задачи

Материал для исследования качества NMT-перевода представляет собой текст юридической, законодательской тематики. Объем выборки для исследования составляет 124 переводческие страницы текста по вышеуказанной области знания. Данный выбор материала обусловлен тем фактом, что такие тексты характеризуются высокой лексической сложностью, обилием дефиницированной терминологии со строгим словарным соответствием на языке перевода, а также высоким уровнем синтаксической «нагроможденности», что в совокупности делает данный тип текстов одним из самых сложных для перевода.

В статье рассматривается система NMT-перевода Google Translate – один из наиболее популярных инструментов как в среде переводчиков, так и рядовых пользователей Интернета. Кроме того, данная система содержит в себе две основные современные NMT-модели – Transformer и seq2seq (см. ниже), что позволяет причислить данный продукт к числу самых передовых на сегодняшний день с точки зрения его технологичности.

Методы исследования эмпирического материала определены целью, задачами объектом и предметом настоящей работы. Сравнительно-сопоставительный метод использовался с целью системного сравнения результирующих предложений для выявления осуществленных структурных и семантических преобразований в процессе перевода. Метод сплошной выборки из специальных текстов и публицистических текстов применялся при отборе эмпирического материала и выявлении стратегий их использования в двуязычной ситуации. Метод структурно-семантического анализа терминологических единиц использовался с целью установления общих и отличительных свойств и признаков структуры и значения лексических единиц. Метод контекстуального анализа способствовал выявлению смысловых различий в результирующих и исходных речевых высказываниях.

Ход исследования

Начнем с обсуждения основных моделей машинного перевода (NMT). Одной из самых популяр-

ных моделей является модель-трансформер. Трансформеры представляют собой архитектуру модели, основанную на механизмах внимания, которая позволяет эффективно обрабатывать контекст и строить глубокие представления текста. Они показали отличные результаты в задачах машинного перевода благодаря своей способности учить долгосрочные зависимости в тексте и эффективно обрабатывать длинные последовательности [Caswell, Liang 2020, p. 38].

Модель трансформера (Transformer) – это мощная архитектура нейронной сети, которая была разработана компанией Google и представлена в статье «Attention is All You Need» [Vaswani, Shazeer, Niki Uszkoreit et al. 2017, p. 117–129]. Трансформер стал одним из прорывов в области обработки естественного языка (NLP) и машинного перевода, превзойдя многие предыдущие методы на задачах, связанных с текстом.

Основным элементом модели трансформера является механизм внимания (attention), который позволяет модели «фокусироваться» на различных частях входной последовательности при генерации выходной последовательности. Это позволяет трансформеру лучше понимать зависимости между словами в тексте и эффективно моделировать долгосрочные зависимости.

Архитектура трансформера состоит из двух основных компонентов: кодировщика (encoder) и декодера (decoder). Кодировщик принимает на вход исходное предложение и создает внутреннее представление (эмбединг) этого предложения. Декодер использует это внутреннее представление для генерации переведенного текста.

Важными особенностями модели трансформера являются использование многоголового внимания (multi-head attention) и позиционно-зависимых нейронных сетей (position-wise feed-forward networks). Многоголовое внимание позволяет модели фокусироваться одновременно на разных аспектах входных данных, а позиционно-зависимые нейронные сети помогают учитывать порядок слов в последовательности.

Трансформеры имеют большую вычислительную мощность и способны обучаться на огромных объемах данных, что делает их эффективными для широкого спектра задач, таких как машинный перевод, суммаризация текста, вопросно-ответные системы и многое другое. Также трансформеры стали основой для разработки различных вариаций, таких как BERT, GPT и других моделей, которые показали впечатляющие результаты на ряде NLP-задач.

Еще одна популярная модель NMT – это seq2seq (sequence-to-sequence). Она состоит из двух рекуррентных нейронных сетей – кодировщика и декодера, которые работают с последовательностями данных. Кодировщик преобразует входные данные во внутреннее представление (векторное представление), которое затем декодирует для генерации выходной последовательности.

Основным принципом работы seq2seq является генерация выходной последовательности пошагово, при этом каждый следующий токен выходной

последовательности предсказывается на основе ранее сгенерированных токенов и контекста исходной последовательности. Для прогнозирования следующего токена в декодере часто используется механизм внимания (attention), который помогает модели фокусироваться на разных частях входной последовательности.

Как видно из вышеприведенного описания, данные модели очень похожи по своему принципу работы. По сути, модель-трансформер является прямым продолжением и усовершенствованной версией модели seq2seq. Для более наглядного представления приведем конкретные отличия двух моделей в следующих пунктах:

1. Архитектура. Модель трансформера (Transformer) представляет собой более сложную и мощную архитектуру нейронной сети, чем seq2seq. Трансформер использует механизм внимания для эффективной обработки входных последовательностей без необходимости рекуррентных или сверточных слоев. В то время как seq2seq состоит из кодировщика и декодера, трансформер включает в себя множество блоков внимания и механизмов с точками внимания.

2. Обработка контекста. Трансформер может обрабатывать более длинные и сложные зависимости в тексте благодаря механизмам внимания. Это позволяет трансформеру лучше улавливать долгосрочные зависимости в тексте, чем seq2seq, что дает ему преимущество в моделировании сложных языковых структур.

3. Эффективность обучения и вывода. Трансформеры обычно учатся быстрее и масштабируются лучше, чем классические seq2seq модели. Это связано с параллельной обработкой входных последовательностей и использованием механизма внимания, что ускоряет обучение и улучшает качество предсказаний.

Таким образом, трансформеры обладают более высокой производительностью и способностью захвата долгосрочных зависимостей в текстовых данных, что делает их более популярными и эффективными для широкого спектра задач обработки естественного языка. В то время как seq2seq остается важным инструментом для некоторых задач, требующих специализированных подходов к генерации последовательностей.

Обе эти модели являются основой для многих современных систем машинного перевода и продемонстрировали свою эффективность в создании человекоподобных переводов.

Перейдем к методам оценки качества машинного перевода. Одной из наиболее распространенных метрик является BLEU (Bilingual Evaluation Understudy). BLEU измеряет совпадение n-грамм между сгенерированным переводом и эталонным переводом, что дает представление о том, насколько хорошо работает модель. Однако BLEU имеет свои ограничения, такие как недостаточный учет контекста и семантики текста. Для учета более широкого спектра языковых особенностей и понимания смысла перевода также используются другие метрики, такие как METEOR и TER.

Важно выбирать подходящую метрику в зависимости от конкретной задачи и контекста, что позволит более точно оценить качество машинного перевода. Вместе с тем оценка качества перевода также может включать субъективные оценки экспертов и пользовательские тесты, чтобы получить информацию о понимании и удовлетворенности полученным переводом.

Другим важным аспектом в развитии машинного перевода является применение нейронных сетей с архитектурой кодировщик-декодировщик, отличающейся от рекуррентных сетей. Так, недавние исследования показывают хорошие результаты применения трансформеров, представляющих собой мощные архитектуры для работы с последовательными данными. Трансформеры позволяют параллельно обрабатывать входные последовательности, что повышает эффективность обучения и улучшает качество перевода.

Еще одним направлением развития в машинном переводе является обучение с подкреплением. Глубокие нейронные сети могут обучаться с учителем на больших объемах текста, что существенно улучшает качество перевода. Также с применением техники обучения с подкреплением модель перевода может самостоятельно улучшать свой результат путем взаимодействия с внешней средой и обратной связи.

Кроме того, для улучшения качества машинного перевода широко применяются методы обучения на параллельных корпусах и использование многоязычных моделей, которые помогают модели учитывать особенности различных языков и повышают ее обобщающую способность.

Второе направление, которое стоит упомянуть, – это развитие моделей глубокого обучения с использованием многоязычных представлений. Такие модели позволяют учиться на параллельных корпусах текстов на разных языках, что позволяет им эффективно обобщать знания и улучшать качество перевода между различными языковыми парами.

Еще одной интересной тенденцией является работа над улучшением способности моделей машинного перевода к обработке редких и сложных языковых конструкций. Это включает в себя разработку специализированных методов предварительной обработки данных, а также дополнительных механизмов внимания и управления контекстом.

Необходимо также отметить активное применение генеративных преобразовательных сетей (GANs) в задачах машинного перевода. Эти модели позволяют генерировать естественно звучащие переводы, сочетая в себе способность генерации текста и обработки контекста. Использование GANs открывает новые возможности для создания более точных и естественных переводов.

Поговорим подробнее о некоторых метриках оценки качества машинного перевода, таких как BLEU (Bilingual Evaluation Understudy), METEOR (Metric for Evaluation of Translation with Explicit Ordering) и TER (Translation Error Rate).

1. BLEU (Bilingual Evaluation Understudy) – это одна из самых популярных метрик оценки качества машинного перевода. BLEU измеряет схожесть

между автоматически сгенерированными переводами и одним или несколькими эталонными (правильными) переводами. Основная идея метрики BLEU заключается в том, что чем больше совпадений между автоматически сгенерированным переводом и эталонными переводами, тем выше будет значение BLEU. BLEU рассматривает не только точность перевода слов, но и учитывает присутствие n-грамм (наборов слов). BLEU можно выразить численно, и чем выше значение BLEU, тем лучше считается перевод [Papineni 2001, p. 55].

2. METEOR (Metric for Evaluation of Translation with Explicit ORdering) – это еще одна популярная метрика для оценки качества машинного перевода. Особенностью METEOR является то, что она учитывает не только точность и полноту перевода, но также оценивает порядок слов и фраз в переводе. METEOR основана на вычислении precision и recall для слов и фраз в автоматически сгенерированных переводах. Затем вычисляется гармоническое среднее от precision и recall, чтобы определить общую оценку качества перевода [Banerjee 2005, p. 234].

3. TER (Translation Error Rate) – это метрика, оценивающая качество машинного перевода путем подсчета количества ошибок в переводе. Основная идея TER заключается в том, что чем меньше ошибок в переводе, тем выше будет значение этой метрики. Ошибки могут включать в себя вставки, удаления, замены и переупорядочивание слов в переводе. TER позволяет оценить частоту ошибок и неточностей в машинных переводах и предоставляет индикатор качества перевода относительно эталонного текста.

Важно отметить, что каждая из этих метрик имеет свои особенности, и ни одна из них не является идеальной. Поэтому для полной и объективной оценки качества машинного перевода часто рекомендуется использовать комбинацию нескольких метрик. Кроме того, большое значение имеет контекст и специфика задачи, для которой производится оценка качества перевода.

Проведем анализ и сравнение различных аспектов перевода текстов юридической тематики с использованием современной системы NMT-перевода Google Translate. Оценим качество перевода, точность передачи специфической терминологии и соответствие текста перевода языковым нормам.

Для оценки качества перевода юридических текстов рассмотрим варианты перевода фрагмента Федерального закона от 27.07.2006 № 152-ФЗ (ред. от 06.02.2023) «О персональных данных».

Оригинальный текст:

«1. Законодательство Российской Федерации в области персональных данных основывается на Конституции Российской Федерации и международных договорах Российской Федерации и состоит из настоящего Федерального закона и других определяющих случаи и особенности обработки персональных данных федеральных законов.

2. На основании и во исполнение федеральных законов государственные органы, Банк России, органы местного самоуправления в пределах сво-

их полномочий могут принимать нормативные правовые акты, нормативные акты, правовые акты (далее – нормативные правовые акты) по отдельным вопросам, касающимся обработки персональных данных. Такие акты не могут содержать положения, ограничивающие права субъектов персональных данных, устанавливающие не предусмотренные федеральными законами ограничения деятельности операторов или возлагающие на операторов не предусмотренные федеральными законами обязанности, и подлежат официальному опубликованию.

3. Особенности обработки персональных данных, осуществляемой без использования средств автоматизации, могут быть установлены федеральными законами и иными нормативными правовыми актами Российской Федерации с учетом положений настоящего Федерального закона.

3.1. Нормативные правовые акты, принимаемые в соответствии с частью 2 настоящей статьи, подлежат обязательному согласованию с уполномоченным органом по защите прав субъектов персональных данных в случаях, если указанные нормативные правовые акты регулируют отношения, связанные с осуществлением трансграничной передачи персональных данных, обработкой специальных категорий персональных данных, биометрических персональных данных, персональных данных несовершеннолетних, предоставлением, распространением персональных данных, полученных в результате обезличивания. Срок указанного согласования не может превышать тридцать дней с даты поступления соответствующего нормативного правового акта в уполномоченный орган по защите прав субъектов персональных данных.

4. Если международным договором Российской Федерации установлены иные правила, чем те, которые предусмотрены настоящим Федеральным законом, применяются правила международного договора.

5. Решения межгосударственных органов, принятые на основании положений международных договоров Российской Федерации в их истолковании, противоречащем Конституции Российской Федерации, не подлежат исполнению в Российской Федерации. Такое противоречие может быть установлено в порядке, определенном федеральным конституционным законом» (Статья 4).

Google Translate:

«1. The legislation of the Russian Federation in the field of personal data is based on the Constitution of the Russian Federation and international treaties of the Russian Federation and consists of this Federal Law and other federal laws defining the cases and features of the processing of personal data.

2. On the basis of and in pursuance of federal laws, state bodies, the Bank of Russia, local government bodies, within the limits of their powers, may adopt regulatory legal acts, regulatory acts, legal acts (hereinafter referred to as regulatory legal acts) on certain issues relating to the processing of personal data. Such acts cannot contain provisions limiting the rights of personal data subjects, establishing restric-

tions on the activities of operators not provided for by federal laws, or imposing obligations on operators not provided for by federal laws, and are subject to official publication.

3. Features of the processing of personal data carried out without the use of automation tools may be established by federal laws and other regulatory legal acts of the Russian Federation, taking into account the provisions of this Federal Law.

3.1. Regulatory legal acts adopted in accordance with Part 2 of this article are subject to mandatory approval by the authorized body for the protection of the rights of personal data subjects in cases where these regulatory legal acts regulate relations related to the cross-border transfer of personal data, processing of special categories of personal data, biometric personal data, personal data of minors, provision,

distribution of personal data obtained as a result of depersonalization. The period of said approval cannot exceed thirty days from the date of receipt of the relevant regulatory legal act by the authorized body for the protection of the rights of personal data subjects.

4. If an international treaty of the Russian Federation establishes rules other than those provided for by this Federal Law, the rules of the international treaty apply.

5. Decisions of interstate bodies adopted on the basis of the provisions of international treaties of the Russian Federation in their interpretation, contrary to the Constitution of the Russian Federation, are not subject to execution in the Russian Federation. Such a contradiction may be established in the manner prescribed by federal constitutional law».

Сводка обнаруженных ошибок и/или неточностей в переводе:

«1. Законодательство Российской Федерации в области персональных данных основывается на Конституции Российской Федерации...»

«1. The legislation of the Russian Federation in the field of personal data is based on the Constitution of the Russian Federation...»

Лексическая ошибка: несмотря на словарное соответствие, корректным вариантом в данном случае является вариант «The laws of the Russian Federation», так как legislation – это, в первую очередь, законотворческий процесс, нежели перечень законов, как подразумевается в исходном предложении.

«2. На основании и во исполнение федеральных законов государственные органы, Банк России, органы местного самоуправления в пределах своих полномочий могут принимать нормативные правовые акты, нормативные акты, правовые акты (далее – нормативные правовые акты) ...»

«2. On the basis of and in pursuance of federal laws, state bodies, the Bank of Russia, local government bodies, within the limits of their powers, may adopt regulatory legal acts, regulatory acts, legal acts (hereinafter referred to as regulatory legal acts) ...»

Синтаксическая ошибка: ввиду дословной структуры перевода в переводном предложении неясно, где начинается подлежащее. Необходимо перестроить первую часть предложения: «State bodies, the Bank of Russia, local government bodies, within the limits of their powers and on the basis of and in pursuance of federal laws, may adopt...». Пропуск союза «и». Несмотря на его отсутствие в тексте оригинала, перечисление однородных членов предложения в английском языке всегда заканчивается с использованием союза and или as well as: «State bodies, the Bank of Russia and local government bodies, within the limits of their powers...»

«...не предусмотренные федеральными законами ограничения деятельности операторов или возлагающие на операторов не предусмотренные федеральными законами обязанности, и подлежат официальному опубликованию...»

«...the activities of operators not provided for by federal laws, or imposing obligations on operators not provided for by federal laws, and are subject to official publication...»

Стилистическая неточность: словарная система английского языка позволяет избежать многократного повторения лексем federal laws, что не было применено в переводе: «...the activities of operators not provided for by federal laws, or imposing obligations not provided thereby, and are subject to official publication...»

«4. Если международным договором Российской Федерации установлены иные правила, чем те, которые предусмотрены настоящим Федеральным законом, применяются правила международного договора».

«4. If an international treaty of the Russian Federation establishes rules other than those provided for by this Federal Law, the rules of the international treaty apply».

Стилистическая ошибка: в контексте официального декларативного текста, необходимо использовать конструкции типа subjunctive mood: «... the rules of the international treaty shall apply».

«5. Решения межгосударственных органов, принятые на основании положений международных договоров Российской Федерации в их истолковании...»

«5. Decisions of interstate bodies adopted on the basis of the provisions of international treaties of the Russian Federation in their interpretation, are not subject to execution in the Russian Federation...»

Синтаксическая ошибка: ввиду дословной структуры перевода в переводном предложении неясно, к чему относится конструкция «in their interpretation». Из синтаксической структуры текста перевода следует, что «интерпретаторами» являются договоры Российской Федерации, нежели Межгосударственные органы, что подразумевается в тексте оригинала. Для внесения ясности необходимо полностью перестроить эту часть предложения: «5. Decisions of interstate bodies adopted on the basis of their interpretation of the provisions of international treaties of the Russian Federation...»

Стилистическая ошибка: Как и в примере выше, необходимо использовать конструкции типа subjunctive mood: «5. Decisions of interstate bodies adopted in their interpretation on the basis of the provisions of international treaties of the Russian Federation, shall not be executed in the Russian Federation».

Как видно из вышеприведенного анализа перевода, NMT-система допустила ряд ошибок и неточностей различных типов при переводе. Однако стоит отметить, что с точки зрения смыслового соответствия существенных ошибок допущено не было. Несмотря на некоторые сложности в восприятии и, как следствие, понимании переводного текста, при повторном прочтении текста и принимая во внимание сложность восприятия исходного текста в том числе, такие неточности нельзя трактовать как смысловые ошибки.

С точки зрения структуры мы наблюдаем полное лексическое и грамматическое соответствие. Как было отмечено, полная грамматическая эквивалентность влечет за собой сложности восприятия текста и некоторые нарушения стилистики текста английского языка. Однако стоит отметить, что для перевода юридического текста, в частности связанного с законодательными и судебными тематиками, дословное воссоздание текста и калькирование его грамматической структуры зачастую является прямым требованием к переводу [Борунов 2022, 1601]. Так, при переводе процессуальных документов, к примеру, какие-либо лексико-грамматические преобразования и вовсе являются недопустимыми – переводной текст в данном случае должен быть строго таким, каким представлен текст на исходном языке.

Стоит отметить, что даже при наличии ошибок, доля текста, которая не потребовала редакторской правки, составила 83,2 %, что позволяет заявить о крайней эффективности данной системы в работе современного переводчика.

Примечательно, что все вышеперечисленные ошибки выходят за пределы строго лингвистические (кроме пропуска союза «и» при перечислении однородных членов предложения. Все ошибки, как лексические (legislation – law; лексические повторения) так и стилистические порождены «непониманием» машиной интенции текста (что именно подразумевается под словом «законодательство», в каком именно контексте и с какой целью применяется та или иная грамматическая форма (речь идет о неточностях в наклонении глагола), что именно подразумевается под выражением «в их интерпретации» (в чьей?)). Так, возвращаясь к предложению № 5, если рас-

сматривать перевод с точки зрения лингвистического (не интенционального) соответствия исходному тексту, то Google-Translate перевел данный сегмент совершенно верно: «5. Решения межгосударственных органов, принятые на основании положений международных договоров Российской Федерации в их истолковании...» – если судить исключительно по синтаксической структуре, действительно неясно, к чему относится конструкция «в их истолковании». Таким образом, мы приходим к выводу, что основным барьером для систем-NMT являются не лингвистические явления, а более глубинные, концептуальные и интенциональные факторы, которые на сегодняшний день так и не были формализованы и, соответственно, не были представлены в виде структур данных, пригодных для алгоритмической обработки с целью их дальнейшей интеграции в языковые системы.

Заключение

В ходе проведенного исследования было выявлено, что, несмотря на некоторые ошибки и неточности, система продемонстрировала высокий уровень эффективности. Особенно важно отметить, что с точки зрения структурных и лингвистических аспектов перевода, система проявила высокую грамматическую и лексическую эквивалентность, что является одним из ключевых требований при переводе юридических документов.

Ошибки, выявленные в процессе исследования, как лексические, так и стилистические, связаны в основном с пониманием интенции текста и контекста использования терминов.

Таким образом, можно сделать вывод, что NMT-системы эффективны в переводе юридических текстов, однако важно учитывать не только лингвистические аспекты, но также концептуальные и интенциональные факторы для более точного и качественного перевода.

Исследование подтвердило значимость развития и совершенствования NMT-технологий и подчеркнуло необходимость дальнейших исследований в области компьютерной обработки концептуальной и интенциональной информации, в чем, помимо прочих дисциплин, могут помочь достижения современной когнитивной лингвистики.

Материалы исследования

Статья 4 – *Статья 4. Законодательство Российской Федерации в области персональных данных*. URL: https://bio.spbu.ru/science/scienceinfo/el_resource.php (дата обращения: 04.04.2024).

Библиографический список

Banerjee, Lavie 2005 – *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*, Ann Arbor, Michigan, 2005, pp. 65– 72. URL: <https://aclanthology.org/W05-0909>.

Bahdanau, Kyunghyun, Bengio 2014 – *Bahdanau Dzmitry, Kyunghyun Cho, Bengio Yoshua. Neural Machine Translation by Jointly Learning to Align and Translate* // *arXiv:1409.0473*. 2014. P. 13. DOI: <https://doi.org/10.48550/arXiv.1409.0473>.

Papineni, Roukos, Ward, Zhu 2001 – *Kishore Papineni, Salim Roukos, Todd Ward, Wei-Jing Zhu*. BLEU: a Method for Automatic Evaluation of Machine Translation // *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*. Morristown, NJ, USA: Association for Computational Linguistics, 2001. P. 311–318. DOI: <https://doi.org/10.3115/1073083.1073135>.

Snover, Dorr 2006 – *Snover M., Dorr B.* A Study of Translation Edit Rate with Targeted Human Annotation // *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers*. Cambridge, Massachusetts, USA, 2006. P. 223–231. URL: <https://aclanthology.org/2006.amta-papers.25>.

Борунов, Улиткин 2022 – *Борунов И.А., Улиткин И.А.* Особенности перевода терминологических единиц в сфере финансов, аудита и инвестиций с английского языка на русский // *Филологические науки. Вопросы теории и практики*. 2022. Т. 15, вып. 5. С. 1599–1606. DOI: <https://doi.org/10.30853/phil20220229>.

Карасев, Артюшина 2011 – *Карасев И.В., Артюшина Е.А.* Системы машинного перевода // *Успехи современного естествознания*. 2011. № 7. С. 117–118. URL: <https://natural-sciences.ru/en/article/view?id=27117&ysclid=m4v25wcinq939792982>; <https://elibrary.ru/item.asp?id=16394393>. EDN: <https://elibrary.ru/nutspt>.

Максименко, Чинина 2014 – *Максименко О.И., Чинина Д.С.* Обзор системы машинного перевода «Google Переводчик» (на примере финского языка) // *Science Time*. 2014. С. 133–139. URL: <https://cyberleninka.ru/article/n/obzor-sistemy-mashinnogo-perevoda-google-perevodchik-na-primere-finskogo-yazyka?ysclid=m4v2cy7159163557504>.

Улиткин 2022 – *Улиткин И.А.* Автоматическая оценка качества машинного перевода научного текста: 5 лет спустя светлой памяти моего учителя Нелюбина Л.Л. // *Вестник Московского государственного областного университета. Серия: Лингвистика*. 2022. № 1. С. 47–59. DOI: <https://doi.org/10.18384/2310-712X-2022-1-47-59>. EDN: <https://elibrary.ru/gtauum>.

Фаустова 2017 – *Фаустова К.И.* Нейронные сети: применение сегодня и перспективы развития // *Территория науки*. 2017. № 4. С. 83–87. URL: <https://cyberleninka.ru/article/n/neyronnye-seti-primenenie-segodnya-i-perspektivy-razvitiya?ysclid=m4v2t5xrx1793391776>.

Шалыпина 1996 – *Шалыпина З.М.* Автоматический перевод: Эволюция и современные тенденции // *Вопросы языкознания*. 1996. № 2. С. 105–117. URL: <https://vja.ruslang.ru/ru/archive/1996-2/105-117>.

References

Banerjee, Lavie 2005 – Banerjee S., Lavie A. (2005) METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments. In: *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*. Ann Arbor, Michigan, 2005, pp. 65–72. URL: <https://aclanthology.org/W05-0909>.

Bahdanau, Kyunghyun, Bengio 2014 – *Bahdanau Dzmitry, Kyunghyun Cho, Bengio Yoshua* (2014) Neural Machine Translation by Jointly Learning to Align and Translate. *arXiv:1409.0473*, p. 13. DOI: <https://doi.org/10.48550/arXiv.1409.0473>.

Papineni, Roukos, Ward, Zhu 2001 – *Kishore Papineni, Salim Roukos, Todd Ward, Wei-Jing Zhu* (2001) BLEU: a Method for Automatic Evaluation of Machine Translation. In: *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*. Morristown, NJ, USA: Association for Computational Linguistics. 2001, pp. 311–318. DOI: <https://doi.org/10.3115/1073083.1073135>.

Snover, Dorr 2006 – *Snover M., Dorr B.* A Study of Translation Edit Rate with Targeted Human Annotation. In: *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers*. Cambridge, Massachusetts, USA, 2006. P. 223–231. URL: <https://aclanthology.org/2006.amta-papers.25>.

Borunov, Ulitkin 2022 – *Borunov I.A., Ulitkin I.A.* (2022) Features of Translating Terminological Units in the Subject Area of Finance, Audit and Investment from English into Russian. *Philology. Theory & Practice*, vol. 15, issue 5, pp. 1599–1606. DOI: <https://doi.org/10.30853/phil20220229>. (In Russ.)

Karasev, Artyushina 2011 – *Karasev I.V., Artyushina E.A.* (2011) Machine translation systems. *Advances in current natural sciences*, no. 7, pp. 117–118. Available at: <https://natural-sciences.ru/en/article/view?id=27117&ysclid=m4v25wcinq939792982>; <https://elibrary.ru/item.asp?id=16394393>. EDN: <https://elibrary.ru/nutspt>. (In Russ.)

Maksimenko, Chinina 2014 – *Maksimenko O.I., Chinina D.S.* (2014) Review of machine translation system «Google Translator» (on the example of the Finnish language). *Science Time*, pp. 133–139. Available at: <https://cyberleninka.ru/article/n/obzor-sistemy-mashinnogo-perevoda-google-perevodchik-na-primere-finskogo-yazyka?ysclid=m4v2cy7159163557504>. (In Russ.)

Ulitkin 2022 – *Ulitkin I.A.* (2022) Automatic evaluation of machine translation quality of a scientific text: five years later. *Bulletin of Moscow Region State University. Series: Linguistics*, no. 1, pp. 47–59. DOI: <https://doi.org/10.18384/2310-712X-2022-1-47-59>. EDN: <https://elibrary.ru/gtauum>. (In Russ.)

Faustova 2017 – *Faustova K.I.* (2017) Neural networks: application today and prospects of development. *Territoriya nauki*, no. 4, pp. 83–87. Available at: <https://cyberleninka.ru/article/n/neyronnye-seti-primenenie-segodnya-i-perspektivy-razvitiya?ysclid=m4v2t5xrx1793391776>. (In Russ.)

Shalyapina 1996 – *Shalyapina Z.M.* (1996) Machine translation: Evolution and contemporary trends. *Voprosy jazykoznanija = Topics in the Study of Language*, no. 2, pp. 105–117. Available at: <https://vja.ruslang.ru/ru/archive/1996-2/105-117>. (In Russ.)