АПЕЛЬСИНОВИЙ БОРИС

«КОМП’ЮТЕРНІ НАВИЧКИ ПЕРЕКЛАДАЧА»

ЗМІСТ

ВСТУП 3

РОЗДІЛ 1 ПЕРЕКЛАДАЦЬКИЙ ПРОЦЕС І ІНСТРУМЕНТИ ЙОГО ЗАБЕЗПЕЧЕННЯ 6

1.1. Процес перекладу 6

1.2. Аналіз вихідних файлів 7

1.3. Переклад 10

1.3.1. CAT-інструменти 10

1.3.2. Перекладацька пам’ять 11

1.3.3. Засоби керування термінологією 14

1.4. Різновиди CAT-інструментів 15

1.5. Автоматичні перевірки 18

РОЗДІЛ 2 ТЕНДЕНЦІЇ В ІНДУСТРІЇ ПЕРЕКЛАДУ. МАШИННИЙ ПЕРЕКЛАД 20

2.1. Сучасні позиції CAT-інструментів 20

2.2. Впровадження машинного перекладу 22

ВИСНОВКИ ДО РОЗДІЛУ 2 25

ЗАГАЛЬНІ ВИСНОВКИ 26

РЕЗЮМЕ 27

ВСТУП

У сучасну епоху глобалізації та інформаційного вибуху, коли обсяг даних, доступних людству, подвоюється кожні 5–10 років, роль засобів автоматизації перекладу та машинного перекладу набуває все більшої важливості. Саме від правильної та ефективної взаємодії перекладача з різними інструментами залежить швидкість поширення інформації та її якість.

Перекладацький процес у сучасному стані потребує від виконавця великої кількості компетенцій, знань і вмінь, щоб створювати в результаті роботи конкурентоспроможний і якісний продукт. До них входять, окрім власне знання мови, вміння працювати з різними типами файлів, які використовуються під час перекладу, володіння спеціальними технічними засобами, здатність виконувати низку перевірок, навички організації та керування проектами тощо.

Унаслідок прийняття закону про забезпечення функціонування української мови як державної зростатиме потреба у кваліфікованих перекладачах програмного забезпечення, веб-сайтів і технічної документації. Кожна із цих категорій має власні особливості та передбачає використання спеціалізованих CAT-інструментів. Тому перекладачам-початківцям, які бажають закріпитися в цій сфері, слід знайомитися з різними сценаріями застосування таких програм та використовувати їх на практиці, наприклад, під час навчального процесу та самостійної роботи.

Теоретичною базою роботи слугують праці дослідників та професіоналів-практиків із галузі перекладу (Ральф Крюґер [13], [14], Алан Мелбі [18], [19], Рафа Ломбардіно [17], Карлос да Сілва Кардозо Тейшейра [26], Ентоні Пім [22], [23]) а також документація до програм із відкритого доступу.

Актуальність дослідження обумовлена, в першу чергу, високим рівнем застосування комп’ютерних технологій у сучасному процесі перекладу. Під час роботи перекладач використовує програми з різним функціоналом, які

призначені для поліпшення, оптимізації та забезпечення якості процесу перекладу. Базове розуміння принципів роботи сучасного перекладацького бюро або перекладача-фрилансера та вміння користуватися програмами автоматизованого перекладу значно полегшує вступ до цієї галузі та скорочує термін навчання в ході роботи. Більш глибокі знання різних CAT-інструментів і прикладних засобів суттєво збільшують продуктивність перекладача, якість його роботи, що призводить до зростання його конкурентоспроможності та, як наслідок, вартості на ринку праці. Вивчення таких програм на базових освітньо-кваліфікаційних рівнях надаватиме випускникам спеціальності 035 Філологія конкурентну перевагу перед випускниками інших ВНЗ.

Мета дослідження — виявлення та узагальнення компетенцій, необхідних перекладачеві для успішної професійної діяльності в сучасному перекладацькому процесі.

Мета передбачає виконання наведених нижче завдань.

1. Визначити основні етапи сучасного перекладацького процесу.
2. З’ясувати компетенції, необхідні перекладачеві на кожному етапі.
3. Описати різні види програм, які використовуються під час перекладу та прийоми роботи з ними.
4. Систематизувати накопичені відомості та знання для можливості їх практичного використання з метою впровадження в навчальний процес, ознайомлення потенційних перекладачів і створення загального уявлення серед цільової аудиторії про сучасний процес перекладу.

Об’єктом дослідження став сучасний перекладацький процес і CAT-інструменти (засоби автоматизованого перекладу), наявні в перекладацькому полі.

Предметом дослідження є компетенції, якими повинен володіти перекладач під час роботи з перекладацькими програмами.

Матеріалом дослідження вибрано прогами різної спеціалізації, що застосовуються учасниками перекладацького процесу: засоби розпізнавання тексту (ABBYY FineReader), пропрієтарні (SDL Trados Studio, memoQ,

Translation Workspace XLIFF Editor) та онлайнові (Smartcat.ai, crowdin.com) CAT-інструменти, засоби перевірки якості (Xbench, Linguistic ToolBox).

Методи дослідження. Для визначення основних етапів перекладацького процесу та застосовуваного програмного забезпечення застосовано метод спостереження. За допомогою аналізу вибраного ПЗ визначався його функціонал, доцільність використання під час роботи з різними типами файлів і проектів. Щоб засвідчити отримані дані та визначити приналежність програми до певної категорії використовувався описовий метод. Зіставний метод застосовувався для порівняння різних програм та визначення їх переваг і недоліків під час певних видів роботи.

Практичне значення дослідження полягає у розширенні знань учасників навчального процесу спеціальності 035 Філологія щодо принципів роботи перекладацької сфери, формуванні орієнтирів для впровадження нових напрямів викладання, інформуванні осіб, зацікавлених у перекладацькій діяльності, щодо володіння необхідними компетенціями перед виходом на ринок праці. Крім того, здобуті результати можуть слугувати базою для більш детального вивчення процесів кожного окремого етапу перекладу або перекладацьких програм.

Робота складається зі вступу, трьох розділів із висновками до них, загальних висновків і списку використаних джерел (30 найменувань). Загальний обсяг роботи — 35 сторінок (обсяг основного тексту — 32 сторінки).

РОЗДІЛ 1
ПЕРЕКЛАДАЦЬКИЙ ПРОЦЕС І ІНСТРУМЕНТИ ЙОГО ЗАБЕЗПЕЧЕННЯ

Наразі майже всі професійні перекладачі використовують комп’ютери та комп’ютерні технології під час своєї роботи [8].

1.1. Процес перекладу

Відтоді як дослідники Хатчінс і Сомерс [10] на початку 1990-х представили свій варіант континуума автоматизації перекладу, перекладацька галузь повільно, але з невпинним прискоренням відійшла від традиційного перекладу людьми до автоматизації цього процесу. Тим часом досягнуто такого високого ступеня автоматизації, що перекладацький процес дедалі більше осмислюється як форма взаємодії людина-машина [20] або перекладач-машина [6]. Для перекладачів це передбачає кардинальну зміну в їхньому робочому процесі та необхідних навичках. Передумови для посилення автоматизації перекладу надходять з двох різних напрямків; з одного боку — це постійний розвиток нових та потужних перекладацьких інструментів, а з іншого — накопичення та доступ до великих обсягів електронних перекладацьких даних.

Перекладацькі інструменти у різних їх формах можуть застосовуватися на всіх етапах сучасного процесу технічного перекладу. Однак це залежить від перекладацько-підготовчої діяльності, як-от створення корпусів паралельних текстів, сполучення вихідних та цільових текстів або термінологічний аналіз технічних документів під час етапу фактичного перекладу, а також пост-редагування.

Процес перекладу складається з кількох послідовних етапів, які відрізняються за своєю специфікою один від одного та потребують застосування різних інструментів автоматизованого перекладу. За типом

завдання їх можна розподілити на 6 видів: аналіз вихідних файлів, власне переклад, редагування, вичитування, виконання автоматичних перевірок і верстання кінцевих документів [9].

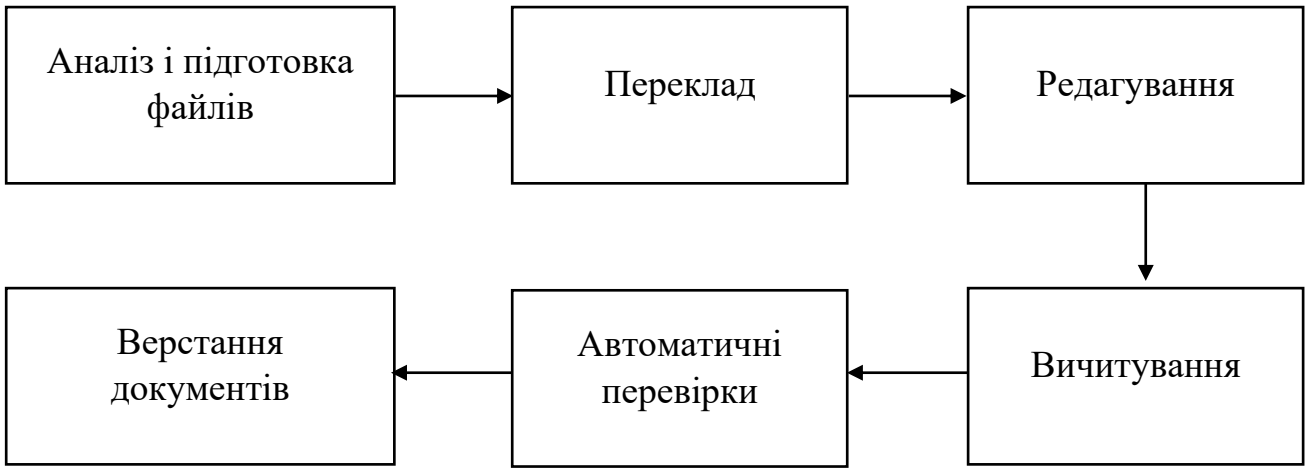

Рис. 1.1. Послідовність етапів перекладу

Кожен етап потребує використання певних комп'ютерних програм. З метою формування уявлення, що відбувається на кожному з етапів і які саме програми використовуються, розглянемо кожен із них окремо.

1.2. Аналіз вихідних файлів

Замовники та клієнти, які мають досвід співпраці з бюро перекладів, одразу надсилають файли, готові до перекладу. Бюро перекладів або перекладачеві потрібно лише пересвідчитися в їх сумісності з програмним забезпеченням, встановленим на робочих станціях, за необхідності — повторно підрахувати слова/символи та відсотки збігів. Під час роботи з новим типом файлу вперше слід також виконати так званий псевдопереклад, тобто перевірити, чи можливо взагалі згенерувати без помилок файл для експорту, якщо потрібно доставити замовнику не двомовні робочі файли, а вже кінцевий документ. До того ж використання псевдоперекладу має сенс під час роботи над програмним забезпеченням, щоб завчасно розуміти, чи достатньо місця у

відведених полях для тексту, наприклад, кирилицею, а також переконатися, чи всі фрагменти тексту імпортовано належним чином [21].

Часто замовник, у якого немає досвіду співпраці з перекладацькими агенціями, надсилає для перекладу файл, не звертаючи уваги на його формат. Однак чи не найважливіше розуміти, яких зусиль потребує файл для обробки різними програмами. Частий випадок, наприклад, — відсканові документи у форматі зображень або PDF. У такому разі необхідно спершу розпізнати текст, форматувати згідно з оригіналом та перевести в двомовний формат для перекладу.

Безперечно, для цього завдання можна скористатися стандартними програмами набору Microsoft Office, як от Word, Excel або звичайним текстовим редактором. Але якщо взяти до уваги середньостатистичну швидкість друку 200 знаків (або 40 слів) на хвилину [3], то на обробку загальноприйнятої перекладацької сторінки (1800 знаків) знадобиться 9 хвилин без урахування часу на форматування відповідно до оригіналу. Для порівняння, середньостатистична швидкість роботи перекладача становить 250 слів на годину або 8 сторінок на добу [27]. Таким чином, після простого підрахунку стає зрозуміло, що на кожні 10 друкарських сторінок агенції знадобиться додаткові півтори години робочого часу для обробки файлів невідповідного формату. Це призводить до збільшення витрат і, як наслідок, впливає на кінцеву вартість послуг.

Щоб уникнути непотрібних витрат, існує велика кількість як окремих програм, так і мережевих сервісів. Серед них, зокрема:

- Kofax OmniPage — програма, призначена для оптимізації робочих процесів підприємств і організацій;
- ABBYY FineReader — одна з перших програм розпізнавання тексту на ринку, здатна працювати з багатьма форматами, має онлайнову версію з певними обмеженнями;
- Adobe Acrobat Pro DC — рішення від провідної компанії-виробника ПЗ Adobe з великою кількістю функцій, яке може працювати зі хмарою;

• FreeOCR — безкоштовна альтернатива, побудована на рушію Tesseract від компанії Google.

Щоб визначити компетенції, якими повинен володіти перекладач на цьому етапі, розглянемо один із цих інструментів для автоматизації процесу — програму оптичного розпізнавання тексту ABBYY FineReader. Згідно з власними даними компанії ABBYY, її використовує третина компаній із списку Forbes 100 та понад 50 мільйонів осіб із понад 200 країн [28].

Програма ABBYY FineReader дає змогу швидко й чітко розпізнавати відсканований текст або документи з фотографій і перетворює їх в електронні файли, які можна редагувати, або файли PDF із функцією пошуку. Технологія Adaptive Document Recognition Technology сприяє збереженню початкової структури документів, що складаються з кількох сторінок. Це також стосується розміщених у тексті таблиць, приміток, нумерації, змісту, порядку сторінок, положення блоків тексту тощо.

ABBYY FineReader підтримує розпізнавання тексту в таких форматах:

- PDF, включно з PDF/A, PDF/UA;
- зображення TIFF, JPEG, PNG, BMP, PCX, GIF, DjVu, XPS;
- документи, доступні для редагування DOC(X), XLS(X), PPT(X), VSD(X), HTML, RTF, TXT, ODT, ODS, ODP.

Отримані текстові дані доступні для редагування як у самій програмі, так і у вигляді експортованих файлів, які можуть бути в одному з наведених вище форматів, а також у форматів електронних книг EPUB або FB2.

Уже на цьому етапі перекладач має володіти технічними знаннями про можливі типи файлів, а також вміти працювати з ними, щоб заощадити час та тим самим підвищити власну продуктивність або продуктивність команди, якщо мова йде про перекладацьке бюро.

Наступним кроком (або першим, якщо замовник одразу надіслав файли потрібного формату, без потреби їх розпізнавання / редагування) у підготовці до перекладу є створення проектів та конвертація файлів у формат, необхідний

для роботи в одній із наявних у перекладача CAT-програм, або завантаження їх на сервер, якщо робота над ними відбуватиметься в онлайновому редакторі або із загальним доступом кількох перекладачів.

1.3. Переклад

Саме на цьому етапі від перекладача вимагається найбільше компетенцій, оскільки він є основним у цьому процесі, потребує найбільше часу та передбачає активну взаємодію з принаймні одним CAT-інструментом.

1.3.1. CAT-інструменти

Computer-assisted translation або *computer-aided translation* (CAT) є видом перекладу з однієї мови на іншу, за якого людина-перекладач користується комп'ютерним обладнанням задля оптимізації та здійснення перекладу [5]. Сучасні CAT-інструменти розроблені на основі великої кількості засобів, метою яких є прискорення роботи перекладача та підвищення його продуктивності шляхом виконання певних завдань за допомогою комп'ютера. Назва CAT-інструменти узагальнено позначає сучасні перекладацькі технології. Найяскравішим їх різновидом є «робоча станція перекладача», концепція, розроблена в кінці 1970-х років дослідниками Мартіном Кеєм і Аланом К. Мелбі [19]., яка є актуальною до сьогодні. Робоча станція складається з інтегрованого комп'ютерного середовища для перекладу файлів у електронному форматі, у центрі якого перебуває перекладач. Перші CAT-інструменти з'явилися у вигляді систем керування термінологією на початку 1990-х років [18]. Відтоді вони значно розвинулися та отримали додаткові функції, як от менеджмент проектів і оцінка якості, а також можуть працювати з різними рушіями машинного перекладу.

Розглянемо деякі основні функції, щоб виявити їх роль у автоматизації системи перекладу та визначити компетенції, потрібні перекладачеві для їх належного застосування.

1.3.2. Перекладацька пам’ять

Необхідною функцією сучасної системи автоматизованого перекладу є застосування перекладацької пам’яті — двомовної бази даних перекладених вихідних текстів і їхніх відповідників цільовою мовою, яка пропонує схожі наявні сегменти під час перекладу нових текстів [12].

Перекладацька пам’ять ділить вихідний текст на одиниці, які називаються *сегменти.* Речення або аналогічна одиниця (наприклад, заголовок, елемент списку, посилання тощо) з вихідного тексту вважається окремим сегментом. Під час роботи перекладача над документом програма з перекладацькою пам’яттю по черзі відображає кожен вихідний сегмент та пропонує повторно використати попередні переклади, якщо в базі даних наявний такий самий або схожий вихідний сегмент. Якщо його немає, перекладач може самостійно ввести власний варіант перекладу для нового сегмента. Після завершення роботи над сегментом переклад зберігається до бази даних, і програма переходить до наступного сегмента. Перекладацька пам’ять переважно є простою базою даних, що містить поле із сегментом вихідною мовою, поле із сегментом цільовою мовою, а також поле з метаданими, як от дата створення, дата останнього використання, назва перекладача / проекту / файлу тощо. Іншим варіантом створення перекладацької пам’яті є співставлення існуючих перекладених документів, замість їх поступового утворення й розширення під час роботи [7].

Такі можливості спричинили виникнення нових перекладацьких стратегій, які включають нелінійну обробку та переклад текстів [22], що задовольняє вимоги до глобального поширення вмісту. Новоутворені тексти стають нелінійними, оскільки вони «запозичені» із вже існуючих.

Міжнародний стандарт ISO 30042:2008 для формату перекладацької пам’яті — Translation Memory eXchange (TMX), який застосовується у більшості програм автоматизованого перекладу. Проте досить поширеними є також інші формати, як от SDLTM і MQTM. Хоча програми, які використовують перекладацькі пам’яті не формату TMX, мають засоби їх перетворення, під час підключення користувачу варто звертати увагу, який саме формат необхідний для вибраної програми. Використання несумісної перекладацької пам’яті зазвичай неможливе, тому необхідно слідкувати, щоб виконувалися вимоги до формату.

Збіги з перекладацькою пам’яттю, залежно від точності, поділяються на кілька типів, наведених нижче.

Без збігу. Коли у перекладацькій пам’яті немає сегменту, відсоток збігу якого перевищує встановлену межу. Для SDL Trados Studio, наприклад, ця межа за промовчанням складає 70 %, для memoQ — 50%.

Неточні збіги (≤ 99 %). Сегменти, які дещо відрізняються від збережених у перекладацькій пам’яті. Якщо відмінність незначна й полягає у різній пунктуації в кінці, тегах, пробілах або числах, значення буде близьке до 99 %. Значення 85--95 % свідчить про відмінність в одне слово, 75–84% — два слова, менше 75 % — понад два слова.

Повні або точні збіги (100 %). Вихідний текст сегмента точно такий, як і збіг із перекладацької пам’яті. Але його контекст або не співпадає, або в метаданих немає відомостей про нього.

Контекстні збіги (позначаються 101 % у програмах memoQ і Smartcat та *CM* у програмі SDL Trados Studio). Вихідний текст сегмента точно такий, як і збіг із перекладацької пам’яті, крім того, сусідні сегменти збігаються з контекстом, збереженим у метаданих перекладацької пам’яті. Тобто наступний і попередній сегменти або заголовок документа ідентичні до контекстної інформації. У структурованих документах, наприклад, XML, HTML або розділених табуляцією (Excel, CSV, TXT тощо) роль контексту може виконувати позначення ідентифікатора (ID).

Ідеальні або подвійні контекстні збіги (позначаються 102 % або *PM*). Є різновидом контекстних збігів, за якого вихідні файли радше порівнюються з попереднім набором відповідних документів, ніж із перекладацькою пам’яттю. Збіг вважається ідеальним, якщо його зміст, метадані та ідентифікатори співпадають з попереднім файлом. Застосовуються для відновлення перекладу або частини документа, які повністю збігаються з початковим документом [17], [1].

Слід зауважити, що залежно від контексту смислове навантаження однакових сегментів може відрізнятися, тому навіть повні збіги потребують перевірки перекладачем, особливо якщо вони не контекстні або не ідеальні. Найчастіше таке відбувається, якщо текст містить багато сегментів, що складаються з одного багатозначного слова. Наприклад, німецьке слово *Anschluss* може одночасно позначати як підключення електричних дротів, так і з’єднання трубопроводу або комутаційний порт.

Окрім відображення збігів власне сегментів, усі CAT-інструменти мають функцію конкордантного пошуку в перекладацькій пам’яті, яка дає змогу переглядати, наприклад, переклад окремих слів і словосполучень, використання власних назв, стиль мови у сегментах із попередніх проектів тощо. Найбільш доцільне використання цієї функції під час одночасної роботи кількох перекладачів над одним проектом зі спільною перекладацькою пам’яттю в хмарі або під час перекладу.

Отже, для належного використання перекладацької пам’яті необхідно розуміти принцип її побудови та взаємодії з програмою автоматизованого перекладу. Обов’язково треба володіти інформацією, який формат пам’яті необхідно підключати до вибраної програми. Не менш важливо знати, які позначення можуть застосовуватися до запропонованих збігів і як ними користуватися, щоб не знижувати якість перекладу.

1.3.3. Засоби керування термінологією

Програмне забезпечення для керування термінологією надає перекладачеві засоби автоматичного пошуку у певній термінологічній базі даних відповідників перекладу до слів, що з'являються в документі, або автоматично відображає терміни у вікні інтерфейсу програми, або відкриває запис у базі за допомогою гарячих клавіш. Деякі програми (SDL Trados Studio, memoQ, Déjà Vu) мають додаткові комбінації клавіш, за допомогою яких перекладач може під час перекладу додавати нові пари термінів до бази даних. Більшість систем автоматизованого перекладу мають вбудовані засоби перевірки термінології в інтерактивному режимі та в режимі команд. З їх допомогою можна визначити, чи була використана правильна комбінація термінів у різних сегментах перекладацької пам'яті вибраного проекту. Крім того, існують незалежні системи керування термінологією, які можуть забезпечити функціональність робочого процесу, візуальну систематику, працювати як засіб перевірки термінології (терміни, які використані неправильно, позначаються відповідно), і підтримують різні типи додаткових даних, як от зображення, відео чи звук.

Термінологічні словники можуть бути різних форматів, що залежить безпосередньо від програми. Найпоширеніші з них — звичайні таблиці Excel. Головна перевага полягає в тому, що за необхідності їх можна конвертувати до іншого формату, який підтримується певною CAT-програмою. Це здійснюється за допомогою, спеціальних інструментів, наприклад, програми керування термінологією від компанії SDL, MultiTerm. Крім того, термінологічну базу Excel можна завантажити до онлайнового CAT-інструмента, що неможливо з більшістю інших форматів.

Попри те, що міжнародним стандартом для формату систем керування термінологією є TBX (TermBase eXchange) [11], поширеними також є термінологічні бази форматів .CSV, .TXT, .XML. Крім того, ключові гравці на ринку програм автоматизації перекладу, як от Trados Studio та memoQ

використовують власні формати для термінологічних баз: SDLTB та MTB відповідно.

Таким чином перед перекладачем постає необхідність володіння компетенціями щодо створення, наповнення, редагування, упорядкування використання, конвертування та актуалізації термінологічних баз різних форматів, які є необхідною умовою забезпечення однорідності перекладу.

1.4. Різновиди CAT-інструментів

На ринку наявна велика кількість різних CAT-інструментів, кожен із яких має власний набір функцій, призначений для певного типу завдань і працює з різними форматами файлів. Попри те, що деякі з цих програм охоплюють широкий спектр завдань, одного засобу автоматизації перекладу як правило недостатньо. Перекладачеві необхідно володіти принаймні кількома з них, оскільки перекладацькі агенції співпрацюють із багатьма клієнтами, які можуть використовувати різне ПЗ, або навіть один і той же замовник надає проекти для виконання одразу в кількох системах.

Це може бути, наприклад, переклад технічної документації до програми в SDL Trados Studio та паралельна локалізація цієї ж програми в SDL Passolo. Або ж переклад у програмі Translation Workspace XLIFF Editor та виконання автоматичних перевірок за допомогою інструмента Linguistic ToolBox.

Таким чином, ще перед початком роботи перекладачеві необхідно розуміти, яке програмне забезпечення доведеться використовувати, його особливості, приклади доцільного застосування тощо. Що вище компетенція перекладача щодо добору та використання належної програми для здійснення перекладу й автоматичних перевірок, то краще його продуктивність за рахунок заощадження часу під час роботи в добре знайомих інструментах.

Доступні CAT-інструменти та перекладацькі системи можна представити у вигляді наведеної нижче таблиці.

Таблиця 1

CAT-інструменти

CAT-інструменти	Платні	Безкоштовні
Онлайнові	Memsource Wordbee XTM Cloud Crowdin	Smartcat MateCat Wordfast Anywhere
Окремі програми	SDL Trados Studio memoQ Wordfast (Classic і Pro) Memsource (пропр. верс.) Déjà Vu Across	OmegaT CafeTran Espresso

SDL Trados Studio — найбільш використовувана професійними перекладачами та агенціями універсальна система автоматизованого перекладу та керування проектами [25]. Має багатий функціонал і поширена по всьому світі. Підтримує понад 70 різних форматів файлів, має вбудовану функцію розпізнавання документів PDF, що дає змогу заощадити в певних випадках кошти на додаткових інструментах, наприклад ABBYY FineReader.

SDL Trados Studio містить інтегровані в робочий процес засоби машинного перекладу та пост-редагування. Якщо для одиниці перекладу немає збігу, програма підставить у цільовий сегмент його машинний переклад, який можна потім відредагувати.

Головним недоліком SDL Trados Studio є надмірна складність освоєння для деяких користувачів, оскільки велика кількість функцій доволі ситуативні та зазвичай не використовуються на практиці під час повсякденних перекладів.

Kilgray memoQ — головний конкурент SDL Trados Studio на ринку, оскільки має аналогічний функціонал, продуктивність і початкову вартість. Відповідно, програма також складна для вивчення як і аналог від компанії SDL.

Обидві з наведених вище програм можуть працювати з білінгвальними файлами одна одної, а також відкривати їх для перевірки та вичитки.

Smartcat — онлайнова система автоматизованого перекладу, яка містить усі основні функції звичайний CAT-інструментів: пам’ять перекладів, керування термінологією, машинний переклад, функцію одночасної роботи кількох перекладачів над одним документом. Цільова аудиторія — перекладачі та компанії. Smartcat надається у форматі SaaS (програма як послуга). Доступ здійснюється за допомогою мережі Інтернет.

Crowdin — веб-орієнтована платна платформа для керування перекладами. Пріоритетний напрямок роботи — локалізація ПЗ. Система орієнтована на розробників мобільних додатків, веб-сайтів, документації, ігор тощо. Унікальною функцією Crowdin є краудсорсинг, тобто можливість запросити користувачів для допомоги з перекладом і локалізацією свого продукту іншими мовами.

Crowdin також окрім звичних перекладацьких форматів (XLIFF, XML тощо) підтримує переклад програмного забезпечення, файлів із вихідним кодом і застосунків для мобільних платформ Android, iOS, Windows 10 Mobile, тощо.

Узагальнити переваги та недоліки використання тих чи інших програм можна у наведений нижче спосіб.

Переваги окремих програм автоматизованого перекладу:

— можливість роботи без необхідності підключення до мережі Інтернет;

— відсутність проблем із серверами: програма працює, доки увімкнуто пристрій користувача.

Недоліки окремих програм автоматизованого перекладу:

— необхідно встановлювати ПЗ на свій пристрій;

— ліцензія часто обмежена одним пристроєм;

— більшість ПЗ працює лише на ПК із ОС Windows;

— безпека й цілісність документів залежить виключно від пристрою, його втрата або пошкодження може знищити всі файли;

— тривалий час очікування виправлень і оновлень.

Переваги онлайнових програм автоматизованого перекладу:

— не потрібно встановлювати на комп’ютер;

— можна працювати за допомогою браузера з будь-якого середовища та пристрою з підключенням до Інтернету;

— переклад зберігається в реальному часі, тому вірогідність втрати даних зменшена;

— простіше організувати одночасну роботу кількох перекладачів над одним проектом;

— своєчасні оновлення та виправлення.

Недоліки онлайнових програм автоматизованого перекладу:

— постійний ризик зламу та викрадення даних із мережі (проте це предмет дискусії, оскільки таке можливо й на окремих програмах);

— висока залежність від стабільного підключення до Інтернету.

1.5. Автоматичні перевірки

Програми для перевірки якості перекладу (QA-програми від Quality Assurance) — це інструменти для пошуку поширених помилок у текстах перекладів за визначеними параметрами. Вони працюють виключно з білінгвальними файлами, які створюються в результаті роботи CAT-програм. Їх завдання — аналіз вихідного і перекладеного тексту в певному сегменті та реєстрація всіх підозрілих співпадінь. Після завершення роботи користувач отримує звіт із знайденими помилками та вирішує, чи це справді помилки, чи просто хибні спрацювання.

Велика кількість розробників систем автоматизованого перекладу мають у своїх програмах такі засоби. До них відносяться SDL Trados Studio, memoQ, Wordfast, Across та інші. Крім того, існує багато окремих інструментів, найпопулярнішими серед яких на сьогодні є Xbench, QA Distiller, Verifica, ErrorSpy, Linguistic Toolbox.

Параметри кожної перевірки можна налаштувати відповідно до власних потреб або побажань клієнта та зберегти в окремий файл конфігурації, який можна використовувати у разі тривалої співпраці з кількома замовниками, проекти яких вимагають перевірок різного типу [2].

До стандартних основних перевірок належать такі:

— пошук неоднорідного перекладу однакових вихідних сегментів або ідентичного перекладу різних;

— різні знаки в кінці сегментів;

— неправильне перенесення одиниць виміру, дробових знаків тощо;

— відсутність цифр, тегів або абревіатур у цільовому сегменті;

— значно відрізняється довжина тексту;

— переклад терміна відрізняється від зазначеного в глосарію;

— відсутні дужки, неправильні лапки тощо;

— помилки правопису.

Правильний добір перевірок дає змогу перекладачеві/редактору заощадити час і зосередити увагу на перевірці змісту та доведенні тексту до відповідності вимогам замовника.

ВИСНОВКИ ДО РОЗДІЛУ 1

У сучасному перекладацькому процесі активно застосовуються різні комп’ютерні технології, які призначені підвищити продуктивність роботи. Кожен із визначених етапів потребує від перекладача, крім мовних, велику кількість технічних компетенцій, що стосуються обробки даних для перекладу, власне перекладу, перевірок, внесення правок і контролю якості. Щоб підтримувати свою конкурентоспроможність, перекладач має володіти багатьма програмами, які призначені для виконання перекладу та супутніх завдань, як от попередньої обробки файлів, власне перекладу, створення та форматування кінцевих файлів, а також різних перевірок. Не менш важливі

знання принципів та алгоритмів роботи використовуваних програм, щоб більш чітко розуміти можливі похибки та негайно їх виправляти.

РОЗДІЛ 2
ТЕНДЕНЦІЇ В ІНДУСТРІЇ ПЕРЕКЛАДУ. МАШИННИЙ ПЕРЕКЛАД

Компанія-лідер у перекладацькій індустрії SDL провела наприкінці 2015 року опитування під назвою Translation Technology Insights, метою якого було оцінити вплив швидкого темпу розвитку технологій, що змінив принципи роботи перекладачів. [24]

2.1. Сучасні позиції CAT-інструментів

Опитування проведено серед 2784 компаній і фрилансерів у 115 країнах. 72 % респондентів дотримуються думки, що вони втратять конкурентну перевагу, якщо відмовляться від використання CAT-інструментів для підвищення перекладацької продуктивності. Попри те, що це опитування засвідчує необхідність і першочерговість якості в процесі перекладу, кінцевою метою діяльності перекладача є остаточний результат, який має відповідати вимогам до термінів здачі проектів та бути економічно обґрунтованим.

Саме тому технологічні інструменти все частіше й інтенсивніше використовуються в індустрії. Приблизно 90 % респондентів опитування, вказаного вище, використовували у своїй роботі системи автоматизованого перекладу. Вони також указали, що інструменти для підвищення перекладацької діяльності посідають перше місце серед способів задовольняти попит, який постійно збільшується. Тобто респонденти надають перевагу цим інструментам перед поліпшенням робочих процесів, розширенням штату працівників та залученням сторонніх ресурсів, як-от фрилансерів, партнерських

перекладацьких агенцій тощо. Майже три чверті з них вважають, що не зможуть утримати конкуренцію без використання цих інструментів.

Варто зауважити, що лише третина опитаних представників перекладацької індустрії (15 % фрилансерів і 21 % перекладацьких агенцій) користувалися для співпраці хмарними рішеннями або моделлю «програма як послуга», коли доступ до служби здійснюється за допомогою Інтернету, а користувач платить не за володіння програмою, а за її використання. Серед ймовірних причин такого низького показника респонденти зокрема назвали такі:

- ризики роботи в хмарі з конфіденціальними файлами;
- недовіра до хмарних інструментів зберігання файлів;
- нижча ефективність хмарних рішень порівняно із вбудованими інструментами.

На сьогоднішній день трохи більше чверті (28 %) опитаних осіб вважають рішення на хмарних платформах кращими за їх класичні реалізації. Проте слід зауважити, що 53 % цих самих респондентів дотримуються думки, що через 5 років хмарні інструменти досягнуть рівня їх вбудованих аналогів та навіть перевищать його. Це є підтвердженням того, що в перекладацькій індустрії, як і в будь-якій іншій, досить швидко запроваджуються нові, ефективніші технології.

Нещодавно наявні методи віддаленої роботи не мали змоги підтримувати та забезпечувати розподіл завдання на невеликі частини та його одночасне виконання кількома підрядниками, а затрати часу на обробку файлів були відчутно більшими. Сучасна робота перекладача будується на кращому взаємозв’язку, обробці нових форматів файлів для вмісту різного типу (зокрема XML) та використанні програмного забезпечення, яке варіюється від звичайних поштових клієнтів та таблиць до інструментів автоматизованого перекладу, машинного перекладу та засобів організації робочого процесу.

Хоча кожна зміна — логічний і обґрунтований крок уперед, загалом індустрія дуже швидко і суттєво змінюється, щоб задовольнити вимоги до рівня

розподіленої роботи якомога швидше. А це значить, що за кілька років перекладачам доведеться освоювати нові засоби та методи роботи, до яких потрібно готуватися заздалегідь.

На сьогодні значна частина процесу перекладу базується на володінні сучасними перекладацькими інструментами для їх ефективного застосування. Зокрема це стосується різноманіття типів файлів, з якими доводиться працювати перекладачеві або перекладацькій агенції. Так, як більшість із них не в змозі впливати на те, у якому вигляді вони отримують від замовника вміст для перекладу, основною перевагою як перекладача, так і інструмента автоматизованого перекладу є здатність працювати з файлами різних типів, включаючи звичайні текстові формати та незвичайні (наприклад XML) без потреби використовувати додаткові набори програмного забезпечення для кожного типу. Здатність перекладача працювати з файлами різних форматів безпосередньо впливає на кількість замовлень та, як наслідок, його ринкову вартість як виконавця.

Таким чином, у сфері CAT-інструментів відбувається поступовий рух до використання хмарних рішень, розподілу великих завдань на менші та кооперації кількох перекладачів під час роботи над проектом.

2.2. Впровадження машинного перекладу

Розробка та комерційне впровадження термінологічних баз даних з 1960-х років і перекладацької пам’яті наприкінці 1980-х — початку 1990-х років поклали початок накопиченню перекладацьких даних у цифровому форматі, що сприяло прискоренню робочих процесів. Великі обсяги даних (*Big Data*) із низьким порогом доступу тепер могли інтегруватися в робочий перекладацький процес. Прикладом таких баз даних цифрового перекладу є великі перекладацькі пам’яті, як-от загальнодоступна перекладацька пам’ять директорату Європейської Комісії, перекладацька пам’ять хмарної системи перекладу *MateCat* або хмара даних *TAUS*, за допомогою якої члени *Спільноти*

користувачів автоматизованого перекладу можуть надавати та/або обмінюватися великими зразками перекладів. Варто також згадати великі термінологічні бази даних, такі як *Inter-Active Terminology for Europe (IATE)* Європейського Союзу або термінологічна база даних компанії *Microsoft* на її мовному порталі, а також великі одномовні корпуси паралельних текстів, такі як *British National Corpus* або двомовні перекладацькі корпуси, наприклад *Europarl-korpus*. Крім того, інтегровані системи перекладу, як-от *Trados Studio, memoQ* або *Memsource* тепер пропонують велику кількість засобів, що полегшують інтеграцію електронних словників, зокрема *Linguee, Dict* або *Leo,* у робоче середовище перекладача [13].

Популярною є думка, що для успіху машинного перекладу спершу потрібно вирішити проблему розуміння природної мови [15]. Тому в цій галузі за останні десятиліття відбувся особливо помітний прогрес у напрямі досліджень, що стосуються штучного інтелекту. Завдяки нейронному машинному перекладу виникла потужна нова архітектура, яка вже випередила наявні системи машинного перекладу за показниками якості, і яка, безумовно, все частіше використовуватиметься у щоденній роботі перекладачів у майбутньому [14].

Сучасні системи машинного перекладу розроблені на основі великих корпусів текстів та покладаються на обчислювальну потужність для побудови статистичних моделей перекладу. Що більше вихідний корпус текстів та правильніше задано правила, то більш точний буде результат роботи рушія машинного перекладу.

У нових системах як правило застосовується гібридний і нейронний тип машинного перекладу. Гібридний рушій поєднує переваги як статистичного підходу так і підходу на основі правил [4]. Нейронний машинний переклад значно розвинувся завдяки технології глибинного навчання. А в 2018 році компанія Microsoft повідомила про паритет із перекладом людиною, що стало історичною віхою в індустрії [16].

Процес постредагування після систем машинного перекладу з високою ефективністю може здійснюватися фахівцями-спеціалістами які володіють лише обмеженими знаннями вихідної мови, але дуже добре обізнані з темою тексту, який необхідно перекласти. Лише після цього готовий текст повинен оброблятися перекладачем, щоб відповідати необхідному мовному рівню. Тому перекладачеві більше не потрібно володіти технічними знаннями з теми. Необхідно лише мати достатньо високий рівень знань мови [23]. Також важливо розуміти, за допомогою якого рушія було виконано переклад, що обробляється. Різні типи рушіїв мають свої особливості генерування цільового тексту. Варто пам’ятати, що незважаючи на технологічний процес, машинний переклад не розуміє смислового навантаження тексту, тому результат часто може відрізнятися за значенням від вихідного, особливо якщо вживаються художні засоби, сталі вирази або усталена гра слів. Те, що на перший погляд здається логічно впорядкованим реченням, насправді може бути повною нісенітницею.

ВИСНОВКИ ДО РОЗДІЛУ 2

Поступовий розвиток машинного перекладу та його впровадження до перекладацького процесу стало причиною появи нового виду діяльності — постредагування, яка, попри свою схожість із роботою редактора, вимагає від користувача зовсім інших компетенцій. Особа, що працює з машинним перекладом, має розуміти, результат роботи якого рушія вона обробляє, бути обізнаною щодо частих і типових помилок, властивих машинному перекладу загалом і його окремим видам, а також бути здатною оцінювати, коли доцільно не використовувати запропонований машинний переклад, а натомість самостійно перекласти наново фрагмент, із яким виникли проблеми.

ЗАГАЛЬНІ ВИСНОВКИ

У ході роботи на основі наявних даних від перекладачів-професіоналів і дослідників визначено основні етапи перекладацького процесу. Для кожного з етапів визначено компетенції, вміння та навички, необхідні перекладачеві для успішної роботи з інструментами автоматизованого перекладу та додатковими засобами для забезпечення якості тексту та його відповідності встановленим вимогам.

Сучасний перекладач має вміти працювати в різних CAT-програмах, користуватися їх функціями, зокрема перекладацькою пам’яттю, пошуком конкорданс, термінологічною базою та вбудованими засобами перевірки якості. Необхідно також отримувати практичні навички постредагування, оскільки в галузі перекладу зростає використання машинного перекладу.

Професія перекладача не обмежується знанням мови. Володіння технічними засобами для перекладу надає конкурентну перевагу потенційному кандидату на відповідну посаду. Високий рівень цих знань також підвищує продуктивність роботи та якість перекладених письмових текстів.

Отримані дані можуть бути використані для актуалізації навчальних планів студентів з метою покращення їхньої підготовки до майбутнього працевлаштування. Крім того, описані програми та перекладацькі інструменти доцільно впровадити до навчального процесу з дисциплін «Переклад і редагування фахових текстів» та «Переклад науково технічної літератури».

РЕЗЮМЕ

Комп’ютерні навички, необхідні перекладачеві, досліджено на прикладі застосування програмного забезпечення під час перекладацького процесу. Визначено компетенції, якими має володіти перекладач на кожному із етапів роботи. Описано типи програм і інструментів, які використовуються під час обробки та перекладу документів. Досліджено тенденції розвитку та актуальність вивчення принципів роботи з машинним перекладом.

Використано такі методи дослідження: метод спостереження (визначення етапів перекладацького процесу та використовуваного ПЗ); аналіз (з’ясування доцільності використання ПЗ); описовий метод (опис функціоналу програм); зіставний метод (порівняння різних програм та визначення їх переваг і недоліків).

Встановлено, що перекладачеві необхідно розуміти принципи роботи та застосування CAT-інструментів, термінологічних баз, перекладацької пам’яті, машинного перекладу, програм для обробки файлів, а також засобів перевірки якості.

Ключові слова: *CAT-інструмент, QA-програма, SDL Trados, процес перекладу, компетенції перекладача, перекладацька пам’ять, термінологічна база.*

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. About Translation Memory Matches // SDL Help. URL: http://producthelp.sdl.com/sdl%20trados%20 studio/client_en/Edit_View/TMs/EVWorkingwithTMsAbout_Translation_Memory_ Matches.htm (дата звернення: 07.12.2019).

2. ApSIC Xbench Documentation. URL: https://docs.xbench.net/user-guide/overview/https://docs.xbench.net/user-guide/overview/ (дата звернення: 07.12.2019).

3. Average typing speed infographic // Ratatype. – 2019. URL: https://www.ratatype.com/learn/average-typing-speed/ (дата звернення: 07.12.2019).

4. Boretz A. AppTek Launches Hybrid Machine Translation Software / Adam Boretz // Speech Technology. – 2009. URL: https://www.speechtechmag.com/Articles/News/News-Feature/AppTek-Launches-Hybrid-Machine-Translation-Software-52871.aspx (дата звернення: 07.12.2019).

5. Bowker L. Computer-Aided Translation Technology: A Practical Introduction. / Lynne Bowker. – Ottawa: University of Ottawa Press, 2002. – 185 с.

6. Bundgaard K. (Post-)Editing – A Workplace Study of Translator-Computer Interaction at Textminded Danmark A/S / Bundgaard Kristine – Århus, 2017.

7. CAT Tools vs. Machine Translation: What's the Best Method? // Asian Absolute. – 2015. URL: https://asianabsolute.co.uk/blog/2015/12/21/cat-tools-vs-machine-translation-whats-the-best-method/ (дата звернення: 07.12.2019).

8. Christensen T. Translation-Memory (TM) Research: What Do We Know and How Do We Know It? / T. Christensen, A. Schjoldager. // Hermes. – 2010. – №44. – С. 89–101

9. Development and implementation of quality management systems in translation agencies // Technolex Translation Studio. – 2015. URL: https://technolex.com/articles/development-and-implementation-of-quality-management-systems-in-translation-agencies.html (дата звернення: 07.12.2019).

10. Hutchins J. An Introduction to Machine Translation / J. Hutchins, H. Somers. – Cmabridge: University Press, 1992. 351 с.

11. ISO 30042:2019. Management of terminology resources — TermBase eXchange (TBX). 2019. 42 с. (Міжнародна організація зі стандартизації)

12. Karsch B. Terminology Managementand MT / Barbara Karsch. // Circuit. – 2012. – №117. С. 11–14.

13. Krüger R. Technologieinduzierte Verschiebungen in der Tektonik der Translationskompetenz / Ralph Krüger. // trans-kom. – 2018. – №11. С. 104–137.

14. Krüger R. Von Netzen und Vektoren – Neuronale maschinelle Übersetzung / Ralph Krüger. // MDÜ. – 2017. – №63. С. 38–44.

15. Lehrberger J. Machine Translation: Linguistic Characteristics of MT Systems and General Methodology of Evaluation / J. Lehrberger, L. Boreau. – Amsterdam: John Benjamins Publishing, 1988. 240 с.

16. Linn A. Microsoft reaches a historic milestone, using AI to match human performance in translating news from Chinese to English / Allison Linn // Microsoft Blog. – 2018. URL: https://blogs.microsoft.com/ai/chinese-to-english-translator-milestone/ (дата звернення: 07.12.2019).

17. Match rates from translation memories and LiveDocs corpora // memoQ Documentation. URL: https://docs.memoq.com/current/en/Things/things-match-rates-from-translation-m.html (дата звернення: 07.12.2019).

18. Melby A. Interview to Jost Zetzsche / Alan K. Melby. – 2013. URL: http://www.internationalwriters.com/BigWave/BigWaveAKM.htm (дата звернення: 07.12.2019).

19. Melby A. Multi-level translation aids in a disbributed system / Alan K. Melby. // Linguistic Series. – 1982. – №47. С. 215–220.

20. O'Brien S. Translation as human-computer interaction. / Sharon O'Brien. // Translation Spaces. – 2012. – №1. С. 101–122.)

21. Pseudo-translation plugin settings // memoQ Documentation. URL: https://docs.memoq.com/current/en/Places/pseudo-translation-plugin-sett.html (дата звернення: 07.12.2019).

22. Pym A. The Moving Text: Localization, translation, and distribution / Anthony Pym. – Amsterdam, 2004. 166 с.

23. Pym A. Translation skill-sets in a machine-translation age / Antony Pym. // Meta. – 2013. – №58. С. 487–503.

24. Research Study 2016: Translation Technology Insights // SDL. – 2017. URL: https://www.sdl.com/software-and-services/translation-software/research/.

25. Tabor J. CAT tool use by translators: what are they using? / Jared Tabor. – 2019. URL: https://go.proz.com/blog/cat-tool-use-by-translators-what-are-they-using (дата звернення: 07.12.2019).

26. Teixeira C. C. The Impact Of Metadata On Translator Performance: How Translators Work With Translation Memories And Machine Translation : дис. докт. філ. наук / Teixeira Carlos da Silva Cardoso – Tarragona, 2014. 225 с.

27. Translation Agency: Behind the Scenes // Technolex Translation Studio. – 2016. URL: https://technolex.com/articles/za-kulisami-byuro-perevodov.html (дата звернення: 07.12.2019).

28. What is ABBYY | Key Facts // ABBYY Official Website. URL: https://www.abbyy.com/en-ee/company/key-facts/ (дата звернення: 07.12.2019).