

RU

Постредактирование англо-русского машинного перевода: проблемы, методы и оптимизация

Хромова А. А., Лукманова Р. Р.

Аннотация. Цель исследования – выработка рекомендаций для оптимизации англо-русского машинного перевода собственно-научных текстов естественно-научной направленности на этапе его постредактирования, ориентированных на улучшение качества машинного перевода. В статье впервые осуществлена систематизация методов постредактирования машинного англо-русского перевода текстов, функционирующих в естественнонаучной области знания – нейробиологии, произведен разноаспектный анализ процесса постредактирования машинного перевода. Сказанное свидетельствует о научной новизне исследования. Приведены примеры успешного постредактирования, обсуждены его перспективы в части совершенствования систем машинного перевода. Полученные результаты показали, что при постредактировании приходится справляться с тем, что онлайн-сервисы генерируют некорректные синтаксические конструкции или порождают терминологические единицы, смысл которых отличается от смысла единиц на языке оригинала. Важная задача постредактора – улучшение восприятия текста, она может быть выполнена, к примеру, с помощью реконструкции предложений.

EN

Post-editing of English to Russian machine translation: Issues, methods, and optimization

Khromova A. A., Lukmanova R. R.

Abstract. The aim of the study is to develop recommendations for optimizing English to Russian machine translation of scientific texts in the field of natural sciences during the post-editing stage, focusing on enhancing the quality of machine translation. The article provides a systematic categorization of post-editing methods for machine translation from English to Russian of texts in the field of natural sciences, specifically neurobiology, and conducts a multi-faceted analysis of the post-editing process of machine translation. This indicates the scientific novelty of the research. Successful post-editing examples are presented, discussing its prospects for improving machine translation systems. The results revealed that during post-editing, challenges arise as online services generate incorrect syntactic structures or introduce terminological units which meanings differ from those in the original language. An important task for the post-editor is to enhance the text's comprehension, which can be achieved, for instance, through sentence reconstruction.

Введение

Авторы работы фокусируют внимание на машинном переводе (МП) – активно развивающейся, не вполне изученной отрасли изысканий в области искусственного интеллекта, привлекающей большое внимание исследователей и специалистов лингвистического профиля и компьютерных наук. Одним из ключевых аспектов МП является постредактирование, направленное на улучшение качества МП и обработку переводного текста в соответствии с требованиями конечного пользователя.

Актуальность данного исследования обуславливается его нацеленностью на выработку рекомендаций для оптимизации англо-русского МП на этапе его постредактирования. Внедрение данных рекомендаций позволит улучшить качество МП – особого вида перевода, прочно вошедшего в жизнь человека и нуждающегося в постоянном изучении. Несомненно, за последние два десятилетия его качество заметно улучшилось, однако данная динамика является недостаточной: до настоящего времени МП (как результат) обнаруживает неточности преимущественно на грамматическом и лексико-семантическом уровнях, в связи с этим невозможно говорить о достижении его адекватности.

Для реализации поставленной цели исследования необходимо решить следующие задачи:

- на основе научной литературы, посвященной разработке различных аспектов МП, определить содержание термина «постредактирование», охарактеризовать его типы;
- выявить и систематизировать основные методы и приёмы постредактирования текстов, генерируемых переводческими онлайн-сервисами;
- проанализировать и сравнить результаты работы сервисов онлайн-перевода DeepL и Google Translate с целью выработки рекомендаций для постредакторов текстов научной области «Нейробиология».

Важно учитывать, что степень адекватности МП коррелирует с целью субъекта перевода – в зависимости от этого выбирается тип постредактирования – лёгкое или полное. Постредактирование, будучи завершающим этапом МП, проводимое на стыке языкознания и компьютерной лингвистики, – продукт междисциплинарного взаимодействия. В данном случае речь идёт о реализации интердисциплинарной стратегии МП – образовании нового общего научного инструментария, методологии и знания, возникающих как в лингвистике, так и в теории и практике программирования.

МП – несовершенная технология, и это несовершенство проявляется неоднозначно: для одного предложения он может дать точный и контекстуально приемлемый перевод, но в следующем может допустить серьёзную смысловую ошибку, нарушить стилистическое единство, оставить логический пробел или, наоборот, добавить лишний оттенок значения. Необходимость применения постредактирования обуславливается главной задачей пользователя: если МП нужен лишь для понимания сути текста, то исправлять такие ошибки, возможно, и не нужно, но если МП применяется для создания текста, предназначенного для открытого опубликования, широкого распространения и т. п., то постредактирование – обязательный этап комплексного процесса МП. Выявление таких ошибок, их доработка или исправление называется постредактированием. Данный термин использовался уже на ранних этапах развития систем МП, когда технология была несколько медленнее, текст отправлялся в электронном виде в систему МП. После перевода готовый материал возвращался отправителю, который занимался редактированием продукта автоматического перевода. Данный термин также используется для описания временных процессов, в которых исправление ошибок происходит одновременно с автоматическим переводом.

Материалом исследования послужили тексты на английском языке естественно-научной направленности, опубликованные в журнале “Nature”, и варианты их перевода на русский язык, сгенерированные двумя онлайн-сервисами – Google Translate и DeepL.

Для достижения поставленной цели авторами были применены различные методы – лингвистические и общенаучные. Так, на первом этапе исследовательской работы из текста на исходном языке для анализа были выбраны единицы перевода, выявлены их текстовые функции: в той или иной речевой ситуации слово как единица языка оказывается связанным системными отношениями с другими словами данного текста, т. е. попадает в ситуативную зависимость или ряд зависимостей от условий текста. Эти зависимости составляют иерархию контекстов, от минимального (соседнее слово) до максимального (весь текст или сверхтекстовые связи). На этом этапе был использован метод герменевтической интерпретации единиц на языке оригинала, он дал возможность применить механизм интерпретации исходной единицы, построить свой смысл её перевода, таким образом сохранить смысл при постредактировании. На втором этапе изысканий был проведён сопоставительно-переводческий анализ единиц перевода – фрагментов текста на исходном языке и единиц, сгенерированных онлайн-сервисами Google Translate и DeepL. Данный метод применён на лексико-семантическом и грамматическом уровнях. Третий этап работы предполагал сравнительный анализ результатов, сгенерированных нашими онлайн-переводчиками. На последнем, четвёртом, этапе исследования проведено обобщение полученных результатов и сформированы рекомендации для оптимизации англо-русского МП на этапе его постредактирования.

Теоретической базой работы послужила публикация Йошуа Бенджио (LeCun, Bengio, Hinton, 2015), известного своими исследованиями в области глубокого обучения нейронных сетей и улучшения качества МП, включая разработку новых моделей и подходов; Кёнхюн Чо работает над оптимизацией и улучшением процесса постредактирования МП с использованием автоматического ранжирования и генерации редакций (Cho, 2016); Филипп Кон занимается исследованиями в области статистического МП и его улучшения через глубокое обучение. Учёный разработал известный алгоритм модели статистического МП и постоянно работает над его развитием и оптимизацией (Koehn, 2021). Кристофер Мэннинг активно исследует естественный язык и его применение в МП. Он разрабатывает новые модели и алгоритмы, основанные на глубоком обучении, для улучшения результатов МП и постредактирования (Manning, 2022). Бенжамин Мари является экспертом в области нечёткого соответствия и построения словарей для МП. Он занимается поиском методов улучшения соответствий между исходным текстом и переводом, а также увеличения точности и согласованности перевода (Marie, Fujita, Rubino, 2021). Мы перечислили лишь некоторые из известных исследователей в области МП и постредактирования. Их последние открытия и исследования помогают лучше понять и оптимизировать процесс МП и постредактирования.

В отечественном языкознании первые системные попытки развития данного направления появились во второй половине XX в. Так, в этой связи значимыми являются работы таких авторов, как В. Ю. Розенцвейг (1964) – «Основы общего и машинного перевода», А. А. Кибрик (1970) – «Модель автоматического анализа письменного текста (на материале ограниченного военного подязыка)», З. М. Шаляпина (1975) – «К проблеме построения формальной модели процесса перевода», О. С. Кулагина (1979) – «Исследования по машинному

переводу», Ю. Н. Марчук (1983) «Проблемы машинного перевода», Ю. Д. Апресян (1984) «Лингвистическое обеспечение системы французско-русского автоматического перевода ЭТАП-1», Г. В. Чернов – «Машинный перевод и прикладная лингвистика. Проблемы создания системы автоматического перевода» (1986), Н. Д. Андреев (1986) – «Основные направления работы экспериментальной лаборатории машинного перевода», Н. Н. Леонтьева (1989) – «База знаний и автоматический перевод (проект многоязычной информационно-справочной системы)» (Утробина, 2022, с. 403).

Системы МП достигли значительного прогресса с точки зрения качества перевода благодаря развитию нейронных сетей и методов глубокого обучения, но, несмотря на этот прогресс, МП требует дальнейшего совершенствования. Ключевая его проблема – недостаточная точность и качество или, оперируя специализированными переводческими терминами, – степень эквивалентности и адекватности перевода. Зачастую в текстах на языке перевода, создаваемых машинными системами, выявляются ошибки лексико-грамматического, лексико-семантического свойства, из внимания упускаются мелкие детали и особенности языка перевода. Чаще всего это проявляется при работе с длинными и грамматически сложными текстами и может стать препятствием для эффективного коммуникативного взаимодействия и использования МП в различных сферах, таких как бизнес, наука, международные отношения и др. В контексте обозначенной проблемы постредактирование становится важным этапом МП. Постредактирование представляет собой процесс ревизии и внесения корректив в МП с целью улучшения его качества для использования в конкретной коммуникативной ситуации. Человеческая интервенция является необходимым этапом работы с системами МП: редактор может дополнить и улучшить результаты работ машинной системы перевода, исправив ошибки, улучшив структуру предложений и приспособив перевод к стилистическим и культурным особенностям языка перевода.

Опытные редакторы, имеющие как лингвистические, так и профессиональные компетенции в конкретных областях научного знания, в процессе постредактирования могут сделать материал более ценным, улучшив качество перевода и придав переведённому тексту естественность, плавность и логическую последовательность. В этом смысле постредактирование вносит необходимые дополнения, исправления и доработки в результаты МП, делая их более точными, понятными и соответствующими требованиям коммуникации.

Постредактирование – это процесс двуязычной обработки текста. Человек, осуществляющий постредактирование, обычно именуется постредактором, однако целесообразность такого наименования спорная, особенно если учесть, что рабочая среда профессионального переводчика предполагает использование автоматизированных средств перевода, объединяющих такие технологии, как память переводов, управление терминологией и МП (Kennu, 2022). По сути, задача по-прежнему сводится к переводу и редактированию, однако специфика работы проявляется в том, какими технологическими средствами или поддержкой пользуется переводчик в конкретный момент.

При постредактировании переводчик должен понять смысл единицы исходного текста и единицы перевода, полученной от системы МП. Затем он должен выявить все ошибки онлайн-сервиса, разработать стратегию их исправления и внедрить эту стратегию. Системы МП могут создавать различные типы ошибок, начиная от грамматических и синтаксических и заканчивая ошибками в лексическом или терминологическом выборе, а также стилистическими неточностями. Типы и количество ошибок, которые могут присутствовать в тексте, подготовленном системой МП, зависят от многих факторов, таких как языковая пара, стилевая принадлежность контента, лексическое наполнение или, например, методы, использованные для обучения системы МП.

Для оценки качества постредактирования можно использовать различные методы, включая экспертные оценки или методы автоматической оценки, такие как BLEU-метрика или TER-метрика. Комбинирование различных методов может помочь получить более полную картину относительно качества постредактирования.

Практическая значимость исследования состоит в том, что его материал может быть полезен пользователям программ, осуществляющих МП. Полученные результаты имеют дидактическую ценность и могут быть применены в преподавании таких дисциплин, как теория и практика перевода, специальный перевод, технический перевод. Кроме того, материалы работы могут быть включены в изыскания ученых-компаративистов, в большей степени переводоведов, специалистов в области теории и практики компьютерного программирования.

Обсуждение и результаты

Одна из основных задач МП – обеспечение перевода текстов, содержащих большое количество информации, с одного языка на другой, ключевым фактором при этом является быстрая скорость перевода, в связи с этим выделяют два основных уровня постредактирования: 1) *лёгкое постредактирование* и 2) *полное постредактирование*. Лёгкое постредактирование включает минимальные исправления и выполняется быстро. Полное постредактирование подразумевает исправление всех ошибок, обнаруженных в результате МП, и требует больше времени. Оба уровня постредактирования обычно предполагают более быстрое выполнение перевода, чем при работе без использования компьютерных средств. Международная организация по стандартизации (International Standards Organization, ISO) разработала стандарт по постредактированию – ISO 18857:2017 (Translation services – Post-editing of machine translation output – Requirements). Он квалифицирует лёгкое

постредактирование как «процесс, направленный на достижение простого понимания текста, без стремления создать продукт, сравнимый с переводом, выполненным человеком», а полное постредактирование – как «процесс, направленный на получение продукта, сопоставимого с переводом, выполненным человеком» (ISO 18857:2017). Шарон О’Брайен справедливо замечает, что данные дефиниции предполагают наличие определённых концептуальных трудностей (Kennu, 2022, p. 107).

Рассмотрим постредактирование обоих уровней подробнее. **Лёгкое постредактирование** предполагает: 1) стремление к семантически корректному переводу; 2) сохранение языковой нормы языка перевода; 3) проверку того факта, что в процессе перевода никакая информация не была случайно добавлена или опущена; 4) редактирование тех фрагментов переводного текста, которые содержат сниженную лексику (за исключением тех случаев, когда речь идет о намеренном использовании таковой в художественных целях); 5) отсутствие необходимости вносить в текст перевода стилистические правки и корректировать структуру единицы перевода, чтобы улучшить его стиль (Kennu, 2022).

Полное постредактирование обеспечивает: 1) стремление к грамматически, синтаксически и семантически правильному переводу; 2) адекватный перевод терминологии, в случае если он невозможен, непереверждённые термины включаются в список «Не переводить»; 3) проверку того факта, что в процессе перевода никакая информация не была случайно добавлена или опущена; 4) редактирование тех фрагментов переводного текста, которые содержат сниженную лексику (за исключением тех случаев, когда речь идет о намеренном использовании таковой в художественных целях); 5) правильность форматирования текста (Kennu, 2022).

Набор рекомендаций для постредактирования, представленный выше, сформирован, по сути, на основе ключевого принципа – максимального использования исходного перевода. Переводчик всегда имеет возможность проигнорировать результаты МП – удалить их и самостоятельно перевести оригинальную единицу. Очевидно, что в редких случаях, когда результат работы онлайн-переводчика оказывается неудовлетворительным (содержит ошибки на разных уровнях языка), специалисты так и поступают, полагая, что можно создать качественный перевод, потратив при этом меньше времени, чем при постредактировании. Однако надо сказать, что с развитием нейромашинного перевода качество исходного перевода значительно возросло, в результате чего постредактирование стало в значительной мере полезным и применяется чаще: исследования показывают, что постредактирование фактически более эффективно с точки зрения временных затрат (Guerberof, 2013).

В большинстве случаев от выбора уровня постредактирования (лёгкого или полного) зависит качество перевода. Так, с одной стороны, лёгкое постредактирование связано с «достаточно хорошим качеством» или «просто понятным» текстом, т. е. текстом, который должен быть точно переведён, но который не обязательно должен быть естественным или соответствующим заданной стилистике текста. С другой стороны, полное постредактирование связано с «качеством, аналогичным или равным переводу, созданному переводчиком». В данном случае мы сталкиваемся с некоторыми трудностями, поскольку изначально предполагается, что «человеческий перевод» всегда соответствует высоким стандартам, то есть является эталонным, однако в некоторых случаях это не так (Guerberof, 2013). С тем, чтобы лучше понять сложную взаимосвязь между уровнями постредактирования и уровнями качества перевода, проанализируем работу сервисов онлайн-перевода Google Translate и DeepL.

Сервис *DeepL* – онлайн-переводчик, разработанный в 2016 году на основе нейросетевых технологий. Он использует свёрточные нейронные сети, специальную архитектуру искусственных нейронных сетей, которые обучаются на основе базы данных онлайн-словаря и системы контекстуального поиска Linguee. Разработчики отмечают, что в настоящее время DeepL является самым точным переводчиком в мире среди всех онлайн-переводчиков. Это подтверждается регулярным слепым тестированием, в котором DeepL демонстрирует лидерскую позицию по сравнению с *Google, Yandex u Amazon* (DeepL. <https://www.deepl.com/whydeepl>).

В 2016 году компания *Google* запустила систему нейронного МП (Neural Machine Translation / NMT), способную преодолеть многие недостатки традиционного статистического МП (Statistical Machine Translation / SMT), на котором ранее основывалась система онлайн-перевода с использованием внутренних оценок, проведённых на материале набора изолированных простых предложений. *Google NMT* снизил количество ошибок перевода в среднем на 60% по сравнению с системой фразового производства *Google SMT* на этапах перевода с английского на французский и немецкий языки (Wu, Schuster, Chen et al., 2016, p. 5).

Цель, заявленная в настоящем исследовании, достигается нами на материале английского и русского языков. Мы фокусируем внимание на текстах академического (собственно-научного) подстиля, учитывая при этом его жанровую дифференциацию и работая на материале научной статьи, опубликованной в журнале “*Nature*” (King, 2016). Научное сочинение Э. Кинга “*Rise of Resilience*” посвящено изучению работы мозга в стрессовой ситуации, механизмам негативного влияния стресса на организм и способам противостояния ему.

“*Confronted with a life-threatening situation, hormones and neurotransmitters prep us for action. Specific stress hormones – cortisol in primates, corticosterone in most rodents – are released, some of which surge across the blood-brain barrier. The human brain has two types of receptor for cortisol. The hippocampus (which is pivotal for memory) and the amygdala (the centre for emotions) contain lots of the high-affinity receptors, and are, therefore, activated by slight rises in the hormone. The frontal lobe, which is involved in executive planning and control, has only the low-affinity receptor, and is activated later, after the tide has risen. And, as Lupien and colleagues found, both memory formation and recall in adults can be influenced by cortisol*” (King, 2016).

Рассмотрим примеры работы переводческих онлайн-сервисов. В текстах выделены фрагменты, которые различаются в переводах выбранных нами систем – DeepL и Google Translate (далее – DL и GT соответственно).

DeepL:

При возникновении ситуации, угрожающей жизни, гормоны и нейромедиаторы готовят нас к действию. Выделяются специфические гормоны стресса – кортизол у приматов, кортикостерон у большинства грызунов, – некоторые из которых проникают через гематоэнцефалический барьер. В человеческом мозге есть два типа рецепторов для кортизола. Гиппокамп (играющий ключевую роль в памяти) и миндалина (центр эмоций) содержат большое количество высокоаффинных рецепторов, поэтому они активируются при незначительном повышении уровня гормона. Лобная доля, которая участвует в исполнительном планировании и контроле, имеет только низкоаффинные рецепторы, и активируется позже, после того, как прилив усилится. И, как обнаружили Люпьян и его коллеги, кортизол может влиять как на формирование памяти, так и на ее запоминание у взрослых.

Google Translate:

Столкнувшись с опасной для жизни ситуацией, гормоны и нейротрансмиттеры готовят нас к действию. Высвобождаются специфические гормоны стресса – кортизол у приматов, кортикостерон у большинства грызунов, некоторые из которых преодолевают гематоэнцефалический барьер. Человеческий мозг имеет два типа рецепторов кортизола.

Гиппокамп (отвечающий за память) и миндалевидное тело (центр эмоций) содержат множество высокоаффинных рецепторов и, следовательно, активируются небольшим повышением уровня гормона. Лобная доля, которая участвует в исполнительном планировании и контроле, имеет только низкоаффинный рецептор и активируется позже, после подъема прилива. И, как обнаружили Лупьян и его коллеги, кортизол может влиять как на формирование памяти, так и на воспроизведение памяти у взрослых.

Отличие переводов в небольшом отрывке, предложенном нами к переводу, достаточно сильное. DL, на наш взгляд, представил русский вариант текста в научном стиле, близком к оригиналу, но всё-таки с небольшими недочётами. Во-первых, термин *amygdala* (миндалевидное тело) он предлагает перевести как *миндалина* или *амигдала* – посредством транслитерации. Эквивалентный и адекватный перевод здесь не был достигнут. «Миндалина» – термин, имеющий широкое значение, его употребление здесь не вполне корректно. Таким образом, онлайн-переводчик применил переводческую трансформацию – генерализацию: вместо термина «миндалевидное тело» употребляется «миндалина», что искажает смысл единицы на языке оригинала, а значит, делает перевод неадекватным, неточным.

«...кортизол может влиять как на формирование памяти, так и на ее запоминание у взрослых» – часть предложения сформулирована неграмотно, существительное *запоминание* является не только неподходящей лексической единицей, но и противоположной по смыслу.

Перевод GT оказался наименее адекватным. Представим некоторые неточности и несоответствия. Во-первых, отметим, что переведённый текст нельзя квалифицировать как сугубо научный, в отличие от текста перевода, сгенерированного первым сервисом. Абзац начинается с деепричастного оборота *столкнувшись с опасной для жизни ситуацией*, который, судя по грамматической конструкции и лексическому составу, должен присоединяться к какому-то одушевленному существительному. Адекватным можно считать перевод, предложенный DeepL: *при возникновении ситуации, угрожающей жизни*. Далее во фразе *высвобождаются гормоны* глагол переведён эквивалентно оригинальному *released*, но не адекватно, в этом случае справился DeepL, он перевёл его как *выделяются*, что соответствует лексическим нормам русского языка. Термин *нейромедиатор* адекватно перевёл DL с английского *neurotransmitter*, GT предлагает транслитерацию – *нейротрансмиттер*.

При переводе фразы *which is pivotal for memory* GT теряет важное прилагательное *pivotal* (основной, важный) и предлагает вариант *отвечающий за память*, что неточно описывает работу гиппокампа. DL перевёл фразу адекватно – *играющий ключевую роль в памяти*, сохранив смысл и цель перевода.

Адекватный перевод фразы *the tide has risen* представлен у DL – *прилив усилится*, вариант GT – *после подъема прилива* – некорректный с точки зрения эквивалентности и адекватности перевода.

Таким образом, если сравнивать варианты перевода, которые предлагают два сервиса, то DL, в сравнении с GT, адекватнее переводил отдельные лексические единицы, без критических ошибок, связанных с нормами языка перевода, в зависимости от принадлежности к функциональному – научному – стилю.

В пределах научного стиля, в нашем случае – его академического подстиля, «научная статья занимает центральное место в конструировании и распространении знаний в научном сообществе, представляя собой относительно закрытый жанр, будучи адресованной небольшой специализированной аудитории. В рамках научного дискурса она представляет собой основной способ обмена знаниями и результатами научных достижений между учёными» (Светаилов, 2022, с. 2301). Таким образом, постредактор выступает в роли посредника между автором научной статьи на английском языке и реципиентом МП на русском языке, который является специалистом в области нейробиологии, поэтому степень ответственности постредактора очень высока, и его профессиональная задача состоит в том, чтобы сделать результат работы онлайн-переводчика адекватным. В проанализированном нами тексте были выявлены проблемы МП терминов нейробиологии, что недопустимо, учитывая специфику области применения его результатов. В большинстве случаев применение МП и постредактирование значительно экономят время, затраченное переводчиком, однако при работе с сугубо научными текстами временные затраты увеличиваются, поскольку качество переводного текста в данном случае – основной параметр, который должен улучшить постредактор.

Анализ текстов, сгенерированных двумя онлайн-сервисами, показал наличие в них смысловых и формальных ошибок, и в этой связи представим собственную постредакторскую версию данных текстов:

«При столкновении с опасной для жизни ситуацией гормоны и нейромедиаторы готовят нас к действию. Выделяются специфические гормоны стресса (кортизол – у приматов, кортикостерон – у большинства грызунов), некоторые из которых проникают через гематоэнцефалический барьер. В человеческом мозге есть два типа рецепторов для кортизола. Гиппокамп (центральный орган памяти) и миндалевидное тело (центр эмоций) содержат большое количество высокоаффинных рецепторов и, следовательно, активируются при незначительном повышении уровня гормона. Лобная доля, участвующая в организации планирования и контроля, имеет только низкоаффинные рецепторы и активируется позже – после того, как прилив уже наступил. Как выяснили Люпьян и его коллеги, на формирование и восстановление памяти у взрослых людей может влиять кортизол».

Приведём другой пример работы сервисов МП.

“Stress affects our relationships with others (and socializing is itself an agent in cognitive health). Nestler has devised a ‘bully mouse’ scenario (pictured) in which, for 5 to 10 minutes a day over 10 days, a normal mouse is placed in a cage that is already occupied by a larger, more aggressive strain of mouse that intimidates the incomer. At all other times, the mice are kept close enough to see and smell one another, but they are separated by mesh. Nestler’s team found that afterwards, some of the bullied mice avoid all social contact, even with smaller, non-aggressive mice...” (King, 2016).

DeepL:

«Стресс влияет на наши отношения с окружающими (а общение само по себе является фактором когнитивного здоровья). Нестлер разработал сценарий “мыши-задиры” (на рис.), в котором обычная мышь на 5-10 минут в день в течение 10 дней помещается в клетку, уже занятую более крупной и агрессивной разновидностью мыши, которая запугивает пришедшую. В остальное время мыши находятся достаточно близко, чтобы видеть и чувствовать запах друг друга, но они разделены сеткой. Команда Нестлера обнаружила, что после этого некоторые из запуганных мышей избегают любых социальных контактов, даже с более мелкими и неагрессивными мышами...».

Google Translate:

«Стресс влияет на наши отношения с другими людьми (и общение само по себе является фактором когнитивного здоровья). Компания Nestler разработала сценарий “мыши-хулигана” (на фото), в котором на 5-10 минут в день в течение 10 дней нормальную мышь помещают в клетку, которая уже занята более крупной и агрессивной разновидностью мышей, пугающей животных. входной. В остальное время мышей держат достаточно близко, чтобы видеть и чувствовать запах друг друга, но они разделены сеткой. Команда Нестлера обнаружила, что после этого некоторые из запуганных мышей избегают любых социальных контактов, даже с более мелкими и неагрессивными мышами...».

В данном случае также отмечены те части, которые различаются в текстах – продуктах деятельности двух сервисов. Настоящий пример показывает, насколько по-разному две системы, основанные на нейронном методе, отличаются в предлагаемых ответах, при этом следует учитывать то, что в этом тексте нет терминов или специальной лексики. Отметим, постредактирование необходимо, поскольку ни одна из систем не смогла сгенерировать адекватный перевод. Информация в переводе в обоих случаях донесена верно, без искажений, но языковые конструкции и лексическое наполнение текста не воспроизводят естественность языка перевода.

DL репрезентирует основные идеи автора исходного текста средствами научного стиля речи не в каждом случае. К примеру, в одном случае терминологическое сочетание *normal mouse* переведено им на русский язык как *обычная мышь*, однако корректным будет перевод *обыкновенная мышь*, а в другом случае (в пределах одного абзаца) онлайн-переводчик правильно передаёт значение исходной единицы *bully mouse* – *мышь-задира*: смысл фрагмента текста при этом сохраняется, поскольку далее объяснено, что мышь «запугивает прибывшую».

GT переводит текст эквивалентно, но не адекватно. Сервис предложил единицу *людьми* вместо *окружающими* и интегрировал лишнее уточнение *компания* перед фамилией *Нестлер*, кроме того, сама фамилия сохранена в тексте на исходном языке в латинской графике, то есть сервис применил переводческую трансформацию – генерализацию. Он преобразовал термин *bully mouse* как *мыши-хулигана*, что не вполне точно передаёт значение единицы на исходном языке. Для передачи исходных конструкций GT достаточно часто применяет страдательный залог, что усложняет восприятие текста. Онлайн-переводчик не сумел адекватно перевести часть предложения *more aggressive strain of mouse that intimidates the incomer...*, а это показывает, что сервис не совсем понимает контекст: ...агрессивной разновидностью мышей, пугающей животных. входной.

Представим собственный вариант постредакторской работы, выполненной на основе текстов, сгенерированных онлайн-сервисом:

«Стресс влияет на наши отношения с другими людьми (а общение само по себе является фактором когнитивного здоровья). Нестлер разработал сценарий “мыши-хулигана” (см. фото), в котором на 5-10 минут в день в течение 10 дней обыкновенную мышь помещают в клетку, уже занятую более крупной и агрессивной разновидностью мыши, которая запугивает вновь пришедшую. В остальное время мыши находятся достаточно близко, чтобы видеть и чувствовать запах друг друга, но они разделены сеткой. Команда Нестлера обнаружила, что после этого некоторые из запуганных мышей избегают любых социальных контактов, даже с более мелкими и неагрессивными мышами...».

В данном случае возможно лёгкое и полное постредактирование – в зависимости от цели. При лёгком постредактировании нет необходимости менять предложения и отдельные слова – для общего понимания информации достаточно перевода GT. В случае, если информация, содержащаяся в исходном тексте, необходима не только для личного ознакомления, но, например, для коммерческого использования или научного

воспроизведения, то к тексту GT следовало бы применить полное постредактирование (оно занимает больше времени, чем лёгкое, но снижает риски, связанные с семантическими потерями в процессе перевода), а к тексту DeepL – лёгкое.

Важно отметить, что качество текста, сгенерированного онлайн-сервисом, напрямую влияет на обучаемость модели МП и обуславливает трудоёмкость работы постредактора. Так, обучение модели основывается на предоставлении ей большого количества правильно переведённых параллельных текстов, чтобы она (модель) могла научиться связывать соответствующие фразы и выражения: чем точнее и разнообразнее данные, используемые для обучения модели, тем лучше результаты МП, и, следовательно, тем меньше работы, связанной с постредактированием, требуется впоследствии.

Вышесказанное позволяет сделать некоторые рекомендации, которые, как мы полагаем, могут существенно повысить эффективность работы постредакторов:

1. *Цель постредактирования.* Следует сфокусировать внимание на том, с какой целью выполняется постредактирование. Необходимо учитывать целевую аудиторию продукта постредакторской работы и контекст его применения (личное ознакомление, использование переводного текста для определённой группы реципиентов и т. д.).

2. *Выполнение перевода.* Следует произвести сопоставительно-переводческий анализ оригинального текста и текста, сгенерированного онлайн-сервисом, затем осуществить сравнительный анализ текстов на переводном языке – результатов работы разных онлайн-переводчиков.

3. *Процедура постредактирования.* В зависимости от цели, следует подвергнуть процедуре лёгкого и полного постредактирования результаты работы онлайн-сервиса. Учитывая стилистическую принадлежность исходного текста, определить, какие правки следует внести в переводной текст.

Заключение

Цель настоящего исследования заключалась в выработке рекомендаций для оптимизации англо-русского МП собственно-научных текстов естественно-научной направленности на этапе его постредактирования, ориентированных на улучшение качества МП. В результате проведённой работы мы пришли к следующим выводам. Оценка МП с опорой на его онтологические признаки – адекватность и эквивалентность – ключевой первоначальный этап улучшения его результатов. В случае англо-русского постредактирования переводных текстов необходимо учитывать особенности обоих языков и выполнять коррекцию переводного текста так, чтобы достичь его точности на грамматическом и лексико-семантическом уровнях – сохранить смысл и стиль оригинального текста. Одним из ключевых аспектов процесса постредактирования является обнаружение и исправление лексических и грамматических ошибок, проявляющихся на уровне синтаксиса и лексики, когда машинные переводчики предлагают некорректные синтаксические конструкции или генерируют терминологические единицы, смысл которых отличается от смысла единиц языка оригинала. При постредактировании необходимо провести аудит переводного текста на наличие подобных ошибок и внести соответствующие исправления.

Кроме того, учитывая цель постредактирования и информацию о целевой аудитории, на которую направлен перевод, важно, чтобы постредактор подробно проанализировал стилистические особенности оригинального текста и его широкий контекст, т. е. ситуацию, в которой результат работы будет воспроизводиться. МП чаще всего нацелен на передачу только основного смысла, и нередко онлайн-сервис не учитывает значимые особенности исходного языка, например стилистическую принадлежность.

К дополнительным, но не менее важным задачам постредактирования следует относить улучшение восприятия текста и его структурирование. Машинные переводчики могут в ряде случаев производить громоздкие и неестественные предложения, сложные для восприятия. С тем, чтобы сделать текст более лёгким для понимания, можно осуществить реконструкцию предложений.

В рамках настоящего исследования, проведённого на материале научной статьи по нейробиологии, реализованы следующие исследовательские процедуры: 1) квалифицированы два основных метода постредактирования – лёгкое и полное; 2) с помощью двух сервисов (DeepL и Google Translate) выполнен перевод оригинальных (англоязычных) единиц на русский язык; 3) выделены те фрагменты переводного текста, которые, на наш взгляд, не следует считать адекватными тексту на языке оригинала; 4) произведено лёгкое или полное постредактирование результатов генерации онлайн-сервисов.

В результате было выявлено, что обоим онлайн-сервисам сложнее всего справляться с переводом терминологических единиц с одного языка на другой и передачей стилистических особенностей оригинального текста. Кроме того, показано, что сложность представляет передача грамматических конструкций, специфичных для английской и русской научной речи.

Постредактирование результата МП играет важную роль в связи с качеством перевода и его дальнейшим использованием представителями разной целевой аудитории. Фокусирование постредактора на таких аспектах работы, как выявление ошибок, полученных после генерации, учёт стилистической дифференциации текстов и контекста, а также нацеленность на облегчение параметров восприятия текста, позволяет сделать результат МП адекватным оригинальному тексту. Важно отметить, что постредактирование МП является

итерационным процессом. Это означает, что его следует проводить в несколько этапов: определить целевую аудиторию переводческого продукта, цель его дальнейшего применения, выполнить перевод посредством онлайн-сервиса, осуществить процедуру постредактирования. Во время каждой итерации необходимо анализировать результат предыдущего постредактирования в том случае, если их было несколько, и вносить соответствующие корректировки в постредакторский текст. На наш взгляд, постредактирование следует квалифицировать как самый эффективный механизм улучшения качества результата МП.

Для успешного постредактирования МП сугубо научных текстов важно иметь специфическое знание и опыт в области, на которую направлен перевод. Компетенции в конкретной профессиональной сфере позволяют интерпретировать оригинальный текст и производить постредактирование качественно. Главным образом наше суждение связано с распознаванием и точным переводом специальной лексики и терминологии, которые могут быть не знакомы машине.

Перспективы дальнейшего исследования мы видим в более детальном изучении методологии МП текстов академического подстиля английского и русского языков естественно-научной направленности, а также в выработке широкого спектра рекомендаций для разработчиков сервисов, осуществляющих МП, и постредакторов текстов МП.

Источники | References

1. Андреев Н. Д. Основные направления работы экспериментальной лаборатории машинного перевода // Hutchins W. J. Machine Translation: Past, Present, Future. N. Y., 1986.
2. Апресян Ю. Д. Лингвистическое обеспечение системы французско-русского автоматического перевода ЭТАП-1. IV. Французский синтаксический анализ // Предварительные публикации ПГЭПЛ. М.: Институт русского языка АН СССР, 1983. Вып. 159.
3. Кибрик А. А. Модель автоматического анализа письменного текста (на материале ограниченного военного подъязыка). М.: Изд-во МГУ, 1970.
4. Кулагина О. С. Исследования по машинному переводу. М.: Наука, 1979.
5. Леонтьева Н. Н. База знаний и автоматический перевод (проект многоязычной информационно-справочной системы) // Тезисы докладов международного семинара по машинному переводу «ЭВМ И ПЕРЕВОД 89» (г. Тбилиси, 27 ноября – 2 декабря 1989 г.). М., 1989.
6. Марчук Ю. Н. Проблемы машинного перевода. М.: Наука, 1983.
7. Машинный перевод и прикладная лингвистика. Проблемы создания системы автоматического перевода / под ред. Г. В. Чернова. М., 1986.
8. Розенцвейг В. Ю. Основы общего и машинного перевода. М., 1964.
9. Светайлов Б. В. Лингвистические и социокультурные особенности хеджирования в научном дискурсе (на материале текстов англоязычных научных статей экономической направленности) // Филологические науки. Вопросы теории и практики. 2022. Т. 15. Вып. 7.
10. Утробина А. А. Компьютерная лингвистика и машинный перевод: об истории становления // Вестник Башкирского университета. 2022. Т. 27. № 2.
11. Шаляпина З. М. К проблеме построения формальной модели процесса перевода // Теория перевода и научные основы подготовки переводчиков: мат. всесоюз. науч. конф.: в 2-х ч. М.: МГПИИЯ им. М. Тореза, 1975. Ч. II.
12. Cho K. Neural Machine Translation by Jointly Learning to Align and Translate // arXiv. 2016. Art. 1409.0473. <https://doi.org/10.48550/arXiv.1409.0473>
13. Guerberof A. A. What Do Professional Translators Think about Post-Editing? // Journal of Specialised Translation. 2013. Iss. 19.
14. Kenny D. Machine Translation for Everyone: Empowering Users in the Age of Artificial Intelligence (Translation and Multilingual Natural Language Processing 18). Berlin: Language Science Press, 2022.
15. King A. Neurobiology: Rise of Resilience // Nature. 2016. Vol. 531. <https://doi.org/10.1038/531S18a>
16. Koehn Ph. Neural Machine Translation // Machine Translation. 2021. Vol. 35. <https://doi.org/10.1007/s10590-021-09277-x>
17. LeCun Y., Bengio Y., Hinton G. Y. Deep Learning // Nature. 2015. Vol. 521. <https://doi.org/10.1038/nature14539>
18. Manning C. D. Human Language Understanding & Reasoning // Dædalus. 2022. Vol. 151 (2).
19. Marie B., Fujita A., Rubino R. Scientific Credibility of Machine Translation Research: A Meta-Evaluation of 769 Papers // Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, August 1-6. 2021. <https://doi.org/10.18653/v1/2021.acl-long.566>
20. Wu Y., Schuster M., Chen Zh., Le Q. V., Norouzi M., Macherey W., Krikun M., Cao Y., Gao Q., Macherey K., Klingner J., Shah A., Johnson M., Liu X., Kaiser L., Gouws S., Kato Y., Kudo T., Kazawa H., Stevens K., Kurian G., Patil N., Wang W., Young C., Smith J., Riesa J., Rudnick A., Vinyals O., Corrado G., Hughes M., Dean J. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation // arXiv. 2016. Art. 1609.08144. <https://doi.org/10.48550/arXiv.1609.08144>

Информация об авторах | Author information**RU****Хромова Анна Андреевна¹****Лукманова Рената Разифовна²**, к. филол. н., доц.^{1,2} Уфимский университет науки и технологий**EN****Khromova Anna Andreevna¹****Lukmanova Renata Razifovna²**, PhD^{1,2} Ufa University of Science and Technology¹ anautro@mail.ru, ² renata89373319690@gmail.com**Информация о статье | About this article**

Дата поступления рукописи (received): 16.11.2023; опубликовано online (published online): 27.03.2024.

Ключевые слова (keywords): машинный перевод; англо-русский перевод; постредактирование; лёгкое постредактирование; полное постредактирование; machine translation; English to Russian translation; post-editing; light post-editing; full post-editing.