

**Галина ВЕСЕЛОВСЬКА,**

*orcid.org/0000-0003-2896-0460*

*кандидат технічних наук,*

*доцент кафедри інформаційних технологій*

*Херсонського національного технічного університету*

*(Херсон, Україна) galina.veselovskaya@gmail.com*

**Світлана РАДЕЦЬКА,**

*orcid.org/0000-0001-7308-8179*

*кандидат педагогічних наук,*

*доцент кафедри теорії та практики галузевого перекладу*

*Херсонського національного технічного університету*

*(Херсон, Україна) rad.svete1976@gmail.com*

## МАШИННИЙ ПЕРЕКЛАД: ЙОГО ТИПОЛОГІЯ, ПЕРЕВАГИ ТА НЕДОЛІКИ

У статті розглянуто можливі класифікації машинного перекладу, виокремлено найбільш чітку типологію, а також визначено, які саме переваги та недоліки в кожного типу машинного перекладу. Міжкультурну комунікацію складно уявити без використання перекладу, але придбання компетенції перекладача вимагає багато часу і зусиль. Тому важко переоцінити актуальність вивчення і вирішення проблем, пов'язаних із машинним перекладом, і важливість його практичного застосування в подоланні мовного бар'єру.

Терміном «машинний переклад» (скорочено МП, англійською мовою «Machine Translation», або «MT») називають дію, коли одна природна мова перекладається на іншу з використанням для цього спеціального програмного забезпечення. Під час дослідження було виявлено, що існує кілька класифікацій машинного перекладу (за кількістю мов, за спрямованістю перекладу, за роллю, яку відіграє людина в процесі МП). Проте ми розглянули найбільш вживану типологізацію – поділ на дві основні групи: машинний переклад, заснований на правилах, та статистичний. Окремо виділені гібридні системи, які покликані поєднувати в собі найбільш ефективні риси систем, заснованих на правилах, і статистичні системи. У дослідженні описано ці чотири типи машинного перекладу, їх особливості, причини та випадки вживання. До того ж вказано, які саме програми належать до того чи іншого типу машинного перекладу. Також у статті вказано на переваги та недоліки кожного з чотирьох типів машинного перекладу.

На основі дослідженого було висновувало, що нині дедалі більшої популярності набувають гібридні підходи, покликані з'єднати в собі плюси класичних і статистичних підходів. Нині системи МП не придатні для роботи з текстами, що містять велику кількість складносурядних і складнопідрядних речень і якісно працюють в основному на рівні словосполучення.

**Ключові слова:** машинний переклад, гібридна система, статистична система, пам'ять перекладів, система заснована на правилах.

**Halyna VESELOVSKA,**

*orcid.org/0000-0003-2896-0460*

*Candidate of Technical Sciences,*

*Associate Professor at the Department of Information Technologies*

*Kherson National Technical University*

*(Kherson, Ukraine) galina.veselovskaya@gmail.com*

**Svitlana RADETSKA,**

*orcid.org/0000-0001-7308-8179*

*Candidate of Philological Sciences,*

*Associate Professor at the Department of Theory and Practice of Translation*

*Kherson National Technical University*

*(Kherson, Ukraine) rad.svete1976@gmail.com*

## MACHINE TRANSLATION: ITS TYPOLOGY, ADVANTAGES AND DISADVANTAGES

*The article identifies the possible classifications of machine translation, highlights the clearest typology, and identifies the advantages and disadvantages of each type of machine translation. Intercultural communication is difficult to imagine without the use of translation, but acquiring the competence of a translator requires a lot of time and effort. Therefore, it is difficult to overestimate the relevance of studying and solving problems related to machine translation, and the importance of its practical application in overcoming the language barrier.*

*The term "Machine Translation" (abbreviated MT, in Ukrainian "Машинний переклад" or "МП") refers to the action when one natural language is translated into another one using special software. During the research it was found that there are several classifications of machine translation (by the number of languages, by the direction of translation, by the role that a person plays in the process of MT). However, we have considered the most commonly used typology – the division into two main groups: machine translation based on rules and statistical. Hybrid systems are singled out, which are designed to combine the most effective features of rule-based systems and statistical systems. The study describes these four types of machine translation, their features, causes and uses. In addition, it is specified which programs belong to a particular type of machine translation. The article also points out the advantages and disadvantages of each of the four types of machine translation.*

*Based on the study, it was concluded that currently becoming increasingly popular hybrid approaches designed to combine the advantages of classical and statistical approaches. At the moment, MT systems are not suitable for working with texts that contain a large number of complex and complex sentences and work well mainly at the phrase level.*

**Key words:** machine translation, hybrid system, statistical system, translation memory, rule-based system.

**Постановка проблеми.** У час інформаційного вибуху і глобалізації всіх видів комерційної діяльності світ гостро потребує компетентних перекладачів. Зіткнувшись із величезними обсягами документів, складеними чи перекладеними на дедалі більшу кількість різних мов, люди приходять до висновку про необхідність використання машинних ресурсів для перекладу. Усі види машинного перекладу неминуче стають глобальною індустрією в цій сфері.

Проте не всі перекладачі обізнані у використанні цього виду перекладу, класифікації та типології системного машинного перекладу, що може призвести до неточностей та затримок у роботі. Саме тому питання типологізації машинного перекладу є нині дуже актуальним.

**Аналіз досліджень.** Теоретичною базою цього дослідження послужили роботи таких дослідників із питань машинного перекладу: Л. Н. Беляєвої, І. А. Борисової, А. Молчанова, Н. К. Рябцевої, М. Коста-юса, М. Хеврне та ін.

**Метою статті** – розгляд можливих класифікацій машинного перекладу, виокремлення найбільш чіткої типології, а також визначення того, які саме переваги та недоліки є в кожного типу машинного перекладу.

**Виклад основного матеріалу.** Машинний переклад (МП) – це галузь наукових досліджень, експериментальних розробок і систем машинного перекладу (СМП), які вже функціонують, в яких основна частина процесу перекладу з однієї природної мови іншою виконується комп'ютером.

Класифікацію систем машинного перекладу можна зробити за різними критеріями (Беляєва, 2007). Наприклад, можна виділити системи:

а) за кількістю мов (бінарні здійснюють переклад в одній парі мов, а багатомовні працюють із кількома мовами);

б) за спрямованістю перекладу (в одному і багатьох напрямках, якщо мова перекладу і мова оригіналу можуть мінятися місцями залежно від вимог користувача);

Залежно від того, яку роль відіграє людина в процесі МП, іншими словами, за ступенем автоматизації, зазвичай виділяють три типи СМП:

- 1) повністю автоматичні системи МП;
- 2) МП-системи, машинний переклад за участю людини;
- 3) ТМ-системи, переклад здійснюється людиною з використанням комп'ютера.

Повністю автоматичні системи машинного перекладу є скоріше нездійсненою мрією, ніж реальною ідеєю. Усі системи машинного перекладу працюють за участю людини в тій чи іншій мірі. Щоб комп'ютер міг перекласти текст, йому потрібна допомога предредактора, який тим чи іншим чином попередньо обробляє текст перекладу, інтерредактора, який бере участь у процесі перекладу, та постредактора, який виправляє помилки і недоліки в перекладеному машиною тексті (Рябцева, 1986). ТМ-системи іноді називають ще «пам'яттю перекладів». Вони є скоріше просто зручним інструментом, ніж елементом автоматизації.

Інший варіант класифікації систем МП, що прийшов з області корпусної лінгвістики – це поділ на підходи, в яких використовуються паралельні корпуси і, відповідно ті, в яких вони не використовуються. Системи, що використовують корпуси, далі діляться залежно від основної

стратегії перекладу – на системи, засновані на прикладах (ЕВМТ), і статистичні системи (SMT) (Hearne, 2011).

Найпростіший і найпоширеніший варіант класифікації – це поділ на два основних типи систем МП (Молчанов, 2013):

а) засновані на правилах (rule-based machine translation, RBMT);

б) статистичні.

Окремо розташовані гібридні системи, які покликані поєднувати в собі найбільш ефективні риси систем, заснованих на правилах, і статистичних систем.

Перший тип, який ми розглянемо – це пам'ять перекладів (Translation memory) (засіб автоматизованого перекладу, який не є системою машинного перекладу, проте користується великою популярністю). Друга назва – накопичувач перекладів. Його слід згадати, тому що він ефективно використовується в сукупності з програмами-перекладачами (PROMT, Lingvo).

Пам'ять перекладів – це, по суті, набір вже перекладених «одиниць перекладу» (translation unit). «Одиниця перекладу» частіше називається сегментом тексту, і кожен такий сегмент є найчастіше одним реченням, або рідше – цілим абзацом або частиною складносурядного речення.

Іншими словами накопичувач перекладів – це база даних, в якій зберігаються пари «сегмент – переклад». Якщо сегмент вихідного тексту повністю збігається з сегментом у базі накопичувача перекладів, він підставляється в переклад. Однак за певних розбіжностей між цими двома сегментами потрібне подальше втручання перекладача для внесення змін.

Ефективність використання пам'яті перекладу в сукупності з програмами-перекладачами пояснюється тим, що в початковому тексті можуть бути присутні сегменти, яких немає в базі накопичувача.

Переваги такого типу МП в тому, що пам'ять перекладу забезпечує однаковість перекладу термінів в однакових фрагментах. Загалом підвищується ефективність перекладу текстів великих обсягів. Підвищується якість перекладу документів – різних договорів і специфікацій. Системи пам'яті перекладів необхідні, коли кілька перекладачів працюють над одним проектом, збільшується їх продуктивність (Молчанов, 2013).

Однак недоліків у процесі використання систем пам'яті перекладів більше, ніж переваг. Види текстів, з якими використання систем накопичувачів перекладів буде ефективним, обмежені. Кожна програма вимагає адаптації, звикання, навчання,

це стає недоліком, коли робота йде з різними програмами пам'яті перекладів. У самому тексті одна помилка може поширитися на весь проект, а також зв'язок між реченням, запропонованим програмою, і наступним реченням (або навіть текстом), може бути зовсім відсутнім. Для ефективної роботи з накопичувачем перекладів потрібен перекладач високої кваліфікації.

Другим типом, розглянутим у нашому дослідженні, є системи машинного перекладу, засновані на правилах (rule-based machine translation), – термін, що позначає системи машинного перекладу, засновані на лінгвістичній інформації про вихідну мову і мову перекладу зі словників та граматики. Системи машинного перекладу, засновані на правилах, охоплюють базис семантичних, синтаксичних, морфологічних закономірностей кожної мови. Подібні системи машинного перекладу називають класичними. Історично ці системи з'явилися першими, їх актуальність до сих пір не втрачена, ця система нині використовуються найбільш широко.

За допомогою цих даних вихідний текст перетворюється на текст перекладу по реченню. Система робить величезну роботу, формує зв'язний переклад речення, ґрунтуючись на інформації про кожне слово, узгоджуючи між собою всі члени речення. Цей алгоритм роботи частково аналогічний роботі людини-перекладача: аналіз тексту, пошук варіантів перекладу в своїй «словниковій базі», застосування правил, тобто спочатку аналіз тексту, після чого з'являється варіант перекладу. Дуже часто такі системи протиставляються статистичним системам машинного перекладу. Принцип роботи подібних систем – зв'язок структур вихідного тексту і тексту перекладу (Борисова, 2014).

Перевагами систем заснованих на правилах є граматична і синтаксична точність, стабільність результату, можливість настройки на певну предметну область.

До недоліків систем, заснованих на правилах, зараховують необхідність створення, підтримки і оновлення лінгвістичних баз даних, трудомісткість створення такої системи, а також її порівняно високу вартість.

Третім досліджуваним нами типом МП, є статистичний машинний переклад – це різновид машинного перекладу тексту, заснований на порівнянні великих обсягів мовних пар.

Мовні пари – це тексти, що містять речення в одній мові і відповідні їм речення в іншій, можуть бути як варіантами написання двох речень людиною – носієм двох мов, так і перекладом із вихід-

ної мови мовою перекладу, виконаним людиною. Таким чином, статистичний машинний переклад має властивість «самонавчання». Чим більше в розпорядженні є мовних пар і чим точніше вони відповідають одна одній, тим кращий результат статистичного машинного перекладу.

Найяскравіший приклад, що допомагає зрозуміти, як працює ця система машинного перекладу, – Google Translate. У процесі роботи система аналізує величезні словникові бази парних фрагментів (фраз з двох-трьох слів) – оригінал фрагмента і його переклад. Програма обчислює найбільш ймовірну послідовність слів вихідної мови, яку вона вважає відповідною перекладу вихідного тексту. На відміну від традиційних систем перекладу (саме систем, заснованих на правилах), статистична програма не враховує у своїй роботі граматичні правила, що, безумовно, відбивається на якості перекладу. Нерідкі випадки, коли замість зв'язкового перекладу статистичний перекладач видає абсолютно неузгоджені речення, позбавлені будь-якого сенсу (Молчанов, 2013).

До недоліків цієї системи зараховують недостатній обсяг словникових баз. Наприклад, сервіс Google не завжди справляється з перекладом імен та назв, виникають складнощі з узгодженням членів речення. Цілком ймовірно, що в майбутньому розробникам вдасться підвищити якість перекладів, виконуваних цим методом машинного перекладу.

З очевидних плюсів варто виділити швидке налаштування (порівняно з системами, заснованими на правилах) і самонавчання (участь експерта під час налаштування системи можна звести до мінімуму).

Найбільша система статистичного машинного перекладу, що працює з українською мовою, – це, звісно, Google Translate.

**Четвертий** тип машинного перекладу – це гібридні системи. Існує думка, що корінь переважної більшості проблем машинного перекладу лежить у невідповідності систем мов, а успіхи в розробці програмного забезпечення для перекладу текстів не пов'язані з поганою роботою програмістів або комп'ютерних техніків, а є результатом поганого опрацювання цієї проблеми з лінгвістичної сторони (Борисова, 2014). Результатом цієї ідеї стали розвиток і розробка гібридних підходів.

Гібридні підходи стають дедалі більш популярними, оскільки вони поєднують кращі якості підходу, заснованого на правилах, і статистичного підходу (Costa-jussà, 2015).

Машинний переклад – це міждисциплінарна галузь знань, а до вирішення завдань машинного перекладу можна підходити з різних позицій, використовуючи дані лінгвістики або статистики. Саме існування різних підходів уможливило створення гібридних методів. Гібридні технології фокусуються на тому, щоб взяти всі кращі якості сучасних підходів. Нині найбільш поширений варіант гібридного перекладача – це підключення правил до вже сучасної системи статистичного перекладу (SMT).

Проте проводяться також дослідження, які фокусуються на поліпшенні роботи систем, заснованих на правилах (RBMT), за допомогою додаткової статистичної інформації. Нині лінгвісти, інженер-програмісти і фахівці з області ІТ активно співпрацюють в області МП у процесі спільних семінарів, проводять експерименти і розробляють архітектуру гібридних систем перекладу (Costa-jussà, 2015).

Говорячи про гібридні підходи, варто також згадати про нову унікальну технологію АВВУУ Compreno, яка спочатку розвивалася як переклад за правилами. Нині вона являє собою багатofункціональну лінгвістичну технологію (Burukina, 2014). Система складається з двох основних і низки додаткових компонентів. Перший компонент – це універсальне дерево понять або універсальна семантична ієрархія. Усі слова в парі мов є листям на цьому дереві, між ними задаються відносини, інформація про семантичну сполучуваність. Другий компонент – це синтаксичний аналізатор, який визначає структуру речення і відносини між його словами. Для отримання точного синтаксичного аналізу використовуються семантичні дані про значення слів, які зберігаються в семантичному компоненті. Крім цього, Compreno використовує статистичні методи для зняття лексико-семантичної омонімії та оцінки ймовірності виникнення різних елементів лінгвістичного опису в текстових корпусах.

**Висновки.** На основі викладеного ми можемо стверджувати, що дедалі більшої популярності набувають гібридні підходи, покликані з'єднати в собі плюси класичних і статистичних підходів. Нині системи МП не придатні для роботи з текстами, що містять велику кількість складносурядних і складнопідрядних речень і якісно працюють в основному на рівні словосполучення. Саме тому ця тема є гарним підґрунтям для вивчення.

### СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Беляева Л. Н. Лингвистические автоматы в современных гуманитарных технологиях : Учебное пособие. Санкт-Петербург, 2007. 243 с.
2. Борисова И. А. К опыту постредактирования на материале англо-русского перевода с помощью автоматических систем Google translate и Promt. *Вестник МГЛУ*. 2014. № 13 (699). С. 53–59.
3. Молчанов А. Статистические и гибридные методы перевода в технологиях компании ПРОМТ. Москва, 2013. 154 с.
4. Рябцева Н. К. Информационные процессы и машинный перевод. Лингвистический аспект. Москва, 1986. 168 с.
5. Burukina I. Translating implicit elements in RBMT. *Translating and the Computer*. 2014. № 36. P. 182–193.
6. Costa-jussà M. Latest trends in hybrid machine translation and its applications. *Computer Speech & Language*. 2015. № 32(1). P. 3–10.
7. Hearne M. Statistical Machine Translation: A Guide for Linguists and Translators. *Language and Linguistics Compass*. 2011. № 5. P. 205–226.

### REFERENCES

1. Beliaeva L. N. (2007) Lingvisticheskie avtomaty v sovremennykh humanitarnykh tekhnologiiakh: Uchebnoe posobie. [Linguistic Machines in modern humanitarian technologies: Textbook]. SPb. 243 p. [in Russian].
2. Borisova I. A. (2014) K opytu postredaktirovaniia na materiale anhlo-russkoho perevoda s pomoshchiu avtomaticheskikh system Google translate i Promt. [To the experience of post-editing based on the material of the English-Russian translation using the automatic systems Google translate and Promt]. *Vestnik MHLU*. № 13 (699). P. 53–59. [in Russian].
3. Molchanov A. (2013). Statisticheskie i hibridnye metody perevoda v tekhnolohiiakh kompanii PROMT. [Statistical and hybrid translation methods in PROMT technologies]. M., 154 p. [in Russian].
4. Riabtseva N. K. (1986) Informatsionnye protsessy i mashynnyi perevod. Linhvisticheskii aspekt. [Information processes and machine translation. Linguistic aspect.] M., 168 p. [in Russian].
5. Burukina I. (2014) Translating implicit elements in RBMT. *Translating and the Computer*. № 36. P. 182–193.
6. Costa-jussà M. (2015) Latest trends in hybrid machine translation and its applications. *Computer Speech & Language*. № 32(1). P. 3–10.
7. Hearne M. (2011) Statistical Machine Translation: A Guide for Linguists and Translators. *Language and Linguistics Compass*. № 5. P. 205–226.