

УДК 81'336

DOI: 10.32342/2523-4463-2019-2-18-25

Н.М. СОПИЛЮК,

*кандидат філологічних наук,
доцент кафедри романської філології та перекладу
Чернівецького національного університету імені Ю. Федьковича*

І.О. ЦАРЕНКО,

*кандидат економічних наук,
магістрант кафедри романської філології та перекладу
Чернівецького національного університету імені Ю. Федьковича*

КОМПАРАТИВНИЙ АНАЛІЗ СИСТЕМ МАШИННОГО ПЕРЕКЛАДУ ЕКОНОМІЧНОГО ДИСКУРСУ (на прикладі французько-українських мовних пар)

У статті продемонстровано компаративний аналіз систем машинного перекладу франкомовного економічного дискурсу українською мовою, враховуючи зростаючі тенденції до зовнішньоекономічної діяльності вітчизняними підприємствами та посилення кооперації в межах ринку праці. Компаративістика охоплює сферу машинного перекладу з точки зору розвитку комп'ютерної лінгвістики та зростання попиту на сервіси з перекладацьких послуг. Дослідження зосереджено на аналізі перекладу французько-української мовної пари економічного дискурсу за допомогою існуючих систем машинного перекладу. Проаналізовано історію виникнення перших систем машинного перекладу. Визначено основні підходи до класифікації таких систем, зокрема, абсолютно автоматизований машинний переклад, машинний переклад за участю людини, переклад, який виконує людина з використанням EOM, та інший підхід – системи, які працюють завдяки використанню правил, статистичні системи машинного перекладу та гібридні системи.

Проведено опитування серед респондентів, щоб визначити найбільш поширені системи серед економістів-практиків та викладачів економічних дисциплін. У результаті, встановлено, що найпоширенішими системами машинного перекладу є «Google-Translate», «Яндекс.Перевод» та «PROMT».

Шляхом лінгвістичного аналізу (на морфологічному, лексичному та синтаксичному рівнях) виявлено переваги та недоліки машинного перекладу онлайн-системами Google Translate, Яндекс та PROMT.

Серед основних переваг машинного перекладу виділено такі: швидкість; доступність; досить широкий спектр вибору еквівалентних відповідників. Серед недоліків відзначено такі: дослівний переклад; непередбачуваність у прийнятті рішень і виборі еквівалента; при перекладі система машинного перекладу не враховує широкий контекст. Визначено причини виникнення перекладацьких помилок та необхідність використання постредагування для поліпшення якості перекладу.

Ключові слова: машинний переклад, економічний дискурс, онлайн-системи машинного перекладу, оцінка якості машинного перекладу, синтаксичний рівень, морфологічний рівень, лексичний рівень.

В статтю продемонстрований компаративний аналіз систем машинного перекладу франкомовного економічного дискурсу на український мовний матеріал з урахуванням тенденції до зовнішньоекономічної діяльності вітчизняних підприємств та посилення кооперації в межах ринку праці. Компаративістика охоплює сферу машинного перекладу з точки зору розвитку комп'ютерної лінгвістики та зростання попиту на сервіси перекладацьких послуг. Дослідження зосереджено на аналізі перекладу французько-української мовної пари економічного дискурсу за допомогою існуючих систем машинного перекладу.

существующих систем машинного перевода. Проанализирована история возникновения первых систем машинного перевода. Определены основные подходы к классификации таких систем, в частности, абсолютно автоматизированный машинный перевод, машинный перевод при участии человека, перевод, выполняемый человеком с использованием ЭВМ, и другой подход – системы, работающие благодаря использованию правил, статистические системы машинного перевода и гибридные системы.

Проведен опрос среди респондентов с целью определения наиболее распространенных систем среди экономистов-практиков и преподавателей экономических дисциплин. В результате установлено, что наиболее распространенными системами машинного перевода являются «Google Translate», «Яндекс.Перевод» и «PROMT».

Путем лингвистического анализа (на морфологическом, лексическом и синтаксическом уровнях) выявлены преимущества и недостатки машинного перевода онлайн-системами Google Translate, Яндекс и PROMT.

Среди основных преимуществ машинного перевода выделены следующие: скорость; доступность; достаточно широкий спектр выбора эквивалентных соответствий, тогда как среди недостатков отмечены следующие: дословный перевод; непредсказуемость в принятии решений и выборе эквивалента; при переводе система машинного перевода не учитывает широкий контекст. Определены причины возникновения переводческих ошибок и необходимость использования постредактирования для улучшения качества перевода.

Ключевые слова: машинный перевод, экономический дискурс, онлайн-системы машинного перевода, оценка качества машинного перевода, синтаксический уровень, морфологический уровень, лингвистический уровень.

Учас прискороеного розвитку науки та технологій, поглиблення процесу інформатизації та діджиталізації усіх сфер суспільного життя не лишилась осторонь і сфера перекладознавства. Саме процес комп'ютеризації наприкінці 1990-х років вплинув на появу кардинально нового напрямку в лінгвістиці, а саме – комп'ютерної, серед проблемних аспектів якої одне з досить важливих місць займає машинний переклад.

Впродовж останнього часу машинний переклад набув досить значної популярності у всіх сферах життя – від повсякденної до професійної, адже поглиблення зовнішньоекономічної діяльності підприємств, розширення зв'язків вплинуло на появу товарів та послуг іноземного виробництва, розширило можливість доступу до широкої бази контентів іноземною мовою тощо. Таким чином, потреба розуміння сутнісного змісту іноземного контенту стала досить важливою.

Проблемам дослідження аспектів особливостей машинного перекладу присвятили праці вчені-лінгвісти різних країн, зокрема, А. Бут і У. Локк [1], Т. Скоробогатова [3], Ю. Марчук [2], Ж. Мацак [3], А. Міщенко [4], В. Хатчинс [7] та ін. Наприклад, праці А. Міщенко [4] та П. Кьоена [8] було присвячено дослідженню статистичних та гібридних систем машинного перекладу. Проте найбільшу цінність з точки зору емпіричності щодо мети нашого дослідження мають наукові доробки П. Хроменкова [5], Д. Арнольда [6] та С. Ніренбурга [10] стосовно якісного рівня вихідного результату систем машинного перекладу.

Окрім фахівців-філологів, вирішенням проблем, пов'язаних із машинним перекладом також займаються представники відомих компаній, таких як SYSTRAN Software Inc, Alis Technologies Inc., Toshiba Corp., Compu Serve, Logos Corp., Globalink Inc., Fujitsu Corp., TRADOS Inc., PROMT, SAP AG та ін.

Проте аналіз особливостей машинного перекладу документів економічного дискурсу досліджено недостатньою мірою, що й обумовило вибір мети цього дослідження. Крім того, з огляду на те, що більшість досліджень пов'язані з аналізом мовної пари англійська-українська, нами було обрано іншу вхідну мову, а саме, французьку. Отже, основною метою дослідження є проведення компаративного аналізу наявних систем машинного перекладу та висвітлення основних помилок при перекладі економічних текстів.

Точкою відліку появи машинного перекладу слід вважати Джорджтаунський експеримент 1954 р., результатом якого стала публічна демонстрація результату перекладу за допомогою електронно-обчислювальної техніки. Експеримент мав на меті здійснити переклад 60 речень з російської мови на англійську, вибір мовних пар пов'язаний з періодом холодної війни. Враховуючи той факт, що речення були простими, вихідний результат вважався успішним.

Загалом етимологія поняття машинного перекладу не має однозначного перекладу, проте більшість дослідників погоджуються, що це процес перекладу вихідного тексту з використанням спеціального програмного забезпечення, встановленого на електронно-обчислювальній техніці.

З огляду на різноманітність підходів до тлумачення поняття, логічним є те, що не існує єдиного підходу до класифікації систем машинного перекладу.

Одним із загальноприйнятих підходів до класифікації є поділ систем за рівнем автоматизованості, який у 1990 р. було запропоновано Л. Чайлдсом. Він передбачає зокрема такі категорії машинного перекладу.

FAMT (Fully-automated machine translation) – абсолютно автоматизований машинний переклад передбачає, що до системи вноситься текст однією мовою, остання, опрацювавши його, виводить вихідний результат іншою мовою.

Основними перешкодами цього типу систем машинного перекладу є в першу чергу велика ймовірність дослівного перекладу ідіоматичних конструкцій та неправильне визначення частин мови. Вважається, що основною мірою якості систем повністю автоматизованого машинного перекладу є можливість вирішення проблеми розуміння природної мови.

HAMT (Human-assisted machine translation) – машинний переклад за участю людини передбачає, що остання здійснює опрацювання тексту, скорочуючи довгі речення, усуваючи неясності та можливий двозначний переклад, складні конструкції. Зазвичай людина, яка використовує цей тип повинна володіти граматиною та мати значний запас слів.

MAHT (Machine-assisted human translation) – переклад, який виконує людина з використанням ЕОМ. Цей тип перекладу повністю виконує людина, а електронно-обчислювальна машина здійснює виключно перевірку термінології, пошук відповідностей у словнику та порівняння з іншими перекладами. Досить часто такі системи перекладу називають ТМ-програмами (від англійського словосполучення «translation memory» – пам'ять перекладу) і вони поширені виключно серед професіональних перекладачів [3].

Ще одним підходом до класифікації є правила, а саме їх використання у процесі перекладу.

Відповідно до цієї класифікації прийнято розрізняти:

1. Системи, які працюють завдяки використанню правил (Rule-Based Machine Translation – RBMT). Такі системи працюють за правилами, які були розроблені професійними експертами та передбачають фіксацію вихідної мови цільовою. Безперечно залучення людини є перевагою та відображається у високій якості вихідного продукту, проте, з іншого боку, це призводить до необхідності постійного оновлення таких систем, а також збільшує часовий проміжок процесу перекладу;

2. Статистичні (Statistical Machine Translation – SMT) системи машинного перекладу – системи, які використовують комп'ютерні алгоритми при здійсненні перекладу, що значно підвищує його якість. Статистичні системи містять «базу даних» із слів та фраз, які автоматично було вивчено та запозичено з двомовних паралельних речень. Основною перевагою статистичних систем машинного перекладу є рівень їх автоматизованості при побудові нової системи, при цьому наявна можливість машинного навчання, в результаті чого, більш швидка дія та нижча вартість обчислювальної потужності, яка необхідна для побудови та експлуатації моделей, в основу яких покладено статистичний алгоритм. Однак варто відзначити наявність значного недоліку – «ефекту розбавлення даних», який спричинено дефіцитом даних, що підходять для здійснення «навчання».

3. Гібридні (Hybrid Machine Translation – HMT hybrid) системи машинного перекладу поєднують у собі переваги двох попередніх та доповнені новими технологіями та алгоритмами, які спрямовані на поліпшення якості та подолання часових обмежень. Проте варто відзначити притаманні таким системам витрати у зв'язку з появою додаткових складнощів керування [4].

З огляду на стрімкий розвиток технологій, останніми роками виник абсолютно новий алгоритм, який здійснює переклад на більш якісному рівні, ніж попередні, а саме, дає змогу усунути недоліки традиційних систем машинного перекладу, які здійснюють його «пофразово», це стало можливим за рахунок того, що в основу системи було покладено штучну нейронну мережу – нейронний машинний переклад (NMT).

Далі вважаємо за необхідне здійснити аналіз результату перекладу економічних документів з французької мови на українську за допомогою систем машинного перекладу.

У зв'язку з тим, що нині існує широкий спектр таких систем, нами було проведено дослідження серед респондентів, щоб визначити найбільш поширені системи серед економістів-практиків та викладачів економічних дисциплін.

Таке опитування було здійснено шляхом проведення анкетування. В результаті нами було опитано 156 респондентів та виявлено такі результати: 61 особа використовують онлайн-перекладач Promt (39,1 %), 48 респондентів користуються послугами перекладача Google-Translate (30,8 %), 32 особи – машинним перекладачем Яндекс (20,5 %), 13 осіб користуються іншими системами машинного перекладу (8,3 %) і лише (1,3 %) 2 особи використовують друковані словники та здійснюють переклад вручну (рис. 1).

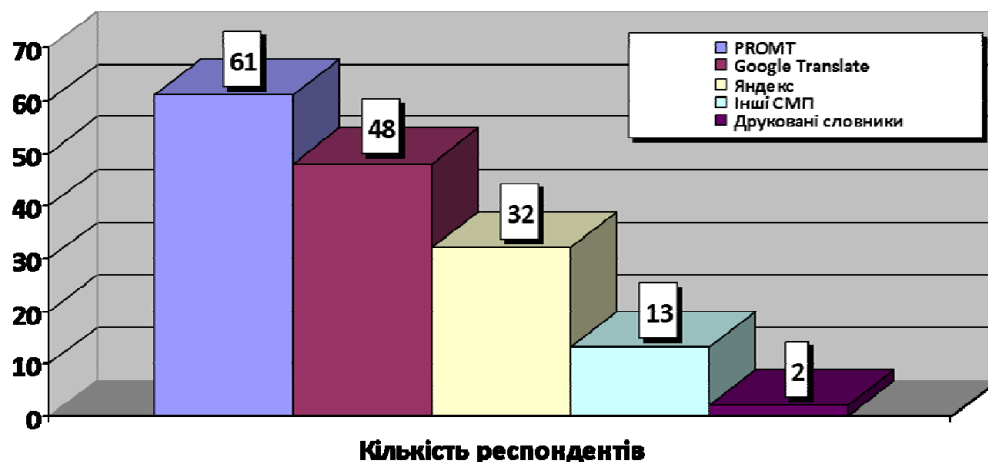


Рис. 1. Результати опитування респондентів щодо систем машинного перекладу

У результаті встановлено, що найпоширенішими системами машинного перекладу серед економістів є «Google-Translate», «Яндекс.Перевод» та «PROMT». Далі вважаємо за необхідне провести безпосередньо аналіз результатів їх роботи. Матеріалом для дослідження було обрано публікації на офіційному сайті Світового банку – *Mettre fin à la pauvreté, investir dans les opportunités* та *Le rapport annuel du Haut Conseil de stabilité financière* [9].

При здійсненні перекладу фрагмента економічного документа було отримано такі результати.

Оригінал:

AU TOTAL, 62,3 MILLIARDS DE DOLLARS de prêts, dons, prises de participation et garanties en faveur de pays partenaires et d'entreprises privées. Le montant total inclut les projets multirégionaux et mondiaux. La répartition par région tient compte de la classification des pays par la Banque mondiale.

Google:

ВСЬОГО 62,3 млрд дол. США в позиках, пожертвах, капітальних вкладеннях та гарантіях для країн-партнерів та приватних компаній. Загальна сума включає мультирегіональні та глобальні проекти. Розбиття за регіонами враховує класифікацію країн Світовим банком.

Yandex:

В цілому 62,3 млрд дол. Загальна сума включає в себе міжрегіональні та глобальні проекти. В розбивці по регіонах враховується класифікація країн Світового банку.

Promt:

У загальній складності, 62,3 мільярда доларів позичок, дотацій, участі і гарантій на користь партнерських країн і приватних підприємств. Загальна сума включає multiregionalnyue і світові проекти. Розподіл регіоном приймає до уваги класифікацію країн світовим банком.

Авторський переклад:

Країнам-партнерам та приватним компаніям були надані позики, гранти, інвестиції в акціонерний капітал та гарантії на загальну суму 62,3 млрд дол. США. Загальна сума включає міжрегіональні та глобальні проекти. Розподіл по регіонах відображає прийняту у Світовому банку класифікацію країн.

У ході проведення аналізу та виявлення помилок були помічено такі закономірності.

Граматичні трансформації на словотвірному рівні

Використання великої літери:

Розподіл регіоном приймає до уваги класифікацію країн світовим банком. / Розподіл по регіонах відображає прийняту у Світовому банку класифікацію країн.

Відповідно до правопису української мови установи державного значення пишуться з великої літери, що не враховує при перекладі машина.

Граматичні трансформації на морфологічному рівні

Категорійна заміна відмінка:

Розподіл регіоном приймає до уваги класифікацію країн світовим банком. / Розподіл по регіонах відображає прийняту у Світовому банку класифікацію країн.

При машинному перекладі маємо додаток в орудному відмінку, при авторському – у місцевому.

Розподіл регіоном приймає до уваги класифікацію країн світовим банком. / Розподіл по регіонах відображає прийняту у Світовому банку класифікацію країн.

Машинний переклад використовує орудний відмінок, що призводить до помилкового сприйняття змісту речення (сприймаємо, що центральний поділ здійснюється регіоном, а не по його частинах), а в авторському перекладі – місцевий відмінок.

Категорійна заміна числа:

62,3 мільярда доларів позичок / Країнам-партнерам та приватним компаніям було надано позики

Машинний переклад здійснює переклад іменника в однині у жіночому роді, а при авторському – у множині.

Граматичні трансформації на синтаксичному рівні

Вилучення:

У загальній складності, 62,3 мільярда доларів позичок, дотацій, участі і гарантій на користь партнерських країн і приватних підприємств. / Країнам-партнерам та приватним компаніям було надано позики, гранти, інвестиції в акціонерний капітал та гарантії на загальну суму 62,3 млрд. дол. США

При авторському перекладі вилучено іменник, адже так нагромадженість речення стає значно меншою, а зміст від цього не зазнає проблемного розуміння.

Зміна синтаксичного зв'язку у реченні:

У загальній складності, 62,3 мільярда доларів позичок, дотацій, участі і гарантій на користь партнерських країн і приватних підприємств. / Країнам-партнерам та приватним компаніям було надано позики, гранти, інвестиції в акціонерний капітал та гарантії на загальну суму 62,3 млрд. дол. США

При машинному перекладі маємо односкладне речення називного типу, а при авторському – односкладне безособове, що, по-перше, допомагає читачу зрозуміти зміст речення, по-друге, використання безособових речень є ознакою наукового та офіційно-ділового стилів мовлення.

Структура речення при машинному перекладі порушує прямий порядок слів в українській мові, тому ми вважаємо за доцільне змінити його на прямий, адже так читач зможе зрозуміти зміст речення в повному обсязі, оскільки при машинному перекладі незрозумілий суб'єкт, який виконує дію.

Зміна порядку слів у реченні:

Розподіл регіоном приймає до уваги класифікацію країн світовим банком. / Розподіл по регіонах відображає прийняту у Світовому банку класифікацію країн.

Неадекватність перекладу:

Загальна сума включає multiregionalnyye і світові проекти. / Загальна сума включає міжрегіональні та глобальні проекти.

Неспроможність системи здійснити переклад слова «міжрегіональні» можна пояснити лише відсутністю останнього у словнику вищезазначеної системи машинного перекла-

ду. Така ситуація може виникати й з іншими словами, тому вважається, що переклад, який виконує людина, є більш досконалим та усуває ймовірність виникнення подібних ситуацій.

Одержані в результаті детального лінгвістичного аналізу дані дають змогу стверджувати таке: машинний переклад (незалежний від людини) на сьогоднішній день є неможливим, однак результати процесу здійснення перекладу системами машинного перекладу є зручними та корисними для здійснення поверхневого ознайомлення з змістовим сенсом за умови, що текст використовується як сигнальна інформація і не потребує ретельного редагування.

Для повноцінного ж перекладу необхідне постредагування тексту з урахуванням виявленої типології помилок машинного перекладу. Слід відзначити, що нами було виокремлено лише найбільш поширені помилки машинного перекладу, і цей список може бути, безсумнівно, продовжений для текстів економічного профілю.

Таким чином, за підсумками аналізу машинного перекладу обраних нами матеріалів було зроблено такі висновки.

По-перше, було виявлено переваги машинного перекладу, а саме: швидкість; доступність; досить широкий спектр вибору еквівалентних відповідників, що дає змогу переструктурувати речення.

Переклад зрозумілий, дотримується еквівалентність перекладу на рівні мети комунікації, не було виявлено серйозного спотворення сенсу. Однак слід зауважити: специфічні аббревіатури не можуть бути правильно перекладені за допомогою машинного перекладу.

Варто зазначити, що для кожної мови спостерігаються свої особливості. Так, наприклад, якість перекладу для мовної пари французька–українська значно нижча, ніж для мовних пар, які є більш спорідненими між собою.

Серед основних недоліків машинного перекладу слід відзначити такі.

1. Дослівний переклад. Незважаючи на те, що система машинного перекладу вдається до трансформацій, у більшості випадків переклад буквальний, у ньому зберігаються синтаксичні конструкції оригіналу.

2. Непередбачуваність у прийнятті рішень і виборі еквівалента. Відзначається велика кількість випадків, коли вибір тієї чи іншої лексичної одиниці не мотивований або нерелевантний, що призводить до порушення лексичної сполучуваності або спотворення змісту речення, що пояснюється роботою алгоритмів системи нейронного машинного перекладу, яка приймає рішення про вибір того чи іншого слова.

3. При перекладі система машинного перекладу не враховує широкий контекст. Переклад здійснюється на рівні речення, тому в одному тексті одне і те ж слово або словосполучення може бути перекладено по-різному. Усі ці фактори знижують якість перекладеного тексту. Крім того, було виявлено, що для машинного перекладача існують різні обмеження. Наприклад, наявність у тексті оригіналу слів мовою, відмінною від заданої, ускладнює переклад, і в результаті цього такі слова залишаються не перекладеними. Ще одним обмеженням для систем машинного перекладу є рідкісні слова, наприклад, імена людей.

Таким чином, з введенням систем нейронного машинного перекладу якість машинного перекладу значно зросла. Завдяки таким системам машинний переклад став більш «природним», знизилася кількість помилок, що веде до скорочення часу, необхідного для постредагування тексту і виправлення помилок. Такий переклад викликає більшу довіру, і в майбутньому перекладачі та звичайні користувачі, наприклад, спеціалісти в економічній сфері, все частіше будуть вдаватися до автоматичного перекладу.

Розглянувши типові помилки на прикладах, ми дійшли висновку, що жодна з систем не є досконалою на цьому етапі, всі вони потребують доопрацювання. Перспективою подальших наукових розвідок є пошук шляхів усунення найбільш поширених помилок систем машинного перекладу при здійсненні перекладу економічного дискурсу.

Список використаної літератури

1. Бут А.Д. Историческое введение / А.Д. Бут, У.Н. Локк // Машинный перевод. – М.: Издательство иностранной литературы, 1957. – С. 15–32.
2. Марчук Ю.Н. Проблемы машинного перевода / Ю.Н. Марчук. – М.: Наука, 1983. – 201 с.

3. Мацак Ж.І., Скоробогатова Т.І. Машинний переклад та його специфіка / Ж.І. Мацак, Т.І. Скоробогатова. – 2009. – [Електронний ресурс]. – Режим доступу: http://www.rusnauka.com/31_ONBG_2009/Philologia/54653.doc.htm (останнє звернення 12.08.2019).
4. 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». – 2013. – № 1051. – С. 172–180
5. Хроменков П.Н. Анализ и оценка эффективности современных систем машинного перевода: автореф. дисс. ... канд. филол. наук / П.Н. Хроменков. – М., 2000. – 170 с.
6. Arnold D. Machine Translation: An Introductory Guide / D. Arnold. – London: NCC Blackwell, 1994. – 240 p.
7. Hutchins W.J. Current and potential applications of machine translation. – 2012. – [Електронний ресурс]. – Режим доступу: <http://www.hutchinsweb.me.uk/Aslib-2012.pdf> (останнє звернення 12.08.2019).
8. Koehn P. Statistical Machine Translation / P. Koehn. – Cambridge: Cambridge University Press, 2010. – 433 p.
9. Mettre fin à la pauvreté, investir dans les opportunités та Le rapport annuel du Haut Conseil de stabilité financière. – 2019. – [Електронний ресурс]. – Режим доступу: <https://www.banquemonde.org/fr/about/annual-report#anchor-annual> (останнє звернення 12.08.2019).
10. Nirenburg S. Progress in Machine Translation / S. Nirenburg. – Amsterdam: IOS Press, 1992. – 223 p.