

Типові помилки систем машинного перекладу

19. Філологія та журналістика

19.1 Філологія

Ключові слова: машинний переклад, системи перекладу, помилки.

Одним зі способів комп'ютерного моделювання природної мови є системи програмного забезпечення перекладу – комп'ютерні програми, які використовуються для перекладу інформації з однієї природної мови іншою. Такі програми називають системами машинного перекладу [10].

У мовознавстві під системами машинного перекладу розуміють такий тип програмного забезпечення, який здійснює обробку та переклад тексту з однієї природної мови на іншу з максимальним збереженням змісту та структури оригіналу [7]. Проте робота алгоритму системи перекладу не підлягає прямому лінгвістичному аналізу, тому у прикладному мовознавстві предметом аналізу стають машинні переклади текстів [5].

Зараз у центрі уваги вчених, що досліджують системи машинного перекладу перебуває аналіз структури і функціонування мовних одиниць як результатів інтелектуальної діяльності людини (М. П. Кочерган), а також їх моделювання у системах комп'ютерної обробки мови (Ю. М. Марчук, Л. Л. Нелюбін). Також широко поширені дослідження систем машинного перекладу, як одного з основних напрямів теорії штучного інтелекту (В. П. Бурдаєв, Г. А. Гайна, В. Г. Доля, О. В. Зубов).

Сучасні системи здійснюють переклад тексту з урахуванням морфологічних, синтаксичних та семантичних даних, втім якість перекладу не завжди є високою через граматичні та лексичні помилки. Звичайно постає питання як саме класифікувати помилки та визначити наскільки коректним є переклад. Але і тут науковці стикаються з проблемою, оскільки переклад розцінюють згідно з його адекватністю та еквівалентністю, а трактування цих термінів є досить розмитим.

Серед інших можна виділити наступні трактування поняття адекватності перекладу: 1) близькість оцінок змісту текстів їхніми адресатами або відповідність поставленій перед перекладачем меті [9]; 2) відповідність перекладу як процесу певним комунікативним умовам [11]; 3) відтворення змісту і форми оригіналу засобами іншої мови, що є метою художнього перекладу [10].

Те ж саме стосуються і поняття еквівалентності, поміж низки інших є такі визначення цього поняття: 1) відповідність тексту перекладу оригіналу [11]; 2) смислова спільність прирівнюваних один до одного одиниць мови й мовлення [10]; 3) збереження відносної рівності змістової, смислової, семантичної, стилістичної і функціонально-комунікативної інформації, що міститься в оригіналі та перекладі [11].

Звичайно потрібно не забувати, що як еквівалентність, так і адекватність є поняттями відносними. При оцінюванні машинного перекладу не потрібно ставити йому ті ж самі вимоги, що і професійному (тобто виконаному фахівцем) перекладу, оскільки його мета інша. Зважаючи на сучасну глобалізацію, машинний переклад надає нам можливість приблизно ознайомитися з необхідною нам інформацією, що доступна лише на іншомовному ресурсі. Так, Ю.М. Марчук називає машинний переклад «чорновим варіантом», це досить вдале зауваження, тому що його мета полягає не у наданні «ідеального» літературного перекладу, а у передачі головної думки [4].

У своїй роботі ми вирішили взяти три безкоштовні загально доступні системи машинного перекладу – *Google*, *Yandex* та *Pragma* та перевірити їх на наявність граматичних та лексичних помилок при перекладі з англійської на українську мову. Задля цього ми використали порівняння різних наявних методик, зокрема аналіз за типом допущеної помилки та зіставлення варіантів перекладу для виявлення його адекватності та коректності. Для пошуку найкращого відповідника також були застосовані елементи лексикографічного, компонентного та трансформаційного аналізів. Таким чином, вказане поєднання методик відповідає різноплановим завданням, поставленим у роботі.

Google Перекладач, або **Google Translate** – сервіс компанії Google, що дозволяє автоматично перекладати слова, фрази та web-сторінки з однієї мови на іншу. Google використовує власне програмне забезпечення для перекладу. Його українська версія з'явилася у 2008 на 11 стадії розвитку сервісу. На жаль, не всі мови мають прямий переклад. Деякі мови перекладені на англійську через проміжну мову (яка у всіх випадках тісно пов'язана з потрібною мовою, але більш широко поширена). Таких мов є 8, серед них українська. У випадку української мови проміжною є російська [1].

Система **Google Перекладач** має функцію перекладу веб сторінки цілком, також **Google Перекладач** ніколи не позиціонував себе як ресурс для перекладачів, а його розробники завжди акцентували увагу на тому, що він може допомогти читачеві зрозуміти загальний зміст тексту іноземною мовою, але не надає точних перекладів.

Яндекс. Перекладач – сервіс компанії Яндекс, призначений для перекладу частини тексту або сторінки на іншу мову. Перекладач вбудовано у Яндекс. Браузер, тож він автоматично пропонує перекласти текст, написаний іноземною мовою. Тестова версія програми з'явилася ще у 2011, тоді для перекладу було доступно 3 мови, а саме англійська, українська та російська. Станом на листопад 2016 система працює з більше ніж 47 мовами [6].

Pragma – це система перекладу, що існує у двох версіях. Електронний словник, який не став предметом нашого дослідження з двох важливих причин, по-перше його не можна віднести до систем, по-друге він не є безкоштовним. Онлайн версія, яку ми розглядаємо у своїй роботі, є безкоштовною та вважається системою машинного перекладу [8].

За основу дослідження було покладено групи типових помилок систем машинного перекладу. У довільному порядку ми виділили декілька речень різних за стилем та галузями, які потенційно містять складну для машинного перекладу лексику, та перевірили, чи були помилки при їх перекладі українською.

1. Неправильне вирішення проблем лексичної полісемії

*Revised wing has a 25% larger area, the **spread** is increased by 4.88 m, which increased fuel capacity by 30% [12].*

Системи Google Перекладач та Yandex вибрали варіант **спред**, хоча розглянувши варіанти перекладу слова без контексту у системі ми виявили правильний відповідник **розмах**, при цьому система Yandex оперує більш широким словниковим банком, що налічує 31 варіант перекладу лексеми, в той час як Google Перекладач – 21.

Система Pragma припустилася, такої ж помилки, а також подала ще два неправильні відповідники: *area* – **площа**, а не **область**, а *spread* – **поширення**, а не **розмах**.

Адекватним варіантом перекладу могло б стати речення: *Перероблене крило має на 25% більшу площу, розмах збільшений на 4,88 м, що дозволило збільшити запас палива на 30%.*

*It currently **mines** iron ore from its one open-pit quartzite field through a process of drilling and blasting, and by the removal of overburden to external dumps [13].*

Під час перекладу цього речення система Pragma переклала *mines* – **мінує**, хоча правильним лексичним відповідником є **видобуває**. Розглянувши варіанти перекладу слова без контексту у системі, ми отримали той самий переклад – **мінує**.

2. Неправильне вирішення проблеми лексично-граматичної омонімії

Згадане вище речення *It currently **mines** iron ore from its one open-pit quartzite field through a process of drilling and blasting, and by the removal of overburden to external dumps [13].*

Програма Google Translator неправильно визначила частиномовну належність дієслова *mines* та переклала його іменником **шахти**. Правильний варіант – **видобуває**.

Система Yandex не допустила помилки та переклала термін правильно – **видобуває**. Система не тільки переклала термін правильно, а й не здійснила інших помилок.

Адекватний переклад речення: *В даний час вона видобуває залізну руду з одного кар'єрного поля кварцитів за допомогою буріння та підривання, а також шляхом перенесення покривних порід на зовнішні відвали.*

3. Використання русизмів

Ця проблема пояснюється використанням російської мови як проміжної ланки при перекладі з англійської мови на українську.

*Carpet is a common name for the group textile floor coverings, consisting tufted carpets and **needlepunched carpets** [16].*

Система Google переклала **needlepunched carpets** – *іглопробівніє килими*. Правильний варіант *голкопробивні килими*.

У перекладі Yandex помилок не виявлено.

Коректним перекладом є: *Килим – це загальна назва для групи текстильних покриттів для підлогових, вони бувають тафтингові та голкопробивні.*

4. Дослівний переклад словосполучень

*Under the Charter, the Security Council has **primary responsibility** for the maintenance of international peace and security [15].*

Системи Google Translator та Yandex виконали адекватний переклад: *Відповідно до Статуту Рада Безпеки несе головну відповідальність за підтримання міжнародного миру та безпеки.*

Pragma запропонувала варіант перекладу фрази **primary responsibility**, як *первинна відповідальність*, але правильний переклад – *головна відповідальність*.

Також при перекладі речення *Carpet is a common name for the group textile floor coverings, consisting tufted carpets and **needlepunched carpets*** Pragma допустила помилку цього типу, запропонувавши переклад *голчасті ударні килими*. Влучний варіант перекладу подано вище.

5. Переклад вузькоспеціальних, непоширених аббревіатур

Проблема відсутності перекладу терміну є актуальною, адже цьому сприяють два чинники: швидкий розвиток галузей і відповідно поява нових термінів, а також не достатньо швидке впровадження нових термінів.

It would be able to produce coal at an average production level of 37.1 mmtpy for 23 years [14].

Google та Pragma не дали перекладу аббревіатури залишивши її *mtpry*, хоча відповідником є *млн тон за рік*.

Yandex при перекладі зробив іншу вельми неочікувану помилку переклавши *mtpry*, як *буде потрібно близько*.

Адекватний переклад речення: *Через 23 роки він зможе виробляти в середньому по 37.1 млн тон вугілля за рік*.

Розглянувши типові помилки на прикладах, ми дійшли висновку, що жодна систем не є досконалою на даному етапі, і всі вони потребують доопрацювання.

Одна з проблем до розгляду у роботі була: як позбавитися таких помилок? Провівши дослідження, ми схильні вважати, що поки це не можливо через відсутність алгоритму аналізу та програмування, але розробники систем роблять все можливе, аби їх мінімізувати. Так, наприклад, Google Перекладач пропонує користувачу надавати свої варіанти перекладу та вказувати на помилки.

Підсумовуючи результати, ми вважаємо, що переклад, який надають нам системи, можна використовувати лише при бажанні ознайомитися з інформацією, що доступна виключно іноземною мовою. Також ми дійшли висновку, що цей переклад не варто використовувати як повноцінний, навіть розробники систем акцентують на цьому увагу. Для того, щоб перетворити отриманий переклад на якісний та адекватний, потрібно виконати постредагування, виправити лексичні та граматичні помилки.

Хоча у роботі ми не ставили собі за мету визначити, яка з програм робить кращий переклад, а лише мали на меті показати типові помилки систем машинного перекладу, втім аналіз адекватності та коректності запропонованих одиниць різними системами машинного перекладу вказує, що: 1) програми, які використовують близькоспоріднену проміжну мову (напр. Yandex), роблять найменшу кількість помилок лексико-граматичного та лексико-семантичного характеру; 2) програми, що не використовують близькоспоріднену мову (напр.

Pragma), видають більше помилкових варіантів при перекладі; 3) можливе існування, але потребує подальшого детального дослідження, зв'язку між розширеністю лексичної бази та адекватністю перекладу.

Дана робота окреслює перспективи майбутніх досліджень у цій сфері. Доцільною видається подальша співпраця між розробниками програмного забезпечення та спеціалістами у галузі прикладної лінгвістики для внесення запропонованих варіантів перекладу до систем задля покращення якості перекладу.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Гугл Перекладач [Електронний ресурс]. – Режим доступу : <https://uk.wikipedia.org/wiki/Google-перекладач>
2. Зубов А. В. Основы искусственного интеллекта для лингвистов : [учеб. пособие] / А. В. Зубов, И. И. Зубова. – М. : Университетская книга; Логос, 2007. – 320 с.
3. Кочерган М. П. Загальне мовознавство / Михайло Петрович Кочерган. – К. : ВЦ “Академія”, 2008. – 464 с.
4. Марчук Ю.Н. Проблемы машинного перевода / Юрий Николаевич Марчук. – М. : Наука, 1983. — 232 с.
5. Новиков В. А. Трансфер в современных системах машинного перевода : дис. ... канд. филол. наук: 10.12.21 / Виктор Алексеевич Новиков. – М., 2000. — 194 с.
6. О сервисе Яндекс. Перевод [Електронний ресурс]. – Режим доступу : <https://translate.yandex.ua/help>.
7. Особливості машинного перекладу [Електронний ресурс]. – Режим доступу : <http://ukrbukva.net/page,4,116910-Osobennosti-mashinnogo-perevoda.html>
8. Перекладач Pragma 6.x [Електронний ресурс]. – Режим доступу : <http://translate.ua/uk/pragma-6x?showall=1>

9. Селіванова О. О. Сучасна лінгвістика : напрями та проблеми : [підручник] / Олена Олександрівна Селіванова. – Полтава : Довкілля-К, 2008. –712 с.
10. Стахмич Ю. С. Адекватність та еквівалентність перекладу в контексті комп'ютерної лінгвістики / Ю. С. Стахмич // Вісник Житомирського державного університету імені Івана Франка : [наук. журн.] / [відп. ред. Н. А. Сейко]. – № 66. –Житомир : Вид-во ЖДУ ім. І. Франка, 2013. – С. 235-238.
11. Швейцер А. Д. Теория перевода : Статус, проблемы, аспекты / Александр Давидович Швейцер. – М. : Наука, 1988. — 216 с.

ДЖЕРЕЛА ІЛЮСТРАТИВНОГО МАТЕРІАЛУ

12. In the future along with Boeing 737-800 NG [Електронний ресурс]. – Режим доступу : <http://www.utair.ua/en/news/4265783.html>.
13. Mining Division [Електронний ресурс]. – Режим доступу : http://www.metinvestholding.com/en/activity/raw_iron.
14. The relationship between.. [Електронний ресурс]. – Режим доступу : https://books.google.com.ua/books?id=ufsxAQAAMAAJ&pg=PA81&lpg=PA81&dq=would+be+able+to+produce+coal+at+an+average+production+level+of+37.1+mmtpy+for+23+years.&source=bl&ots=b4A_3nhVJx&sig=cui0tCit0v9q9f3SKK886QIzfzo&hl=uk&sa=X&ved=0ahUKEwiF1N-I28DQAhVFkiwKHQXyA_AQ6AEIGDAA#v=onepage&q=would%20be%20able%20to%20produce%20coal%20at%20an%20average%20production%20level%20of%2037.1%20mmtpy%20for%2023%20years.&f=false
15. The UN Safety [Електронний ресурс]. – Режим доступу : <http://www.un.org/en/peacekeeping/operations/rolesc.shtm>
16. What is the difference... [Електронний ресурс]. – Режим доступу : http://carpets.sintelon.rs/What_Is_the_Difference_between_Tufted_and_Needlepunched_Carpets-27-1-72