

ПРОБЛЕМИ АВТОМАТИЗОВАНОГО ПЕРЕКЛАДУ

Статтю присвячено розгляду механізмів, за допомогою яких системи автоматизованого перекладу здійснюють переклад, і проблемам, з якими стикаються для досягнення функціонально-прагматичної адекватності. З отриманих результатів можна зробити висновок, що причинами неадекватності перекладу виявилися помилки на синтаксичному, фразеологічному, лексико-семантичному і морфологічному рівнях мови.

Ключові слова: автоматизований переклад, функціонально-прагматична адекватність, синтаксичний, фразеологічний, лексико-семантичний і морфологічний рівні мови.

Статья посвящена исследованию механизмов, при помощи которых системы автоматизированного перевода осуществляют перевод, и проблемам, с которыми сталкиваются для достижения функционально-прагматической адекватности. Исходя из полученных результатов можно сделать вывод, что причинами неадекватности перевода послужили ошибки на синтаксическом, фразеологическом, лексико-семантическом и морфологическом уровнях языка.

Ключевые слова: автоматизированный перевод, функционально-прагматическая адекватность, синтаксический, фразеологический, лексико-семантический и морфологический уровни языка.

The article is focused on the study of the automatic machine translation systems performance mechanism and the challenges arising in achieving functional and pragmatic adequacy. The results of the research have revealed that inadequate translation has been caused by certain errors on the syntactic, phraseological, lexical-semantic and morphological language levels.

Key words: the automatic machine translation, functional and pragmatic adequacy, syntactic, phraseological, lexical-semantic and morphological levels.

Характерними тенденціями розвитку сучасного суспільства є посилення процесів глобалізації та інформатизації, що призводять до збільшення й спрощення інтернаціональних зв'язків і контактів. Єдиною перешкодою на цьому шляху залишається мовний бар'єр,

подолання якого за допомогою традиційного перекладу не розв'язує проблеми повністю, оскільки обсяг перекладів у світі постійно зростає, і навіть збільшення кількості перекладачів не в змозі повністю задовольнити потреби в перекладах.

Однією з можливостей подолання мовного бар'єра у спілкуванні є застосування систем автоматизованого перекладу (САП) як засобу підвищення продуктивності перекладацької праці [1].

Спроби оцінки якості автоматизованого перекладу мають місце у сучасних перекладознавчих дослідженнях у роботах Л. О. Андрієнко [1], М. С. Блехмана [2], В. К. Войнова, Ю. Н. Марчук, З. М. Шалляпина, проте слід констатувати, що всі вони містять різні методологічні підходи до питання оцінки якості або такі підходи відсутні взагалі.

Об'єктом дослідження є системи автоматизованого перекладу “Pragma 6” та “GoogleTranslate”, що вивчаються, досліджуються й оцінюються, з урахуванням інформації щодо алгоритмів роботи систем автоматизованого перекладу.

Предметом дослідження є рівень адекватності перекладів, виконаних за допомогою систем автоматизованого перекладу “Pragma 6” та “GoogleTranslate-online”.

Матеріалом дослідження слугували укладені власноруч тестові збірки речень з максимальною кількістю мовних конструкцій, які, ймовірно, становлять складності для САП. Усього проаналізовано 300 речень (з науково-технічної статті “UKRAINE SPATIAL: LAUNCHES INTO FUTURE”, узятої з сайту www.bbc.com) та варіанти їхнього перекладу на українську мову, зроблені професійним перекладачем та системами автоматизованого перекладу **Google Translate** та **Pragma 6**.

На сучасному етапі розвитку САП визначають два напрямки:

- 1) створення САП з опорою на використання універсальної мови;
- 2) створення електронних словників лінгвістичних одиниць, а також автоматизованих термінологічних банків даних.

Вони характеризуються деяким поєднанням результатів двох підходів на рівні нових моделей, тобто моделювання дій людини-перекладача і створення як великих за обсягом словників загальнолексичного значення, так і специфічних словників субмов [3].

У зв'язку з тим, що онлайн перекладачі Google Translate Beta компанії Google і СМП Pragma 6.x компанії Trident Software на сьогодні найбільш відомі і використовують найсучасніші словники і

технології перекладів, ми будемо використовувати для дослідження онлайн версії даних перекладачів, що знаходяться відповідно на сайтах <http://translate.google.com> і <http://www.trident.com.ua/ukr/online.php>.

Системи автоматизованого перекладу Google Translate та Pragma 6. забезпечують повний цикл перекладу всього документа: введення початкового тексту, переклад на іншу мову, редагування, форматування й збереження перекладеного тексту.

Слід зазначити, що якщо розглядати процес перекладу з однієї природної мови на іншу як вид мовленнєвої діяльності, то ми можемо встановити такі відповідності у функціях, які виконує САП і перекладач:

- лінгвістичні проблеми тексту розв'язує перекладач;
- моделювання перекладу на рівні підбору еквівалентів зі словника виконує САП;
- припустимий ступінь семантичної оцінки точності описів понять комунікативної спрямованості перекладу виконується перекладачем [74].

Як одиниця перекладу в даному дослідженні використовується *текст*. Він розглядається як набір елементів абстрактної множини M , між якими встановлені деякі види відносин. У текстах на природних мовах ці відношення між елементами виражають закодовану інформацію, закладену в текст його творцем.

У процесі розуміння тексту ми явно чи неявно встановлюємо додаткові значеннєві відношення між словами, а також вносимо в текст нові елементи. Тим самим у процесі розуміння (електронного аналізу) утвориться новий текст T_2 , що містить множину M_1 , що складається з множини M тексту 1 і додаткових множин декодованих відношень перекладу (керування, узгодження, однорідність, еквівалентність).

Необхідно відзначити певну відмінність між цими текстами: перший текст (T_1) є текстом, даним у безпосередньому спостереженні, а другий (T_2) — деяка конструкція, що описує процес розуміння і процес утворення синтаксичних зв'язків. Синтаксична схема тексту T_1 визначає структуру синтаксичних відносин вихідного тексту T_1 , що в кінцевому тексті T_2 не виражені. Синтаксична структура відбиває те загальне, що є в однаково синтаксично побудованих текстах. Тому текст T_1 можна вважати рівним тексту перекладу T_2 на іншій природній мові за умови, коли дотримано такі положення:

- синтаксичні конструкції цих текстів відповідають одна одній чи принаймні збігаються в деяких істотних частинах;
- задана множина елементів алфавіту, складових речень тексту T1 у заданій синтаксичній схемі мають чітку відповідність із заданими елементами множини тексту T2;
- текст T1 є продуктом породження тексту T2.

Умовою породження еквівалентної синтаксичної схеми між текстами T1 і T2 є еквівалентність окремих частин цих схем (наприклад, підмет у вихідному тексті повинен відповідати підмету в тексті перекладу) і загальна значеннева еквівалентність, обумовлена як сумарна відповідність тексту перекладу стосовно тексту оригіналу. Таким чином, перше завдання перекладу — встановити схему зв'язків, за допомогою яких можна створити синтаксичну схему, еквівалентну вихідній синтаксичній схемі. При цьому необхідно уточнити, що в даному випадку під синтаксисом ми розуміємо не тільки формальне проходження елементів у просторі, але й значеннєві відношення слів одне до одного, тобто відношення залежностей. У цьому разі мовою комп'ютера під текстом слід розуміти таке. Текст є деяким набором формальних символів, що несуть певний значеннєвий зміст.

Складність визначення єдиних методів оцінки якості перекладу полягає у першу чергу в тому, що при оцінці, як і при будь-якій іншій складній людській діяльності, до єдиної шкали маємо вводити досить різні фактори: повноту, точність, адекватність перекладеного тексту; мету перекладу, тип тексту; форму та умови здійснення перекладу (усний, письмовий, синхронний); напрямок; об'єм тексту, його складність.

Слід окремо зазначити, що будь-яка якісна оцінка перекладу ґрунтується на оцінці стосовно якоїсь норми. В Україні використовується Стандарт асоціації перекладачів України, який містить вимоги щодо показників якості перекладацьких послуг. Зокрема, показником якості перекладу, згідно з цією нормою, буде його відповідність вихідному тексту за змістом, смислом, стилістикою та оформленням. Переклад не повинен мати граматичних помилок, має бути повним та адекватним, термінологія та скорочення повинні відповідати галузевій належності вихідного тексту [4].

Все це здебільшого стосується перекладу традиційного. Щодо машинного перекладу, то при оцінці якості кінцевого продукту на перший план виходить його прийнятність для користувача [1]. Традицій-

но під цим розуміється продукт, прийнятний за якістю з точки зору певного інформаційного завдання.

Як свідчить аналіз теоретичних робіт з перекладу [5], традиційно оцінка якості перекладу являла собою процес порівняння тексту оригіналу та перекладу. Результатом цього порівняння була констатація ступеня адекватності: адекватний, вільний переклад, нормативний/ненормативний, підрядник.

Щодо адекватності АП, то ми керувались підходом, запропонованим А. Д. Швейцером та Л. С. Бархударовим [6], які акцентують увагу саме на функціонально-прагматичному аспекті даної категорії. При виявленні ступеня адекватності машинного перекладу на перший план виходить прагматичний фактор орієнтації на реципієнта перекладу та виконання поставлених завдань. Таким чином, адекватність АП визначається нами як адекватність відтворення основної прагматичної функції оригіналу.

У практиці машинного перекладу критерій зрозумілості є найбажанішим, оскільки надає вичерпну інформацію щодо змісту висловлювання. Зрозумілість тексту визначається його прозорістю та чіткістю для середньостатистичного читача з обмеженими фаховими знаннями без доступу до оригіналу на основі безпосередньо отриманої інформації. Методика визначення вищезазначеного полягає в підрахунку правильних та хибних відповідників у тексті машинного перекладу та їх кількісному порівнянні. Аналіз та класифікація помилок відбувалась на основі лінгвістичної типології. На *лінгвістичному* рівні, який полягає в підрахунку та класифікації помилок, був досліджений характер подібності оригіналу в лексичному, граматичному, синтаксичному і семантичному планах.

Інструментом для категоризації помилок при автоматизованому перекладі текстів науково-технічного напрямку у даному дослідженні є тестові збірки, за допомогою яких можна дослідити, які з лінгвістичних феноменів успішно опановуються системою та які призводять до помилок. Збірки складаються з складносурядних, складнопідрядних і простих не поширених речень цільової мови, що містять сталі словосполучення, фразеологічні одиниці, інфінітивні конструкції та інфінітивні звороти та ін. (герундій, модальні дієслова) для тестування систем машинного перекладу, кожна з яких містить певний мовний феномен, набір феноменів або взагалі певні мовні галузі для перевірки лексичних, синтаксичних та семантичних можливостей системи. Результати представлені в таб. 1

**Кількість позитивних та негативних перекладів за допомогою
САП Google-on-line порівнянню з Pragma 6**

| Мовна одиниця, що була проаналізована | Кількість речень | Google on line | | Pragma 6 | |
|---------------------------------------|------------------|----------------|----------|----------|---------|
| | | + | – | + | – |
| Сталі словосполучення | 26 реч. | 3 реч. | 23 реч. | 4 реч. | 22 реч. |
| Прийменники | 31 реч. | 8 реч. | 23 реч. | 6 реч. | 25 реч. |
| Фразеологічні одиниці | 30 реч. | 4 реч. | 26 реч. | 10 реч. | 20 реч. |
| Керування і узгодження | 19 реч. | 5 реч. | 14 реч. | 4 реч. | 15 реч. |
| Інфінітив | 20 реч. | 3 реч. | 17 реч. | 4 реч. | 16 реч. |
| Інфінітивні звороти | 24 реч. | 4 реч. | 20 реч. | 4 реч. | 20 реч. |
| Складносурядні речення | 17 реч. | 16 реч. | 1 реч. | 17 реч. | 0 реч. |
| Складнопідрядні речення | 24 реч. | 6 реч. | 18 реч. | 6 реч. | 18 реч. |
| Прості непоширені речення | 25 реч. | 25 реч. | 0 реч. | 25 реч. | 0 реч. |
| Часові форми дієслова | 29 реч. | 9 реч. | 20 реч. | 9 реч. | 20 реч. |
| Сполучуваність слів | 13 реч. | 3 реч. | 10 реч. | 3 реч. | 10 реч. |
| Питальні речення | 4 реч. | 1 реч. | 3 реч. | 1 реч. | 3 реч. |
| Займенники | 8 реч. | 2 реч. | 6 реч. | 1 реч. | 7 реч. |
| Інше (герундій, модальні дієслова) | 30 реч. | 4 реч. | 26 реч. | 2 реч. | 28 реч. |
| Всього | 300 речень | 93 реч. | 207 реч. | 96 | 204 |
| | 100 % | 31 % | 69 % | 32 % | 68 % |

Розглянемо випадки допущення помилок комп'ютером при перекладі текстів науково-технічного характеру на усіх рівнях мови для виявлення рівня семантико-синтаксичної адекватності перекладу. Помилки на даних мовних рівнях можуть допускатися або при розумінні вихідного тексту або при викладі тексту перекладу. Визначати мовні рівні, на яких допущена помилка, ми можемо тільки за проявами цих помилок у тексті перекладу. А саме — при викладі тексту перекладу. При цьому слід зазначити, що допущення помилки при розумінні вихідного тексту обов'язково тягне за собою допущення помилки при викладі тексту перекладу. Таким чином, виявлення помилки на будь-якому мовному рівні в тексті перекладу є наслідком допущення її на цьому ж мовному рівні в процесі перекладу (розуміння тексту на одній мові й подальшого його викладу на іншій). Виходячи з вищесказаного, визначимо прояви та можливі причини допущення помилок при перекладі на всіх рівнях мови.

Помилка на **морфологічному** рівні допускається у випадках, коли комп'ютер не справляється з визначенням морфем при віртуальному членуванні слова. Ці помилки є типовими, при цьому комп'ютер або переносить слово в його оригінальному написанні в текст перекладу, або транслітерує:

“In a postindustrial society information becomes the basic industrial resource where as in “pre-industrial” and “industrial” societies the main resources were raw materials and energy respectively.”

“Інформація postindustrial суспільства стає основним індустриальним ресурсом, моді як в “pre-industrial” і “індустріальний” суспільство головний ресурси були сирими матеріалами і енергією відповідно” (Pragma 6).

Скорочення-аббревіатури “*GMES*” і “*ГМЕС*” в реченні “*This motto is being materialized through the intergovernmental initiative GEOSS and its European component GMES (Global Monitoring for Environment and Security)*” транслітеруються:

“Цей девіз у цей час відбулося через міжурядові ініціативи ГЕОСС та її європейські компонент ГМЕС (Глобальний моніторинг в інтересах охорони навколишнього середовища та безпеки)” (Google Translate).

Помилка на **словотворчому** рівні може допускатися в тих випадках, коли комп'ютер не може визначити, з яких саме морфем складається слово, або підібрати потрібні морфемі для утворення слова, при цьому слово не узгоджується з іншими в першу чергу граматично. Наприклад, при перекладі речення “*The new program is also supposed to yield significant social effects*” Google Translate не впорався з підбором правильної форми слова “*повинен*” (“*Нова програма також повинен принести значні соціальні наслідки.*”), яке є присудком і не узгоджується з підметом. Правильна форма цього слова — “*повинна*”.

Помилки на **лексико-семантичному** рівні припускаються при нерозумінні значення слова комп'ютером. При даному виді помилки полісемічне слово має правильну граматичну форму, але не вписується в зміст найближчого контексту. Так, наприклад, при перекладі речення “*Innovative development is not only a technological issue*” Pragma 6 переклала слово “*issue*” як “*результат*” (*Інноваційний розвиток — не лише технологічний результат*), хоча правильний варіант перекладу — “*проблема*” (“*Інноваційний розвиток — це не лише технологічна, а й соціальна проблема*”).

Помилка на **фразеологічному** рівні виникає, коли комп'ютер не може підібрати відповідний фразеологічний еквівалент через відсут-

ність такого в базі даних. При цьому він не здатен ні визначити, ні правильно відтворити фразовий еквівалент. Тоді весь вислів має неправильний зміст і часто не вписується в контекст пропозиції. Відносини слів усередині фрази можуть бути правильними: фраза “у світанку космічної ери” при перекладі Pragma 6 “*У світанку космічної ери, цієї ролі навіть не передбачалася*” англійського речення “*At the dawn of the space era, this role was not even forecast.*” правильно перекладається як “на початку космічної ери”, як це переклала людина (Зуважимо, що така роль космонавтики не була передбачена на початку космічної ери), або “на зорі космічної ери”, як це переклав Google Translate (“На зорі космічної ери, ця роль була навіть не прогноз”).

Помилка на **синтаксичному** рівні може бути викликана не розумінням смислу речення, не розумінням відносин елементів більш низьких рівнів, змісту всього тексту, предмета тексту і т.д. Наприклад, Pragma 6 при перекладі речення “*What are the priorities of Ukraine’s astronautics?*” граматично правильно побудувала речення при перекладі, при цьому вірно підібравши словесні і фразові еквіваленти: “*Чим є пріоритети української астронавтики?*”. Однак синтаксис англійського речення не відповідає варіанту переклада, що є суттєвим порушенням адекватності.

Таким чином, при вищенаведеній статистиці перекладу речень текстів причинами неадекватності перекладу речень он-лайн перекладачем Google Translate з’явилися помилки на синтаксичному, фразеологічному, лексико-семантичному, словотворчому і морфологічному рівнях мови. Результати можна побачити у таблиці 2.

При цьому варто враховувати, що, виходячи з теорії ієрархії рівнів мови, ми не враховували помилки на рівнях мови, якщо в складі елемента рівня вже присутня помилка на більш низьких рівнях мови.

Таблиця 2

Кількість помилок, допущених при перекладі за допомогою Google-on-line порівнянно з Pragma 6

| Рівні мовлення | Кількість помилок, допущених при перекладі за допомогою | |
|---------------------|---|----------|
| | Google-on-line | Pragma 6 |
| Синтаксичний | 66 | 80 |
| Фразеологічний | 106 | 120 |
| Лексико-семантичний | 146 | 106 |
| Словотворчий | 54 | 146 |
| Морфологічний | 40 | 26 |

Таким чином, в одному реченні максимальна кількість помилок на рівні мови може бути не більше кількості елементів на цьому ж рівні. Наприклад, тільки одна синтаксична помилка в реченні, яку ми врахували, якщо в реченні не знайдено помилок на більш низьких рівнях мови.

У ході аналізу було помічено, що Pragma 6 частіше формує речення, правильні з позиції граматики, хоча іноді це є причиною збільшення ступеня дезінформації користувача через допущені помилки на більш низьких рівнях мови.

З отриманих результатів можна зробити висновок, що якість перекладу обох систем приблизно однакова і, якщо цей текст призначено для розуміння тематики та головної інформації, то можна стверджувати, що обидва перекладача продемонстрували досягнення функціонально-прагматичної адекватності на рівні приблизно 50 % Google Translate та 45 % Pragma 6 (виходячи з того, що саме речення із помилками, віднесеними до типу спотворення, суттєво впливають на рівень розуміння).

Google Translate частіше підбирає правильні лексичні і фразеологічні еквіваленти, хоча гірше справляється з побудовою граматично правильних речень. Це можна пояснити тим, що в базі даних цієї системи більша кількість англо-українських відповідників.

У ході аналізу було помічено, що Pragma 6 частіше формує речення, правильні з позиції граматики, хоча іноді це є причиною збільшення ступеня дезінформації користувача через допущені помилки на більш низьких рівнях мови.

Google Translate частіше підбирає правильні лексичні і фразеологічні еквіваленти, хоча гірше справляється з побудовою граматично правильних речень. Це можна пояснити тим, що в базі даних цієї системи більша кількість англо-українських відповідників.

Також те, що контекст статистичної системи Google Translate становить лише 5 слів, пояснює складність побудови системою граматично правильних речень. Також необхідно відзначити той факт, що при перекладі довгих речень, якщо Google Translate допустив помилку на початку, він може не допустити помилку в кінці, причому з більшою ймовірністю, ніж Pragma 6, яка в таких випадках спотворює все речення, намагаючись узгодити першу частину з помилкою з другою частиною. При цьому частіше всього помилка переноситься також і на другу частину.

Підрахунок правильних та хибних відповідників тестової збірки речень, перекладених за допомогою систем автоматизованого пере-

кладу, у їх кількісному порівнянні виявив, що найбільшу кількість помилок обидві досліджувані системи роблять відносно перекладу таких синтаксичних структур: складнопідрядні речення, а також неособисті форми дієслова у функції доповнення та обставини та відповідних інфінітивних, причастних і герундіальних оборотів.

Безсумнівно, причиною більшості помилок стало нерозуміння тексту системою АП, так як розуміння вихідного тексту є первинним етапом у процесі перекладу, а, відповідно, й основою правильного викладу його на іншій мові. Оскільки людина має матеріал, який дозволяє адекватно викладати смисловий зміст певною мовою, наприклад, в першу чергу граматику, і взагалі будь-яка інформація про мову, керуючись якою, людина правильно формує свої повідомлення, володіючи інформацією, яку їй потрібно повідомити, то цілком очевидно, що проблема перекладу полягає в першу чергу у складності створення такої інформаційної системи, щоб вона могла витягти з тексту оригіналу хоча б приблизно стільки ж, скільки отримує з нього людина.

До найбільш типових помилок, характерних для автоматизованого перекладу, можна віднести такі: неправильний переклад прийменника; неправильний переклад багатозначного іменника, дослівний переклад ідіоматичного виразу; неправильне визначення частини мови, неправильний переклад дієслівних форм, прийменників, зворотів.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Андрієнко Л. О. Проблеми розвитку машинного перекладу на сучасному етапі // Гуманітарний вісник. Серія: Іноземна філологія: — Черкаси: ЧДТУ, 2005. — Число дев'ять. — С. 348–351.
2. Блехман М. С. Комп'ютерна лінгвістика та системи “розуміння” тексту [WWW document]. URL http://www.franko.lviv.ua/ujcl/2_2st.htm.
3. Каничев М. Шаг в автоматизации перевода // Мир ПК. — 2001. — № 10. — С. 84–87.
4. Стандарт Асоціації перекладачів України. Послуги з письмового та усного перекладу. Загальні правила і вимоги до надання послуг. — К.: СТТУ АПУ, 2000. — 37 с.
5. Миньяр-Белоручев Р. К. Теория и методы перевода / Рюрик Константинович Миньяр-Белоручев. — М.: Московский лицей, 1996. — 208 с.
6. Швейцер А. Д. Эквивалентность и адекватность перевода / А. Д. Швейцер // Тетради переводчика. — М.: Междунар. отношения, 1989. — Вып. 23. — С. 31–39.

Стаття надійшла до редакції 25.09.13