

УДК 8.81.13'25

ПРОБЛЕМИ АВТОМАТИЗОВАНОГО ПЕРЕКЛАДУ ДІЛОВОЇ КОРЕСПОНДЕНЦІЇ ЗА ДОПОМОГОЮ ВІЛЬНИХ ПРОГРАМНИХ ПРОДУКТІВ

Башманівський О. Л.

У статті проаналізовано проблеми автоматизованого перекладу ділової кореспонденції за допомогою безкоштовних у доступі програм-перекладачів. На прикладі вільного програмного продукту – Google Translator Toolkit встановлено алгоритм роботи перекладача, визначено перелік труднощів, які виникають у цьому процесі. З'ясовано, що незважаючи на свою відкриту простоту і доступність у використанні, сервіс Google Translator Toolkit є досить потужним засобом перекладу. Незаперечним є той факт, що автоматизований переклад хоча і полегшує роботу перекладача прискорюючи її, проте поки що не може повноцінно його замінити і вимагає постійного редагування.

Ключові слова: ділова кореспонденція, автоматизований переклад, вільний програмний продукт, трансформація структур, корпусна лінгвістика, лінгвістичні алгоритми.

ПРОБЛЕМЫ АВТОМАТИЗИРОВАННОГО ПЕРЕВОДА ДЕЛОВОЙ КОРРЕСПОНДЕНЦИИ С ПОМОЩЬЮ СВОБОДНЫХ ПРОГРАММНЫХ ПРОДУКТОВ

Башмановский А. Л.

В статье проанализированы проблемы автоматизированного перевода деловой корреспонденции с помощью бесплатных в доступе программ-переводчиков. На примере свободного программного продукта – Google Translator Toolkit установлено алгоритм работы переводчика, определен перечень проблем, которые возникают в этом процессе. Установлено, что несмотря на свою открытую простоту и доступность в использовании, сервис Google Translator Toolkit является достаточно мощным средством перевода. Неоспоримым является тот факт, что автоматизированный перевод хотя и облегчает работу переводчика ускоряя ее, однако пока не может полноценно его заменить и требует постоянного изменения.

Ключевые слова: деловая корреспонденция, автоматизированный перевод, свободный программный продукт, трансформация структур, корпусная лингвистика, лингвистические алгоритмы.

THE PROBLEMS OF AUTOMATED TRANSLATION OF BUSINESS CORRESPONDENCE USING FREE SOFTWARE

Bashmanivskiy O. L.

In the article the problem of automated translation of business correspondence using free software is analyzed. On the example of free software – Google Translator Toolkit installed the algorithm translator, a list of difficulties that arise during this process. There was found that despite its open simplicity and availability to use the service Google Translator Toolkit is a quite powerful tool for translation. The undeniable fact is that automated translation although it facilitates the work of the translator by accelerating it, but so far can not fully replace it and require constant editing.

Key words: *business correspondence, automatic translation, free software, the transformation of structures, corpus linguistics, linguistic algorithms.*

Постановка проблеми. Процес інтеграції в європейське суспільство охопив практично всі сфери нашого життя, і хоч не так швидко, проте відбуваються певні зміни в нашій країні. Поглиблення стосунків з європейськими державами в різних галузях передбачає покращення комунікації, проте це і є основною проблемою подальшого розвитку інтеграційних процесів. Радянське минуле залишило нам російсько-українську двомовність, проте значну некомпетентність у знанні європейських мов. Англійська мова, здавалося, могла б вирішити цю проблему мовного вакуума, проте застарілі методики і цілі навчання зробили величезні перепони в комунікації, зокрема у діловому листуванні.

Ведення сучасного бізнесу передбачає постійне співробітництво на всіх рівнях – обмін виробничою інформацією, погодження, планування, вирішення конфліктних ситуацій. Успішність просування міжнародних бізнес-проектів залежить не тільки від професійних особливостей певної діяльності, але й від комунікативних компетенцій, які реалізуються у здатності використовувати певні мовні засоби для досягнення поставлених цілей. Стилїстика офіційно-ділового стилю формувалася якщо не століттями, то десятиліттями і слугує для врегулювання певних

людських взаємин: відносини між державою і населенням, між країнами, між підприємствами, установами й організаціями, між індивідумом і суспільством [1].

Євроінтеграційні процеси в нашій державі ставлять одним із першочергових завдань переклад документації державного патентного фонду України, який складається з масиву паперових і електронних матеріалів, це передбачає створення термінологічних стандартів української мови та двомовних словників. На сьогодні українська мова має статус державної на світовому інформаційному ринку, що певною мірою ізолює нашу державу від світового контенту. Важко переоцінити актуальність і складність цих проблем, проте людство весь час працює над розробкою універсального комунікатора, а з винаходом першого комп'ютера одним з перших завдань для нього було налагодження перекладу текстів.

Метою нашої статті буде дослідити проблеми автоматизованого перекладу ділової кореспонденції на прикладі вільного програмного продукту Google Translate, найбільш вживаного серед користувачів мережі Internet.

Розглянемо чинники, які роблять програми машинного перекладу такими популярними. По-перше, користувачі застосовують їх при абсолютному незнанні мови, необхідності швидкого ознайомлення з великим обсягом лінгвістичного матеріалу, потребі зробити певну оцінку цього тексту. Більшість автоматизованих перекладачів досить вдало справляються з цією задачею, проте існують випадки, коли вони можуть навіть заплутати користувача – тексти юридичного характеру, для яких необхідна висока ступінь відповідності лексем перекладеного тексту оригіналу. По-друге, програми автоматизованого перекладу застосовуються у разі необхідності швидкого перекладу Web-сторінок, з метою швидкого розуміння змісту інформації, для подальшої її обробки. У цьому випадку програми-перекладачі є найбільш сприйнятливими як для користувачів зі знаннями цієї мови, так і для не знавців. По-третє, програми машинного перекладу застосовуються професійними перекладачами для створення чорнового матеріалу перекладу. У цьому випадку ці програми в декілька разів можуть збільшити продуктивність

професійного перекладача, якому залишається проводити редагування вихідного машинного тексту. І нарешті, програми автоматизованого перекладу є незамінними у разі проведення ділової переписки з іноземними партнерами, де ці програми виконують основну функцію – одержання необхідної інформації для користувача.

Ми можемо відмітити, що в українському перекладацькому бізнесі документознавства спостерігається явне відставання використання сучасних комп'ютерних технологій. Підвищенню якості перекладу ділового листування мали б сприяти технології Translation Memory і інформаційних систем, що автоматизують виробничий процес, проте, на жаль, тільки лідери ринку професійно використовують ці продукти. Тим часом, у західному світі використання технології Translation Memory уже давно стало звичним фактом, особливо при перекладі проектів об'ємних масивів.

Аналіз останніх досліджень. Проблемами автоматизованого перекладу займаються як комерційні корпорації, так і дослідники багатьох наукових напрямків. Проте найбільша увага вчених сфокусована саме в галузі лінгвістики, це ми можемо спостерігати на прикладі таких праць, як фундаментальна монографія Ф. Джорджа «Основи кібернетики» [3], Дж. Вудера «Science without properties», О. К. Жолковського «О правилах семантического анализа», Ю. М. Марчука «Проблемы машинного перевода», Г. С. Цейтін, М. І. Откупщиковой «Система анализа текста с процедурным представлением словарной информации» [4] та інших, в яких визначено основні принципи і проблеми комп'ютерного перекладу. Для повноцінного аналізу нашої проблеми слід залучити також наукові доробки в галузі корпусної лінгвістики О. Демської-Кульницької «Основи національного корпусу української мови», [2] та інших вчених (праці Б. Алтенберга, М. Баньки, У. Френсиса, Г. Кеннеді, Г. Ліча, А. Баранова, М. Михайлова, Р. Рикова, Л. Ричкової, С. Шарова) [6]. Зараз продовжується робота по створенню корпусів української мови (Національний корпус української мови – в Інституті української мови НАНУ та Український національний лінгвістичний корпус – в Українському мовно-

інформаційному фонді НАНУ). Ці наукові дослідження в Україні ще, на жаль, не набули широкої популярності й перебувають на початковому етапі свого становлення.

Виклад основного матеріалу. Для дослідження сучасних технологій перекладу ділового листування перш за все з'ясуємо, що ж таке ділова кореспонденція і які відмінності вона передбачає.

Ділова кореспонденція – це водночас і джерело інформації, і документ писемної комунікації для досягнення певної поставленої мети. Вона складається від імені юридичної особи та має юридичну силу. Будь-який за змістом документ, котрий пересилається поштою та вирішує певні організаційні питання і правові проблеми, можна віднести до ділового листа. Всю ділову кореспонденцію поділяють на службову (листи для встановлення офіційних службових контактів між суб'єктами підприємницької діяльності, вони зазвичай нескладні за своїм змістом та невеликі за обсягом), та комерційні (листи, які пов'язані з процесом погодження та виконання певних комерційних угод). У загальному вигляді ділові листи можуть бути присвячені наступним питанням: співробітництво з вітчизняними та іноземними фірмами; участь у торговій і комерційній діяльності; співробітництво у рамках консорціуму; питання пов'язані з підготовкою та реалізацією контракту; формам розрахунку і умовам платежу; командировання спеціалістів; підготовка іноземних і національних кадрів; реклаमाції і врегулювання претензій; влаштуванню на роботу; медичному обслуговуванню та іншим питанням [1].

Може здатися, що ділове листування нічим не відрізняється від звичайної кореспонденції, проте використання сталих фразеологічних одиниць, типових для цього стилю, характерна традиційність засобів вираження, наявність значної кількості архаїчних слів і виразів дуже ускладнює процес перекладу і спілкування між партнерами. У стилі ділової кореспонденції лексеми вживаються переважно в базових предметно-логічних значеннях (винятками є похідні предметно-логічні значення, які в цій галузі є термінологічними). Слід відмітити також й іншу особливість стилю ділової мови – відсутність зображальних

засобів у текстах (неприпустимість застосування метафор чи метонімії та інших засобів).

Слід зазначити, що у діловому листуванні, зокрема англійською мовою, наявні певні особливості: текст поділяється на абзаци без відступів; у верхньому лівому куті листа вказуються адреси відправника і адресата; дата відправлення зазначається трьома рядкам нижче, або у верхньому правому куті листа; текст розташований у центральній частині листа і, зазвичай, розпочинається з причини звернення; закінчується висловлюванням подяки і вітаннями; у разі, якщо діловий лист не є електронним, то підпис автора є обов'язковим і ставиться між вітаннями й іменем відправника. Ця сформована стандартизація мови ділових паперів забезпечує той ступінь комунікативної точності, що надає документу міжнародну юридичну чинність. Для стилю ділової кореспонденції існує певний набір традиційних мовних кліше чи штампів: «greeting» (Dear Sir, Dear Madam, Dear Sirs etc. I answer to your letter... «the body» (I beg to inform you, with a view to), «the closing paragraph» (Thank you... sincerely yours, sincerely, kind regards, with best regards, best wishes yours, very truly yours, cordially yours, yours faithfully, Thanks & Regards etc. Також слід пам'ятати, що будь-яка фраза, будь-яке речення повинні мати тільки одне лексичне значення і конкретне тлумачення. З цією вимогою допускаються повторення, що не припустимо в літературно-художньому стилі [5].

Для проведення машинного перекладу існують певні різновиди перекладацьких комп'ютерних систем: системи, що працюють, як правило, з великими масивами текстів і видають користувачу «чорновий» переклад, який можна використовувати для розуміння загального змісту тексту, або відредагувати і отримати остаточний документ; системи, котрі являють собою аналог традиційних «паперових» словників: користуючись ними, користувач може знайти переклад терміну, що його цікавить. На сьогоднішній день найбільш відомими системами, які здатні перекладати тексти з англійської на українську мову є Google Translate, Pragma Online, Bing Translator та інші. Всі вони використовують статистичний підхід до перекладу.

Поступово програмне забезпечення автоматизованого перекладу пройшло еволюцію від словників до технологій Translation Memory, принцип яких базується на використанні попередніх перекладів-корпусів, які зберігаються в двомовній базі даних – пам'яті перекладів. Найвідомішими представниками програм цього типу є Google Translator Toolkit, Wordfast Pro, SDL Trados, Omega T та інші.

На прикладі використання системи Google Translator Toolkit ми проаналізуємо принципи роботи і основні недоліки, які присутні в цьому програмному продукті, які зазвичай є спільними для аналогічних продуктів на нашому ринку. Відразу можна стверджувати, що незважаючи на свою відкриту простоту і доступність у використанні, сервіс є досить потужним засобом посереднього перекладу. Система забезпечує не тільки збереження та обмін документів, а й дозволяє використовувати глобальну пам'ять перекладів, водночас не потребує встановлення якихось додаткових елементів. І хоча зараз вона не є популярною серед професіоналів, проте для більшості користувачів вона є незамінним вікном у світ міжнародної комунікації.

Система автоматизованого перекладу Google Translate відбирає еквівалентні, варіантні та трансформаційні відповідники шляхом використання алгоритмічних засобів граматичного аналізу, звертаючись до наявних у ньому двомовних словників, які містять необхідну інформацію морфологічного, синтаксичного чи семантичного рівнів. Трансферна зона цієї автоматичної системи здатна обробляти семантичну структуру тексту. За умови відсутності еквівалента машина не завжди враховує необхідні лінгвістичні та екстралінгвістичні фактори: стилістичну приналежність відповідників, контекстуальне оточення, мовні норми, сполучуваність слів тощо. Google Translate – безкоштовний онлайн-сервіс машинного перекладу, який дозволяє автоматично перекладати слова, тексти, електронні документи, веб-сторінки багатьма мовами світу. Для певних мов користувачам система пропонує варіанти перекладів, наприклад для термінів і багатозначних слів [7].

Google Translator не є цілісною програмою, навпаки, даний веб-сервіс складається з багатьох компонентів і служб,

які взаємодіють при виконанні роботи. Недосконалістю цієї системи є певна примітивність автоматизованого перекладу – програма просто здійснює підстановку тих лексем, які існують в її базі даних. Одним з головних завдань Google Translate є виконання чорнового перекладу, а Google Translator Toolkit допомагає провести подальше редагування, запису в пам'ять перекладів, надання доступу іншим користувачам, тощо. Переклад здійснюється з використанням статистичного підходу до машинного перекладу, що дозволяє системі «навчатися» і самовдосконалюватися: коли Google Translate створює переклад, він шукає зразки в сотнях мільйонів документів, порівнює мільйони корпусів, щоб надати користувачу найкращий переклад. Знайшовши зразки в текстах корпусів, перекладених іншими людьми, Google Translate може робити інтелектуальні припущення щодо правильності перекладу. Цей процес пошуку зразків у великих обсягах тексту і називається «статистичним машинним перекладом». Оскільки переклади генеруються машинами, не всі вони ідеальні. Чим більше текстів певною мовою перекладають люди, а Google Translate їх аналізує, тим кращою буде якість перекладу. Для покращення якості роботи програми необхідні корпуси паралельних текстів і колекції пам'яті перекладів. Google Перекладач може виконувати оборотний переклад у будь-якій мовній парі, що підтримує система, але у більшості випадків реально виконує переклад через англійську, більш того, для української використовується як проміжна російська. Цим дуже погіршується якість перекладу, це добре помітно при перекладі з польської на українську, що здійснюється через англійську та російську [8].

Google Translate успішно інтегрується в браузер, завдяки спеціальним розширенням, що встановлюються додатково. Для Mozilla Firefox існує дуже багато плагінів, щодо Google Chrome, то дані компоненти вже встановлені у нових версіях. Існує також спеціальний клієнт для ОС Windows, що реалізований у безкоштовній та комерційній версії, яка відрізняється кількома компонентами. Сервіс також працює з популярними зараз мобільними операційними системами, такими як Android та IOS. Для мобільних пристроїв була створена спеціальна функція

голосового вводу, яка підтримує 15 мов (української, на жаль, поки що немає), а система може «прочитати» переклад на 23 мовах за допомогою синтезу мовлення [8].

Google Translate – перспективна система для перекладу, тому що постійне збільшення пам'яті перекладів і поповнення двомовних корпусів забезпечує кращу якість перекладів. Її перевагою є також те, що вона може самовдосконалюватися за рахунок збільшення баз даних. Однак статистичний метод не є ідеальним, тому система такого типу не зможе надати повноцінного перекладу.

Програми-перекладачі текстів більш-менш задовільно складають змістовну структуру майбутнього перекладу використовуючи, як правило, дослівну обробку тексту на матеріалі готових корпусів, причому з кількох можливих значень слів обирається частіше за все загальновживане. Така система роботи навряд чи зможе задовольнити фахівця певної галузі.

Робота з двомовними корпусами ускладнюється специфікою досліджуваних мов – англійська та українська мови належать до різних структурних типів мов (українська – до флективної, а англійська – до аналітичної). Велика кількість розбіжностей у будові мов та у наборів їхніх граматичних форм і конструкцій, які не дозволяють створити певні універсалії. Такі граматичні явища як модальні дієслова, артиклі, часові форми Continuous та Perfect, формальний додаток «it» не мають відповідностей в українській мові, а в англійській мові відсутні категорії роду і відмінку іменників і прикметників тощо. Відмінності притаманні також і побудові речень, тому що на відміну від української мови, англійській мові притаманний прямий порядок слів [5].

Аналізуючи найбільш суттєві помилки автоматизованого перекладу необхідно вказати на неправильний переклад дієслівних форм, відсутність відповідних лексем у словнику, неправильну структуру речень, визначення частини мови програмою, дослівний переклад фразеологічних зворотів та інших ідіоматичних виразів. Серед незначних помилок, які не впливають на розуміння змісту тексту і досить легко виправляються під час редагування, є переклад особового займенника, неправильний вибір лексеми багатозначного іменника, переклад прийменників

і використання артиклів. До «комп'ютерних помилок» необхідно також віднести морфологічну невизначеність – деякі англійські іменники і дієслова мають однакове написання та транскрипцію, а морфологічну належність визначає або артикль (для іменників), або частка «to» (для дієслів). Відсутність артикля чи частки «to» робить переклад повністю незрозумілим. Прийменники також викликають безліч помилок при автоматизованому перекладі, тому що значення багатозначного прийменника можна встановити тільки з урахуванням конкретного контексту, а відсутність в електронному словнику вказівок на правильне керування, зокрема дієсловами чи віддієслівними іменниками, робить для програми єдиний можливий переклад автономний – загалом повністю неправильний (In the accurate analysis – у точному аналізі – при точному аналізі). Пасивні конструкції в процесі машинного перекладу в більшості повністю спотворюються – відбувається неправильне розпізнавання омонімії та не враховуються можливості поєднання дієприкметника з модальними дієсловами – ці помилки потребують повної корекції редактора як у заміні недоконаного виду дієприкметника на доконаний, так і повної трансформації виразу до мови перекладу. Проблеми, які постійно існують в традиційному-авторському перекладі: невірне узгодження слів у реченні, неправильне визначення роду, відмінку, множини іменника, прикметника, числівника, визначення особи дієприкметника, дієвідміни дієслова наявні також і в автоматизованому перекладі.

Значна кількість фразеологічних зворотів, яка існує в обох мовах, створює великі проблеми користувачам, спотворюючи зміст речень при автоматизованому перекладі. І враховуючи, що технічна документація містить велику кількість різних фразеологічних зворотів загальнолітературного характеру, які не внесені до електронного словнику, в процесі машинного перекладу вони перекладаються дослівно (Of the same type – того ж типу – однакового типу), також в більшості помилковим є переклад англійських усталених зворотів, зокрема, зловживання прийменником «of» родового відмінка замість того, щоб знаходити більш звичні варіанти [5].

Висновки. Враховуючи вищесказане, слід констатувати, що програма-перекладач, що не має екстралінгвістичних знань про

оточуючий світ, не володіє предметною інформацією, не завжди правильно аналізує синтаксичні конструкції багатозначності, а отже, не може правильно перекласти численні технічно-юридичні словосполучення, перекладаючи їх за першим словниковим відповідником. Безсумнівним є факт необхідності обов'язкового подальшого редагування документа людиною-перекладачем, який знає не лише мову, а й орієнтується в предметній галузі. Безсумнівно, що програмний перекладацький продукт полегшує і прискорює роботу перекладача, проте поки що не може повноцінно його замінити. Проте, враховуючи темпи розвитку цієї галузі, можемо сподіватися на швидке зростання якості вільних перекладацьких програмних продуктів на нашому ринку.

Список використаних джерел

1. Биби́к С. П. Сюта́ Г. М. Ділові документи та правові папери / С. П. Биби́к, Г. М. Сюта – Харків: Фоліо 2006. – 493 с.
2. Демська-Кульчицька О. Основи національного корпусу української мови / О. Демська-Кульчицька. – Монографія. – Київ, 2005. – 218 с
3. Джордж Ф. Основы кибернетики / Ф. Джордж. – М. : Радио и связь, 1984. – 272 с.
4. Історія машинного перекладу – Вікіпедія [Електронний ресурс] // Wikipedia. – 2012. – Режим доступу : http://uk.wikipedia.org/wiki/Історія_машинного_перекладу
5. Корунець І. В. Теорія і практика перекладу (аспекти перекладу) / І. В. Корунець – Вінниця 2003. – 448 с.
6. Корпусна лінгвістика / [Широков В. А., Бугаков О. В., Грязнухіна Т. О. та ін.] – К.: Довіра, 2005. – 471 с.
7. Макоєд Н. О. Формування у майбутніх інженерів умінь перекладу фахових текстів із застосуванням комп'ютерних технологій : дис. ... канд. пед. наук : 13.00.04 / Макоєд Наталя Олексіївна. – Одеса, 2002. – 243 с.
8. Garcia, I.; Stevenson, V. Google Translator Toolkit. Free web-based translation memory for the masses. – Multilingual (September 2009). – P. 16–19.

References

1. Bibik S. P. Syuta G. M. Dilovi dokumenty` ta pravovi papery` / S. P. Bibik, G. M. Syuta – Xarkiv: Folio 2006. – 493 s.
2. Dems`ka-Kul`chy`cz`ka O. Osnovy` nacional`nogo korpusu ukrayins`koyi movy`. Monografiya. – Ky`yiv, 2005. – 218 s

3. Dzhordzh F. Osnovy ky`bernety`ky` / F. Dzhordzh. – M. : Rady`o y` svyaz`, 1984. – 272 s.
4. Istoriya mashy`nnogo perekladu – Vikipediya [Elektronny`j resurs] // Wikipedia. – 2012. – Rezhy`m dostupu : [http://uk.wikipedia.org/wiki/Istoriya\\_mashy`nnogo_perekladu](http://uk.wikipedia.org/wiki/Istoriya_mashy`nnogo_perekladu)
5. Korunecz` I. V. Teoriya i prakty`ka perekladu (aspekty` perekladu). / I. V. Korunecz` – Vinny`cya 2003. – 448 s.
6. Korpusna lingvisty`ka / [Shy`rokov V. A., Bugakov O. V., Gryaznuxina T. O. ta in.] – K.: Dovira, 2005. – 471 s.
7. Makoyed N. O. Formuvannya u majbutnix inzheneriv umin` perekladu faxovy`x tekstiv iz zastosuvannyam komp`yuterny`x tehnologij : dy`s. ... kand. ped. nauk : 13.00.04 / Makoyed Natalya Oleksiyivna. – Odesa, 2002. – 243 s.
8. Garcia, I.; Stevenson, V. «Google Translator Toolkit. Free web-based translation memory for the masses». Multilingual (September 2009). – P. 16–19.

ДАНІ ПРО АВТОРА

Башманівський Олексій Леонідович, кандидат педагогічних наук, доцент кафедри слов'янських та германських мов
Житомирський державний університет імені Івана Франка, 10003 Житомир, вул. Короленка, 31, кв. 2, тел. 09843717034, e-mail: bashmanivskiy@ukr.net

ДАННЫЕ ОБ АВТОРЕ

Башмановский Алексей Леонидович, кандидат педагогических наук, доцент кафедры словянских та германских языков
Житомирский государственный университет имени Ивана Франка, 10003 Житомир, ул. Короленко, 31, кв. 2, тел. 09843717034, e-mail: bashmanivskiy@ukr.net

DATA ON AUTHOR

Bashmanivskiy Oleksiy L., candidate of pedagogical sciences, associate professor of Department of the Slavic and the Germanic Languages
*Zhytomyr Ivan Franko State University
31, Korolenko St., App. 2, Zhytomyr, Ukraine, 10003
tel. 09843717034,
e-mail: bashmanivskiy@ukr.net*