

*Пушик Н. В.,**старший викладач кафедри перекладу та філології
Університету Короля Данила**Горда В. В.,**асистент кафедри перекладу та філології
Університету Короля Данила*

ТЕХНІКА МАШИННОГО ПЕРЕКЛАДУ

Анотація. Вступивши в цифрову епоху, яка характеризується здатністю вільно приймати та передавати великі обсяги інформації, мова в різних видах і формах стала найважливішим засобом спілкування. У процесі глобалізації із швидким розвитком Інтернету та інтеграцією світової економіки виникла нагальна потреба подолати мовні бар'єри. Щоб сформувавши необхідні знання про об'єкти чи явища, людина вдається до пошуку різного роду інформації та перекладів отриманих текстів, здійснюваних як людьми, так і технічними засобами. Переклад текстів людьми не завжди може задовільнити особисті та суспільні запити, тому використання машинного перекладу стало неминучою тенденцією. Стрімкий розвиток науки та техніки сприяв появі й широкому застосуванню машинного перекладу, під яким розуміють процес перетворення однієї природної мови на іншу за допомогою комп'ютера. Машинний переклад має тісний взаємозв'язок з обчислювальною лінгвістикою та розумінням природної мови. Ця стаття містить аналіз методів машинного перекладу, що використовуються для багатомовного перекладу щодо ефективності та простоти використання. Описано чотири основні типи запитів на машинний переклад. Обґрунтовано сильні та слабкі сторони різних методів машинного перекладу, а саме: перекладу, здійсненого на основі правил, статистичного, нейронного та гібридного перекладів. Проаналізовано внесок машинного перекладу для спрощення його виконання. Також наведені приклади перекладацьких систем. З'ясовано, що машинний переклад виконує функцію допоміжного інструмента в роботі перекладача, оптимізує процес перекладу, обробляє великі обсяги інформації, тому і результати машинного перекладу неодмінно підлягають редагуванню з метою запобігання виникнення помилок у кінцевому тексті. Наголошується, що з часом машинний переклад буде щоразу досконалішим та якіснішим, а людина без знання іншої мови матиме змогу в режимі онлайн перекладати певний текст чи розуміти свого співрозмовника, однак, на наше переконання, жодна машина не замінить людину із її унікальністю.

Ключові слова: машинний переклад, техніка перекладу, гібридний машинний переклад, статистичний переклад, корпусний машинний переклад, нейронний переклад, переклад, здійснений на основі правил.

Постановка проблеми. Жодна інша жива істота не має такої системи спілкування, яку можна порівняти з людською мовою. Мову можна розглядати як основний засіб вивчення культури та цивілізації. Те, як і в який спосіб ми формуємо та висловлюємо свої думки, ідеї, впливає на кінцевий результат нашої діяльності, на якість життя. Своя та інша культура пере-

даються і пізнаються завдяки мові. Люди демонструють вражаючу здатність комунікувати не тільки один з одним, а й з ЕОМ, що відбувається за допомогою системи звуків, жестів та символів. Таким чином вони полегшують собі збереження та пошук інформації, а також розширюють можливості машини, навчаючи її.

Допитливий людський характер і пристрасть до подорожей по всьому світу вимагають взаємодії різних культур між собою, засобом для досягнення якої є людська мова. Представники різних культур спілкуються різними мовами. Тому важливо, щоб люди перекладали та тлумачили певні тексти правильно, з метою кращої координації дій у повсякденному житті, ведення бізнесу і т.д. Із вдосконаленням технологій комп'ютерні системи полегшили складний процес перекладу та сприяють досягненню позитивних результатів за мінімально короткої проміжок часу, хоча ці системи не дають точних перекладених віршів, але достатньо і відповідної інформації, яка може бути використана фахівцями для розуміння природи інформації, що міститься в документі. Інструменти, такі як Babelfish та Google Translator, є прикладами таких систем. Для автоматизації процесу перекладу розроблено численні методики, які називаються машинним перекладом, що може бути визначено як завдання автоматичного перетворення однієї природної мови на іншу. Ці автоматизовані системи перекладу використовують найсучасніші технології з широкими словниками та набором лінгвістичних правил, які перекладають тексти з різних мов, не покладаючись на перекладачів-людей.

Аналіз останніх досліджень і публікацій. Нині спостерігається новий прояв інтересу до систем машинного перекладу у зв'язку з розвитком мережі Internet. Мільйони людей, які говорять різними мовами, опинилися в єдиному інформаційному просторі. Слід зазначити, що домінують є англійська мова і є користувачі, які нею не володіють, як, втім, є безліч вебсторінок, написаних не англійською. Заради полегшення перегляду сторінок у мережі Інтернет незнайомою користувачеві мовою з'явилися доповнення до браузерів, які здійснюють негайний переклад вибраних користувачем фрагментів потрібної вебсторінки. Варто лише виділити частину тексту «мишкою» і перенести її на спеціальну панель або натиснути покажчиком на спеціальну кнопку меню. Прикладом систем машинного перекладу є система Web Trans Site фірми Промт, створена на базі програми Stylus, яка підключається як до браузера Netscape Navigator, так і до браузера Microsoft Internet Explorer. «Promt 98» – професійна 32-розрядна система для автоматичного перекладу документів, що успадкувала одно-

часно лінгвістичні й інтерфейсні розробки систем машинного перекладу сімейства Stylus. Ця система складається з кількох модулів, пов'язаних між собою так званним інтегратором. До її складу входить модуль FileTranslator для пакетної обробки великої кількості документів. Для цього необхідно вказати системі файли, які потрібно перекласти і запустити у процес перекладу. Можна формувати чергу перекладу як безпосередньо в модулі File Translator, так і відправляти на переклад документи, відкриті в системі «Промт». Модуль WebView – браузер, який дає змогу одержати синхронний переклад HTML-сторінок при роботі в мережі Інтернет. Модуль QTrans використовується для перекладу невеликих неформатованих текстів із буфера обміну чи введених на клавіатурі. «Promt 98» дозволяє виконати переклад для трьох мовних пар у двох напрямках: англійська – російська, німецька – російська і французька – російська. Система підтримує формати: .txt, .doc, .rtf, .wri, .htm; має інтуїтивно зрозумілий інтерфейс. «Promt 98» вмонтована в Word і Excel, вона використовує системи перевірки правопису Lingvo Corrector, «Пропис», «Орфо» та ін., дає змогу користувачу вибрати варіант перекладу серед кількох запропонованих.

Система «Плай» – універсальний російсько-український перекладач. Забезпечує миттєвий переклад будь-яких редакторських виправлень у вихідному тексті, має інтелектуальну технологію введення нових слів і словосполучень у словники користувача, забезпечує повне збереження структури документа, створення списку зарезервованих слів, а також представляє словники синонімів українських і російських слів, російсько-український електронний словник і вмонтовані лінгвістичні модулі системи перевірки правопису і граматики «Рута».

Система «Pars» характеризується такими особливостями, як одночасна робота багатьох користувачів, проста процедура поповнення словників, що дає змогу налаштувати систему на потрібні тематики. Система надає користувачам такі можливості: задавати словники потрібної тематики та встановлювати їхню пріоритетність.

У майбутньому варто очікувати лише збільшення кількості серверів-перекладачів.

Є низка типових помилок, які виникають у процесі машинного перекладу. До граматичних проблем перекладу можна зарахувати неправильне розпізнавання типів зв'язків між членами речення, що своєю чергою порушує їх порядок у реченні, заміну одного члена речення іншим, також комп'ютерному перекладачу досить складно дається розпізнавання форми роду та відмінку.

До лексичних проблем, перш за все, належать неправильний принцип відбору словникових відповідників, повна або часткова неможливість перекладу, невірно перекладена термінологія.

Нині інтенсивно проводяться дослідження в галузі перекладу усної мови. Компанія IBM, яка вважається лідером у цій галузі, випустила програмний пакет ViaVoice 4.1, який дає змогу комп'ютеру сприймати до 140 слів на хвилину неперервного диктування. Попередні версії цієї програми передбачали лише дискретний спосіб введення мови. Однак розпізнавання мови не означає розуміння її змісту. Слід розрізняти розпізнавання машинної мови і перетворення її на текст, її використання у вигляді команд і справжнє розуміння змісту, як це робить

людина. Останнє вимагає знання комп'ютером усього обсягу стилістичних та семантичних конструкцій, правил використання слів та висловів. У процесі перекладу усного мовлення виникає низка складних ситуацій, адже варто звернути увагу й на те, що люди висловлюють свою думку, як їм заманеться, не звертаючи уваги на те, наскільки це грамотно і літературно. Іноді навіть те, що висловилось вголос, не відповідає тому, що малося на увазі, тому навчити машину розуміти людей – задача неймовірно складна.

Метою статті є вивчення методів машинного перекладу, які використовуються для багатомовного перекладу щодо ефективності та простоти використання, просторово-часової складності.

Виклад основного матеріалу. У сучасному глобальному світі людині доступні великі обсяги інформації, яка викладена різними мовами. З метою формування знань вона не тільки шукає й виокремлює інформацію, а й усвідомлює та аналізує її зміст.

Щоб пришвидшити та полегшити переклад інформації з однієї мови іншою, людина вдається до машинного перекладу, під яким розуміють переклад текстів (усних чи письмових) з однієї мови на іншу за допомогою спеціальних комп'ютерних програм. У процесі машинного перекладу програма аналізує алгоритми й на їхній основі відбувається переклад тексту. Варто зазначити, що значно точніше відтворюються технічні тексти, на відміну від літературних, оскільки в них не міститься «двозначності», також точніше відтворюються тексти споріднених мов, оскільки в них побудова речень схожа [1, с. 82].

Найчастіше використовують машини для пошуку чотирьох основних типів запитів.

Перший тип запити традиційний, це запит на якісний переклад, якого зазвичай очікують від перекладачів, де роботи представлені у вигляді друкованих текстів (найпоширенішим видом є публіцистичні твори).

Другим за порядком є запит на переклади з дещо нижчим рівнем якості, які призначені для користувачів, які хочуть з'ясувати суть змісту певного документа і зробити це якнайшвидше.

Третій тип запити – це переклад між учасниками особистого спілкування (телефонного чи письмового листування).

Четверта область застосування – переклад у багатомовних інформаційних системах, що передбачає пошук, вилучення інформації, доступ до бази даних тощо.

Перший тип запити демонструє використання машинного перекладу для тиражування, поширення текстів. Вперше такий переклад був здійснений у 1960-х роках. Машина створює переклад, який має переглядатися або «редагуватися» перекладачами-людьми, щоб досягти якості, яка потрібна. Цей переклад ще можна назвати «чорновим». Однак деякі системи машинного перекладу були розроблені для роботи з дуже вузьким діапазоном текстів та мовного стилю, і вони можуть вимагати незначної підготовки чи перегляду текстів або зовсім не вимагати їх. В останні роки використовувалися системи машинного перекладу, доповнені розвитком засобів перекладу (наприклад, бази даних термінології та пам'яті перекладів), інтегровані в авторські та видавничі процеси. Ці «перекладацькі робочі станції» є більш привабливими для перекладачів, тому що вони можуть контролювати комп'ютерні засоби, які вони можуть прийняти чи відхилити в разі потреби.

Другий тип запиту при машинному перекладі використовується для асиміляції самого процесу перекладу. Оскільки системи машинного перекладу не давали та все ще не можуть давати високих показників якісного перекладу, деякі користувачі виявили, що вони можуть витягти те, що їм потрібно, тобто дізнатися з невідредагованого перекладу. З появою дешевших ПК цей тип використання швидко та істотно зріс.

Щодо третього типу запиту на машинний переклад, що використовується для обміну інформацією, ситуація швидко змінюється. Попит на переклади електронних текстів в Інтернеті, таких як вебсторінки, електронні листи та ін., швидко зростає. У цьому контексті про переклад людиною не може бути й мови. Потреба полягає в негайному перекладі для передачі основного змісту повідомлень. Сучасні системи машинного перекладу відіграють «природну» роль, оскільки вони можуть функціонувати віртуально або фактично в режимі реального часу та режимі онлайн. Іншим контекстом для машинного перекладу в особистому обміні є спрямованість на різноманітні дослідження. До них зараховують розвиток систем перекладу розмовної мови, наприклад через телефон під час ділових переговорів.

Четвертий тип програм машинного перекладу – як компонент доступу до інформації системи – це інтеграція програмного забезпечення для перекладу в системи пошуку та пошук повних текстів документів із баз даних (як правило, електронні версії наукових статей), або для пошуку бібліографічної інформації для вилучення інформації з текстів, зокрема з газетних повідомлень і т.д. У цьому напрямі останнім часом фокусується низка проєктів в Європі, які мають на меті розширення доступу всіх членів Європейського Союзу до джерел даних та інформації незалежно від мови джерела.

Алгоритм діяльності в процесі машинного перекладу включає такі фази:

- проводиться морфологічний та синтаксичний аналіз тексту, для кожного речення добираються різні варіанти перекладу, при цьому враховується вірогідність використання;

- на наступному етапі декодер за допомогою моделювання тієї чи іншої мови оцінює, як часто це речення вживається, і вибирає його.

Загалом усі методи машинного перекладу поділяються на такі:

- переклад, здійснений на основі словника. Цей метод перекладу враховує розділи мовного словника, де перекладений текст створюється з використанням відповідного слова або еквівалентних слів у словнику. Це означає, що слова будуть тлумачитися так, як це робить словник, – слово за словом, здебільшого без особливого взаємозв'язку між ними. Пошук слова може виконуватися з аналізом або лематизацією (пов'язаного з формою та структурою). Машинний переклад на основі словника добре підходить для перекладу довгих списків фраз або простих товарів та послуг. Його також можна використати для прискорення ручного перекладу, якщо особа, яка здійснює його, знайома з обома діалектами й, відповідно, здатна виправити структуру мови та мовного використання. Оригінал машинного перекладу з кінця 1940-х рр. та до середини 1960-х рр. зосереджувався на машинному дешифруванні або електронних посиланнях на слова. Цей спосіб все ще корисний для перекладу фраз, проте певною мірою не речень. Більшість підходів до перекладу, які були створені пізніше, використовують двомовні лексикони з синтаксичним принципом [2];

- переклад, здійснений на основі правил. Цей механізм машинного перекладу бере до уваги етимологічні дані про діалекти та структуру речень, що охоплюють первинні семантичні, морфологічні та синтаксичні закономірності кожного діалекту окремо. Маючи речення даних (на якомусь діалекті-джерелі), система перекладу, здійсненого на основі правил, виробляє їх, щоб дати речення (на якомусь об'єктивному діалекті) на передумові морфологічного, синтаксичного та семантичного вивчення як джерела, так і об'єктивних діалектів. Машинний переклад, що базується на правилах, управляє морфологічними, синтаксичними та семантичними даними про вихідний та цільовий діалект. Ці дані використовуються для виготовлення лінгвістичних принципів. Принципова методологія систем перекладу, здійсненого на основі правил, полягає у зв'язку структури цього інформаційного речення зі структурою запитуваного речення, що принципово зберігає їх нове значення;

- машинний переклад, здійснений на основі знань. Цей тип системи об'єднаний навколо лексики «концепція», що представляє домен. Наприклад, система KANT є прикладом системи машинного перекладу на основі знань для багатомовного перекладу. Вона була заснована в 1989 р. для дослідження та розробки масштабних практичних систем перекладу технічної документації. З метою досягнення дуже високої точності перекладу KANT використовує контрольований словниковий запас та граматику для кожної мови-джерела та чіткі, але цілеспрямовані семантичні моделі для кожної технічної галузі [3];

- корпусний машинний переклад. Із 1989 р. методологія машинного перекладу, заснована Корпусом, стала однією з широко досліджуваних галузей машинного перекладу та керувалася різними методологіями, враховуючи високий рівень точності перекладу на три адреси. Якщо потрібна лише одна адреса, потрібно відцентрувати весь текст адреси. Для двох адрес використовують дві відцентровані вкладки тощо;

- статистичний підхід. Статистичний підхід до машинного перекладу був представлений у 1949 р. Уорреном Вівером. У цій методології використовуються статистичні методи для створення перекладеної форми з використанням двомовних корпусів. Статистичний машинний переклад використовує фактичні моделі перекладу, параметри яких впливають із вивчення одномовних та двомовних корпусів. Побудова статистичних моделей перекладу є швидким процесом, однак новачія сильно залежить від наявних багатомовних корпусів;

- статистичний підхід до моделі машинного перекладу на основі слова. Основною одиницею цієї методології є слово. Алгоритми, які ідентифікуються з розташуванням слів, зобов'язані досягати гранично точних перекладів речень. Складні слова, вирази та омоніми сприяють багатогранній якості базового перекладу на основі слів;

- статистичний підхід до моделювання на основі фраз. Вирішальною одиницею цієї моделі є фраза або групування слів. У разі перекладу на основі фраз мета полягає у зменшенні обмежень перекладу на основі слів шляхом перекладу цілих послідовностей слів, де довжина може відрізнятися. Послідовності слів називаються фразами, але, як правило, це не лінгвістичні фрази, а фрази, знайдені за допомогою статистичних методів із двомовних текстових корпусів [4; 5];

- машинний переклад на основі прикладу. Переклад на основі прикладу базується на перегляді або виявленні випадків, що дуже нагадують діалектну пару. Ця ідея «перекладу за ана-

логією» була запропонована Макоото Ногао в 1981 р. Подаються речення на діалекті-джерелі, з якого здійснюється переклад, та порівнюються переклади кожного речення на об'єктивному діалекті з картографічним зображенням. Ці зразки використовуються для розшифровки порівнянних типів речень вихідного діалекту з цільовим діалектом. Принципова причина полягає в тому, що, якщо раніше перекладене речення трапляється ще раз, той самий переклад може бути ще раз правильним [6];

– нейронний машинний переклад. Це новий підхід до машинного перекладу, в якому машини «навчаються» перекладати через одну велику нейронну мережу, яка складається з великої кількості обробних пристроїв за зразком мозку;

– гібридний машинний переклад. Являє собою інтеграцію різних підходів машинного перекладу з можливих варіантів, внаслідок якого користувач отримує якісний переклад, що забезпечує високу швидкість перекладу.

Висновки. Нині статистичний підхід до машинного перекладу вважається найбільш поширеним підходом машинного перекладу, а гібридний – найбільш якісним. Найкращі з усіх методологій можна об'єднати для збільшення результатів. Ідеальним було б отримати граматичну правильність перекладу, здійсненого на основі правил, лексичний відбір статистичного підходу до машинного перекладу та толерантність до несподіваних структур, оскільки така система машинного перекладу, що базується на поєднанні підходів, є цікавим викликом і допоможе розв'язати проблеми в наявних методах машинного перекладу. Головною метою цього гібридного підходу є використання переваг як лінгвістичних правил, так і статистичних методів. Гібридний машинний переклад мотивується тим, що методи гібридизації стали популярними, оскільки вони намагаються поєднати найкращі характеристики високосконалих підходів на основі чистого правила або основи корпусу.

Варто наголосити, що увесь текст, який переклала машина, варто обов'язково піддавати редагуванню. Можна запевнити, що в майбутньому людина без знання іншої мови матиме змогу в режимі онлайн перекладати певний текст чи розуміти свого співрозмовника, спостерігаючи при цьому його міміку.

З часом машинний переклад буде досконалішим та якіснішим, але, на наше переконання, жодна машина не замінить людину з її унікальністю.

Література:

1. Хвощ П.М. Прикладна лінгвістика сьогодні. Київ : Світ, 2004. 210 с.
2. Yang V S-C, Electronic dictionaries in machine translation Encyclopedia of Library and Information Science, 48, 1991. pp. 74–92.
3. Nyberg, E., Mitamura, T., & Carbonell, J.G. (1997). The KANT machine translation system: from R&D to initial deployment. URL: https://www.cs.cmu.edu/~jgc/publication/The_KANT_Machine_Translation_Sys_From_R&D_LISA_1997.pdf (Accessed on 29 May, 2021).

4. Koehn P, Och F J and Marcu D, Statistical phrase based Machine translation (2003). URL: <https://www.aclweb.org/anthology/N03-1017.pdf> (Accessed on 20 May, 2021).
5. Zens R., Och F.J. and Ney H., Phrase based machine Translation. Lecture Notes in Computer Science; Springer, 2002. pp. 35–56.
6. Carbonell J.G. and Brown R. D., Example based machine translation (2004). URL: https://www.academia.edu/39920933/Approaches_to_machine_translation.
7. Бацевич Ф. Основи комунікативної лінгвістики. Київ : Академія, 2004. С. 304–342.
8. Кузьменко Д.М. Комп'ютерна лінгвістика і Web 2.0. Київ : Наукова думка, 2009. 250 с.
9. Мацак Ж.Г., Скоробогатова Т.І. Машинний переклад та його специфіка. URL: http://www.rusnauka.com/31_ONBG_2009/Philologia/54653.doc.htm.
10. Leonardi V. Equivalence in Translation: Between Myth and Reality. URL: <http://www.translationjournal.net/journal/14equiv.htm>.

Pushyk N., Horda V. Machine translation technique

Summary. Entering the digital age, language as a carrier of information has become the most important means of human communication, which is sometimes seen as a barrier to communication between people from different countries. The problem of transforming a language into another one became a problem that still worries humanity. Machine translation is the process of transforming one natural language into another by using a computer. It is a branch of natural language processing, and it is closely related to computational linguistics and natural language understanding. With the rapid development of the Internet and the integration of the world economy, how to overcome the language barrier is becoming a common problem in the international community. Human translation cannot meet the demand of society, so the use of machine translation to help people obtain information has become an inevitable trend. This article contains an analysis of machine translation methods used for multilingual translation in terms of efficiency and ease of use. Four main types of machine translation requests are described. The strengths and weaknesses of different machine translation methods are substantiated, namely: rules-based translation, statistical, neural and hybrid translations. The contribution of machine translation to simplify translation is analyzed. Examples of translation systems are also given. It was found that machine translation serves as an auxiliary tool in the work of the translator, optimizes the translation process, processes large amounts of information, but the results of machine translation must be edited to prevent errors in the final text of the translation. It is observed that over time machine translation will be more perfect and better, a person without knowledge of another language will be able to translate a certain text online or understand his/her interlocutor, but, in our opinion, machine cannot replace a person with his/her uniqueness.

Key words: machine translation, translation techniques, rule-based machine translation, hybrid machine translation, statistic translation, corpus machine translation, neural translation.