

УДК 811.111-26

DOI: 10.31891/2415-7929-2019-17-22-26

МАРТИНЮК О. В.
Хмельницький національний університет

ПОПЕРЕДНЄ ТА КІНЦЕВЕ РЕДАГУВАННЯ ТЕКСТІВ У ПРОЦЕСІ МАШИННОГО ПЕРЕКЛАДУ ЗАСОБАМИ КОМП'ЮТЕРНОГО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ТА ОНЛАЙН-СЕРВІСІВ

У статті розглянуто особливості попереднього редагування та кінцевого редагування (постредагування) текстів для отримання кращої якості машинного перекладу засобами онлайн-сервісів та комп'ютерних програм-перекладачів. Узагальнено рекомендації щодо оптимізації форми представлення текстової інформації перед її обробкою системою машинного перекладу та визначено основні етапи постредагування вихідного тексту перекладу. Окреслено недоліки та переваги машинного перекладу документів.

Ключові слова: машинний переклад (МП), системи МП, програмне забезпечення, онлайн сервіс МП, попереднє редагування, постредагування, вихідний текст.

MARTYNYUK O.
Khmelnitskyi National University

PRE-EDITING AND POST-EDITING OF TEXTS IN THE PROCESS OF MACHINE TRANSLATION BY MEANS OF TRANSLATION SOFTWARE AND ONLINE SERVICES

The article deals with the peculiarities of preliminary text preparation and post-editing to ensure higher quality of machine translation (MT) performed by various translation software and online translation services. MT can be worthwhile when dealing with large volumes of similar texts. Translated documents contain lexical, grammar and syntax errors, though the gist of the article can be discovered for internal use. It has been proven that incorporating an efficient pre-editing process, correct adaptation of MT systems, including "training" in a particular subject area, can considerably simplify the job of both MT systems and editors involved in the post-editing process. Pre-editing has been defined as preparation involved prior to running a particular text through an automatic translation tool. It involves reviewing the content for basic errors, tagging certain content to be translated a certain way (or to not be translated at all), and optimizing the formatting. Post-editing has been described as the process of "fixing" MT output to bring it closer to a human translation standard. Pre-editing can provide better results when the recommendations summarized by the author are followed. They include avoiding misprints and spelling mistakes, minding punctuation marks, using diacritics correctly, observing the case of letters, using simple syntactic constructions with direct word order, using only conventional abbreviations, avoiding the use of slangy expressions, avoiding clichés and colloquial phrases, using infinitive verb forms instead of gerunds, using the active voice instead of passive voice, avoiding compound sentences and homogeneous parts of sentences, writing phrases that you can recycle, removing needless words, etc. It has been emphasized that in most cases, MT requires post-editing by an editor, who normally has special training and is experienced in handling machine-translated texts. In post-editing MT content, the process involves the following steps: a sentence or segment of the MT output is read; the MT output is compared to the source text; post-editor applies some form of quality judgment on each segment, based on explicit instructions; quick decisions are made on whether the translation is correct; the text is improved or re-translated.

Keywords: machine translation (MT), MT systems, software, MT online service, pre-editing, post-editing, output text.

Актуальність дослідження. Надзвичайно швидкий прогрес у галузі розробки електронного обладнання та комп'ютерного програмного забезпечення, а також спроби отримати прийнятний переклад за мінімальний період часу сприяють розвитку машинного перекладу (МП). Машинний переклад, який ще називають *автоматичним* або *миттєвим*, визначається як переклад тексту комп'ютером, без участі людини. Основна перевага машинного перекладу – можливість швидко обробляти великі за обсягом тексти, що зазвичай робить його більш економічним, ніж переклад, зроблений людиною. Тут варто зауважити, що якість МП завжди буде поступатися людському перекладу.

Звичайно, поліпшення якості МП є головним завданням розробників спеціалізованого програмного забезпечення. Однак користувач також може докласти певних зусиль для досягнення прийнятних результатів, оскільки якість МП певною мірою залежить і від якості поданого тексту.

Аналіз останніх джерел. Проаналізувавши наукові праці зарубіжних (М. Черагуй [5], П. Коен [7], М. Окпор [8]) та українських (П. Асоянц [2], В. Масюкевич [3], В. Яковина [3], А. Янковець [4]) дослідників, інформацію з вебсайтів розробників перекладацького програмного забезпечення та власний досвід викладання МП, ми переконалися, що недосконалість програмних засобів МП призводить до наявності лексичних, граматичних та синтаксичних помилок у перекладеному документі, хоча суть тексту можна зрозуміти. Тому МП застосовують передусім у тих випадках, коли документ передбачено для внутрішнього чи особистого використання. Наприклад, для ознайомлення з інформацією вебсайту, статті чи листа іноземною мовою або пошуку повідомлення про певну подію у міжнародних засобах масової інформації. По-друге, МП можна використовувати для технічних та вузькоспеціалізованих текстів, які в подальшому потребують постредагування фахівцями в цій галузі. У такому випадку цільовий текст використовується як

«підрядковий» переклад, на основі якого технічний експерт продукує кінцевий варіант тексту, спираючись на знання предметної галузі.

Багато типів текстових матеріалів не можна перекладати засобами МП. Сюди віднесемо тексти, що містять інформацію, неправильне трактування якої загрожуватиме здоров'ю людини; описи складних пристроїв чи механізмів; документи, пов'язані з юридичною відповідальністю сторін (договори, гарантії, контракти); маркетингові матеріали, де текст має бути переформований у новому культурному контексті або фактично відтворений з нуля, тощо. У таких випадках заощаджений час не виправдовує ризиків.

Загалом, належної якості можна очікувати від МП формалізованих технічних текстів, тоді як літературний переклад виходить за рамки МП.

Постановка проблеми. Незважаючи на розвиток та вдосконалення технології машинного перекладу, професійні людські навички попереднього редагування і постредагування тексту залишаються вкрай необхідними для створення високоякісних перекладів.

Метою нашого дослідження є узагальнення рекомендацій щодо попереднього редагування та постредагування текстової інформації для отримання кращої якості МП засобами онлайн-сервісів та комп'ютерних програм-перекладачів. Об'єктом дослідження є процес МП, що включає попередню підготовку тексту (попереднє редагування), власне МП та етап постредагування. Предметом дослідження є шляхи покращення якості МП, пов'язані зі зміною форми висловлювання у тексті, тобто якості самого тексту.

Виклад основного матеріалу. Попереднє редагування не варто плутати з постредагуванням. Попереднє редагування стосується підготовки тексту до обробки програмним продуктом для автоматичного перекладу. Воно передбачає перегляд змісту для виявлення основних помилок, позначення певного контенту для перекладу тим чи іншим чином (або як неперекладного) та оптимізацію форматування [6]. Ця підготовка тексту до його обробки засобами перекладу може значно покращити кінцевий результат, тому вона вважається важливим етапом процесу автоматичного перекладу. Постредагування – це процес «виправлення» вихідного тексту МП з метою наближення його якості до стандарту перекладу, здійсненого людиною [9]. Ефективне попереднє редагування тексту може значно покращити якість вихідного тексту перекладу, спростити роботу як систем МП, так і редакторів, перед якими стоїть завдання вдосконалити вихідний текст машинного перекладу на завершальному етапі обробки (постредагування) матеріалів.

Попередня підготовка тексту починається на етапі створення його змісту. Для цього розробляють спеціальні стандарти для технічних письменників та авторів, які дотримуються їх з метою забезпечення кращого розуміння змісту та перекладу текстів як засобами комп'ютерної техніки, так і професійними перекладачами.

Проаналізуємо та узагальнимо основні рекомендації щодо попереднього редагування текстової інформації для отримання кращої якості МП засобами онлайн-сервісів та комп'ютерних програм-перекладачів під час перекладу з англійської мови і навпаки [6]. Сюди віднесемо:

1. Уникнення друкарських та орфографічних помилок. Комп'ютерний перекладач не може виправити помилки та розпізнати неправильно надруковані слова. Тут можуть стати у нагоді спеціальні програмні засоби для перевірки правопису, так звані Spellcheckers. Їх можна встановити на персональному комп'ютері, використовувати як безкоштовний онлайн-сервіс чи плагін до текстових редакторів. Часто програми для перевірки правопису дозволяють також перевіряти граматику. Як і онлайн словники, онлайн перевірка правопису дозволяє користувачеві друкувати слова, рядки, параграфи, перевіряючи автоматично правопис українською та іншими мовами.

2. Дотримання правил пунктуації. Пропущені або, навпаки, зайві розділові знаки можуть перешкодити електронному перекладачеві правильно зрозуміти синтаксичну структуру речення. Наприклад, мітки абзацу (¶) програма-перекладач автоматично видаляє, таким чином рядки «зливаються» в один. Тому важливо завершувати речення крапкою (.)

3. Адекватне використання діакритичних знаків. Діакритичний знак – це надрядковий або підрядковий знак при букві, що вказує на вимову, яка відрізняється від вимови звука, позначеного цією ж буквою без знака [1]. Як правило, електронний перекладач не може розпізнати слова, що містять російську букву «ё», або слова з позначками наголосу.

4. Дотримання регістру (розкладки) літер. Малу літеру в слові можна легко перетворити на прописну (велику) літеру, наприклад, на початку речення або в заголовку. Це враховується під час розробки систем МП. І навпаки, прописна літера рідко стає малою. Здебільшого це пов'язано з утворенням нових слів, наприклад, при переході іменника, що є власною назвою, до розряду загальних назв: *xerox* та ін.). Слово «Internet» зазвичай пишеться з великої літери, тому електронний перекладач може не розпізнати його у випадку написання з малої літери, тощо.

Окрім того, існують мови, в яких велика літера змінює частину мови, до якої належить слово. Яскравий приклад – німецька мова, де іменники з великої літери живаються як на початку, так і в середині речення. Порівняємо ці переклади:

- "wie funktioniert das übersetzen mit dem "clipboard"?" – "how does this function translate with "clipboard"?"

- "Wie funktioniert das Übersetzen mit dem "clipboard"?" – "How does the translation with "clipboard" function?"

У реченні німецькою мовою *das Übersetzen* (укр. *переклад*) є іменником, відповідно воно має починатися з великої літери. У першому випадку цей факт не враховано, а тому в реченні англійською мовою *das übersetzen* перекладено дієсловом *translate* (укр. *перекладає*). Таким чином, переклад першого речення є неправильним, тоді як друге речення, де враховано розкладку літер, перекладено правильно.

5. Уникнення зайвих слів. Мова йде про видалення слів, які не розширюють змістового наповнення речення. Наприклад:

He (укр. *Він*) замість *He is a man who* (укр. *Він – чоловік, який*).

6. Повторення іменників замість займенників. Це сприяє адекватному «сприйняттю» змісту речень програмами-перекладачами. У наступному прикладі перше речення відредаговано правильно:

*You must pre-edit the text before using MT engine to translate **this text*** (укр. *цей текст*).

*You must pre-edit the text before using MT engine to translate **it*** (укр. *його*).

7. Вживання означеного артикля (*the*) для позначення іменників в англійській мові. Наприклад:

*Use **the** necessary pre-editing techniques* (відредаговано правильно).

Use necessary pre-editing techniques (відредаговано неправильно)

8. Використання загальноприйнятих абревіатур. Неправильний переклад абревіатури є лише частиною проблеми. Справа в тому, що навіть одне неперекладене слово може перешкодити електронному перекладачеві правильно проаналізувати синтаксичну структуру речення (абревіатури беруть участь у синтаксичних зв'язках поряд із звичайними словами). Правопис деяких абревіатур інколи збігається з написанням звичайних слів, що може мати неприємні наслідки. Наприклад, російська абревіатура «ПО» (анг. *software*, укр. *програмне забезпечення*) пишеться так само, як і російська прийменник «по» (анг. *on*, укр. *на*). У цьому прикладі реєстр літер не відіграє ніякої ролі, оскільки прийменник може бути надрукований прописними літерами, наприклад, у заголовку. Тому з російської мови речення «Я ЧАСТО ИСПОЛЬЗУЮ ЭТО ПО» може бути перекладене англійською мовою як «I FREQUENTLY USE IT ON». З іншого боку, якщо відредагувати речення, уникаючи абревіатури «Я часто использую это программное обеспечение», переклад його англійською мовою буде цілком адекватним: «I frequently use this software».

9. Уникнення сленгових виразів. Звичайно, мова не йде про злочинний сленг, хоча можна припустити, що злочинці можуть використовувати системи МП. У неформальній комунікації навіть законослухняні носії мови часто використовують слова, вирази та конструкції, що не належать до літературної мови. Як приклад, розглянемо речення російською мовою:

*Люди, решите **трабл**! Не могу зарегить **мыло**!*

З одного боку, проблема полягає в тому, що такі сленгові слова з'являються у мовленні раніше, ніж у словниках. З іншого боку, додавання неологізмів до словника не завжди є доцільним. Наприклад, російське слово «мыло» (англ. *soap*) для більшості користувачів систем МП вживається для позначення миючого засобу.

10. Уникнення кліше, ідіом та розмовних фраз. Засоби МП можуть неправильно (часто буквально) передати їх зміст. Тоді речення мовою перекладу не матиме сенсу. Наприклад, англійське речення «*It is easy*» (укр. «*Це легко*») відредаговано правильно. Тоді як речення «*It is a piece of cake*», що містить ідіому «*a piece of cake*» зі значенням «*легко*», засобами машинного перекладу перекладається як «*Це шматок пирога*».

11. Використання простих синтаксичних конструкцій з прямим порядком слів. Наприклад, в англійській мові на перше місце треба поставити підмет чи його групу (*I, you, he, my cat, my chief, son of my girlfriend*). На другому місці має стояти присудок, виражений дієсловом (*want, know, like*).

Не варто надмірно ускладнювати структуру речень. Наприклад, друге речення англійською мовою містить складніші синтаксичні конструкції:

Show that you can organize your thoughts by using a simple sentence structure in your texts – «Покажіть, що ви можете організувати свої думки, використовуючи просту структуру речень у своїх текстах» (відредаговано правильно).

You, in your texts, to show that you can organize your thoughts, should use a simple sentence structure – «Ви в своїх текстах, щоб показати, що ви можете організувати свої думки, повинні використовувати просту структуру речень» (відредаговано неправильно).

12. Уникнення складних речень та однорідних частин речень. У найкращому випадку, кожне речення має передавати одну логічну думку. Це правило, яке однаково застосовується до всіх мов, є найбільш ефективним з усіх.

Необхідно також уникати сполучників (*and, but, which, etc.*) та, коли це можливо, випадків уживання більше одного підрядного простого речення в межах складного. Рекомендований обсяг речення – не більше 25 слів. Кожне речення має бути граматично завершеним.

13. Вживання інфінітивних конструкцій замість герундіальних. Як бачимо з наступного прикладу, перше речення з англійською герундіальною конструкцією «*refusing to tell*» містить помилку у перекладі. Тоді як уживання інфінітивної конструкції «*to refuse to tell*» у другому випадку допомогло правильно передати зміст речення.

«*A woman's idea of keeping a secret is refusing to tell who told it*» – «Ідея жінки зберігати таємницю відмовляється розповідати, хто це сказав».

«*A woman's idea of keeping a secret is to refuse to tell who told it*» – «Ідея жінки зберігати таємницю – це відмовитись розповідати, хто це сказав».

14. Використання активного стану дієслів замість пасивного. Активний стан – це прямий стиль написання, який допомагає уникнути розмитості та неоднозначності. Програмам МП дуже складно перекладати розмиті фрази або слова з подвійним значенням. Наприклад:

«*My first time meeting him will always be remembered*» – «Перший мій зустріч з ним завжди запам'ятається» (відредаговано неправильно).

Якщо замінити пасивну конструкцію активною, переклад стане зрозумілішим, але досі не бездоганним:

«*I will always remember meeting him*» – «Я завжди буду пам'ятати, як його зустрічали».

Однак варто продовжити редагування речення, уникнувши вживання герундія:

«*I will always remember how I met him*» – «Я завжди буду пам'ятати, як я його зустрічав» (відредаговано правильно).

Як бачимо, лише після внесення останніх змін машинний переклад речення українською мовою виявився правильним.

15. Вживання фраз, однакових за структурою. Сюди віднесемо фрази, що мають однакову синтаксичну структуру і неодноразово вживаються в кількох частинах документа. Наприклад:

Follow these three steps to build your MT engine:

1. *Gather your training data.*
2. *Build the MT engine.*
3. *Translate your clients' files.*

Звичайно, розглянуті вище рекомендації не можуть виправити усі недоліки МП. Однак, як показує практика, дотримання цих простих правил та належна адаптація системи МП можуть значно спростити постредагування тексту цільовою мовою.

Постредагування передбачає удосконалення тексту, отриманого у процесі МП, редактором, який зазвичай проходить спеціальну підготовку та має досвід роботи з текстовими документами такого типу. Кількість часу та праці, необхідних для постредагування, є одним із ключових факторів, які слід враховувати під час оцінки економічної ефективності МП. Літературні, рекламні та інші тексти, зазвичай не рекомендовані для МП, також не підлягають постредагуванню. Щоб забезпечити належну якість тексту, наближену до людського перекладу, редакторам може знадобитися заново переписати текст. У цьому випадку переваги МП практично зводяться до нуля.

Процес постредагування тексту МП включає такі етапи:

1. Перечитування речення чи фрагменту тексту МП.
2. Порівняння вихідного тексту МП з оригінальним текстом.
3. Оцінювання якості кожному фрагменту тексту перекладу на основі стандартизованих інструкцій.
4. Прийняття рішення щодо достовірності перекладу того чи іншого фрагмента.
5. Удосконалення або повторний переклад тексту [9].

Машинний переклад з постредагуванням є доцільним під час роботи з великими за обсягом та подібними за формою текстами. Великі за обсягом переклади часто виконуються у стінах перекладацьких бюро, які спеціалізуються у певних предметних галузях. Тому використання ефективних, хоча й часто дорогих систем МП останнього покоління, є економічно виправданим у цих випадках. Однак ні перекладацькі бюро, ні перекладачі-фрилансери не можуть ефективно використовувати програмні продукти для МП без необхідності постредагування.

Висновки. Перш ніж вдаватися до МП, слід сформулювати чітке уявлення про його кінцевий результат та усвідомити недоліки цієї технології. Окрім цього, необхідно враховувати, що системи МП потребують складного налаштування, вдосконалення та «навчання» у певній предметній галузі, без якого вони не виконують покладених на них функцій. Машинний переклад доцільно використовувати під час роботи з великими за обсягом і подібними за формою текстами, що належать до однієї предметної галузі (здебільшого технічної). Поліпшення якості МП можна забезпечити шляхом дотримання проаналізованих вище рекомендацій, які можна застосовувати як при попередньому редагуванні текстів, так і безпосередньо під час їх написання. Постредагування тексту перекладу полягає у «виправленні» вихідного тексту МП з метою наближення його якості до стандарту перекладу, виконаного людиною.

Перспективи використання результатів дослідження. Результати дослідження можуть бути використані під час розробки змісту навчальних дисциплін, пов'язаних із застосуванням систем машинного перекладу текстів, для формування професійних навичок майбутніх філологів у закладах вищої освіти, професійного розвитку фахівців у галузі перекладу та редагування текстової інформації.

Література

1. Академічний тлумачний словник української мови [Електронний ресурс]. – Режим доступу: <http://sum.in.ua/s/diakrytychnyj>.
2. Асоянц П. Г. Автоматизовані технології науково-технічного перекладу / П. Г. Асоянц. – К. : КНЛУ, 2012. – 243 с.

3. Яковина В. Огляд та аналіз метрик оцінювання якості машинного перекладу / В. Яковина, В. Масюкевич // Вісник Національного університету «Львівська політехніка». – 2013. – № 771 : Комп'ютерні науки та інформаційні технології. – С. 101-107.
4. Янковець А. Особливості використання машинного перекладу в умовах професійної підготовки майбутніх перекладачів / А. Янковець // Науковий вісник Ужгородського національного університету. Серія : Педагогіка. Соціальна робота. – 2012. – Вип. 25. – С. 233-236. – Режим доступу: http://nbuv.gov.ua/UJRN/Nvuuped_2012_25_80.
5. Chérageui M. A. Theoretical overview of machine translation / M. A. Chérageui // Proceedings ICWIT. – 2012. –р. 160-169.
6. How to improve the quality of machine translation [Online resource]. – Access Mode: <http://www.online-translator.com/Help/Tips/>.
7. Koehn P. Statistical Machine Translation / P. Koehn. – Cambridge University Press, 2009. – 433 p.
8. Okpor M. Machine translation approaches: issues and challenges / M. Okpor // International Journal of Computer Science Issues (IJCSI). – 2014. – Vol. 11(5). – P. 159-165.
9. Post-editing guidelines [Online resource]. – Access Mode: https://kantanmt.com/documents/Post-Editing_Guidelines.pdf.

References

1. Akademichnyi tlumachnyi slovnyk ukrainskoi movy [Elektronnyi resurs]. – Rezhym dostupu: <http://sum.in.ua/s/diakrytychnyj>.
2. Asoiants P. H. Avtomatyzovani tekhnolohii naukovo-tekhnichnoho perekladu / P. H. Asoiants. – K. : KNLU, 2012. – 243 s.
3. Yakovyna V. Ohliad ta analiz metryk otsiniuvannya yakosti mashynnoho perekladu / V. Yakovyna, V. Masiukevych // Visnyk Natsionalnoho universytetu «Lvivska politehnika». – 2013. – № 771 : Kompiuterni nauky ta informatsiini tekhnolohii. – S. 101-107.
4. Yankovets A. Osoblyvosti vykorystannia mashynnoho perekladu v umovakh profesiinoi pidhotovky maibutnix perekladachiv / A. Yankovets // Naukovyi visnyk Uzhhorodskoho natsionalnoho universytetu. Seriiia : Pedagogika. Sotsialna robota. – 2012. – Vyp. 25. – S. 233-236. – Rezhym dostupu: http://nbuv.gov.ua/UJRN/Nvuuped_2012_25_80.
5. Chérageui M. A. Theoretical overview of machine translation / M. A. Chérageui // Proceedings ICWIT. – 2012. –р. 160-169.
6. How to improve the quality of machine translation [Online resource]. – Access Mode: <http://www.online-translator.com/Help/Tips/>.
7. Koehn P. Statistical Machine Translation / P. Koehn. – Cambridge University Press, 2009. – 433 p.
8. Okpor M. Machine translation approaches: issues and challenges / M. Okpor // International Journal of Computer Science Issues (IJCSI). – 2014. – Vol. 11(5). – P. 159-165.
9. Post-editing guidelines [Online resource]. – Access Mode: https://kantanmt.com/documents/Post-Editing_Guidelines.pdf.

Рецензія/Peer review : 03.06.2019

Надрукована/Printed : 04.07.2019