

ВЛИЯНИЕ ПРЕДРЕДАКТИРОВАНИЯ ТЕКСТА НА КАЧЕСТВО МАШИННОГО ПЕРЕВОДА С РУССКОГО НА ЯПОНСКИЙ ЯЗЫК

Медяникова О.Н. (ТОГУ), Окладников В.Е. (ВЦ ДВО РАН)

Научные руководители – к.ф.н, доцент Кривошеева Е.И. (ТОГУ), начальник отдела Вычислительного центра ДВО РАН Кожевникова Т.В. (ВЦ ДВО РАН)

Введение. Современные системы машинного перевода продолжают совершенствоваться, однако качество перевода остается серьезной проблемой, в особенности для языковых пар с разной грамматической структурой, таких как русский и японский. Один из возможных способов улучшения итогового перевода — предредактирование исходного текста, которое позволяет сделать его более подходящим для обработки машинным переводом [1].

В совместном исследовании Алены Животовой и Виктора Бердоносова (2024) показано, что предредактирование помогает повысить качество перевода с русского на английский [2,3]. В нашей работе мы проверяем актуальность предредактирования для языковой пары русский – японский.

Основная часть. Для проверки гипотезы о влиянии предредактирования на качество машинного перевода с русского на японский было проведено исследование, включающее несколько этапов:

1) Выбор исходных предложений.

В качестве исходных данных было отобрано 18 фраз из фильма «Чучело» режиссера Ролана Быкова. Предложения содержат фразеологизмы, различные грамматические конструкции, стилистические и лексические особенности, потенциально оказывающие влияние на качество машинного перевода.

2) Создание эталонного перевода.

Каждое предложение было переведено на японский язык с участием переводчика. Полученный результат использовался в качестве эталонного образца для оценки качества машинного перевода.

3) Предредактирование исходных предложений.

Русскоязычные предложения были вручную изменены с целью минимизации возможных проблем при выполнении машинного перевода. В рамках предредактирования выполнялись:

- а) Упрощение сложных грамматических конструкций,
- б) Устранение двусмысленности,
- в) Замена идиоматических выражений на более нейтральные формулировки,
- г) Адаптация синтаксиса к особенностям японского языка.

4) Выполнение машинного перевода оригинальных и предредактированных предложений.

Исходные и предредактированные русские предложения были переведены на японский язык с использованием Yandex GPT.

5) Оценка качества перевода с помощью метрик.

Для оценки качества машинного перевода с эталонным переводом использовался метод, основанный на эмбедингах BERT-модели для японского языка:

а) Для каждого переведенного машиной японского предложения (как из оригинального текста, так и из предредактированного варианта) были получены эмбединги с помощью BERT-модели.

б) Затем вычислялось косинусное расстояние между эмбедингами машинного перевода и эталонного перевода. Чем ближе значение к 1, тем выше схожесть с эталонным переводом.

Проведенное исследование показало, что прирост корректности перевода с русского языка на японский составил 3,17%. Однако в некоторых случаях наблюдалось ухудшение результата при предредактировании, что связано со следующими факторами:

1) Потеря фразеологии.

При предредактировании часто устраняются идиоматические выражения и стилистические особенности, которые машинный переводчик не может корректно передать. В человеческом эталонном варианте перевода такие элементы сохраняются, что делает его более точным.

2) Замена устаревших слов на современные аналоги.

В процессе предредактирования архаичные или редкие слова заменялись на более современные. В эталонном переводе сохранялась стилистика исходного текста, что влияло на итоговую оценку схожести.

3) Изменение грамматических конструкций.

При предредактировании использовались более простые или нейтральные синтаксические конструкции, что иногда приводило к снижению экспрессивности перевода. В то время как в человеческом эталонном варианте перевода подобные конструкции сохранялись для передачи эмоциональной окраски.

Выводы. Полученные результаты показывают, что предредактирование может улучшить качество машинного перевода с русского на японский, однако оно не является универсальным решением. В отдельных случаях изменения в исходном тексте приводят к потере смысловых оттенков, что требует дополнительного анализа.

В дальнейшем планируется исследование автоматического предредактирования с использованием нейросетевых моделей, позволяющих выявлять и корректировать «плохо переводимые» конструкции.

Список использованных источников:

1. Jiang W. Pre-editing for Machine Translation / Weimin Jiang // In Proceedings of the 9th Conference of the Association for Machine Translation in the Americas: Government MT User Program. – 2010.

2. Животова, А. А. Оптимизационное предредактирование узкоспециальных русскоязычных текстов для их машинного перевода на английский язык / А. А. Животова, В. Д. Бердоносков // Информационные и математические технологии в науке и управлении. – 2024. – № 2(34). – С. 169-182. – DOI 10.25729/ESI.2024.34.2.016. – EDN JLAWZU.

3. Zhivotova, Alena & Berdonosov, Victor. (2023). Pre-editing Strategy Based on Automatic Evaluation of Translation Complexity to Improve the Quality of Specialized Texts Machine Translation into English. 10.28995/2075-7182-2023-22-1141-1149.