

Чернікова Л.Ф.

УДК 81'255 : 004

## ЛІНГВІСТИЧНА ЯКІСТЬ МАШИННОГО ПЕРЕКЛАДУ

**Анотація.** Переклад являє собою багатогранний об'єкт вивчення. Сьогодні комп'ютер навчений здійснювати чимало трудомісткої роботи – укладати словники різних типів, перекладати стандартизовані тексти, здійснювати коректорську, редакторську роботу, підтримувати масштабні бази даних для найрізноманітніших дослідницьких проектів. Сучасний розвиток усіх сфер людської життєдіяльності становить нові завдання в комунікаційному просторі людства, коли рух інформаційних потоків не знає меж у часі і просторі, роль перекладу невпинно зростає.

**Ключові слова:** комп'ютер, машинний переклад, лінгвістика, комунікація, операційні системи.

**Аннотация.** Перевод представляет собой многогранный объект изучения. Сегодня с помощью компьютера осуществляется немало трудоёмкой работы – составляют словари различных типов, переводят стандартизированные тексты, осуществляют корректорскую, редакторскую работы, поддерживают масштабные базы данных для разнообразных исследовательских проектов. Современное развитие всех сфер человеческой жизнедеятельности ставит новые задачи в коммуникационном пространстве человечества, когда движение информационных потоков не знает границ во времени и пространстве, роль перевода неустанно возрастает.

**Ключевые слова:** компьютер, машинный перевод, лингвистика, коммуникация, операционные системы.

**Summary.** Translation is a versatile object for study. Today a computer can do a lot of tedious work – compose dictionaries of different types, translate standardized texts, carry on corrective and editing work, support relevant database for variety of research projects. Contemporary development of all the spheres of human activity sets new tasks in communicative space of humanity, where information flow oversteps the limits of the time and space the role of translation is increasingly growing.

**Key words:** computer, computer translation, linguistic, communication, operating systems.

Сучасний розвиток усіх сфер людської життєдіяльності ставить нові завдання в комунікаційному просторі людства. Коли рух інформаційних потоків не знає меж у часі і просторі, роль перекладу невпинно зростає. До традиційних видів останнього – письмовий, усний, синхронний, послідовний, з листа та ін. – додається машинний переклад, час практичного використання якого розпочався з 1980-х років [3].

У наш час є достатньо широкий вибір програм, які полегшують працю перекладача, котрі умовно можна підрозділити на дві основні групи: електронні словники (*electronic dictionary*) та системи машинного перекладу (*machine translation system*). Системи машинного перекладу забезпечують послідовний переклад текстів, що враховує морфологічні, синтаксичні та семантичні зв'язки членів речення. Програми перекладу (системи машинного перекладу) з'явилися у відповідь на потреби користувачів в оперативному перекладі різної комерційної, технічної або Інтернет - інформації, яка подана в електронному вигляді. Аналізуючи програми машинного перекладу, потрібно відразу зазначити, що вимоги до них не повинні бути такими ж, як і до перекладу, який виконує людина. Переклад, зроблений комп'ютером, поки що далеко не ідеальний, але текст, отриманий в результаті роботи електронного перекладача, дозволяє у більшості випадків зрозуміти суть документа, який перекладається. Далі цей документ можна корегувати, маючи базові знання іноземної мови та добре орієнтуючись в предметній галузі, до якої належить інформація, що перекладається [4].

Вперше можливість машинного перекладу на практиці передбачив Ч. Беббідж, що у першій половині 19 століття працював над проектом цифрової аналітичної машини – механічного прототипу електронних цифрових обчислювальних машин. Процес перекладу він уявляв так: «В мене перед очима текст, написаний російською, але я збираюсь уявити, що насправді він написаний англійською, але за допомогою доволі дивних знаків. Все, що мені потрібно – це зламати код для того, щоб вилучити інформацію, що міститься у тексті.» Перші патенти на створення перекладацьких машин було видано у середині 30-х років минулого століття. Ідея науковця з Західної Європи полягала у створенні автоматичного двомовного словника на основі перфострічки, але проект росіянина П. Троянського був детальнішим. Винайдений ним пристрій включав двомовний словник, здатний оперувати граматичними особливостями за принципом мови есперанто. Система поділялася на три стадії. Спочатку носій мови мав розподілити слова за їх логічними формами та синтаксичними функціями. Потім машина мала виконати переклад на потрібну мову, а носій – відредагувати висхідний матеріал.

Першу пропозицію машинного перекладу за допомогою комп'ютера було висунуто Уорреном Вівером, дослідником з Фонду Рокфеллера у його меморандумі. Пропозиції базувалися на інформаційній теорії, успіхах у зламуванні кодів протягом другої світової війни та обговореннях універсальних та основних принципів мов. За кілька років після опублікування меморандуму розпочалися серйозні дослідження у багатьох університетах Сполучених Штатів. 7 січня 1954 у Нью-Йорку в головному офісі ІВМ було вперше проведено публічну демонстрацію системи машинного перекладу (МП). Про демонстрацію повідомили в газетах, тож подія отримала широкий розголос. Попри те, що сама система мала лише 250 слів та 49 перекладених на англійську російськомовних речень (головним чином у області хімії) і була доволі примітивною, вона продемонструвала перспективи машинного перекладу, стимулювавши фінансування цього дослідження не тільки у США, а й у всьому світі. Експеримент було визнано успішним, що сповістило про початок ери вагомих капіталовкладень у дослідження машинного перекладу. Автори стверджували, що за кілька років машинний переклад буде повністю втілено в життя.

У ранніх системах використовувалися великі двомовні словники та закодовані вручну правила для визначення порядку слів у висхідному продукті. У результаті цей метод було визнано занадто обмеженим, а

завдяки тогочасному розвитку лінгвістики для покращення якості перекладу було запропоновано дослідження генеративної лінгвістики та трансформаційної граматики. Але в цей час операційні системи вже застосовувались. Військово-повітряні сили США використовували систему, розроблену IBM і Вашингтонським університетом, в той час як на ВПС Італії працювала розробка Джорджтаунського університету. Попри те, що якість продукції була низькою, це задовольняло клієнтів, в основному з точки зору швидкості.

Наприкінці 1950-х Г. Бар, дослідник, що на замовлення США вивчав можливість створення повністю автоматичного якісного перекладу, наголосив на проблемі семантичної двозначності під час машинного перекладу. Розгляньмо наступний приклад: *Little John was looking for his toy box. Finally he found it. The box was in the pen.* Слово «pen» має два значення: прилад, що використовується на письмі і певний контейнер. Для людини значення є очевидним, але машина без «універсальної енциклопедії» ніколи не зможе вирішити цю проблему. Сьогодні проблема семантичної двозначності може бути вирішена шляхом написання висхідних текстів контрольованою мовою, тобто застосовуючи словник, у якому для кожного слова є тільки одне значення. Дослідження 1960-х років у СРСР та Сполучених Штатах сконцентрувались головним чином на російсько-англійській мовній парі. В основному об'єктами перекладу виступали науково-технічні документи, як от статті з наукових журналів. Недбалий переклад був достатнім для розуміння сенсу статей. Якщо стаття стосувалася інтересів безпеки, її надсилали живому перекладачу для повного перекладу, решту ж перекладали автоматично. МП зазнав нищівного удару у 1966 році разом із публікацією звіту ALPAC (дорадчого комітету з автоматичної мовної обробки), що складався з семи вчених, скликаних американським урядом у 1964. Американський уряд був занепокоєний повільним просуванням експерименту попри значні видатки, тож було ухвалено рішення, що машинний переклад був дорожчим, менш точним та повільнішим за людський, і незважаючи на витрати, машинний переклад навряд чи досягне якості людського найближчим часом. Однак, у звіті рекомендували продовжувати дослідження у галузі комп'ютерної лінгвістики та створити автоматичні словники, що допомагали б перекладачам. Публікація звіту вплинула на дослідження машинного перекладу у Сполучених Штатах, меншою мірою – у Радянському Союзі та Великій Британії. Звіт, принаймні, майже повністю припинив будь-які дослідження у Сполучених Штатах майже на десятиліття. Однак, у Канаді, Франції та Німеччині дослідження продовжувались; у 1970 систему Systran використовували ВПС США, і, як наслідок, Комісією європейської економічної спільноти. Систему МЕТЕО, розроблену в Університеті Монреаля, було застосовано у 1977 року в Канаді для перекладу прогнозів погоди з англійської мови на французьку. Система перекладала близько 80 000 слів в день або 30 млн слів на рік, поки не була замінена системою конкурента 30-го вересня 2001. В той час як дослідження 1960-х концентрувалися на проблемі обмеження мовних пар та системі вводу інформації, протягом 1970-х років постала потреба в недорогих системах, що змогли б перекладати діапазон технічних та комерційних документів. Ця вимога заохочувалася посиленням глобалізації та потреби перекладів в Канаді, Європі та Японії. До 1980-х років збільшилось різноманіття та число встановлених систем для машинного перекладу. Збільшилась кількість систем, що працювали на основі електронно-обчислювальних машин, як от Systran і Logos. Поширення мікрокомп'ютерів знаменувало створення дешевого ринку систем машинного перекладу, тому багато європейських, американських та японських компаній не втратили шансів цим скористатися. Схожі системи потрапили на ринки Китаю, Східної Європи, Кореї та Радянського Союзу. Протягом 1980-х жвава діяльність у галузі машинного перекладу розгорнулася в Японії. Завдяку комп'ютеру п'ятого покоління Японія мала намір перестрибнути конкурентів у галузі елементів електронних пристроїв та програмного забезпечення. Багато великих японських компаній було залучено до роботи над створенням англо-японських та японо-англійських штучних перекладачів. У дослідженнях 1980-х років переклад вбачався як певний різновид проміжного лінгвістичного відтворення, що залучає морфологічний аналіз разом з синтаксичним та семантичним. Наприкінці 1980-х відбувся вагомий стрибок у галузі створення нових методів для машинного перекладу. Система, розроблена IBM, базувалася на статистичних методах, інші групи застосовували техніку, що базується на великій кількості перекладів у якості зразків.

Протягом 1990-х років, після успіхів у розпізнанні мови та мовному синтезі, дослідження перейшло у стадію мовного перекладу. Відбувся значний ріст використання машинного перекладу в результаті появи дешевих та потужних комп'ютерів. На початку 1990-х машинний переклад став можливим не тільки на великих базових комп'ютерах, а й на персональних комп'ютерах та автоматизованих робочих місцях [1].

Таким чином, з перших же кроків у машинному перекладі вималювалися три основні інваріантні функціонально-процедурні блоки: 1) аналіз вхідного тексту за допомогою спеціального машинного словника та граматики; 2) перетворення результатів аналізу на інформацію, необхідну для побудови перекладеного еквівалента; 3) синтез вихідного тексту, що використовує два типи інформації: отриману на етапі трансферу та граматику синтезу вихідного тексту. Машинний переклад еволюціонував від максимального спрощення версій ("лексиконних") до версій, "заглиблених" у зміст. Його еволюцію можна представити як "естафету" п'яти послідовних поколінь та, відповідно, типів систем:

1. *Системи послівного перекладу (word-for-word translation), доповнені деякими допоміжними граматичними конструкціями* – системи, що здатні опрацьовувати лише окремі мовні ситуації в конкретних випадках, але не здатні оперувати ними в цілому, тобто на рівні типів, класів, груп тощо.

2. *Структурно-граматичні системи машинного перекладу, що базуються на морфологічних кореляціях між вхідною та вихідною мовами - "морфологічні системи"* – виявляються ефективними для організації перекладу в межах споріднених мов (наприклад, українсько-російський машинний переклад та навпаки).

3. Структурно-граматичні системи машинного перекладу, що спираються на синтаксичні кореляції між вхідною та вихідною мовами - "синтаксичні системи".

Центральною процедурою тут є синтаксичний аналіз вхідної фрази, яка далі трансформується в структурно-синтаксичний каркас вихідної фрази.

4. Структурно-семантичні системи машинного перекладу, що оперують глибинними структурами вхідного та вихідного контекстів. У таких системах передбачається багаторівневе опрацювання мовного матеріалу.

5. Автоматизовані робочі місця (станції) перекладача – інтерактивні словниково-орієнтовані системи машинного перекладу з великими за обсягом і деталізованими термінологічними словниками. Концепція полягає в тому, що системи машинного перекладу принципово неспроможні забезпечити його високу якість, тому вони не можуть цілком замінити користувача-перекладача, але мають допомагати йому.

Отже, наприкінці ХХ сторіччя машинний переклад існує вже не лише як теоретична дисципліна з експериментальним супроводом, але і як технологічна реальність. Проте, навіть при такому рівні розвитку сучасних систем автоматичного перекладу, не завжди вдається правильно перекласти деякі словосполучення у сферах професійного спрямування [3].

При створенні СМП, які базуються на використанні лінгвістичних правил, потрібне знання розпізнавання ознак тексту, що відносяться до сфери прагматики: жанр та стиль (наприклад, це публіцистична стаття, вірш чи документ встановленого зразка); область знання, до якого текст відноситься (розпізнавання термінології); зв'язаність частин тексту, що не завжди описується за допомогою синтаксичних чи лексико-семантичних критеріїв; і т.д. Можна вважати, що для адекватного перекладу автоматична система повинна:

- знати внутрішні структури мов, між якими здійснюється переклад;
- мати ясне уявлення про культуру, історію, мораль, переважні типи мислення народів, що є носіями мови;
- володіти по можливості більшим словниковим запасом, більш-менш структурованим по областях застосування слів (спеціальна термінологія, діалекти, ідіоматика, сленг);
- мати явний чи інтуїтивний тезаурус слів обох мов, тобто по даному слову вміти запропонувати семантичні функції від нього, такі як синонім, антонім, конверсив, класичний атрибут, а також вміти запропонувати похідні частини мови від даного слова, якщо такі існують (добро – добрий – добріше – подобрів і т.п.) [5].

Використання систем машинного перекладу (СМП) з українською та російською мовами є наразі дуже поширеним в Україні. Проте, якщо при перекладі тексту з російської на українську мову користувачеві пропонуються скальковані російські форми з автоматичного словника. Отже, питання покращення таких словників, на наш погляд, є досить важливим. Інакше переклад, який ряснітиме покручами, не матиме жодної цінності. Вичитуючи перекладений з російської мови текст, пересічний користувач без відповідної філологічної підготовки виправляє, як правило, звичайні помилки, пов'язані з узгодженням слів, відмінковим та прийменниковим керуванням. При цьому лексичні помилки часто залишаються непоміченими, оскільки, на думку користувача, якщо слово є в автоматичному словнику, то вони мають бути правильними. Проблема полягає в тому, що навіть академічні словники містять подекуди форми, скальковані з російської мови.

Українська та російська мови близькі за своєю будовою, граматичними конструкціями тощо, тому створювати текстові збірки для оцінювання якості перекладу за допомогою відповідних розбіжностей між цими мовами, на нашу думку, недоречно. Тестові речення мають містити, на наш погляд, такі речення і словосполучення, правильний переклад яких вимагає якісного автоматичного словника та використання певних перекладацьких трансформацій. Ми обрали 1000 тестових речень для російсько-українського перекладу. Для оцінювання цього напряму перекладу ми використовували такі системи машинного перекладу: "Ленгвідж Майстер 98", "Прагма 5.0" та "Плай 5.0" (надалі "ЛМ", "Прагма", "Плай").

При перекладі текстів системами машинного перекладу часто трапляються слова, які відсутні в автоматичному словнику. Як правило, це терміни певної галузі або власні назви. Неперекладене слово суттєво впливає на загальну якість перекладу, оскільки воно випадає із загального аналізу й може призвести до неправильного синтезу вихідного речення. Кількість неперекладених слів також може свідчити про якість системи. Щодо нашої збірки речень, найбіднішим виявився автоматичний словник системи "ЛМ", який не зміг перекласти 43 слова з тесту. Система "Плай" не переклала 24 слова. Кращий показник з цього погляду належить перекладачеві "Прагма" – у тестових реченнях знайшлося лише 9 слів, які невідомі системі. Усім, кому доводилося вивчати іноземну мову, знають, що вживання тих чи інших прикметників часто не збігається в рідній та іноземній мовах. Навіть у таких близьких мовах, як українська та російська, моделі словотворення збігаються не завжди. За цілком слушною думкою С. Караванського, "... кожна мова творить у процесі розвитку свою власну модель паруння прикметників з іменниками, і ця модель не може око-в-око повторитися в іншій мові". Так, наприклад, для перекладу російського слова "жесткий" українці вживають чотири різні прикметники: *твердий* (вагон), *круті* (заходи), *тяжкі* (умови), *суворий* (контроль). Для перекладу слова "жесткий" СМП "Прагма" використовує лише одне слово – "жорсткий". У словнику системи "Плай" є ще один варіант перекладу – "твердий", що збільшує можливість правильного перекладу, але не відповідає всім варіантам вжитку слова "жесткий".

Кожна мова в процесі свого розвитку намагається уникати довжелезних і складних конструкцій. Будь-який мовець з розвиненим чуттям мови не вживатиме таких форм. Короткі та природні форми є окрасою будь-якої мови, вони створюють її неповторну палітру й полегшують спілкування. Так, наприклад, кожна з

трьох систем машинного перекладу перекладає російське словосполучення “должны были” як “повинні були”. А якщо речення має складний присудок, то взагалі отримуємо занадто складну форму: “... резолюції, які повинні були бути включені в загальну постановку” (“ЛМ”, 18). Звичайно, що форма “мали бути включені” буде більш відповідною й звучатиме по-українському.

З трьох систем, які ми оцінюємо в цій статті, найкраще перекладає керування діє- слів “Прагма”, хоча результати перекладу цього граматичного явища в кожній системі досить невтішні.

Крім того, до грубих помилок належать випадки хибного перекладу лише одного слова, переклад якого не потребує будь- якого аналізу та жодних перекладацьких трансформацій. Наведемо приклади таких помилок: “болільник” замість “вболівальник” (“Плай”, 278), “бастували” замість

“страйкували” (“Прагма”, 259). Схожі помилки мають бути виправлені в автоматичних словниках у першу чергу. Їх виправлення ніяк не пов’язане зі зміною алгоритму й не може призвести до погіршення якості перекладу.

Серйозною проблемою для багатьох мовців є переклад на українську мову російських дієприкметників. Як відомо, з погляду стилістики української мови краще уникати частого вживання дієприкметників. Коли йдеться про активні та пасивні дієприкметники теперішнього часу, то вони практично відсутні в сучасній українській мові, що є її лексико-граматичною особливістю, а не вадою. Тестованим нами системам це явище, здається, невідоме. Майже всі російські дієприкметники з добірки тестових речень були перекладені тільки дієприкметниками.

Замість них можна вживати прикметники: “панівні кола”, а не “правлячі” (“Плай”, 817); “життєствердні мотиви”, а не “життєстверджуючі” (“ЛМ”, 823); “всеохопний підхід”, а не “усеохоплюючий” (“Плай”, 820). Досить продуктивною є заміна активних дієприкметників іменниками: не “командуючий” (“Прагма”, 833), а “командувач флотом”; не “завідуючий кафе-дрою” (“ЛМ”, 832), а “завідувач кафедрою”; не “граючий тренер” (“Плай”, 827), а “тренер-гравець”; не “виступаючі відзначили” (“ЛМ”, 323), а “промовці відзначили” тощо.

Розглянемо нарешті сталі словосполучення, вдалий переклад яких, на нашу думку, є основним показником якості систем машинного перекладу для цього напрямку.

Системи машинного перекладу третього покоління володіють достатнім інструментарієм для адекватного перекладу сталих словосполучень. Проте оцінювання та аналіз тестованих речень для російсько-українського напрямку свідчить про те, що з цим видом у перекладацьких систем поки, на жаль, не все гаразд.

Найбільшу групу сталих словосполучень серед тестових речень становлять дієслівно-іменникові словосполучення, частина яких перекладена цілком пристойно: “Суперник чинив гідний *опір*” (“Плай”, 154); “Питання, звичайно, не однозначне, але в цьому він, безумовно, має *рацію*” (“Прагма”, 234); “З метою безпеки, охорона була вимушена *вжити* найсуворіших *заходів*” (“ЛМ”, 314). Проте більшість словосполучень перекладено невдало: “*відволікати увагу*” (“Прагма”, 139) замість “*відвертати увагу*”; “*взяти напрокат* автомобіль” (“Плай”, 317) замість “*випозичити* авто”; “*підняла крик*” (“ЛМ”, 695) замість “*вчинила скарс*”; “*читати про себе*” (“Плай”, 980) замість “*читати подумки*” та ін.

Таким чином, робота над покращенням якості систем автоматичного оброблення текстової інформації, зокрема систем машинного перекладу, потребує багато зусиль та часу, але це не безплідна робота. На нашу думку, тут потрібно застосовувати як квалітативний, так і квантитативний підходи: що більше буде складено добірок з тестовими реченнями для оцінювання якості машинного перекладу, то більше буде досліджена та проаналізована робота системи з певними мовними елементами, що рано чи пізно дасть змогу суттєво покращити рівень якості перекладу. Отже, основною прерогативою роботи розробників СМП у межах цієї мовної пари є наразі, на наш погляд, поліпшення автоматичних словників. Оскільки досягнення ідеальної якості машинного перекладу – завдання надто складне, то переклад може бути недосконалим, але принаймні українським, а не калькованим, щоб запобігти засміченню та дискредитації мови [2].

#### Джерела та література:

1. Hutchins, J. (1986) Machine Translation: past, present, future (Chichester : Ellis Horwood) ISBN <http://www.traduceme.org/profiles/blogs/history-of-machine-translation>
2. Бірюков А.В. Тестування лінгвістичної якості систем машинного перекладу “Плай”, “Прагма” та “Ленгвідж майстер” – Режим доступу : [http://www.nbuv.gov.ua/portal/Soc\\_Gum/Dtr\\_gn/2008\\_4/files/GN\\_04\\_2008\\_Birukov.pdf](http://www.nbuv.gov.ua/portal/Soc_Gum/Dtr_gn/2008_4/files/GN_04_2008_Birukov.pdf)
3. Кісіль М. В. До проблеми машинного перекладу // Актуальні проблеми соціальної комунікації : збірник матеріалів конференції. – Івано-Франківськ : В-во ІФНГУНГ, 2011. – С. 97 – 103.
4. Мацак Ж. Г., ст. викл. Скоробогатова Т. І. Кременчуцький державний університет, Україна Машинний переклад та його специфіка – Режим доступу : [http://www.rusnauka.com/30\\_NIEK\\_2009/Philologia/54186.doc.htm](http://www.rusnauka.com/30_NIEK_2009/Philologia/54186.doc.htm)
5. Чрділелі Т. В., Кожемяченко Н. В., Ю. Г. Хохлов. Інформаційні технології у мовознавчих студіях. – 2010 – Режим доступу : [http://www.nbuv.gov.ua/portal/natural/newtech/2010\\_1/articles/2-8.pdf](http://www.nbuv.gov.ua/portal/natural/newtech/2010_1/articles/2-8.pdf)