

## СИСТЕМИ МАШИННОГО ПЕРЕКЛАДУ: ОГЛЯДОВИЙ АНАЛІЗ

Гречуха Л.О., Кузєбна В.В.

Черкаський державний технологічний університет

Розмежовується автоматизований та машинний види перекладу. Наводяться існуючі класифікації систем машинного перекладу. Аналізуються системи машинного перекладу відповідно методів лінгвістичних досліджень, міри розробленості, міри автоматизованості та з точки зору користувача системи. Пропонується коротка характеристика кожного типу машинного перекладу.

**Ключові слова:** машинний переклад, автоматизований переклад, системи машинного перекладу, перекладацька діяльність, мовний бар'єр.

**Постановка проблеми.** Ми живемо в світі інформаційних технологій, які міцно ввійшли в наше життя. З кожним роком збільшується число користувачів Інтернету, що претендує на роль єдиного інформаційного простору планетарного масштабу. Єдиною перешкодою, яка незримо присутня у всій Мережі, є мовний бар'єр. Ця проблема до тепер так і не знайшла свого кардинального рішення. Спроби впровадження універсальної мови виявилися неефективними. І однією з можливостей подолання мовного бар'єру у спілкуванні залишається переклад. Втім, сьогодні, де інформація змінюється 24 години на добу та широко застосовуються електронні засоби зв'язку, диктує свої умови. У такій ситуації класичний підхід до здійснення перекладу не завжди виправдовує себе, оскільки вимагає великих капіталовкладень і затрат часу. Саме тому доцільно звертатися до систем машинного перекладу як засобу підвищення продуктивності перекладацької діяльності. Розвиток таких систем дозволить оперативно здійснювати переклад інформації та обробляти великі масиви документів у надзвичайно короткі терміни, тобто задовольняти основну вимогу сьогодні: оперативну обробку величезних масивів інформації при мінімальних затратах.

**Аналіз останніх досліджень та публікацій.** Ще від появи ідеї машинного перекладу в першій половині 19 ст., першої комп'ютерної програми для машинного перекладу в 1954 році, над питаннями машинного перекладу працювали та продовжують працювати багато, як іноземних так і вітчизняних дослідників. Серед них: Алекс Вэйбел, Філіп Коін, Франз Джосеф Оч, Данієл Марцу, Б. Лавуа, М. Уайт, Т. Корельський, А.А. Ляпунов, Р.Г. Піотровській, Д.Ю. Панов, В.В. Воронович, І.А. Большаков, Р.Г. Котов, Р.С. Ісламов, А.Г. Фомін, К.К. Колін, А.А. Хорошилов, О.С. Кулагіна, П.А. Соболева, Г.Г. Белоногов, І.І. Русанова, А.Л. Міщенко, Смірнова П.М., Палій С.В., Анісімов В., Блехман М.С., Войнов В.К., Марчук Ю.Н., Шаляпіна З.М. та інші.

Виділення раніше невирішених частин проблеми. Сьогодні існує велика кількість різноманітних систем машинного перекладу. Серед них є комерційні та безкоштовні, локальні та «хмарні», з різноманітними інтерфейсами, обмеженнями та мовами, що підтримуються. Вибір даної теми дослідження обумовлений в першу чергу необхідністю систематизації існуючих класифікацій систем машинного перекладу, що раніше не ставало метою окремих досліджень.

**Мета статті.** Мета дослідження полягає в систематизації підходів до виділення типів машинного перекладу, запропонованих різними дослідниками.

**Виклад основного матеріалу.** При перекладі текстів з однієї природної мови на іншу за участі комп'ютерної програми розрізняють автоматизований переклад та автоматичний, або машинний переклад. Машинний переклад (МП) – це переклад текстів (письмових та усних) з однієї природної мови на іншу за допомогою комп'ютера. Автоматизований переклад (англ. Computer – Aided Translation) – переклад текстів на комп'ютері з використанням комп'ютерних технологій. Від машинного перекладу останній відрізняється тим, що увесь процес перекладу здійснюється людиною, комп'ютер лише допомагає їй зробити готовий текст або за менший час, або кращої якості. На сьогодні існує кілька класифікацій систем машинного перекладу, до огляду яких ми й звертаємося.

Отже, сучасні системи машинного перекладу (СМП), згідно класифікації, запропонованої А.Д. Бакуловим, Н.Н. Леонтьєвою, яка сфокусована на методах лінгвістичних досліджень, підрозділяються на системи прямого типу, трансферні системи перекладу та системи машинного перекладу семантичного типу [1]. Відтак, системи прямого перекладу – найчисельніші, оскільки почали створюватися ще в 1950-60-і роки для фіксованих пар мов. У цих системах словник і синтаксис вхідної мови аналізуються лише в тій мірі, в якій це необхідно для ідентифікації правильних виразів вихідної мови та порядку розташування слів. Системи перекладу з використанням мови-посередника (семантичні), служать для відображення «сенсу» вхідного тексту, який перетворюється в семантичні і синтаксичні уявлення, загальні для декількох вихідних мов. Цей метод застосовується зазвичай при необхідності перекладу вихідного тексту на кілька мов. Системи перекладу з трансфером більш складні, ніж попередні типи, оскільки мови-посередники застосовуються двічі – перший раз при перекладі з вхідної мови, другий – при перекладі на вихідну мову. У цьому випадку стає необхідним додатковий етап перекладу – з мови-посередника вхідного мови на мову-посередник вихідного мови. У такий спосіб досягається більш глибокий лінгвістичний аналіз і синтез. Переважна більшість сучасних комерційних СМП відноситься до систем трансферного типу. Системи прямого типу нині практично не використовуються, ставши

надбанням історії розвитку СМП. Системи, ґрунтовані на використанні баз знань і мови-посередника, на сьогодні є системами майбутнього та знаходяться на етапі лабораторних досліджень і являються, по суті, експериментальними.

З точки зору користувача системи машинного перекладу можуть поділятися на три основних типи: інформативні, професійні та персональні. Інформативні системи призначені для допомоги тим, кому потрібен доступ до інформації на іноземній мові і хто готовий користуватися «грубим», але досить зрозумілим перекладом. Такі системи, як правило, мають словники великого обсягу, але не спираються на новітні досягнення в лінгвістиці і програмуванні. Професійні дозволяють створювати нотатки для перекладу професійним перекладачам. Такі системи тепер використовуються все рідше, – як правило, при великому обсязі поточної перекладацької роботи, що виконується одночасно багатьма фахівцями в одній предметній області. Найчастіше в цих ситуаціях перекладачів постачають автоматичними словниками, тезаурусами з інтерактивним доступом або системами, які отримали назву «пам'ять перекладача». Персональні системи машинного перекладу зазвичай «працюють в діалозі» з користувачем і можуть давати задовільний переклад (якість якого все ж залежить від того, наскільки автор володіє вихідним мовою).

Слід також враховувати класифікацію СМП щодо міри їх розробленості. Відповідно до поділу, запропонованого Н.Н. Леонтьєвою, З.М. Шаляпіною і Е.І. Корольовим, ми підрозділяємо СМП на наступні типи [4]:

- дослідницькі прототипи (викладений принцип дії системи безвідносно до умов її функціонування);
- експериментальні системи (системи, що пройшли перевірку на експериментальних тестових масивах);
- промислові системи, що працюють з реальними текстами, мають певні завадостійкі, якісні і швидкісні характеристики;
- комерційні системи – системи, що мають усі властивості промислових систем, призначені для комерційного продажу і такі, що є закінченим самодостатнім продуктом, що має властивості товару.

За мірою автоматизованості СМП можуть бути автоматичними і автоматизованими (у яких ряд функцій перекладу лишається за людиною). Детальніша класифікація за рівнем автоматизованості перекладу була запропонована у 1990 р. Ларрі Чайлдсом [2]. Відповідно до цієї класифікації, виділяють наступні категорії МП: *FAMT* (Fully-automated machine translation) – повністю автоматизований машинний переклад; *HAMT* (Human-assisted machine translation) – машинний переклад за участю людини; *MANH* (Machine-assisted human translation) – переклад, здійснюваний людиною з використанням комп'ютера.

Повністю автоматизований переклад (*FAMT*) полягає в тому, що в систему автоматизованого перекладу вводиться текст однією мовою. Цей текст опрацьовується і виводиться текст іншою мовою. Реалізація такого виду автоматичного перекладу зустрічається з великою кількістю пере-

шкод, найпоширенішими серед яких є дослівний переклад ідіоматичних виразів та неправильне визначення частин мови. Часто стверджують, що успіх повністю автоматизованого перекладу потребує, в першу чергу, вирішення проблеми розуміння природної мови.

Машинний переклад за участю людини (*HAMT*) можливий тільки в умовах штучно обмеженої, як за словниковим запасом, так і за граматиною, мови. Не тільки комп'ютер перекладає весь документ, але й людина, яка усуває при цьому неясні, двозначні, складні конструкції (так звана «контрольована мова»). Це відбувається завдяки тому, що людина заздалегідь вкорочує довгі речення або коли людина повинна сама обрати значення слова.

При *MANH* увесь процес перекладу здійснює людина, а комп'ютер автоматично перевіряє термінологію, здійснює пошук необхідних слів у словнику і порівнює отримане з іншими перекладами. Такі програми-перекладачі називають *TM*-програмами (від *translation memory* – пам'ять перекладу). Цю категорію програм застосовують професійні перекладачі. Основу *TM*-програм складають спеціалізовані словники, які відповідають тематиці тексту, що перекладається. При перекладі використовуються конструкції та значення слів і стійких словосполучень, обрані професійним перекладачем і занесені в словникові системи, а отриманий текст піддається редагуванню фахівцем-перекладачем. Словники і вже перекладені фрагменти текстів, що запам'ятовуються в *TM*-системі, можуть бути повторно використані у великих колективних проєктах.

Залежно від підходу, на основі яких створені системи машинного перекладу, вони також поділяються на [3]: системи, які працюють завдяки використанню правил (*rule-based*), статистичні (*statistic*) та гібридні (*hybrid*).

Машинний переклад на основі правил (*Rule-based machine translation*) характеризується використанням і створенням лінгвістичних правил вручну. Ефективність систем машинного перекладу на основі правил визначається якістю двомовних словників та точністю заданих правил, а їх створення потребує довготривалої роботи [3].

Перші системи машинного перекладу створювалися для конкретних пар мов і ґрунтувалися на складних процесах моделювання мови, основу яких становили методи аналізу, трансферу, синтезу й інтерлінгви. Системи машинного перекладу першого покоління працювали за методом прямої заміни слів мови оригіналу словами мови перекладу. Системи другого покоління аналізували структури мови оригіналу, а потім на основі трансферу синтезували їх в еквівалентні структури мови оригіналу. Третім поколінням були системи машинного перекладу на основі формальної мови-посередника – інтерлінгви. Ця концепція передбачала перетворення слів на мову-посередника, яка є універсальною мовою, створеною для системи, незалежною від залучених у процес перекладу мов. Цей підхід використовує два методи: аналізу і синтезу.

Системи машинного перекладу на основі правил ґрунтуються на різних рівнях лінгвістичного опрацювання мовної пари [3]:

1. Морфологічному: лематизація лексичних одиниць, пошук лексичних одиниць у словнику, аналіз морфем, розпізнавання контекстного граматичного класу лексичних одиниць, відмінків, флексій тощо.

2. Синтаксичному: розпізнавання типів синтаксичних структур, реляційних зв'язків між окремими елементами синтаксичної структури тощо.

3. Семантичному: виокремлення лексичного значення багатозначних лексичних одиниць та афіксів, визначення їхньої семантичної функції, синтез їхньої синтаксичної однозначності на основі семантичного аналізу.

Системи машинного перекладу на основі правил не потребують доступу до баз паралельних текстів, їх можна налаштовувати, що поліпшує якість перекладу спеціалізованих текстів [6].

Системи на основі правил можуть мати справу з багатьма мовними явищами і зручні в супроводі. Проте винятки в граматиці додають певних труднощів, що потребує розроблення нових алгоритмів і покращення раніше створених [8].

Статистичний машинний переклад (Statistical machine translation) – різновид машинного перекладу, де переклад генерується на основі статистичних моделей, параметри яких є похідними від аналізу двомовних корпусів тексту. Статистичний машинний переклад має властивість «навчання». Чим більше в розпорядженні є мовних пар і чим точніше вони відповідають одна одній, тим кращим буде результат статистичного машинного перекладу. Під поняттям «статистичного машинного перекладу» мається на увазі загальний підхід до вирішення проблеми перекладу, який заснований на пошуку найімовірнішого перекладу речення з використанням даних, отриманих з двомовної сукупності текстів. Як приклад двомовної сукупності текстів можна назвати парламентські звіти, які являють собою протоколи дебатів в парламенті. Статистичний машинний переклад протистав-

ляють системам машинного перекладу, заснованим на правилах.

Ідея статистичного машинного перекладу випливає з теорії інформації. Для систем статистичного перекладу характерне використання масивів текстів, представлених одночасно двома мовними версіями (так звані паралельні корпуси). Чим більший об'єм паралельного корпусу, а також чим якісніший переклад текстів, що містяться в ньому, тим краще перекладає статистична система. В якості теоретичної основи технології статистичного машинного перекладу використовується модель, що базується на теоремі Байєса. Дана модель надає можливості покращити переклад, використовуючи найбільш частотні вживання слів різними мовами, враховуючи потім відповідні частоти при перекладі документа.

Принцип роботи системи статистичного машинного перекладу полягає в наступному: перед перекладом програма аналізує великий корпус двомовних текстів. Цей процес забезпечує те, щоб слова і граматичні форми знаходились разом на обох мовах, залежно від частоти використання та змістовної близькості. Таким чином, генерується словник та граматичні правила і на цій основі перекладаються тексти [7]. Метод статистичного машинного перекладу широко розповсюджений, тому що він не ставить за передумову ніяких знань відповідної мови.

Гібридні системи машинного перекладу створюються на базі наявних систем машинного перекладу на основі правил з додаванням до них статистичних методів. Таке додавання дещо покращує якість перекладів, особливо при недостатньому обсязі вхідних даних, що використовуються при побудові індексу машинного перекладача.

**Висновки і пропозиції.** Таким чином, кожен перекладач професійний чи ні, може вибрати ту систему машинного перекладу, яка буде відповідати його особистим потребам та вподобанням.

## Список літератури:

1. Бакулов А.Д., Леонтьева Н.Н. Теоретические аспекты машинного перевода. В кн.: Искусственный интеллект. В 3 кн. Кн.1. Системы общения и экспертные системы: Справочник / Под ред. Э.В. Попова. – М.: Радио и связь, 1990. – С. 201–216.
2. Мацак Ж.І., Скоробогатого Т.І. Машинний переклад та його специфіка – Режим доступу: [http://www.rusnauka.com/31\\_ONBG\\_2009/Philologia/54653.doc.htm](http://www.rusnauka.com/31_ONBG_2009/Philologia/54653.doc.htm)
3. Міщенко А.Л. Машинний переклад у контексті сучасного науково-технічного перекладу / А.Л. Міщенко // Вісник ХНУ ім. В.Н. Каразіна. – Серія «Романо-германська філологія. Методика викладання іноземних мов». – № 1051, 2013 – С. 172–180.
4. Палій С.В. Метод інкорпорування системи машинного перекладу в інформаційно-організаційне середовище підготовки іноземців. – Режим доступу: <http://urss.knuba.edu.ua/files/zbirnyk-15/143-151.pdf>
5. Пістунов І.М., Амеліна С.М. Технології комп'ютерного перекладу: Навч. посібник. – Дніпропетровський державний аграрний університет, 2008. – 107 с.
6. Проблемы машинного перевода // Компьютера. – 2002. – № 21. – С. 26–37.
7. Ревзин И.И., Розенцвейг В.Ю. Основы общего и машинного перевода / И.И. Ревзин, В.Ю. Розенцвейг. – М.: Высшая школа, 1964. – 243 с.
8. Смірнова Т.В. Переваги та недоліки застосування машинного перекладу. – Режим доступу: <http://readera.org/article/perevahye-ta-nedoliky-zastosuvannja-mashyennoho-perekladu-10182877.html>

**Гречуха Л.А., Кузедна В.В.**

Черкасский государственный технологический университет

## **СИСТЕМЫ МАШИННОГО ПЕРЕВОДА: ОБЗОРНЫЙ АНАЛИЗ**

### **Аннотация**

Разграничиваются автоматизированный и машинный типы перевода. Наводятся существующие классификации систем машинного перевода. Анализируются системы машинного перевода относительно методов лингвистического исследования, меры разработанности и меры автоматизации, с точки зрения пользователя системы. Предлагается короткая характеристика каждого вида машинного перевода.

**Ключевые слова:** машинный перевод, автоматизированный перевод, системы машинного перевода, переводческая деятельность, языковой барьер.

**Hrechukha L.O., Kuzebna V.V.**

Cherkassy State Technological University

## **SYSTEMS OF MACHINE TRANSLATION**

### **Summary**

Computer-aided and machine translations are differentiated. Existed kinds of machine translation classifications are taken into consideration. Systems of machine translation based on methods of linguistic research, development and automated levels, and on the point of view of a user. Short characteristics of all machine translation types are given.

**Keywords:** machine translation, computer-aided translation, systems of machine translation, translation process, language barrier.