

Сергій Фокін
м. Київ, Україна

Проект навчальної програми автоматизованого перекладу

Abstract. *In spite of a noticeable CAT-tools diversity, among Ukrainian translators there is noticeable knowledge gap concerning the functionality and the rational ways of their usage in individual or collective projects. While modern neural translation yields nowadays much better results, the CAT tools seem to be rather an instrument to control the translation consistency and quality, a role to be stressed in translators' curricula. The basic functionality could be introduced through a particular software with simplified interface and methods accessing the TM aimed to develop the awareness of the utility rather than learning a particular UI.*

Key words: *CAT-tools didactics, TM, machine translation, hybrid translation, translation quality control automation*

Розробка програм автоматизованого перекладу, заснованого на використанні комп'ютерної перекладацької пам'яті, стала карколомним проривом у цифровій філології. Користувачі, які раніше сміялися з перлів комп'ютерного перекладу й переказували анекдоти, з повагою відзначали послідовність і точність перекладу окремих частини тексту за допомогою програм, базованих на комп'ютерній перекладацькій пам'яті (ТМ). Оскільки переклад пропонувався лише для збігів (повних або часткових) з фрагментами, що зберігалися у пам'яті, на виході користувач отримував не повний переклад, а лише для повних або часткових знайдених збігів з фрагментами оригіналу, тому й переклад цей назвали «автоматизований», а не «автоматичний». Поширилася й англomовна назва «Cat tools», яка вже без перекладу тиражується в Україні у розмовному, а подекуди – й офіційному просторах. Оскільки питомий український термін – програми автоматизованого перекладу, тут і надалі живимо абревіатуру ПАП.

Звичайно, повні або майже повні збіги траплялися далеко не для всіх вихідних речень, і, з одного боку, ці програми було раціонально використовувати лише для проєктів з великою кількістю однотипних текстів та з часто повторюваними реченнями. Якщо ж таких проєктів не було, не виникало й потужного стимулу опанувати ці програми. Перекладачів зупиняла висока ціна ліцензії, певна складність в опанування інтерфейсу, який відрізнявся в розробках з подібним функціоналом. Як і в будь-якому новому програмному забезпеченні, у ПАП виникали технічні помилки, і перекладачі ділилися досвідом, як долати перешкоди [1].

Звичайно, помилки виникали, як нам відомо, і в інших поширених програмах: інтернет-оглядачах, текстових редакторах, поштових менеджерах тощо.

Розв'язання проблем у програмах автоматизованого перекладу, щоправда, було ускладнено через їхню меншу поширеність, меншу кількість досвідчених користувачів та видпків успішного досвіду в розв'язанні проблем. З одного боку, ефективність цих програм була високою лише у великих перекладацьких проєктах, а, з іншого боку, висока ціна, незнайомий інтерфейс, неочікувані помилки, міфи про неетичність, некоректність і безперспективність використання комп'ютера в перекладі, відсутність оперативних порад щодо розв'язання технічних проблем призводило до того, що перекладачі відмовлялися від опанування ПАП. При цьому у простих безоплатних програмах використовувався статистичний машинний переклад, який не давав бажаних результатів. Але комп'ютерні технології перекладу розвивалися, і завдяки впровадженню нейромережових алгоритмів якість перекладу суттєво поліпшувалася. Хоча, як відзначають фахівці, не ідеальна, все ж, якість сучасного машинного перекладу стала порівнюваною подекуди з людським перекладом посередньої якості. Що раніше викликало сміх, змусило перекладачів серйозно замислитися щодо власного працевлаштування й обміркувати, чи не стануть машини їхніми серйозними конкурентами.

Перед фахівцями галузі перекладу поставала дилема: якщо вже можна отримати готовий переклад непоганої якості простим копіюванням, який сенс опанувати ПАП, переважно платні програми, які ще й до того ж потребують спеціальної підготовки. Подібні твердження лунають і серед корифеїв галузі. У своїй статті «Примерк Cat-tools» Й. Гуммель, засновник компанії “Trados”, зазначає: «Протягом десятиліть автоматизований переклад (CAT tools), базований на пам'яті перекладу речень, був стандартним інструментом для виходу на світовий ринок. Хоча спочатку САТ-інструменти були розроблені з урахуванням персональних комп'ютерів середини 90-х, і були пропозиції щодо зміни основної моделі даних, базова архітектура САТ лишилася без змін. На тлі грандіозних досягнень нейронного машинного перекладу (NMT) програми автоматизованого перекладу тепер видаються застарілими» [2].

Автор статті вважає, що все ж це твердження слід розглядати як певне емоційне перебільшення. У той самий час інші розвідки свідчать, що зацікавлення професіоналів цими знаряддями невпинно зростає. Звичайно, раціональним є гібридний підхід: якщо у програмі використовуються переклади для збігів, що вже наявні в перекладацькій пам'яті, а для інших застосовується нейромережовий переклад, береться найкраще від обох підходів. Однак на 2023 рік лише до 12 % програм поєднували нейромережовий переклад до САТ-tools [3]. Тому комусь могло здаватися, що на цьому тлі ПАП у чистому вигляді (без гібридного підходу) віджили своє, що насправді не так. За останній рік, згідно з дослідженням *European Language Industry Survey*, автоматизований переклад став лідером інноваційних технологій перекладу, при цьому деякі компанії розробляють власне програмне забезпечення, інколи суто для внутрішнього використання [4].

Для такого зростання є декілька передумов: розробники програм активно реагують на зміни ринкової кон'юнктури і диверсифікують функції, а роль ПАП тепер полягає не лише в забезпеченні читабельного прийняттого цільового тексту, а й у досягненні однотонності (гомогенності, послідовності) використання термінології, чого в нейромережевому перекладі, звичайно, домогтися важко. Особливої актуальності набуває питання уніфікації термінології у колективних перекладацьких проєктах, зокрема ж у зв'язку з поширенням різних варіантів перекладу непрофесійними перекладачами.

Поза тим, деякі програми автоматизованого перекладу здатні допомогти в автоматизації перевірки якості перекладу та врахувати необхідні адаптації з огляду на локалізацію, що також не забезпечується нейромережевим перекладом. Приміром, може відстежуватись відповідність відтворюваних термінів тим еквівалентам, що занесені до глосарія. У разі локалізації сайтів або програмного забезпечення деякі програми здатні полегшити завдання перекладача, приховавши технічні частини HTML-коду, залишаючи лише фрагменти тексту природною мовою для перекладу. Слід відзначити такі корисні опції як контроль кількості символів фрагмента цільового тексту, перевірку вживання великих літер, числівників, пунктуації кінця речень, коректного відтворення тегів тощо. На сайтах спільнот перекладачів ряснують дописи, статті, розвідки про зростання значущості програм автоматизованого перекладу, і ми не схильні думати, що це пов'язано з явною чи прихованою рекламою розробленого продукту, однак з етичних і легальних міркувань тут не згадуємо назви програм, аби не виокремлювати одні розробки з-поміж інших.

Отже, програми автоматизованого перекладу не виглядають аж ніяк застарілими: вони є надійною опорою в перевірці якості, у керуванні великими перекладацькими проєктами, а також під час здійснення локалізації. Тим не менш відзначимо, що може зупиняти перекладача та майбутнього фахівця (здобувача) на шляху опанування ПАП: 1) необхідність отримання ліцензії; 2) складність опанування інтерфейсу та функцій; 3) нерозуміння призначення та оптимальних умов використання ПАП; 4) технічні збої й несподівані помилки під час використання програм; 5) поліпшення якості машинного нейромережевого перекладу, який здійснюється у програмах з максимально простим інтерфейсом; 6) брак часу для навчання з використання ПАП; 8) можливість за потреби довідатися самостійно завдяки численним навчальним відоматеріалам (більшість перекладачів спільноти ProZ, що взяли участь в опитуванні, відзначали, що витратили на навчання до 30 годин, а 48% – до 10 годин [5]); 8) амбітні прагнення вийти на ринок усного перекладу, для здійснення якого немає нагальної потреби в опануванні ПАП.

Активні користувачі відзначають як плюси, так і мінуси автоматизованого перекладу, як зазначено у цьому дописі: «Для виконання окремих завдань ці знаряддя справді чудові. (...) Але вони ж і дозволяють замовникові (передусім спеціалізованим агенціям) пропонувати замовлення за менші кошти.

І мені вирішувати – чи погоджуватися на це, чи ні, також з огляду на подекуди безглузду організацію виробничого процесу, коли тобі надсилають файл з проектом листа, аби отримати попередній переклад 'Dear Sir' і 'Regards' на додачу з якимись недолугими розривами у назві компанії. Але розумний баланс все ж можливий, і залежить він значною мірою від встановлених бізнес-стосунків та здорового глузду клієнта» [6].

З власного досвіду можемо відзначити, що в одному з онлайн-сервісів деякі функції просто не працювали, поки користувач не авторизувався за допомогою облікового запису. При цьому програма не видавала повідомлень про помилку чи попереджень про необхідність авторизуватися – про шляхи розв'язання доводилося здогадуватися, так само, як і в роботі з деякими іншими функціями програм. Приміром, при створенні глосарія в більшості середовищ у назві колонки слід правильно поставити позначення мови і країни, при цьому дуже легко помилитися з великими і малими літерами, також слід обрати лише один формат файлу, врахувати, щоб не було зайвих пробілів, комірок без перекладу, але у разі помилок так само не з'являлися повідомлення про характер/тип помилки. Подібних проблем, що призводять до втрати маси часу й зусиль користувача, а інколи навіть унеможливають використання програми, можна було б уникнути за допомогою додавання до коду кількох простих умов.

Згідно з даними опитування, 93% перекладачів ProZ відзначають, що опанували бодай одну з таких програм [7], однак це фахівці, активні учасники спільноти, які вже обрали свій професійний шлях, у той час як також є суттєва кількість пасивних учасників. У той самий час, студенти ще перебувають у стані вибору професії. Якщо ці знаряддя можна опанувати за 10-30 годин, як уже зазначено, для молодих людей це не становитиме перешкоду, отже, студенти не поспішають вивчати програми суто про всяк випадок. Хоча у нас немає даних з проведення опитувань, констатуємо, що під час вибору спеціалізацій у 2023 році студенти ОС «Магістр» кафедри теорії та практики перекладу романських мов імені Миколи Зерова з трьох різних спеціальностей не обрали вибіркового предмету, власне присвячений машинному й автоматизованому перекладу, очевидно, з зазначених вище міркувань. Вибір програмного забезпечення для навчального процесу – складний процес, у якому має брати участь група фахівців та відповідальний із закупівель. А за умови наявності безоплатних програм, можливо, і не буде виправданим придбання платного ліцензійного забезпечення.

Враховуючи ті чи інші перешкоди в набутті ліцензій, виборі програми, яка оптимально підійде для навчання, складності опанування і суттєві розбіжності як у функціях, так і інтерфейсах, пропонуємо проєкт програми, яка слугуватиме насамперед для навчання й ілюстрації першочергово функцій ПАП: автоматизувати використання раніше перекладених фрагментів (речень, їхніх частин, термінів), допомагати у менеджменті колективного перекладацького проєкту шляхом забезпечення одноманітності використовуваних перекладацьких відповідників, створення колективної перекладацької пам'яті та

термінологічного глосарія, автоматизації контролю якості виконуваного перекладу. У той самий час пропонуємо максимально спростити інтерфейс так, аби час навчання становив від 30 до 60 хвилин, за який можна буде опанувати 70-80% функціоналу, наявного у сучасних ПАП. Наразі у програмі не реалізовано можливість опрацьовувати тексти зі складним форматуванням, а менеджмент проєкту обмежено використанням спільної перекладацької пам'яті і глосарія, однак у перспективі є прагнення позбутися цих обмежень.

Функції ПАП у запропонованій програмі:

Функція	Реалізація
Використання перекладацької пам'яті	ТАК
Спрощений перегляд пам'яті (доступний одним кліком)	ТАК
Пропозиція автозаміни наявних у пам'яті перекладів за повними та неповними збігами	ТАК
Можливість поповнення перекладацької пам'яті одним кліком під час виконання проєкту	ТАК
Спрощене поповнення поповнення глосарія новою парою еквівалентів (двома кліками)	ТАК
Статистична оцінка проєкту	ТАК
Автоматична перевірка якості перекладу	ТАК
Перевірка відповідності довжини оригіналу й перекладу	ТАК
Перевірка пунктуації кінця речень	ТАК
Перевірка числівників	ТАК
Перевірка послідовності вживання термінів відповідно до глосарія	ТАК
Автоматичне поповнення перекладацької пам'яті після перекладу	ТАК
Використання віддленої колективної перекладацької пам'яті (доступної усім учасникам онлайн)	ТАК
Переклад текстів різних форматів	НІ
Створення колективного проєкту	НІ (опосередковано здійснюється через ведення спільної перекладацької пам'яті)

Загалом у запропонованій ПАП передбачаються такі спрощення та оптимізації:

1. Застосування гібридного підходу (у разі, якщо програма не знаходить збіги з пам'яттю, використовувати сторонній сервіс нейромережевого перекладу); 2. Можливість вставляти текст оригіналу простим копіюванням у тестове поле; 3. Можливість поєднання перекладацької пам'яті і глосарія в єдиній таблиці поширених форматів xls, xlsx, csv; 4. Можливість завантажити одним кліком перекладацьку пам'ять у форматах xls, xlsx, csv; 5. Можливість однією кнопкою переглядати зміст поточної перекладацької пам'яті і глосарія (а кнопки назвати відповідно до її функції – «Переглянути перекладацьку пам'ять»); 6. Можливість одним кліком додавати в процесі роботи до проєктів *ad hoc* еквіваленти термінів і фрагментів речень з метою їх повторного використання; 7. Перенесення двома кліками еквівалентів терміна до глосарія; 8. Поєднання програми з тестовим завданням для самоперевірки засвоєння функцій і базової термінології; 9. Інтеграція програми автоматизованого перекладу і засобів збагачення комп'ютерної перекладацької пам'яті, зокрема, шляхом подрібнення перекладацьких фрагментів.

Наразі пропонується готовий робочий ескіз програми, оскільки алгоритм реалізовано спрощено, суто з ілюстративною метою (мовами PHP, JavaScript, оскільки вона доступна в інтернет-браузерах і не потребує використання сторонніх переважно платних сервісів – серверів і баз даних). Звичайно, наразі це передбачатиме невелику потужність програми, обмеженість обсягу текстів, що перекладаються (успішно тестувалися переклади обсягом до 100 сторінок), що на цьому етапі розробки не дозволяє здійснювати більш масштабні спільні проєкти, однак не завадить в ознайомленні з корисними функціями й сенсом використання програм автоматизованого перекладу, які в такий спосіб можуть бути проілюстровані не лише на застиглих слайдах та в словесному поясненні під час лекцій, а й у динамічному використанні під час практичних занять й отриманні конкретних результатів. Це дасть змогу швидко й ефективно власноруч оцінити переваги й недоліки автоматизованого перекладу, засвоїти їхнє призначення та функції. У разі успішної апробації в навчальному процесі передбачено створення нової потужнішої версії програми з використанням серверних технологій і можливістю дистанційного керування спільних проєктів.

Література

1. Hummel J. Sunsetting CAT-tools. 2018. URL: <https://coreon.com/2018/12/12/sun-setting-on-cat-tools/> 2. Texnolex Translation Studio, 2021. URL: https://technolex.com/uk/newtips.html?breaking_space=&show=24%C2%A0translator%E2%80%99s%20workbench&tag=trados%20studio 3. Tabor J. CAT tools in 2023: part six, 2023. URL: <https://go.proz.com/blog/cat-tools-in-2023-part-six> 4. European Language Industry Survey 2024 URL: <https://elis-survey.org/wp-content/uploads/2024/03/ELIS-2024-Report.pdf> 5. Tabor J. CAT tools in 2023: part five. URL: <https://go.proz.com/blog/cat-tools-in-2023-part-five> 6. Hughes R. Is a translator who does not use CAT tools unworthy of being a successful translator in today's market?, 2022. URL: <https://www.quora.com/Is-a-translator-who-does-not-use-CAT-tools-unworthy-of-being-a-successful-translator-in-todays-market> 7. Tabor J. CAT tools in 2023: part two. URL: <https://go.proz.com/blog/cat-tools-in-2023-part-two>