

СПЕЦИФИКА ПРОЦЕССА ПОСТРЕДАКТИРОВАНИЯ ПРОГРАММ АВТОМАТИЗИРОВАННОГО ПЕРЕВОДА

Бакиева Дженнет Нуриягдыевна

*Студент-магистрант, Московский городской педагогический университет
Российская Федерация, г. Москва, Малый Казенный переулок 5Б, 105064
jennet.bakieva@mail.ru*

Аннотация. Настоящее исследование посвящено анализу качества перевода технических и массмедиаальных текстов, полученных с помощью программ автоматизированного и машинного перевода. На материале переводов новостных сводок и статей BBC и the Guardian и технической документации производителя высокотехнологичных продуктов Samsung с английского языка на русский при помощи ПО SmartCAT предлагается статистическая оценка ошибок, совершаемых программой. В связи с активным развитием повсеместной глобализации и расширением компьютерных технологий потребность в услугах переводчика значительно увеличивается. Помимо этого, современные объемы информации для перевода в разы отличаются от того количества, которое существовало в предыдущие столетия, все время увеличиваясь, при этом времени, затрачиваемого на процесс профессионального перевода, становится все меньше. В связи с нехваткой времени и увеличенной нагрузкой, вызванных современными условиями развития общества, для ускорения перевода и автоматизации некоторых его процессов, которые не требуют повсеместного вмешательства человека, были созданы программы компьютерного перевода. Компьютерный перевод далее подразделяется на машинный и автоматизированный виды. Поскольку данная технология является относительно молодой по сравнению с традиционным видом перевода, ранее такие программы имели ограниченные способы его применения и отличались низким качеством перевода. Например, в переводческих агентствах подобный вид перевода использовался для подстрочного перевода, в то время как в современных условиях благодаря активному развитию сферы компьютерного перевода существуют такие программы, пригодные для полноценного пользования и работы как в профессиональной, так и любительской сфере, на домашнем или корпоративном компьютере, а также в онлайн режиме. В современных реалиях развития переводческой отрасли умение пользоваться программами машинного и автоматизированного перевода уже стало необходимостью, способной значительно облегчить работу переводчика. Для повышения качества такого вида перевода в качестве результата исследования предлагаются решения и исправления ошибок, совершенных программами машинного и автоматизированного перевода. Актуальность исследования обосновывается необходимостью механизации и автоматизации для повышения эффективности работы переводчика и ускорения процессов

перевода в связи с активным расширением объемов информации, необходимой для перевода, а также потребностью в разработке и обновлении высоких технологий, связанных с машинным и автоматизированным переводом. Предлагаемый алгоритм исправления ошибок, как представляется, позволяет избежать неточностей на различных языковых уровнях на этапе постредктирования.

Ключевые слова: *машинный перевод, автоматизированный перевод, постредктирование, программа машинного перевода, программа автоматизированного перевода.*

Для цитирования: Амангельдина Г.Б., Дудович Д.Л. Влияние педагогической оценки на повышение мотивации к обучению и качество образования // Focus on Language Education and Research. 2022. Т. 3, № 1. С. 24-35.

Настоящая статья посвящена проблеме использования информационных технологий в процессе перевода и особенностям переводческого редактирования при работе с программами машинного и автоматизированного перевода. Перевод, выполненный при помощи программ машинного и автоматизированного перевода, в большинстве случаев требует последующее редактирование, особенности которого излагаются в данной статье.

Процесс постредктирования чаще всего ассоциируется с заданием, относящемуся к машинному переводу (МП). Другими словами, постредктирование является процессом исправления, переделывания и редактирования текста, переведенного до этого в программе машинного перевода с языка оригинала на язык перевода [2, с. 26]. Само понятие «постредктирование» принято определять как «термин, относящийся к исправлению машинного перевода лингвистом или редактором» [5, с. 288]. Другое толкование данного термина представляется следующим образом: «исправление уже переведенного текста, нежели перевод с нуля» [6, с. 2].

В данной работе представляется необходимым развести два термина:

«машинный перевод» и «автоматизированный перевод» (применяются сокращения МП и АП соответственно). Так, за Машинный перевод принято считать компьютеризированный процесс перевода текста с языка источника на язык перевода. Такой процесс выполняется по формальным правилам, заданным в виде определенного алгоритма. Обычно результат такого перевода не соответствует нормам адекватного перевода, поэтому требует доработок и правок [7].

Автоматизированный перевод считается переводом, который также осуществляется при помощи высоких технологий, однако такой вид перевода отличается от машинного вмешательством человека в процесс перевода, который управляет компьютерным переводом посредством создания глоссариев или ручных правок текста [4, с. 144]. Процесс создания глоссариев связан с таким видом автоматизированного перевода, как «память переводов». Память переводов является технологией накопления предыдущих вариантов перевода отдельных сегментов текста или «единиц перевода» [10] для последующего повторного использования в переводе. Для поиска таких сегментов или переводных единиц применяются программы, которые получили название вспомогательные программы для перевода [1, с. 148]. Переводчик при работе с такими программами, которые позволяют максимально автоматизировать и облегчать работу, решает ментальные задачи, и автоматизированный перевод находится в руках человека, а не машины. В настоящей статье рассматривается такая программа, как SmartCAT.

Исследования в области постредактирования начали проводиться в 1980-е годы. В конце XX века переводческие проекты, включавшие постредактирование, проходили через программы МП единым комплектом, состоящим из текстов разных жанров и направлений, черновые варианты

которых далее переходили к редакторам или лингвистам. Последние использовали стороннюю программу (например, Microsoft Word, Pages и т.д.) для исправления полученного машинного перевода, занимались так называемым «пассивным постредактированием», поскольку переводчикам приходилось адаптировать такой новый вид перевода под вышеназванные программы, которые для этого не были предназначены. В связи с отсутствием каких-либо инструкций по эффективному постредактированию, работа с машинным переводом каждый раз требовала введение новых стратегий и тактик по выполнению постредактирования. Таким образом, для облегчения работы переводчика были составлены определенные критерии, которым должно было соответствовать постредактирование. Для этой цели в 1999 г. члены Американской ассоциации машинного перевода (АМТА) и Европейской ассоциации машинного перевода (ЕАМТ) создали специальную группу, занимающуюся постредактированием [3, с. 9].

В начале XXI в. среди проведенных исследований были предприняты попытки определить степени постредактирования, которые сводились к следующим пунктам: обычное и быстрое или полное и быстрое. В настоящее время наиболее распространенным является легкое и полное постредактирование. Легкое постредактирование предполагает минимальное вмешательство переводчика, поскольку его целью является перевод в ознакомительных целях. Качество такого постредактирования обычно является низким. Полное постредактирование предполагает большее вмешательство для того, чтобы достичь качества, соответствующего всем требованиям. Считается, что результатом должен быть текст, который вызывает ту же реакцию реципиента на переведенный текст, что и отправителя в тексте оригинала, а также стилистически соответствует исходному тексту [6, с. 3].

Однако эти критерии, как и задачи, которые они предполагают, чаще всего недостаточны, поскольку отсутствуют детальные критерии, которые бы определяли сам процесс постредактирования. В связи с этим, работа переводчика заключается в адаптации исходного текста под формат машинного перевода для более четкого считывания текста программой машинного перевода.

На основании анализа технической документации, представленной в виде инструкции по применению ноутбука производства компании Samsung, а также новостных статей, полученных из официальных сайтов новостных агентств BBC и the Guardian, был выведен определенный алгоритм, который призван облегчить работу переводчика, работающего с программами АП.

Итак, первым этапом работы с программой АП SmartCAT является занесение необходимого текста в программу. В качестве примера была взята инструкция по применению ноутбука Samsung 2020 года производства на английском языке на 102 страницы печатного текста и новостная статья BBC на английском языке, занимающая 10 страниц печатного текста. Обработка такого объема информации у программы АП SmartCAT заняло около 4-х минут, поскольку данный процесс осложняло наличие иллюстраций, сопровождаемых подписями.

Вторым этапом работы с МП является определение неизвестных и неправильно переведенных терминов, слов и фраз. Такие слова, фразы и термины включают в себя акронимы, аббревиатуры, названия компаний или определенные словарные слова, которые машинный перевод не воспринимает. Данный пункт применим по отношению к технической документации, поскольку в данном виде текстов преобладают термины, присущие определенному жанру документа. Однако, как показал анализ инструкции по применению ноутбука от компании Samsung, данный пункт не

стал проблемой для программы АП SmartCAT. Все примененные термины и слова технической направленности получили релевантный и адекватный перевод (ср, the LCD screen – жидкокристаллический экран, Laser Safety Note – Примечание по технике безопасности лазера и т.д.), за исключением некоторых аббревиатур. Например, CMOS battery программа перевела как КМОП батарея, тогда как верным вариантом перевода будет описательный: батарея для BIOS системы, либо можно оставить английский вариант аббревиатуры, поскольку в сфере информационных технологий многие англоязычные термины не получают перевода на русский язык.

В новостной статье проблемой для машинного перевода стало название компании AstraZeneca и вакцины Oxford-AstraZeneca. В одном из представленных предложений программа оставила оба названия без перевода, тогда как в другом случае она транслитерировала на русский язык: АстраЗенека и Оксфорд-Астра Зенека соответственно.

Важным фактором удобства программы АП SmartCAT заключается в подключении глоссариев и систем памяти перевода, которые создаются лично самим переводчиком или предоставляются клиентом для каждого отдельного документа. Для этого предлагается третий этап работы – составление глоссария. После анализа текста на предмет неизвестных и неправильно переведенных терминов необходимо внести данные единицы в конкретные глоссарии, подключенные к документам. Для этого при загрузке документа для перевода в созданный проект в продвинутых настройках можно подключить как память переводов и глоссарии, так и систему машинного перевода.

Так, в глоссарий, подключенный к тексту технической инструкции, были внесены аббревиатуры, с которыми машинный перевод не справился, а именно, оставил без перевода. Например: RTC (real time clock) – ЧРВ (часы

реального времени), FCC (Federal Communications Commission) – Федеральная комиссия по связи, NCRP (National Council on Radiation Protection and Measurement) – Национальный совет по радиационной защите и измерениям (НКРЗ) и т. д. Всего таких аббревиатур на 102 страницы печатного текста оказалось около 30 единиц.

В глоссарий, подключенный к тексту новостной статьи, вошли названия компаний, вакцины, организаций и политических явлений. Например: The European Medicines Agency – Европейское агентство по лекарственным средствам, Brexit – Выход Великобритании из Евросоюза и т.д. Всего, таких названий, которые не имели перевода в выходном тексте, было немного: всего около 9, поэтому и глоссарий вышел небольшим.

Последующим четвертым этапом работы является повторный прогон текста через систему. Однако на данном этапе вступает в силу автоматизированный перевод, поскольку все предыдущие шаги, упомянутые выше, подстроили текст под перевод, который удобен для постредактирования человеком. В общей сложности время, затраченное на выполнение всех описанных шагов, может варьироваться от 2-х до 6-ти часов работы над одним документом. Данный диапазон зависит от размеров документа и степени сложности текста.

Так, на техническую инструкцию ушло 2 часа вычитки текста, нахождения проблемных зон и неизвестных терминов / слов и дальнейшего внесения в глоссарий и финального автоматизированного перевода. Поскольку качество машинного перевода технического текста оказалось довольно высоким, то и количество затраченного времени не является таким большим.

Что касается новостной статьи, то качество машинного перевода в этом случае снизилось по сравнению с переводом технического текста. Это

связано как с самим жанром текста, так и с осложнением названиями компаний и организаций. Всего, на обработку, вычитку текста и составление глоссария на 10 печатных страниц ушло около 3 часов.

Таким образом, результатом этого этапа являются следующие пункты: документ, полученный посредством автоматизированного перевода, на который ушло времени намного меньше, чем если бы он был переведен с нуля; глоссарий, специально составленный для данного документа, но который также может быть использован и для последующих проектов. Такой документ может быть использован либо в ознакомительных целях, либо для внутреннего пользования без широкого использования и продажи клиенту, так как качество данного перевода не соответствует требованиям адекватного перевода.

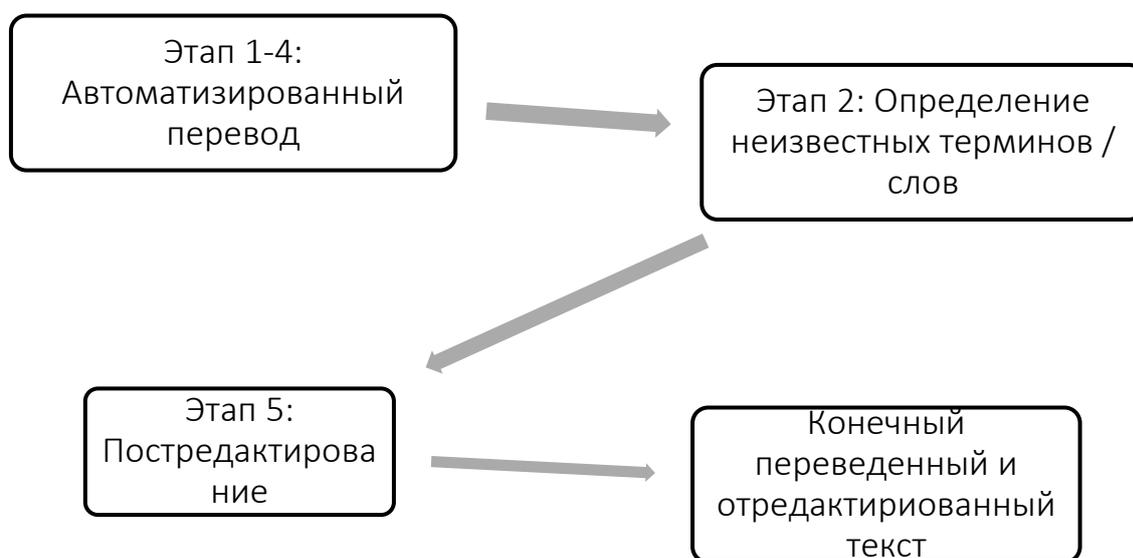
И, наконец, пятым, завершающим, этапом работы с программой АП SmartCAT является полное постредактирование. Документы, предназначенные для широкого пользования и последующей продажи клиентам, требуют высококачественного полного постредактирования. Для удобства перевода программа SmartCAT предлагает деление текста на смысловые абзацы, которое производится автоматически при внесении текста на начальном этапе работы с документом. Таким образом, можно либо менять вариант перевода, предоставленный машинным переводом, либо заняться самостоятельным переводом текста.

Так, технический документ не потребовал полного постредактирования, а легкого, поскольку, как уже неоднократно было отмечено, качество перевода программой машинного перевода соответствует адекватному восприятию текста. Что касается текста новостной статьи, то изменения претерпели синтаксические конструкции, которые машинный перевод представил на русский язык дословно. Ср.:

Оригинал	МП	Постредактирование
At the EU leaders' summit on Thursday they'll be pressing for solutions.	На саммите лидеров ЕС в четверг они будут настаивать на решениях.	В четверг на саммите лидеров ЕС будет поднят вопрос о решении ситуации с вакциной.

Таким образом, все описанные этапы работы с программой АП SmartCAT можно представить в виде следующей схемы, ср.:

Схема 1. Этапы работы с программой АП SmartCAT.



Подводя итоги, представляется необходимым отметить, что применение программы автоматизированного перевода SmartCAT оправдано, поскольку, как показал анализ текстов технической документации и новостной статьи, данный процесс позволяет значительно сократить как время, так и ресурсы на совершение перевода документа, предназначенного для определенного проекта.

Другое положение, которое нашло доказательство в ходе проведенного анализа, показало, что создание системы профессионального машинного перевода, которая способна создавать высококачественный перевод, в настоящее время не представляется возможным, поскольку машинный перевод, который хоть и претерпевает изменения и улучшения из

года в год, по качеству далек от работы живого переводчика, что требует последующего вмешательства человека посредством постредактирования текста. В современных реалиях процесс постредактирования является необходимостью.

В случае перевода технической документации процесс постредактирования облегчается за счет клишированной лексики и шаблонностью, что и показал анализ данного вида текста. В результате проведенного анализа мы пришли к выводу о том, что машинный перевод не создал полный и адекватный вариант перевода, поскольку конечный результат МП требовал гораздо больше доработок и постредактирования по сравнению с тем же процессом, пройденным в случае перевода технического текста.

Автор выражает признательность за консультативную и техническую помощь проф. В. Д. Шевченко.

Библиографический список

1. Попов С. А., Жукова Е. Ф. Современные системы автоматизированного перевода / Великий Новгород: Изд. НовГУ им. Ярослава Мудрого, 2014. 241 с.
2. Allen J. Postediting: an Integrated Part of a Translation Software Program // LanguageInternational. 2001. №13 (2). P. 26-29.
3. Allen J. An Introduction to Using MT Software // Translation: the Guide from Multilingual Computing & Technology. 2005. Vol. 16. №1 (69). P. 8-12.
4. Bowker L. Computer-aided Translation Technology / Ottawa: University of Ottawa Press, 2002. 144 pp.
5. Senez D. The Machine Translation Help Desk and the Postediting Service

- // Terminologie&Traduction. OPOCE. European Commission. 998. P. 289-295.
6. Wagner E. Post-Editing Systran – A challenge for Commission Translators // Terminologie&Traduction. OPOCE. European Commission. 2000. P. 1-6.
 7. Официальный сайт компании ПРОМТ [Электронный ресурс]. – URL: <http://www.promt.ru/> (дата обращения: 10.11.2020).
 8. Официальный новостной портал BBC [Электронный ресурс]. – URL: <http://www.bbc.com> (дата обращения: 21.03.2021).
 9. Официальный сайт компании SmartCAT [Электронный ресурс]. – URL: <https://smartcat.ai/> (дата обращения: 15.09.2020).
 10. Официальный сайт компании Samsung [Электронный ресурс]. – URL: <https://www.samsung.com/us/support/downloads/> (дата обращения: 22.03.2021).
-

SPECIFICS OF THE POSTEDITING PROCESS IN AUTOMATED TRANSLATION PROGRAMMES OUTPUT

Jennet Bakieva

*Masters student at Moscow City University
Russia, Moscow, Maly Kazenny Pereulok 5B, 105064
jennet.bakieva@mail.ru*

Abstract. The paper analyses the text quality made by computer-assisted translation programmes. The paper addresses the issue of quality of machine translation on the example of English news reports and technical manuals translated by means of the computer-assisted translation software SmartCAT from English into Russian and provides the statistics on the mistake evaluation. The research findings suggest how to avoid and correct the mistakes made by the computer-assisted translation software. The research relevancy is supported by the steady need for automation and acceleration of translation services as well as private interest in developing and advancing linguistic software. The outcome of the research is the translation

algorithm that will allow editors to avoid or correct translation errors on different language levels.

Keywords: *machine translation, automated translation, postediting, computer-assisted translation, machine translation programme, automated translation programme*

For citation: Bakieva J. Specifics of the postediting process in automated translation programmes output // Focus on Language Education and Research. 2022. Vol. 3, No. 1. P. 24-35.

References

1. Popov S., Zhukova Ye. *Sovremennye sistemy avtomatizirovannovo perevoda / Veliky Novgorod: iz. NovGU im. Yaroslava Mudrovo, 2014.241 s.*
2. Allen J. Postediting: an Integrated Part of a Translation Software Program // *LanguageInternational*. 2001. №13 (2). P. 26-29.
3. Allen J. An Introduction to Using MT Software // *Translation: the Guide from Multilingual Computing & Technology*. 2005. Vol. 16. №1 (69). P. 8-12.
4. Bowker L. *Computer-aided Translation Technology / Ottawa: University of Ottawa Press, 2002. 144 pp.*
5. Senez D. The Machine Translation Help Desk and the Postediting Service // *Terminologie & Traduction. OPOCE. European Commission*. 998. P. 289-295.
6. Wagner E. Post-Editing Systran – A challenge for Commission Translators // *Terminologie & Traduction. OPOCE. European Commission*. 2000. P. 1-6.
7. PROMT [elektronny resurs]. – Available at: <http://www.promt.ru/> (accessed 10.11.2020).
8. BBC [elektronny resurs]. – Available at: <http://www.bbc.com> (accessed 21.03.2021).
9. SmartCAT [elektronny resurs]. – Available at: <https://smartcat.ai/> (accessed 15.09.2020).
10. Samsung [elektronny resurs]. – Available at: <https://www.samsung.com/us/support/downloads/> (accessed 22.03.2021).