

Материалы международной научно-практической конференции, посвященной 130-летию С. Сейфуллина = С. Сейфуллиннің 130 жылдығына арналған халықаралық ғылыми - практикалық конференциясының материалдары. - 2024. – Ч.ІІ.- С.3-6.

УДК 62.01.1

ПОСТ-РЕДАКТИРОВАНИЕ ИИ: ПЕРСПЕКТИВЫ ПЕРЕВОДА МАЛОРЕСУРСНЫХ ЯЗЫКОВ НА ПРИМЕРЕ КАЗАХСКОГО

Байбақты Е.Н., 2 курс, магистрант

Касекеева А.Б., PhD

Евразийский Национальный университет им. Л.Н. Гумилева

В данной статье рассматривается постредактирование машинного перевода для казахского языка, который относится к категории малоресурсных языков. В исследовании анализируются результаты работы двух моделей - GPT-4o и Gemini 1.5 Flash, с акцентом на их способности к выполнению задач постредактирования. Особое внимание уделяется метрике BLEU, использованной для оценки качества перевода.

Полученные результаты показали, что модель GPT-4o демонстрирует стабильные и точные результаты, следуя инструкциям и обеспечивая минимальные изменения текста. В отличие от неё, Gemini 1.5 Flash столкнулась с проблемами, включая отклонения от инструкций и появление галлюцинаций, что негативно сказалось на точности перевода.

Статья подчеркивает важность разработки и применения современных методов машинного перевода для малоресурсных языков, а также необходимость дальнейшего исследования в этой области, включая использование дополнительных метрик для более полной оценки качества перевода и интеграцию специализированных языковых моделей.

Машинный перевод широко применяется пользователями для понимания иностранных языков, что обуславливает растущий спрос на качественные и точные переводы. Этот вопрос особенно актуален для стран с малоресурсными языками, таких как казахский. В настоящее время наиболее популярными системами машинного перевода являются Yandex Translate и Google Translate. Однако, несмотря на их широкое использование, данные системы нередко допускают ошибки, связанные с грамматикой, пониманием контекста и особенностями структуры языка. В качестве альтернативы был разработан исследователями переводчик Tilmash, который по показателям качества сопоставим с ведущими системами машинного перевода [1].

Для повышения качества перевода применяется метод постредактирования, при котором человек оценивает и корректирует машинный перевод для достижения более точного результата. [2]. Однако данный процесс требует значительных временных и интеллектуальных ресурсов, так как редактор должен обладать хотя бы базовыми знаниями грамматики целевого языка. В свете этих трудностей перспективным решением является использование современных ИИ-моделей, основанных на

архитектуре трансформеров, таких как ChatGPT и Gemini, которые будут подробно рассмотрены в данной статье.

Цель статьи заключается в сравнительном анализе эффективности постредактирования с использованием ИИ-моделей.

В работе [3] рассматривается задача постредактирования машинного перевода для малоресурсных языков, включая казахский и узбекский. Авторы разработали систему постредактирования, основанную на нейронных сетях и методах трансформеров, что позволило улучшить качество перевода для новостных публикаций, законодательных документов и текстов из IT-сферы. Эксперименты показали значительное повышение показателей качества перевода, особенно для русско-казахского и англо-казахского направлений.

Исследование [4] рассматривает возможность использования больших языковых моделей (LLM), таких как GPT-4, для автоматического постредактирования результатов нейросетевых машинных переводов (NMT) для высокоресурсных языков, таких как английский, китайский и немецкий. Хотя исследование фокусируется на высокоресурсных языковых парах, оно показывает потенциал использования GPT-4 для улучшения качества перевода, что может быть полезным для разработки аналогичных подходов и для малоресурсных языков, таких как казахский.

Постредактирование машинного перевода делится на четыре типа, каждый из которых направлен на достижение определенного качества конечного текста. Легкое постредактирование фокусируется на устранении самых очевидных ошибок и обеспечивает базовую понятность текста. Полное постредактирование направлено на создание стилистически безупречного текста, сохраняя при этом общий стиль оригинала. Высокое постредактирование предполагает глубокую коррекцию и доработку текста с целью достижения качества, сопоставимого с человеческим переводом. Слабое постредактирование является наименее глубоким типом и заключается в минимальном вмешательстве в исходный текст [3].

Легкое постредактирование позволяет нам сосредоточиться на оценке способности моделей ChatGPT и Gemini к пониманию контекста и исправлению наиболее очевидных ошибок. В качестве исходного материала мы выбрали следующий промпт: «Ты — эксперт по постредактированию текста на казахском языке, отлично владеющий как английским, так и казахским. Твоя задача — выполнить легкое постредактирование текста, предоставленного в оригинале на английском и его переводе на казахский».

Для эксперимента будет использовано пятнадцать предложений, переведенных с английского на казахский с помощью Google Translate. Они взяты из работы [1]. Затем оригинальные тексты и их переводы будут переданы ИИ, которые выполнят легкое постредактирование. Полученные результаты будут оценены с использованием метрики BLEU [5].

В ходе постредактирования модель Gemini 1.5 Flash не следовала предоставленным инструкциям, что привело к отклонениям от заданной задачи. В частности, модель начала переводить текст на русский язык вместо

выполнения легкого постредактирования казахского текста. Кроме того, вместо предложений с однозначными исправлениями модель предлагала несколько альтернативных вариантов текста. Повторные запросы на уточнение инструкций не приводили к улучшению, и модель продолжала генерировать отклоняющиеся от исходных инструкций варианты. Также наблюдалось множество галлюцинаций, включающих значительные изменения текста с потерей исходного смысла. Основным недостатком заключался в том, что модель забывала первоначальную инструкцию, что требовало постоянного напоминания. Положительным аспектом Gemini 1.5 Flash является возможность неограниченного использования модели.

Модель GPT-4o показала более стабильные результаты, строго следуя инструкциям и предоставляя краткие и точные ответы с минимальными изменениями текста. Галлюцинации в ходе работы с этой моделью не наблюдались. Основным недостатком является ограниченное количество запросов на использование.

Оценка BLEU изображены в таблице 1.

Таблица 1 - Показатели BLEU

	BLEU score
Google Translate	21,28
GPT-4o	20,71
Gemini 1.5 Flash	12,86

Результаты показывают, что существенных улучшений качества перевода не наблюдается. Однако система ChatGPT демонстрирует перспективы для дальнейшего развития, учитывая её способность понимать базовые элементы грамматики и структуры казахского языка. Например, в таблице 2 видно, как модель ChatGPT корректно заменила слово «ойлау» на «нұсқаушы» при переводе термина "suggestive", что более точно передает смысл "внушающий" или "намекающий". Кроме того, замена фразы «куәгерлердің куәлігіне» на «куәгерлердің көрсетпелеріне» значительно улучшила читаемость текста и его соответствие контексту.

Перевод модели Gemini 1.5 Flash содержит несколько ключевых ошибок, влияющих на точность передачи оригинального текста. Во-первых, термин «suggestive» был неправильно переведен как «еске түсіруге бағытталған», что искажает смысл оригинала, так как "suggestive" означает внушающий, а не восстанавливающий память. Во-вторых, изменение структуры предложения ослабляет связь между внушающими вопросами и их влиянием на воспоминания и свидетельские показания. Также использование термина «жадының қайта құрылуына» вместо «реконструкциялық жады» звучит менее естественно и менее точно передает идею реконструктивной памяти.

Таблица 2 - Постредактирование GPT-4o и Gemini 1.5 Flash

Английский	Google Translate	GPT-4o	Gemini 1.5 Flash
Elizabeth Loftus, a prominent cognitive psychologist, has conducted a fascinating program of research on reconstructive memory – investigating how “suggestive” questioning can influence memory and subsequent eyewitness testimony.	Белгілі когнитивтік психолог Элизабет Лофтус реконструкциялық жады бойынша қызықты зерттеу бағдарламасын жүргізді - «ойлау» сұраудың есте сақтауға және одан кейінгі куәгерлердің куәлігіне қалай әсер ететінін зерттейді.	Белгілі когнитивтік психолог Элизабет Лофтус реконструкциялық жады бойынша қызықты зерттеу бағдарламасын жүргізді, онда «нұсқаушы» сұрақтардың есте сақтауға және кейінгі куәгерлердің көрсетпелеріне қалай әсер ететінін зерттейді.	Белгілі когнитивтік психолог Элизабет Лофтус адамның жадының қайта құрылуына қатысты қызықты зерттеу жүргізді. Ол өз зерттеулерінде «еске түсіруге бағытталған» сұрақтардың адамның жадына және кейіннен куәгерлердің көрсетпелеріне қалай әсер ететінін зерттеді.

Анализ проведенного исследования показал, что наибольшую точность в постредактировании перевода с русского на казахский язык продемонстрировала модель GPT-4o. Она стабильно следовала инструкциям и обеспечивала минимальные изменения текста, что делает её подходящей для задач, связанных с обработкой малоресурсных языков, таких как казахский. В то же время модель Gemini 1.5 Flash, несмотря на её гибкость и неограниченное использование, показала низкие результаты из-за отклонений от инструкций и появления галлюцинаций, что снижает её надежность для точного перевода.

Оценка BLEU подтверждает выводы о качестве моделей: GPT-4o получила результат, сопоставимый с Google Translate, тогда как Gemini 1.5 Flash значительно отстала по точности перевода. Данные результаты подчеркивают важность выбора правильной модели для постредактирования в зависимости от поставленных задач и уровня требований к точности перевода.

Будущие исследования могут быть направлены на улучшение постредактирования перевода через интеграцию более специализированных языковых моделей для казахского языка, а также на тестирование моделей на других тюркских языках для расширения их применения.

Список литературы

- 1 Yeshpanov, R., Polonskaya, A., Varol, H. A. (2024). KazParC: Kazakh Parallel Corpus for Machine Translation. In *2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation, LREC-COLING 2024 - Main Conference Proceedings* European Language Resources Association (ELRA). 9633–9644.
- 2 Vieira, LN, Alonso, E., Bywood, L. (2019). Introduction: Post-editing in practice-Process, product and networks. *J. Spec. Transl.*, 31, 2–13.
- 3 Rakhimova, D, Karibayeva, A, Turarbek, A, (2024). The Task of Post-Editing Machine Translation for the Low-Resource Language. *Appl. Sci.*, 14, 486.
- 4 Raunak, V., Sharaf, A., Wang, Y., Awadalla, HH, Menezes, A. (2023). Leveraging GPT-4 for Automatic Translation Post-Editing. In *Findings of the Association for Computational Linguistics: EMNLP 2023*. 12009–12024.
- 5 Papineni, K., Roukos, S., Ward, T., Zhu, WJ. (2002). BLEU: A method for automatic evaluation of machine translation. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. 311–318.
- 6 Kasekeyeva, AB, Batura, TV, Efimova, LV, Murzin, FA, Tussupov, JA, Yerimbetova, AS. Doshtayev, KZh. (2020). Link grammar and formal analysis of paraphrased sentences in a natural language. *Journal of Theoretical and Applied Information Technology*, 98(10), 1724–1736.