



УДК 81'322.4

EDN VASDBH

<https://doi.org/10.33910/2686-830X-2024-6-3-172-183>

Постредактирование машинного перевода как студенческий проект (на материале художественного текста)

К. В. Богданова ¹

¹ Российский государственный педагогический университет им. А. И. Герцена,
191186, Россия, г. Санкт-Петербург, наб. реки Мойки, д. 48

Сведения об авторе

Ксения Витальевна Богданова,
SPIN-код: 6977-6319,
ResearcherID: K-6461-2018,
ORCID: 0000-0003-4660-1373,
e-mail: ksuschannna@yandex.ru

Для цитирования:

Богданова, К. В. (2024)
Постредактирование машинного
перевода как студенческий проект
(на материале художественного
текста). *Исследования языка
и современное гуманитарное
знание*, т. 6, № 3, с. 172–183.
<https://doi.org/10.33910/2686-830X-2024-6-3-172-183>
EDN VASDBH

Получена 3 июля 2024; прошла
рецензирование 19 августа 2024;
принята 4 сентября 2024.

Финансирование: Исследование
не имело финансовой поддержки.

Права: © К. В. Богданова (2024).
Опубликовано Российским
государственным педагогическим
университетом им. А. И. Герцена.
Открытый доступ на условиях
лицензии CC BY-NC 4.0.

Аннотация. В статье рассматривается постредактирование машинного перевода (РЕМТ) как ключевой актуальный навык, который рекомендуется культивировать у начинающих переводчиков на фоне стремительного развития технологий автоматизации перевода. После вводного экскурса в историю машинного перевода (МП), начиная с первых экспериментов середины XX века и заканчивая нейросетями, автор обращается к МП художественной литературы в контексте практических занятий по РЕМТ. Ввиду того, что именно в художественном тексте сконцентрированы основные «слабые места» МП — эмоционально-оценочные коннотации, единицы с несвободной сочетаемостью, имитация живой разговорной речи и прочее — данный функциональный стиль может стать источником богатого наглядного материала для выявления, систематизации и исправления ошибок МП. Автор приводит конкретный случай из собственной педагогической практики: студенты третьего курса бакалавриата (направление подготовки «Перевод и переводоведение») успешно осуществили РЕМТ рассказа Надежды Тэффи «Рубин принцессы» (1905 г.), который был автоматически переведен с русского языка на американский английский. В ходе работы на базе специализированной платформы Phrasе начинающие лингвисты столкнулись с различными ошибками МП, классификация которых представлена в настоящей статье. Кроме нескольких единичных проблем с транслитерацией имен собственных, удалось выделить три группы повторяющихся ошибок в МП художественного текста: пунктуационные ошибки (55 шт.), лексико-семантические ошибки (67 шт.) и грамматические ошибки (13 шт.). Автор сопровождает типичные ошибки МП исправлениями, которые были предложены студентами в ходе РЕМТ, и в отдельных случаях предлагает собственные варианты. Неоднократно подчеркивается связь между вышеперечисленными ошибками и особенностями художественного функционального стиля, что оправдывает включение РЕМТ художественной литературы в число практических заданий для подготовки переводчиков.

Ключевые слова: постредактирование машинного перевода, автоматизация перевода, перевод художественной литературы, ошибки перевода, подготовка переводчиков, перевод с русского языка на английский

Post-editing machine translation as a student project, exemplified by literary fiction

К. В. Богданова 

¹ Herzen State Pedagogical University of Russia, 48 Moika Emb., Saint Petersburg 191186, Russia

Author

Kseniia V. Bogdanova,
SPIN: 6977-6319,
ResearcherID: K-6461-2018,
ORCID: 0000-0003-4660-1373,
e-mail: ksuschannna@yandex.ru

For citation: Bogdanova, K. V. (2024) Post-editing machine translation as a student project, exemplified by literary fiction. *Language Studies and Modern Humanities*, vol. 6, no. 3, pp. 172–183. <https://doi.org/10.33910/2686-830X-2024-6-3-172-183>
EDN VASDBH

Received 3 July 2024; reviewed 19 August 2024; accepted 4 September 2024.

Funding: The study did not receive any external funding.

Copyright: © K. V. Bogdanova (2024). Published by Herzen State Pedagogical University of Russia. Open access under [CC BY-NC License 4.0](https://creativecommons.org/licenses/by-nc/4.0/).

Abstract. The article examines post-editing machine translation (PEMT) as a key relevant skill that needs to be nurtured in aspiring translators amid the rapid development of translation automation technologies. Following a brief historical background of machine translation (MT), from the very first mid-20th century experiments to neural networks, the author turns to literary fiction in the context of PEMT tasks for students. Taking into consideration that literary texts in particular abound in the main MT “pitfalls” — words with emotional and evaluative connotations, idiomatic expressions, emulation of natural colloquial speech, and more — fiction is capable of providing rich illustrative material for identifying, systematizing, and correcting MT errors. The author cites a personal case study, wherein third-year BA students majoring in Translation and Translation Studies successfully performed PEMT on Nadezhda Teffi’s short story, *The Princess’s Ruby* (1905), which had been automatically translated from Russian into American English. While working with the Phrase translation platform, students encountered various MT errors, the classification of which is presented in this study. Apart from a few isolated issues with the transliteration of proper names, this article identifies three major groups of reoccurring errors in the literary text’s MT: punctuation errors (55 instances), lexical and semantic errors (67 instances), and grammatical errors (13 instances). The author accompanies these errors with corrections proposed by the students during PEMT and, in some cases, offers alternatives. The study repeatedly stresses that the recurrence of the aforementioned errors is tied to the unique features of literary fiction, which can justify including PEMT of literary texts into the roster of practical tasks for translator training.

Keywords: post-editing of machine translation, translation automation, literary translation, translation errors, translator training, Russian to English translation

Введение

Постоянный рост числа текстов из различных профессиональных областей, нуждающихся в переводе, требует поиска инструментов для их обработки в максимально сжатые сроки. Именно это позволяет делать машинный перевод (МП) — использование компьютерной техники для преобразования текста на одном естественном языке в текст на другом естественном языке, эквивалентный ему по содержанию (Маматкасимова 2022, 52).

У истоков МП стоят несколько научно-технических разработок: «машина для подбора и печатания слов», спроектированная П. Троянским в 1930-е гг. (История машинного перевода... 2019); меморандум У. Уивера от 1949 г., вдохновленный, в частности, успехами криптографии в годы Второй мировой войны и доказавший осуществимость перевода при помощи ЭВМ (Hutchins 1999); и, наконец, знаменитый Джоржтаунский эксперимент 1954 г. — первая в истории успешная операция по автоматиче-

скому переводу нескольких десятков предложений с русского языка на английский, которую компьютер модели IBM-701 осуществил с опорой на небольшую базу данных из 250 слов и 6 синтаксических правил (Жданова, Ефремова 2017). Именно к этому эксперименту восходит самая ранняя и, несмотря на успех первоначальной демонстрации, наименее продуктивная разновидность МП — перевод на основе правил (Rule Based Machine Translation, RBMT). В основе RBMT лежат четкие грамматические правила для языка оригинала и для языка перевода, встраиваемые, наряду с двуязычным словарем, непосредственно внутрь системы МП (Rule-based... 2024).

После того, как RBMT не оправдал себя, будучи недостаточно точным и при этом требуя значительных временных затрат, были предприняты попытки его усовершенствовать. В 1980-е гг. японские исследователи разработали так называемый «перевод на основе примеров» — Example Based Machine Translation, или EBMT — используя уже не набор грамматических правил

и словарь, а пары типовых предложений на языке оригинала и перевода (Ясаревская 2017). На смену ЕВМТ пришел сначала статистический перевод, в процессе которого компьютер оперирует гораздо большими по объему двуязычными корпусами, и, наконец, гибридный перевод. Как отмечает Л. Н. Беляева, все современные системы МП, в сущности, являются гибридными, сочетая использование «архивов» из предыдущих переводов с реализацией МП «для тех фрагментов текста, которые в этих архивах не найдены» (Беляева 2019, 13). В частности, начиная с 2016 г. в гибридном переводе активно используются нейронные сети — самообучаемые вычислительные алгоритмы, способные имитировать центральную нервную систему живых организмов, в том числе и человека (Иванько и др. 2019).

Итак, сегодня, спустя три четверти века после меморандума Уивера и Джорджтаунского эксперимента, благодаря усилиям лингвистов и математиков, а также стремительному распространению и совершенствованию компьютерных технологий, в распоряжении пользователей находится широчайший спектр инновационных инструментов МП: от специального программного обеспечения (ПО) до онлайн-сервисов для рядовых пользователей и корпоративных клиентов. Как следствие, все большая роль в арсенале переводческих задач отводится процедуре, известной как РЕМТ (post-editing/post-edited machine translation) — т. е. вычитке и исправлению результатов МП (транслята) усилиями лингвиста-человека (Камшилова, Беляева 2023). В прејскуранте переводческих фирм РЕМТ нередко выделяется как самостоятельная услуга (Чақырова 2013). Более того, исходя из личного опыта автора в сфере письменного перевода, отдельные бюро, привлекающие к сотрудничеству лингвистов-фрилансеров, практически полностью отошли от выполнения проектов для своих корпоративных заказчиков методом конвенционального перевода и предлагают своим исполнителям исключительно задачи на РЕМТ. Таким образом, нельзя не заключить, что обучение РЕМТ должно стать частью профессиональной подготовки студентов лингвистических и, в частности, переводоведческих специальностей.

Художественный текст как объект РЕМТ

Потребность в РЕМТ возникает потому, что, невзирая на огромный путь, преодоленный от самого раннего РВМТ до нынешних гибридных

систем, процесс МП по-прежнему несовершенен и чреват характерными ошибками. Так, МП носит «локальный» характер: многие платформы МП разбивают исходный текст на отдельные предложения и даже фрагменты предложений — сегменты, каждый из которых воспринимается как самодостаточная изолированная сущность, оторванная от общего контекста. Это затрудняет анализ связей внутри сверхфразового единства и, тем самым, препятствует корректному соотношению местоимений-заместителей с их антецедентами, вызывает «путаницу» с синонимами и омонимами и т. д. Такой принцип перевода чреват и тем, что фразеологические единицы (ФЕ) в тексте оригинала будут переводиться путем калькирования — буквального воспроизведения внутренней структуры ФЕ лексическими средствами переводящего языка (Беляева 2019; Шукаева, Шидловская 2022). Основная отличительная особенность ФЕ — невозможность «вывести» общее значение единицы из суммы индивидуальных значений ее компонентов — при этом будет игнорироваться.

Кроме того, наличие в тексте труднораспознаваемых слов, таких как не общеизвестные имена собственные или «нестандартная» (архаичная, сниженная, диалектная) лексика, вызывает сбои в анализе предложения машинным алгоритмом и может привести к транслитерации в тех местах, где она не оправдана, или же неправильному определению имен собственных как нарицательных (Беляева 2019). Нельзя забывать и о том, что во многих системах МП существуют ограничения на максимально допустимую длину предложения, при которой синтаксическая структура воспроизводится корректно. Иными словами, ближе к концу развернутых сложноподчиненных конструкций возрастает вероятность «галлюцинаций» — ситуаций, когда алгоритм обнаруживает в анализируемой информации несуществующие элементы и генерирует ложные или бессмысленные выходные данные (What are AI... 2024).

В данной статье рассматривается использование художественного текста в задании на РЕМТ для студентов. Подобный выбор материала может на первый взгляд показаться не вполне обоснованным, т. к. МП в реальной практике нередко используется для работы с документами, отличающимися высокой степенью повторяемости и стандартизованности и выстроенными на устоявшейся терминологии и клишированных оборотах, с которыми современные системы МП справляются весьма успешно, особенно при грамотной интеграции

гlossариев и переводческой памяти. Художественные произведения, напротив, обладают целым рядом характеристик, затрудняющих МП — это не только и не столько длинные предложения (их в изобилии можно встретить и в текстах официально-делового стиля), но и различные стилистические приемы, «языковые игры» с синонимами, омонимами, ФЕ и прочее. Однако мы полагаем, что в силу особенностей художественного текста типичные неудачи МП предстанут в нем наиболее наглядно, что даст студентам, впервые имеющим дело с РЕМТ, прекрасную возможность начать тренироваться в выявлении ошибок и их сопровождении лингвистическими комментариями и анализом.

Заметим также, что в спектр материалов от корпоративных заказчиков, предлагаемых на РЕМТ профессиональным лингвистам, помимо «шаблонной» документации входят и маркетинговые тексты, близкие к художественным, поскольку реклама активно «заимствует» выразительные средства других функциональных стилей и «переподчиняет» их своей основной цели — воздействию на потенциальных покупателей (Морозова 2020, 3). Таким образом, РЕМТ художественных текстов может подготовить студентов и к РЕМТ рекламы.

Наконец, нельзя исключить и возможность того, что экономии средств и человеко-часов будет уделено больше внимания, чем качеству конечного литературного продукта, и в ближайшее время МП будет подвергаться все больше художественных текстов, с учетом текущей популяризации так называемого «вычислительного творчества» (Степаненко и др. 2020). Все вышесказанное делает целесообразным включение упражнений на РЕМТ художественного текста в программу практических занятий по переводу.

Если обратиться к двум типам РЕМТ, которые выделяет Ю. И. Чакырова (Чакырова 2013) и рассматривает А. О. Ушакова (Ушакова 2022), РЕМТ художественного текста относится к полному, или конвенциональному постредактированию. В отличие от неполного постредактирования, которое исследователи также называют «легким» или «быстрым», полное постредактирование подразумевает не просто беглое знакомство с транслятом с целью удостовериться, что конечный текст более или менее понятен читателю, невзирая на «стилистические ошибки и странно звучащие предложения» (Ушакова 2022, 71) — напротив, здесь преследуется цель превратить транслят в точное и законченное сообщение, неотличимое от созданного человеком и соответствующее всем требованиям к высоко-

качественному тексту соответствующего функционального стиля.

Можно заключить, что из перечня навыков постредактора, разработанного А. О. Ушаковой (Ушакова 2022), в РЕМТ художественного текста задействованы следующие: оценка степени собственного вмешательства в транслят; поиск ошибок МП и определение степени их критичности; дальнейшая обработка транслята в соответствии с техническим заданием (ТЗ) (например, при работе с литературным произведением в ТЗ можно включить восстановление художественной образности исходного текста в тех местах, где она была утеряна в силу несовершенства МП). Оценка качества МП также приучает начинающих переводчиков к самоконтролю.

Практический кейс: студенческий проект

Во II семестре 2023/2024 учебного года автор настоящей статьи предложил группе студентов третьего курса, обучающихся по направлению подготовки бакалавров «Перевод и переводоведение» в Институте иностранных языков РГПУ им. А. И. Герцена, осуществить РЕМТ художественного произведения в языковой паре «русский — американский английский». Для студенческого проекта был выбран рассказ «Рубин принцессы», написанный в 1905 г. выдающейся писательницей Н. А. Лохвицкой (1872–1952), более известной под псевдонимом Надежда Тэффи. Рассказ представляет собой субъективное повествование от третьего лица, раскрывающее внутренние переживания главного героя — Евгения Андреевича Руданова, который с глубоким нежеланием, раздражением и даже отвращением общается с дочерью покойного знакомого, одинокой и крайне болезненной девушкой по имени Аглая и прозвищу «Принцесса», пока последняя не умирает, пробудив в Руданове запоздалое чувство вины (Тэффи 1998).

Исходный текст содержит потенциально сложные для транслитерации имена собственные с фонемой «йот» /j/ (Евгений Андреевич, Аглая Николаевна), а также характеризуется лексическим многообразием и яркими речевыми портретами различных персонажей. Данные аспекты, как неоднократно отмечалось выше, входят в «зону риска» МП.

Работа с текстом осуществлялась на базе Phrase (Phrase TMS) — платформы для совместной деятельности переводчиков, редакторов и менеджеров переводческих проектов (Phrase

TMS... 2024). Все глоссарии, предыдущие переводы и другие материалы хранятся в облаке и доступны в любое время с любого устройства, а менеджер проекта — в данном случае в его роли выступал преподаватель — может следить за процессом перевода в режиме реального времени. В число возможностей, предоставляемых пользователям Phrase, также входит инструмент МП — Phrase Language AI.

Объем исходного текста, составляющий 4601 слово (или 27 994 знака с пробелами), является вполне достаточным, чтобы разбить рассказ на фрагменты и назначить на каждый из них студента из группы в качестве исполнителя. Такой режим работы с текстом позволяет симулировать реальный проект, в рамках которого лингвисты переводят разные части текста параллельно друг с другом, а затем результаты их работы сводятся воедино и проходят финальный контроль качества. К моменту создания проекта студенты уже получили определенный опыт работы с Phrase в рамках изучения второго иностранного языка, однако пока не имели возможности использовать Phrase Language AI, что позволило преподавателю интегрировать РЕМТ в учебный процесс как принципиально новое задание.

Характерные ошибки МП художественного текста

Сразу необходимо отметить, что перед началом работы у преподавателя и студентов возникли подозрения, что имена собственные в тексте могут быть транслитерированы неединообразно (например, имя отца героини, Николай Уздольский, в различных сегментах транслята теоретически могло бы быть отражено как *Nikolaj Uzdolskij*, *Nicholas Uzdolsky* и т. п.). В целях контроля над соблюдением единообразия студентам было предложено потренироваться в составлении базы терминов средствами Phrase. Фрагмент данной базы представлен на рисунке 1.

Однако в итоге студенты обнаружили, что Phrase Language AI удалось «не сбиться» при обработке имен собственных, которые были транслитерированы идентично на протяжении всего повествования, несмотря на принцип «локальности» МП и даже на то, что исходное произведение было разбито на несколько файлов. В англоязычном тексте фамилии, имена и отчества наиболее часто упоминаемых персонажей повсеместно переданы следующим образом: *Nikolai Uzdolsky*, *Aglaya Nikolaevna Uzdolskaya*, *Evgeny Andreevich Rudanov*. Однако

☰ Базы терминов / Ruby

Язык	Термин	Статус	Создан
<input type="checkbox"/> en ^{US}	Aglaya	Утверждено	07.05.2024 14:53
<input type="checkbox"/> ru	Аглая	Утверждено	07.05.2024 14:53
<input type="checkbox"/> en ^{US}	Rudanov	Утверждено	07.05.2024 14:50
<input type="checkbox"/> ru	Руданов	Утверждено	07.05.2024 14:50
<input type="checkbox"/> en ^{US}	Uzdolskaya	Утверждено	07.05.2024 14:55
<input type="checkbox"/> ru	Уздольская	Утверждено	07.05.2024 14:55

Рис. 1. Фрагмент базы терминов в интерфейсе Phrase

Fig. 1. Fragment of the term base in the Phrase interface

псевдоним самой писательницы, являющийся самым первым сегментом исходного текста, не был корректно распознан и преобразовался из «Надежда Тэффи» в транслят *Taffy's Hope* (рис. 2).

Таким образом, уже в самом начале выполнения задачи по РЕМТ студентам пришлось столкнуться с ранее упомянутой ситуацией, когда имя собственное неверно распознается как нарицательное. Также можно отметить, что несмотря на отсутствие проблем с транслитерацией Ф. И. О. главных героев, студенты обнаружили ошибку в одном из сегментов, содержащих фамилию второстепенного персонажа — некой Никаноровой, у которой Аглая снимает жилье. Никанорова лично не участвует в событиях рассказа, и ее фамилия упоминается в тексте лишь трижды. В последнем из этих трех случаев алгоритм МП проигнорировал типичное окончание родительного падежа женского рода и преобразовал женскую фамилию в мужскую: таким образом, исходному сегменту «На двери не было больше имени Никаноровой» соответствует транслят *Nikanorov's name was no longer on the door*.

Помимо обозначенных единичных и сразу бросающихся в глаза недочетов МП, в ходе дальнейшей совместной работы с текстом и дискуссий между студентами и преподавателем были выявлены и определенные повторяющиеся, системные ошибки, которые можно объединить в три основные группы.

Пунктуационные ошибки

К первой группе студенты и преподаватель отнесли пунктуационные ошибки. Их общее число достигло 55, и их можно было встретить практически в каждом сегменте транслята, где присутствовали диалоговые реплики. Столь высокая частотность также вызвана локальным характером МП. Поскольку каждое предложение воспринимается алгоритмом как самостоятельный сегмент, открывающая/закрывающая кавычка ставится только в тех местах, где в исходном русскоязычном тексте присутствует тире, указывающее на начало диалога — даже невзирая на то, что каждая реплика может состоять из нескольких предложений. Наглядно такое оформление диалога представлено на рисунке 3.

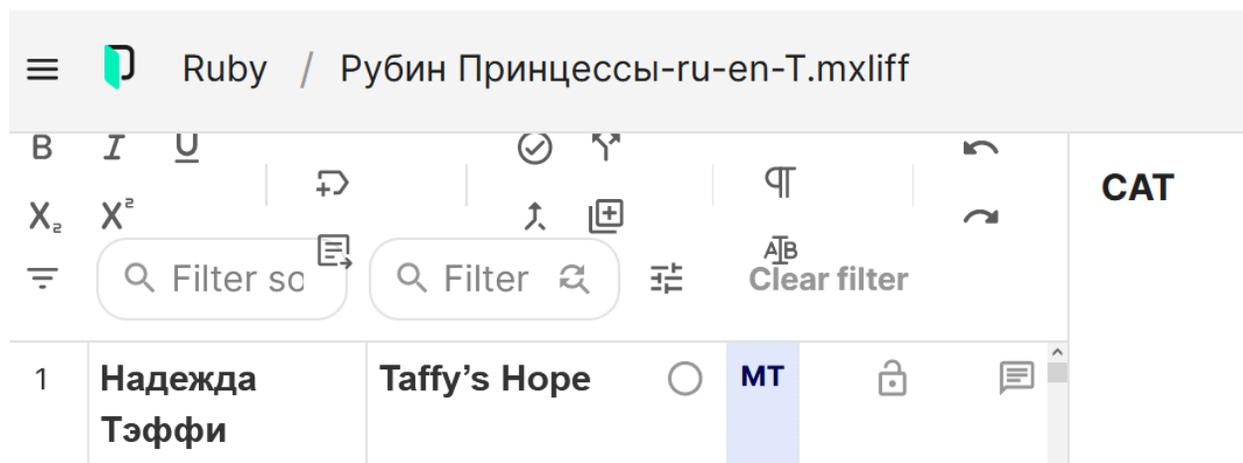


Рис. 2. Первый сегмент художественного текста в интерфейсе Phrase

Fig. 2. The first segment of the literary text in the Phrase interface

30	— Вы не помните меня, Принцесса?	"Don't you remember me, Princess?"	○	MT
31	Я Руданов.	I'm <u>Rudanov</u> .	○	MT
32	Мы раз с вами целый вечер болтали.	We once chatted all evening.	○	MT

Рис. 3. Оформление диалога в интерфейсе Phrase

Fig. 3. Dialogue rendition in the Phrase interface

Отметим, что платформа Phrase оснащена функцией контроля качества (QA Check), помогающей обнаружить в тексте орфографические и пунктуационные ошибки, лишние пробелы и прочее. Однако QA Check не реагирует на пунктуацию, подобную указанной на рисунке 3, т. е. с формальной точки зрения, правила оформления диалога соблюдены: в сегменте № 30 присутствуют кавычки. Таким образом, исправления необходимо вносить вручную, что помогает прорабатывать навык самоконтроля.

Лексико-семантические ошибки

Не менее объемную группу (67 случаев) составили ошибки лексико-семантического характера, которые, в свою очередь, можно подразделить на несколько подгрупп. В частности, при поиске соответствий в языке перевода алгоритм МП склонен игнорировать критически важные семы исходных единиц. Всего в трансляте удалось насчитать 39 таких ошибок.

Так, персонаж Аглаи вводится в повествование как «маленькая худенькая девочка, старательно подчищавшая **резинкой** переписанную рукопись». Ее описание изначально было переведено как *a small thin girl who was diligently cleaning the copied manuscript with an **elastic band***. В данном случае исходная единица «резинка» реализует словарное значение «небольшой кусок особой резины, предназначенный для стирания, написанного на бумаге», однако алгоритм МП подобрал иноязычное соответствие для данной единицы в словарном значении «резиновая нитка, тесемка или кольцо, растягивающиеся при натяжении» (Кузнецов 1998). В процессе РЕМТ студентами были предложены варианты перевода *rubber* или *eraser*. К аналогичным ошибкам можно отнести и МП исходного сегмента «вы **сочиняете**» (в значении «выдумываете») как *you are **composing*** (исправлен студентом на *you are **making it up***; также приемлемым является перевод *you are **fibbing***) или МП «[в рукописи] ровно тридцать листов» как *exactly thirty **leaves*** (у студента — *exactly thirty **sheets***).

Не всегда успешно передается средствами МП и характерная для художественного функционального стиля лексика с эмоционально-оценочными коннотациями. Примером здесь может послужить словосочетание «узенькая **кроватька**» из описания комнаты Аглаи. Ему соответствует транслят *a narrow **crib***, очевидно некорректный по ряду причин.

Обращаясь к классификации разновидностей информации в тексте, предложенной И. Р. Гальпериным (Гальперин 2007), приходим к следующему заключению: лексическая единица «кроватька» содержит в себе не только явную,

содержательно-фактуальную информацию, или СФИ (указание на небольшой размер кровати), но и скрытую, содержательно-подтекстовую информацию (СПИ), которую можно извлечь из окружающего контекста. Аглая живет весьма бедно, и Руданов рассматривает аскетичный интерьер ее комнаты со смешанными ощущениями жалости и омерзения — именно такие эмоционально-оценочные коннотации заложены в единицу «кроватька». Очевидно, что алгоритм «обнаружил» лишь СФИ, следствием чего стала «галлюцинация»: в трансляте появилась не просто кровать небольшого размера, а кровать для маленького ребенка (*crib*), какие-либо отсылки к которой отсутствуют в исходном повествовании. Вариант перевода *tiny narrow bed*, предложенный студентами, можно счесть приемлемым, однако как альтернативу можно было бы использовать перевод наподобие *meager narrow bed*, поскольку прилагательное *meager* несет в себе отрицательную оценку.

Встречаются и сегменты, где в трансляте была подобрана лексика с определенным эмоционально-оценочным компонентом, однако характер оценки не соответствует оригиналу. Сравним, например, следующее высказывание из оригинального текста: «Руданов **грустно усмехнулся** и провел рукой по своим густым седеющим волосам» с МП: *Rudanov **grinned sadly** and ran his hand through his thick graying hair*. Толковый словарь Merriam-Webster определяет глагол *to grin* как *to draw back the lips so as to show the teeth, especially in amusement or laughter* ('широко, **весело** улыбаться, показывая зубы') (Merriam-Webster Dictionary 2024). Иными словами, в семантику глагола заложены ассоциации с хорошим настроением, что никак не соответствует смыслу исходного высказывания; словосочетание *grinned sadly* можно счесть оксюмороном. В процессе РЕМТ было предложено заменить *grinned* на *chuckled*; в том же словаре Merriam-Webster находим: *to chuckle is to laugh inwardly or quietly* ('смеяться негромко, про себя') (Merriam-Webster Dictionary 2024). Наличие семы *quietly* ('тихо, негромко'), несомненно, способствует достижению большей переводческой адекватности. Допустимым вариантом перевода может стать и *smiled wryly* — словосочетание, использующееся в англоязычной художественной литературе в схожих контекстах, о чем, например, свидетельствуют результаты поиска по корпусу COCA (Corpus of Contemporary American English 2024). На рисунке 4 можно увидеть один из 138 примеров, представленных в COCA: сцену из художественного произведения, где герой с горечью и сожалением напоминает героине, как давно они не ужинали вместе.



Source information:

Source	FIC: Analog Science Fiction & Fact
Date	2016
Publication information	; New York Vol. 136, Iss. 11, (Nov 2017): 69-76.
Title	The Desolate Void
Author	Werkheiser, Jay

Expanded context

? " # " Hmm? " She distractedly dropped her frilly napkin onto the remnants of lobster shell on her plate. # " We need to talk. " # The quiet din of nearby mealtime conversations faded into the background. Those words. " What is it? Your funding didn't get approved? " # Tomas sighed sadly. " No, that's not it. " # " What then? Your research isn't- " # " It's not about my God damned research. " His voice raised as the sentence went on. He ducked his head and said softly, " It's about us. " # Her heart thumped. " Us? " Then it hit her. " We haven't gone out like this in a long time. " # " I did try. " He smiled **wryly**. " But you never made time. " # " My work is important. " # " More important than us. " It wasn't a question.

Рис. 4. Фрагмент художественного текста из поисковой выдачи в корпусе COCA по запросу *smil* wryly*

Fig. 4. Fragment of a work of literary fiction from the COCA search results returned for the *smil* wryly* query

Говоря о неверной передаче семантики при МП, нельзя не обратить внимание и на ошибки, возникшие в ходе «обработки» ФЕ и иных единиц с несвободной сочетаемостью компонентов. В рамках проекта студенты обнаружили 14 таких ошибок. Пожалуй, наиболее примечательным примером, вызвавшим у студентов эмоциональный отклик ввиду откровенно комичного несоответствия оригинала и перевода, стала сцена, где Даша, служанка квартирной хозяйки, возмущенно реагирует на обвинения в краже со стороны Руданова и Аглаи: «Она [Аглая] сама неряха, а **на людей грех наводит!**». Как и следовало ожидать, в процессе МП исходное выражение было калькировано, что, учитывая несвободную сочетаемость его компонентов, привело к явному искажению смысла: *She is a slob herself, but she brings sin on people*. Чтобы восстановить исходный смысл и повысить качество транслата, студент, ответственный за работу с соответствующим фрагментом текста, остановился на более адекватном описательном переводе: *She herself is a slob, but accuses others of it*.

Кроме единиц с устойчивым компонентным составом, МП не поддались и речевые ошибки, характеризующие речь персонажей низкого сословия, таких как вышеупомянутая служанка Даша и подмастерье портного. В частности, Даша произносит слово «брильянты» как «бральянты» («Никаких я **бральянтов** не брала», «Где у меня тут бральянты лежат?»), а подмастерье использует широко известный диалектизм «пинжак» («Хозяин нашел [это] в вашем **пинжаке**»). В транслате реплики персонажей

превращаются в *I didn't take any **bralance**, Where do I have **bralance** here?* и *The owner found it in your **pinjac*** соответственно. Стремление сохранить исходную орфографию путем транслитерации в данном случае не имеет смысла, поскольку подобный перевод не только приводит в замешательство англоязычного читателя, но и игнорирует прагматический аспект — создание яркого портрета персонажа через идиолект. Студенты, осуществлявшие РЕМТ, заменили транслитерированные новообразования на общелитературные лексические единицы: *I didn't take any **diamonds**, Where do I have **diamonds** here?, My master found it in your **jacket***. Это облегчает понимание текста, однако для восстановления художественной образности предпочтительно было бы прибегнуть к переводческой компенсации и ввести в текст диалектизмы, характерные для английского языка, причем не обязательно на лексическом уровне (*I didn't take no diamonds, Where would I have them diamonds?, My master's done found it in your jacket*).

Лексико-семантические ошибки, вызванные неверным соотношением местоимения и антецедента, также были выделены в отдельную подгруппу из 11 примеров. В основном алгоритм «сбивался» при попытке определить, является ли антецедент одушевленным (т. е. требующим местоимение *he* или *she*) или неодушевленным (требующим местоимение *it*). Например, Аглая говорит о своем фамильном сокровище — заглавном «рубине Принцессы»: «**Он** остался от моей матери». При МП связь с антецедентом «рубин» установлена не была, в результате чего

получаем: *He was left by my mother*. При работе над данным сегментом студент успешно отредактировал транслят, не только использовав верное местоимение, но и преобразовав грамматическую структуру для легкости чтения: *I got it from my mother*.

Грамматические ошибки

Наконец, кроме пунктуационных ошибок и нескольких подтипов лексико-семантических ошибок, в процессе МП были допущены и грамматические ошибки. Их удалось обнаружить в 13 случаях, и они, как правило, заключаются в неверном выборе видо-временной формы глагола.

Повествование в рассказе ведется в прошедшем времени, однако в тех сегментах, где в авторской речи в оригинале опущено сказуемое, образовавшуюся лакуну в англоязычном тексте заполняет глагол в Present Simple, что нарушает согласование видо-временных форм в сверхфразовом единстве и выбивается из общей нарративной канвы. Например, вернемся к описанию скудной обстановки в комнате Аглаи: «Руданов **оглядел** комнатку Принцессы. <...> На подоконнике чернильница. <...> В углу <...> — платяя ...». В трансляте односоставные предложения преобразованы в двусоставные с использованием конструкции *there is*: *Rudanov looked around the Princess's room*. <...> *On the windowsill there is an inkwell*. <...> *In the corner <...> there are dresses*. После того, как студент осуществил РЕМТ, транслят приобрел следующий вид: *Rudanov looked around the Princess's little room*. <...> *On the windowsill there was an inkwell*. <...> *In the corner <...> hung dresses*. Видо-временные формы согласованы, во втором предложении конструкция *there is* заменена на глагол *to hang* ('висеть') в видо-временной форме Past Simple, что оправдано, поскольку в исходном тексте также упоминаются «вбитые в стену гвозди».

Машинный алгоритм неверно интерпретирует и односоставные предложения в диалогах. Так, встретив Руданова на работе и приглашая его в гости, Аглая говорит: «**Я дома** с девяти часов». В МП используется видо-временная форма Present Perfect (*I've been home since nine o'clock*), указывающая на то, что героиня уже вернулась с работы в девять и до сих пор находится дома. Однако это не согласуется с исходным контекстом: героиня сообщает о своей каждодневной рутине, что требует видо-временной формы Present Simple, корректно использованной студентом в ходе РЕМТ: *I come home at nine o'clock*.

Допустимые варианты МП

В заключение заметим, что несмотря на многочисленные ошибки, освещенные и классифицированные выше, некоторые фрагменты художественного текста были переведены вполне приемлемо. Это прежде всего касается самых «элементарных» высказываний, которые имеют синтаксическую структуру простых двусоставных предложений с минимальным числом второстепенных членов, содержат в основном СФИ и могут быть с легкостью интерпретированы даже в отрыве от более широкого контекста. Например: «Принцесса уже не плакала» — *The Princess was no longer crying*; «Она молча покачала головой» — *She shook her head silently* и т. п. Перевод таких высказываний может быть интегрирован в конечную версию целевого текста без изменений. Понимание, в каких случаях транслят не требует излишнего вмешательства, даже при полном постредактировании, также входит в число ценных навыков, необходимых специалисту по РЕМТ.

Будущие перспективы

В дальнейшем представляется продуктивным предложить студентам для постредактирования другие художественные произведения, включая как и рассказы Надежды Тэффи, где — в противоположность лирическому тону «Рубина принцессы» — в большей степени проявляется ее идиостиль как писательницы-сатирика, так и работы других авторов, где демонстрируются уникальные художественные приемы, плохо поддающиеся МП. В результате может быть собран обширный наглядный материал — фрагменты исходного текста, их МП, комментарии и исправления, внесенные студентами и преподавателем — систематизация которого способна лечь в основу будущего практического пособия по РЕМТ.

Список сокращений

- ФЕ — фразеологические единицы
- МП — машинный перевод
- ПО — программное обеспечение
- СПИ — содержательно-подтекстовая информация
- СФИ — содержательно-фактуальная информация
- ТЗ — техническое задание

List of Abbreviations

COCA — Corpus of Contemporary American English
 EBMT — Example Based Machine Translation
 MT — machine translation
 PЕMT — post-editing/post-edited machine translation
 RBMT — Rule Based Machine Translation

Благодарности

Автор выражает искреннюю благодарность студентам, обучающимся по специальности «Перевод и переводоведение», в частности, старостам групп: Старостиной Юлиане Олеговне, Шабакоевой Анастасии Тимуровне, Непочатых Софье Владимировне.

Acknowledgments

The author is happy to acknowledge the contribution of Translation and Translation Studies students, particularly group representatives: Yuliana Starostina, Anastasia Shabakayeva, and Sofya Nepochatykh.

Конфликт интересов

Автор заявляет об отсутствии потенциального или явного конфликта интересов.

Conflict of Interest

The author declares that there is no conflict of interest, either existing or potential.

Источники

Тэффи, Н. А. (1998) *Собрание сочинений: в 5 т. Т. 2. Неживой зверь*. М.: Лаком, 382 с.
Phrase TMS: The leading translation management system. (2024) [Online]. Available at: <https://phrase.com/platform/tms/> (accessed 29.06.2024).

Словари

Кузнецов, С. А. (ред.). (1998) *Большой толковый словарь русского языка*. СПб.: Норинт, 1534 с.
Corpus of Contemporary American English. (2024) [Online]. Available at: <https://www.english-corpora.org/coca/> (accessed 29.06.2024).
Merriam-Webster Dictionary. (2024) [Online]. Available at: <https://www.merriam-webster.com/> (accessed 26.06.2024).

Литература

Беляева, Л. Н. (2019) Машинный перевод в работе переводчика: практический аспект. *Вестник ПНИПУ. Проблемы языкознания и педагогики*, № 2, с. 8–20. <https://doi.org/10.15593/2224-9389/2019.2.1>
 Гальперин, И. Р. (2007) *Текст как объект лингвистического исследования*. 5-е изд. М.: КомКнига, 144 с.
 Жданова, Т. А., Ефремова, Г. А. (2017) Автоматизация перевода: прошлое и настоящее. *Вестник Хабаровского государственного университета экономики и права*, № 1 (87), с. 46–50.
 Иванько, А. Ф., Иванько, М. А., Колесникова, О. Д. (2019) Информационные нейронные сети. *Научное обозрение. Технические науки*, № 4, с. 11–16. <https://doi.org/10.17513/srts.1250>
 История машинного перевода: от гипотез Лейбница и Декарта — до мобильных приложений и облачных сервисов. (2019) *PROMT*. [Электронный ресурс]. URL: <https://www.promt.ru/press/blog/istoriya-mashinnogo-perevoda-ot-gipotez-leybnitsa-i-dekarta-do-mobilnykh-prilozheniy-i-oblachnykh-se/> (дата обращения 20.06.2024).
 Камшилова, О. Н., Беляева, Л. Н. (2023) Машинный перевод в эпоху цифровизации: новые практики, процедуры и ресурсы. *Terra Linguistica*, т. 14, № 1, с. 41–56. <https://doi.org/10.18721/JHSS.14105>
 Маматкасимова, В. А. (2022) Общие принципы машинного перевода. *Вестник науки и образования*, № 6-2 (126), с. 23–25.
 Морозова, Т. В. (2020) Жанр рекламы в функционально-стилистическом освещении. *Ученые записки Новгородского государственного университета*, № 4 (29), с. 1–4. [https://doi.org/10.34680/2411-7951.2020.4\(29\).12](https://doi.org/10.34680/2411-7951.2020.4(29).12)
 Степаненко, А., Каменщиков, С., Суетин, Н. (2020) Искусственный интеллект в современном искусстве. *Инновационный центр «Сколково»*. [Электронный ресурс]. URL: <https://sk.ru/news/iskusstvennyy-intellekt-v-sovremennom-iskusstve/> (дата обращения 22.06.2024).
 Ушакова, А. О. (2022) Постредактирование машинного перевода технического текста. *Евразийский гуманитарный журнал*, № 3, с. 69–76.

- Чакырова, Ю. И. (2013) Постредактирование в транслатологической парадигме. *Вестник ПНИПУ. Проблемы языкознания и педагогики*, № 8, с. 137–144.
- Шукаева, Ж. Ж., Шидловская, И. А. (2022) Особенности машинного перевода немецких фразеологизмов на русский язык. В кн.: *Филологические чтения. Материалы Всероссийской научно-практической конференции с международным участием*. Оренбург: Изд-во Оренбургского государственного университета, с. 246–251. [Электронный ресурс]. URL: https://elibrary.ru/download/elibrary_48654711_61289346.pdf (дата обращения 22.06.2024).
- Ясаревская, О. Н. (2017) Модели статистического машинного перевода. *XXI век: итоги прошлого и проблемы настоящего плюс*, № 5-6 (39-40), с. 171–176.
- Hutchins, W. J. (1999) Warren Weaver memorandum, July 1949. *MT News International*, no. 22, pp. 5–6.
- Rule-based machine translation. (2024) *Machine translate*. [Online]. Available at: <https://machinetranslate.org/rule-based-machine-translation> (accessed 21.06.2024).
- What are AI hallucinations? (2024) *IBM*. [Online]. Available at: <https://www.ibm.com/topics/ai-hallucinations> (accessed 22.06.2024).

Sources

- Phrase TMS: The leading translation management system*. [Online]. Available at: <https://phrase.com/platform/tms/> (accessed 29.06.2024). (In English)
- Teffi, N. A. (1998) *Sobranie sochinenij: v 5 t. T. 2. Nezhivoj zver' [Complete works: In 5 vols. Vol. 2. Unliving beast]*. Moscow: Lakom Publ., 382 p. (In Russian)

Dictionaries

- Corpus of Contemporary American English*. (2024) [Online]. Available at: <https://www.english-corpora.org/coca/> (accessed 29.06.2024). (In English)
- Kuznetsov, S. A. (ed.). *Bol'shoj tolkovyj slovar' russkogo yazyka [A large explanatory dictionary of the Russian language]*. Saint Petersburg: Norint Publ., 1534 p. (In Russian)
- Merriam-Webster Dictionary*. (2024) [Online]. Available at: <https://www.merriam-webster.com/> (accessed 26.06.2024). (In English)

References

- Belyaeva, L. N. (2019) Mashinnyj perevod v rabote perevodchika: prakticheskij aspekt [Machine translation in a translator workflow: Practical view]. *Vestnik PNIPI. Problemy yazykoznanija i pedagogiki — PNIPI Linguistics and Pedagogy Bulletin*, no. 2, pp. 8–20. <https://doi.org/10.15593/2224-9389/2019.2.1> (In Russian)
- Chakyrova, Yu. I. (2013) Postredaktirovanie v translatoologicheskoj paradihme [Post-editing in the translation studies paradigm]. *Vestnik PNIPI. Problemy yazykoznanija i pedagogiki — PNIPI Linguistics and Pedagogy Bulletin*, no. 8, pp. 137–144. (In Russian)
- Galperin, I. R. (2007) *Tekst kak ob'ekt lingvisticheskogo issledovaniya [Text as an object of linguistic study]*. 5th ed. Moscow: KomKniga Publ., 144 p. (In Russian)
- Hutchins, W. J. (1999) Warren Weaver memorandum, July 1949. *MT News International*, no. 22, pp. 5–6. (In English)
- Istoriya mashinnogo perevoda: ot gipotez Lejbnitsa i Dekarta — do mobil'nykh prilozhenij i oblachnykh servisov [History of machine translation: From hypotheses by Leibniz and Descartes to mobile applications and cloud services]. (2019) *PROMT*. [Online]. Available at: <https://www.promt.ru/press/blog/istoriya-mashinnogo-perevoda-ot-gipotez-leybnitsa-i-dekarta-do-mobilnykh-prilozhenij-i-oblachnykh-se/> (accessed 20.06.2024). (In Russian)
- Ivanko, A. F., Ivanko, M. A., Kolesnikova, O. D. (2019) Informatsionnye nejronnye seti [Information neural networks]. *Nauchnoe obozrenie. Tekhnicheskie nauki — Scientific Review. Technical Sciences*, no. 4, pp. 11–16. <https://doi.org/10.17513/srts.1250> (In Russian)
- Kamshilova, O. N., Belyaeva, L. N. (2023) Mashinnyj perevod v epokhu tsifrovizatsii: novye praktiki, protsedury i resursy [Machine translation in the age of digitalization: New practices, procedures and resources]. *Terra Linguistica*, vol. 14, no. 1, pp. 41–56. <https://doi.org/10.18721/JHSS.14105> (In Russian)
- Mamatkasimova, V. A. (2022) Obshchie printsipy mashinnogo perevoda [General principles of machine translation]. *Vestnik nauki i obrazovaniya — Bulletin of Science and Education*, no. 6-2 (126), pp. 23–25. (In Russian)
- Morozova, T. V. (2020) Zhanr reklamy v funktsional'no-stilisticheskom osveshchenii [The genre of advertising in the functionally-stylistic lighting]. *Uchenye zapiski Novgorodskogo gosudarstvennogo universiteta — Memoirs of NovSU*, no. 4 (29), pp. 1–4. [https://doi.org/10.34680/2411-7951.2020.4\(29\).12](https://doi.org/10.34680/2411-7951.2020.4(29).12) (In Russian)
- Rule-based machine translation. (2024) *Machine translate*. [Online]. Available at: <https://machinetranslate.org/rule-based-machine-translation> (accessed 21.06.2024). (In English)

- Shukaeva, Zh. Zh., Shidlovskaya, I. A. (2022) Osobennosti mashinnogo perevoda nemetskikh frazeologizmov na russkij yazyk [Specifics of machine translation of German idioms into Russian]. In: *Filologicheskie chteniya. Materialy Vserossijskoj nauchno-prakticheskoy konferentsii s mezhdunarodnym uchastiem [Philological readings. Proceedings of All-Russian scientific conference with international participation]*. Orenburg: Orenburg State University Publ., pp. 246–251. [Online]. Available at: https://elibrary.ru/download/elibrary_48654711_61289346.pdf (accessed 22.06.2024). (In Russian)
- Stepanenko, A., Kamen'shchikov, S., Suetin, N. (2020) Iskusstvennyj intellekt v sovremennom iskusstve [Artificial intelligence in modern art]. *Innovatsionnyj tsentr Skolkovo*. [Online]. Available at: <https://sk.ru/news/iskusstvennyy-intellekt-v-sovremennom-iskusstve/> (accessed 22.06.2024). (In Russian)
- Ushakova, A. O. (2022) Postredaktirovanie mashinnogo perevoda tekhnicheskogo teksta [Technical text machine translation post-editing]. *Evrazijskij gumanitarnyj zhurnal — Eurasian Humanitarian Journal*, no. 3, pp. 69–76. (In Russian)
- What are AI hallucinations? (2024) IBM. [Online]. Available at: <https://www.ibm.com/topics/ai-hallucinations> (accessed 22.06.2024). (In English)
- Yasarevskaya, O. N. (2017) Modeli statisticheskogo mashinnogo perevoda [Revisiting statistical machine translation models]. *XXI vek: itogi proshlogo i problemy nastoyashchego plyus — XXI Century: Resumes of the Past and Challenges of the Present plus*, no. 5-6 (39-40), pp. 171–176. (In Russian)
- Zhdanova, T. A., Efremova, G. A. (2017) Avtomatizatsiya perevoda: proshloe i nastoyashchee [Translation automation: Past and present]. *Vestnik Khabarovskogo gosudarstvennogo universiteta ekonomiki i prava — Bulletin of Khabarovsk State University of Economics and Law*, no. 1 (87), pp. 46–50. (In Russian)