

**СИСТЕМА АВТОМАТИЧНОГО КОРЕГУВАННЯ АНГЛІЙСЬКО-УКРАЇНСЬКОГО КОМП'ЮТЕРНОГО ПЕРЕКЛАДУ ДЛЯ ТЕХНІЧНИХ ТЕКСТІВ В ГАЛУЗІ АВТОМАТИЗАЦІЇ ТЕХНОЛОГІЧНИХ ПРОЦЕСІВ**А.О. Стопакевич<sup>1</sup>, А.М. Тігарєв<sup>1</sup>, О.Р. Романюк<sup>1</sup>, О.А. Стопакевич<sup>2</sup>

---

<sup>1</sup>Державний університет інтелектуальних технологій та зв'язку  
1, Кузнечна, Одеса, 65029.<sup>2</sup>Національний університет «Одеська Політехніка»  
1, Шевченка пр. Одеса, 65044.  
email: stopakevich@gmail.com

---

Метою роботи є розробка системи автоматичного корегування перекладених комп'ютером текстів зі специфічною термінологією, що притаманна науковим та технічним текстам в галузі автоматизації технологічних процесів. Приведений аналіз причин чому комп'ютерні перекладачі не можуть досягти високої якості англо-українського перекладу технічних текстів в зазначеній галузі. Зроблено висновок, що в межах підходу, який використовують сучасні комп'ютерні перекладачі, якість таких перекладів покращена бути не може. Проведено аналіз досвіду корегування перекладів професійними перекладачами, наявних метрик оцінки процесу та результатів корегування комп'ютерних перекладів, доступних програмних рішень для роботи з текстами, написаними українською мовою. Зроблено висновок, що для комп'ютерного англо-українського перекладу єдиним практично значимим підходом до оцінки його якості є вимірювання кількості роботи, що необхідно виконати професійному перекладачеві для того, щоб текст задовольняв літературним нормам. Аналіз наукових текстів, які були перекладені DeepL показав, що кількість такої роботи може бути істотно зменшена, оскільки помилки, які робить цей перекладач мають систематичний характер. Таким чином, аналізуючи помилки, які робить комп'ютерний перекладач, можна сформулювати універсальні правила корекції для всіх перекладених певним перекладачем текстів галузі, які можуть виконуватись програмним застосунком автоматично. Ефективність підходу продемонстрована на прикладі розробки правил, що виходять з аналізу результатів перекладу двох наукових статей. Показано, що заміна приблизно 5% слів в комп'ютерному перекладі істотно підвищує його якість.

**Ключові слова:** корегування, комп'ютерний, переклад, автоматизація, технологічних, процесів, NLP, програма, python, термінологія, deepl.

**Вступ.** Українська термінологія в галузі автоматизації була сформована на базі радянської термінології, визначення якої звичайно фіксувались в державних стандартах та інших документах. Після завершення радянського періоду робота в напрямку розробки сучасної новітньої термінології в галузі автоматизації майже не ведеться. Більшість діючих ДСТУ в галузі автоматизації були введені на початку незалежності й фактично їх відмінність від радянських полягала тільки в тому, що вони були написані українською мовою. Таким чином в ДСТУ майже всі зафіксовані в радянських ГОСТ англomовні переклади термінів с залишилися без змін. Проблемою цих перекладів є те, що в сучасній англійській мові приведені відповідники не застосовуються, а застосовується своя достатньо специфічна термінологія. Крім того, ця термінологія, як і галузь у цілому, постійно розвивається і коли спеціалісти стикаються з новими термінами, то без чіткої мовної термінологічної бази це призводить до утворення англійського термінологічного “суржику”, розвиток якого активно спостерігається в середовищі українських ІТ-спеціалістів.

Наприклад, в [1] вся термінологія про автоматизовані системи подається в основному за ДСТУ 2226-93. Й, хоча, українські терміни відповідають еволюції української мови з періоду незалежності, в тому числі зафіксовані в стандарті освіти за 151 спеціальністю, їх приведені англійські відповідники значною мірою не використовуються в сучасній англійській мові.

Порівняно невелика кількість наявних в інтернеті точних перекладів між текстами за галуззю автоматизації українською та англійською мовою призводить до того, що якість перекладу текстів комп'ютерними перекладачами за цією галуззю є низькою. Причин тут три. Перша причина - відносно мала кількість наявних ідентичних перекладів текстів за галуззю (особливо це стосується літератури з теорії керування). Друга причина - наявність надмірної кількості варіантів написання терміну для певного поняття (наприклад, передаточна, передатна, передавальна функція), які зафіксовані в друкованій літературі та зміна нормативного тлумачення багатозначних термінів (наприклад, управління і керування). В результаті перекладачі використовують в тексті їх всі разом, розглядаючи як синонімічні. Третя причина - різниця між традиціями створення термінів в мовах, що призводить до того, що кількість слів в англійській мові менша за українську. Наприклад, ключовим словом для автоматичного керування є дієслово *to control*, який суміщає два поняття, які українською мовою мають різний сенс, а саме - "контролювати" та "регулювати". Узагальнена статистика по ряду перекладених нами текстів галузі автоматизації свідчить, що в символах тексти українською мовою займають в два рази більший обсяг, ніж їх оригінали англійською.

Приведені вище причини, що призводять до неякісного комп'ютерного перекладу, не перестануть бути актуальними в найближчий час, оскільки їх розв'язок полягає в формуванні істотної кількості якісних й термінологічно провірених перекладів, за якими зможуть навчитись коректному перекладу комп'ютерні перекладачі (КП). Сучасні КП, такі як Google Translate (GT) і Microsoft Translate (MT), виникли з появи інформаційної та розрахункової можливості емпірично вивести з наявних в інтернеті джерел правила перекладу. Надалі, статистичний підхід еволюціонував до навчання нейромереж, які намагаються при підборі відповідників врахувати певний сенсовий контекст блоку тексту.

Зворотною стороною всіх технологій, що виходять з обробки емпіричного масиву даних, є відсутність гарантії якості кінцевого результату. Результат кожного запиту є непередбачуваним для розробників системи й може змінитися після наступного перенавчання. А рішення щодо застосування нових результатів навчання приймається на базі статистичних досліджень змін поведінки. Виходячи з цього, комп'ютерні перекладачі не можуть, за умов збереження поточного підходу до їх побудови, повноцінно замінити людину, особливо в терені вузькопрофільних перекладів та перекладів юридичних та фінансових документів.

Незважаючи на зазначені недоліки, КП побудовані за вказаним підходом, стали щоденним інструментом багатьох людей. Не даючи професійний переклад, вони дають, звичайно, переклад достатній для розуміння загального сенсу тексту людиною, що часто є для споживача задовільним результатом.

Поява можливості для звичайних користувачів інтернету швидко та безкоштовно отримати переклад тексту, а для бізнесу перекладати необмежену кількість документів, призвела до появи нової сфери досліджень в галузі перекладів - дослідження процесу корегування комп'ютерних перекладів (ККП, РЕМТ, *Post-editing of machine translation output*). Найважливішою практичною задачею таких досліджень є пошук шляхів скорочення часу на корегування згенерованих КП перекладів виходячи зі спостережень за діями, що виконують зі згенерованим перекладом професійні перекладачі.

В цій роботі пропонується підхід до реалізації системи автоматичного корегування текстів зі специфічною термінологією галузі автоматизації. Працездатність підходу проілюстрована за допомогою розробленого програмного застосунку, що виконує токенізацію текстових блоків та заміну їх за правилами, які формуються емпіричним чином в результаті спостереження за помилками, які припускає при перекладі КТ. Як виявляється, ці помилки мають систематичний характер, а, отже, й можуть бути систематично усунуті.

**Мета роботи.** Сформувані з застосуванням сучасних технологій обробки тексту новий підхід до побудови системи автоматичного корегування комп'ютерних перекладів текстів наукової та технічної літератури в галузі автоматизації. Особливістю підходу є застосування доступних безкоштовних програмних NLP технологій. Підтвердити результативність сформованого підходу шляхом розробки та дослідження роботи програмного застосунку при використанні його для корегування перекладів наукових статей.

**Задачі роботи.** Проаналізувати досвід корегування перекладів безкоштовно доступних комп'ютерних перекладачів професійними перекладачами, викладений в сучасній науковій літературі. Зробити висновок про те, який перекладач доцільно взяти за основу для перекладів.

- 1) Проаналізувати наявні метрики оцінки процесу та результатів корегування комп'ютерних перекладів, зробити висновки про доцільність їх застосування при розв'язку поставленої задачі.
- 2) Проаналізувати доступні програмні рішення для перевірки та корекції текстів українською мовою.
- 3) Сформувані підхід до побудови системи автоматичного корегування комп'ютерних перекладів технічних та наукових текстів в галузі автоматизації та запропонувати робочий варіант його програмної реалізації.
- 4) Дослідити потенціал застосування підходу в розробленому програмному застосунку під час розв'язку реалістичних задач комп'ютерного перекладу текстів наукових статей в галузі автоматизації.

#### **Аналіз досвіду граматичного аналізу та корегування комп'ютерних перекладів професійними перекладачами.**

Історія сучасних КП починається з 2006 р., коли Google прийняв рішення розробляти свій спеціальний перекладач - Google Translate (GT). Перша версія перекладача застосувала суто статистичний підхід й результати перекладу були гіршими, ніж у поширених тоді перекладачів, що були побудовані за класичним принципом - за правилами. В 2010 р. після кількох років роботи над проектом, розробники перекладача досягли рівня задовільності перекладу, що став не гіршим за переклад інших розповсюджених тоді КТ, Досягши позитивного результату, Google інтегрує свій перекладач з браузером Chrome та Android.

В 2011 р. було проведено систематичне дослідження процесу перекладу з англійської на датську з застосуванням GT [2]. Групою з 8 датчан різних професійних здібностей блоки англійського тексту були перекладені як вручну, так і кореговані після застосування GT. Оцінені спеціалістами результати показали, що майже рівноймовірно, що найкращим перекладом буде або ручний або скорегований переклад. Час ручного перекладу та час корегування перекладу виявився майже однаковим. А якість кінцевого перекладу не була корельовано з часом корекції. Це пояснювалось тим, що перекладач може дати як прийнятних переклад, так й зовсім не прийнятний. Звичайно, як показують результати роботи [3] неприйнятний переклад професійні перекладачі витирають повністю й пишуть новий з чистого аркуша.

Принципове покращення GT, а згодом й інших аналогічних КП пройшло після заміни статичної технології навчання на технологію нейромереж. Приблизно

в 2016 році GT та Microsoft Translator (MT) проводять таку заміну й в них в 2017р. з'являється новий конкурент - DeepL (DL), який робить акцент на технологію глибинного навчання. В 2020 р. розробники DL презентували звіт що переклади, зроблені саме DL визнаються професійними перекладачами як найкращі в переважній кількості випадків в порівнянні з GL, ML та перекладачем Amazon. Додавання в вересні 2022 р. української мови в перелік доступним мов DL разом з більш природною та диференційованою результативною лексикою перекладу робить, на наш погляд, DL найкращим для англо-українського перекладу як технічних текстів, так й більш загальних.

В роботі [4] запропонована класифікація помилок КТ: “лексичні, синтаксичні, морфологічні, орфографічні, пунктуаційні та інші, вторинні, помилки прагматичного/культурного характеру, а також ті, що не можна чітко класифікувати за попередніми категоріями” й за цією класифікацією зроблено дослідження німецько-українського перекладу. Результати перекладу казки та публіцистичної статті далекі від ідеалу, хоча публіцистичний текст був перекладений дещо краще. Кількість помилок вимірюється десятками, при чому в них є такі, що повністю спотворюють сенс перекладу.

В роботі [5] розглянута задача англо-індонезійського перекладу анотацій наукових статей з використанням GT. Автори розділяють помилки перекладу на 13 категорій : граматичні помилки, термінологічні помилки, пропуски, синтаксичні помилки, помилки перекладу, помилки буквального/вірності, помилки вживання, пунктуаційні помилки, помилки додавання, двозначності, помилки словоформ, помилки написання з великої літери та орфографічні помилки. В роботі продемонстрована нездатність GT зрозуміти контекст всього тексту за межами речення, що призводить до грубих помилок. Аналогічне дослідження тайського-англійського перекладу анотацій проведено в пізнішій роботі [6], в якій автори приходять до висновку що переклад GT не відповідає стандартами академічного перекладу.

Результати показового групового експерименту 2019 р. щодо дослідження феномену професійної корегування комп'ютерного перекладу з англійської на польську мову приведені в роботі [7]. Автори провели статистично достовірний експеримент з групою магістрантів - професійних перекладачів. Була розглянута задача перекладу різних текстів загальним обсягом трудомісткості перекладу приблизно в 40 хв з англійської мови на польську. Розглядався як переклад з чистого тексту, так і корекція перекладу після того, як його згенерували GT, MT та DL. Новим результатом цього дослідження, в порівнянні з експериментом з роботи [2] 2011 р., стала чітка статистична залежність економії часу відносно повістю ручного перекладу від застосовуваного КТ : економія часу з DL склала ~ 25%, MT ~ 12, GT ~ 6 %.

Нами не знайдені результати досліджень процесу корегування перекладів технічних та наукових текстів. Проблема таких досліджень в тому, що “професійний перекладач” тут не може бути розглянута як достатня групова категорія. Якщо професійний викладач не володіє певною спеціальною темою та термінологією, то зв'язок між часом перекладу з “нуля” та часом перекладу після “перекладача” буде явно складнішим, як і результат такого корегування. Нами знайдені роботи, в яких проводиться статистичний аналіз розподілення різних типів помилок в комп'ютерному перекладі, однак такі результати не дуже цікаві, оскільки істотно відрізняються в залежності від направлення переведу та стилістики тексту. Значно більший практичний інтерес має виявлення систематичності помилок, які роблять комп'ютерні перекладачі, при певному напрямку переведу в певних умовах, однак таких досліджень не знайдено.

Підсумовуючи, зробимо висновок що застосування DL для розглянутої в статті задачі є найбільш доцільним, виходячи з зафіксованих результатів експериментів, які показали, що його застосування забезпечує найбільше прискорення корегування тексту. Це забезпечується за рахунок застосування власної архітектури нейронної мережі, що здатна самонавчатись не тільки за рахунок обробки загальнодоступних точних перекладів між різними мовами, але й на основі текстів, що не мають перекладів. Основною відмінною рисою DL від GT та MT є спроба виявлення та відтворення стилістики тексту. Перекладач може враховувати не тільки жанри, але й, наприклад, діалект мови. Так, на відміну від GT, DL може відрізнити британській та американські англійські мови, а також європейську португальську й бразильську португальську. Зворотною стороною такої переваги є й недоліки - DP має найменшу в порівнянні с GT та MT кількість підтримуваних мов та найменшу швидкість перекладу.

**Аналіз проблеми оцінки якості комп'ютерних перекладів та результатів їх корегування.** Кількість потенційно можливих перекладів одного тексту на іншу мову є достатньо великою. Множина однакових за сенсом перекладів тим не менш не може бути прийнята як повністю еквівалентна. Певне викладення може бути більш зрозумілим, а інше - більш лаконічним, одне - звучати більш сучасно, інше - більш офіційно. КП не формують переклад, намагаючись досягти подібні цілі, тому зараз критерій якісного комп'ютерного перекладу доцільно сформулювати лише в негативній постановці - як відсутність помилок перекладу (наприклад в межах класифікації з 13 пунктів роботи [8]). Проблема визначення метрик якості перекладу стискається з тим, що потрібно проводити порівняння з певним "вірним" перекладом або орієнтуватись на системи автоматичної перевірки граматики та стилю тексту. Перший підхід є цікавим для дослідження процесу перекладу та корегування перекладу при різних вхідних умовах. Такими чином, спостерігаючи за "трудовитратами" перекладачів ми робимо певні висновки. Другий підхід показує нам певні некоректності й дослідження полягає в оцінці їх серйозності та кількості необхідних заміни для того, щоб досягти безпомилковий переклад.

Приведемо найбільш поширені метрики оцінювання переведеного тексту (гіпотези) відносно еталону (ручного перекладу людиною). Найбільш популярною метрикою є BLEU, яку запропонувала фірма IBM. ідея якої полягає в розрахунку точності співпадіння слів чи ланцюжків слів з еталоном з урахуванням штрафу за малу величину тексту. Проблемаю цієї метрики є те, що збільшення показника не обов'язково означає покращення тексту. Наприклад, в одному реченні лише одне слово не вірне, але воно спотворює сенс усього речення. Метрика NIST - це модифікація BLEU, яка зважає слова чи ланцюжки слів за інформаційністю. Міра інформаційності слова приймається оберненою до частотності застосування слова. Обидва варіанти не підходять для оцінки перекладу словотворчими мовами, такими як українська чи словацька. Англійська мова переважно є аналітичною, це означає відсутність різниці між називним та знахідним відмінками, наявність жорстких порядків слів тощо. Українська мова переважно характеризується синтетичною морфологією. Це забезпечується численні форми та морфеми, словотворчі суфікси, що змінюють основи слів, та модифікації слів, що виражають різні граматичні категорії (наприклад, рід, число, відмінок) переважно за однією формальною ознакою. Тому для оцінки української мови доцільна лише метрика METEOR. METEOR - це модифікація BLEU, яка розглядає як еталон не фіксовану структуру, а множину структур, яка містить всі можливі синоніми та відмінки слів. Можливо також подальше ускладнення метрики шляхом застосування одночасно декількох мір точності: влучності, повноти та *F*-фактору.

Найбільш поширеними метриками, що виходять з трудовитрат є PErpT та TER. PErpT це - нормалізований до кількості символів час перекладу, а translation

Error Rate (TER) - кількість необхідних правок блоку тексту людиною для того, щоб текст став відповідним. Деякі дослідники як признак трудовитрат також розглядали тести. Ідея тут полягає в спостереженні скільки слів треба зафіксувати очима, що пов'язано теж з певною ментальною роботою. Приклади програм, які можуть бути застосовані для таких досліджень - це Translog-II [9] та PosEduOn [10]

Виходячи зі складності формування всіх можливих вірних перекладів в українській мові, доцільно застосовувати як критерій якості машинного перекладу тільки критерії витрачених на корегування трудовитрат, такі як PErTrT та TER.

**Аналіз доступних програмних рішень для перевірки та корекції текстів українською мовою.** В першу чергу відмітимо перелік джерел awesome-ukrainian-nlp, розміщений на github Крім посилань на ключові інструменти, моделі й бібліотеки, перелік містить збірник гігабайтів оброблених й необроблених текстів українською мовою, паралельних перекладів, словників тощо. Це саме ці матеріали, які є базою всіх NLP рішень українською мовою.

Серед програмних рішень, що перевіряють орфографію текстів українською мовою та можуть запропонувати коректні виправлення, слід обов'язково відмітити систему LanguageTool. Ця система була розроблена в 2003 р. й досі набирає популярність. Система побудована класичним чином з застосуванням правил, що робить її порівняно невибагливою до ресурсів. Система безкоштовна, тому в принципі її можна інсталивати собі на сервер, а також застосовувати десктопної перевірки орфографії в десктопних версіях Office. Система перевірки української мови має більше 1000 правил, при чому система враховує всі особливості правопису 2019 р. Важливо підкреслити, що система слідкує за чистотою й сучасністю мови, тому система правил передбачає виключення т.зв. барбаризмів (суржику та інших неправильних слів та словосполучень) й застарілих термінів. В роботі [11] була концептуалізована та опробувана чорнова версія системи, яка позиціонується як потенційно альтернативне до LanguageTool рішення, базою якого є технології NLP. Показано, що система може не тільки виявляти, але й автоматично виправляти певні помилки.

У обидвох найбільш популярних NLP-бібліотеках для структурного аналізу текстів (токенізації, лематизації тощо) - Spacy і NLTK - реалізована підтримка української мови. В бібліотеці Spacy ця підтримка реалізована на високому рівні. Так, для системи Spacy регулярно оновлюються українські бази чотирьох розмірів, найбільша з яких займає обсяг на сам написання біля 400 МБ. При застосуванні NLTK позитивний результат отримати можна, але розробник може зустрітись з рядом проблем, обумовлених недостатньою підтримкою української мови, деякі з яких проілюстровані в роботі [12].

Існуючі рішення щодо морфології текстів для української мови є зараз достатньо недосконалими. Тому такі задачі як провідняти словосполучення, а тим паче замінити одне словосполучення на інше в тому ж відмінку й числі не мають гарантовано працюючого методу розв'язку. В роботах [11, 12] автори вказують на доступність бібліотеки rymorphu. Однак базою цієї бібліотеки є російська мова (українська розглядається лише як експериментальна), розвиток бібліотеки припинено й ця бібліотека розв'язує свою задачу не спираючись на правила українського правопису достатньою мірою (тим паче не спирається й на норми правопису 2019 р.). Аналогічні результати й у іншій подібній, але комерційній бібліотеці - morpher.

Таким чином, підсумовуючи сказане можемо заключити, що тонізація текстів українською мовою може бути проведена на високому рівні й перевірка правопису теж. Проблемним місцем є відмінювання слів й особливо словосполучень. Повністю відсутні готові рішення, які дозволяють встановити повну відповідність між словами в українському тексті та словами в його перекладі

англійською мовою (чи навпаки). Не розв'язаною проблемою є виділення авторських позначень та введених змінних з наукових текстів будь-якою мовою. З більш докладний оглядом основних програмних лінгвістичних технологій, що можуть бути застосовані при розробці застосунків для обробки текстів українською мовою та рядом інших європейських мов рекомендуємо ознайомитись в роботі [11].

**Розробка підходу до побудови системи автоматичної корекції комп'ютерних перекладів технічних та наукових текстів в галузі автоматизації.** Для того, щоб сформувавши підхід були проаналізовані результати отриманих за допомогою DL англо-українських перекладів ряду наукових статей, присвячених розв'язанню задач автоматизації технологічних процесів хіміко-технологічного типу. Аналіз результатів перекладу показав, що характер помилок в таких текстах має достатньо систематичний характер. Таким чином, знаючи контекст тексту, англійський текст та згенерований переклад, можливо проводити автоматичну корекцію за допомогою спеціально сконструйованої системи правил. Критерієм в такому підході буде мінімум невірних замінів. Введення додаткових правил звичайно не має істотно змінювати поведінку перекладача, тому переважна кількість правил має бути точковою й обов'язково спиратися на наявність певних ключових слів в англійській версії оброблюваного речення.

Послідовність роботи користувача з системою наступна:

- 1) Завантаження статті (звичайно як PDF файлу)
- 2) Виділення текстової частини з файлу з урахуванням особливості верстки файлу (послідовний текст, двоколонна, з нерівномірними текстовими блоками).
- 3) Автоматична фільтрація системою текстової частини: зайвих реквізитів, підписів рисунків, формул, що записані в окремих рядках тощо.
- 4) Розбивка тексту на невеликі блоки. Кожен рядок тексту (блок) має містити одне чи кілька зв'язаних речень. Якщо в рядку присутній символ переносу, то цей рядок має об'єднуватись з наступним. Користувачеві повідомляється які рядки файлу треба перевірити на наявність в них цілісного блоку тексту.
- 5) Переклад отриманого тексту в DL. В результаті має бути отримано два текстових файли з англійський та українським текстом з повною відповідністю між рядками текстових файлів.
- 6) Запуск основної програми автоматичної корекції текстових файлів й відкриття звіту в браузері. В звіті відображаються запропоновані зміни.
- 7) У випадку виявлення під час перегляду звіту певних блоків, які вимагають корекції, проаналізувати за функціонали сторінки звіту токени проблемного текстового блоку й загальну причину помилки, додати в базу правил необхідні зміни та ввести в програму номери змінених рядків. Оновити сторінку зі звітом в браузері.
- 8) Повторювати п. 7. доки результат перекладу не стане задовільним.
- 9) Якщо отриманий текст виглядає задовільним, то рекомендується остаточно перевірити його за допомогою LanguageTool на предмет невеликих помилок.

Як мову програмування для реалізації системи виберемо Python 3.11. Програма буде мати нескладний текстовий інтерфейс, метою якого буде відпрацювання нових правил. Базовою NLP бібліотекою для програми виберемо Spacy 3. База даних правил буде зберігатись в декількох CSV файлах. Звіт буде формуватись в спеціальній папці документу як HTML файл з застосуванням бібліотеки BeautifulSoup.

**Приклади формування корекційних правил та аналіз результатів, досягнутих програмою при їх застосуванні.** Як приклад розглянемо створення правил за результатами аналізу перекладу двох статей [13,14]. Позначення: **НФ** - невірний

фрагмент перекладу, **ПС** - пояснення до фрагменту, **ПР** - загально сформоване правило заміни.

Основні правила, що були застосовані надані в достатній мірі для розуміння проблеми переводу та запропонованого підходу.

НФ: Closedloop block diagram → Блок-схема замкненого циклу.

ПС: Слово “loop” має перший сенс «цикл». Звичайно КП вибирає вірне слово «контур», але в цьому випадку виявилось замало слів.

ПР: Заміна «замкнений цикл» на «замкнений контур», якщо є closed ?loop”.

НФ: Steady-state error → похибка в усталеному режимі.

ПС: Постійний буквальный переклад при наявності українського відповідника.

ПР: Заміна «похибка в усталеному режимі» на «статична похибка» якщо є steady, state, error

НФ: smaller closed loop damping coefficient → менший коефіцієнт демпфування замкненого контуру.

ПС: Постійний буквальный переклад при наявності українського відповідника. Цей термін зустрічається в українській мові, але в текстах, що присвячені аналізу сигналів.

ПР: Заміна «коефіцієнт демпфування» на «коефіцієнт затухання» (варіант згасання) якщо є damping, coefficient.

НФ: Analyze the complexity of tuning a PI controller to control liquid level - Проаналізувати складність налаштування ПІ-регулятора для контролю рівня рідини.

ПС: В англійській під “control” розуміють як «контроль», так і «регулювання». В українській мові під контролем розуміється сигналізація, тому цей переклад спотворює сенс. DP іноді перекладає liquid level control як контроль рівня рідини, а іноді – як регулювання. Хоча в текстовому блоці присутній ПІ-регулятор, для DP це не стало чомусь в цьому випадку умовою зміни «контроль» на «регулювання».

ПР: Заміна «контролю рівня» на «регулювання рівня» якщо є control, liquid чи level.

НФ: The open loop system has a process transfer function  $gP(s)$  relating the controlled variable  $H$  and the manipulated variable  $F_{out}$ . The transfer function  $gL(s)$  relates the output variable  $H$  and the load disturbance  $F_{in}$ . → Розімкнута система має передатну функцію процесу  $gP(s)$ , що пов'язує керовану змінну  $H$  і керовану змінну  $F_{out}$ . Передавальна функція  $gL(s)$  пов'язує вихідну змінну  $H$  і збурення навантаження  $F_{in}$ .

ПС1: DL застосовує одночасно три форми, вважаючи їх синонімічними: передатна, передавальна та передаточна функція.

ПР1: Треба замінити всі варіанти на певний один варіант, наприклад на «передаточна».

ПС2: DL спотворив сенс, переклавши “controlled variable” і “manipulated variable” як «керована змінна». Подібна помилка має систематичний характер.

ПР2: виправлення має спиратися на те, що після “variable” йде позначення. Якщо в тексті зустрічається «керована змінна» з позначенням, то треба його виділити й перевірити наявність біля позначення в англійській версії “manipulated variable”. Тоді цю «керовану змінну» замінюємо, наприклад, на керуючий вплив.

НФ: with two tanks in series for P control ( $K_C = 2$ ) and PI control ( $K_C = 5$  and  $T_I = 5$  min) → у двох послідовно з'єднаних резервуарах для P-регулювання ( $K_C = 2$ ) і PI-регулювання ( $K_C = 5$  і  $T_I = 5$  хв).

ПС: DL використовує в текстах як «PID-регулятор», так і «ПІД-регулятор». Тут бачимо, що це про всі форми регулятора: «P-регулювання», «PI-регулювання».

ПР: При наявності «P/PI/PID control» замінити P/PI/PID на кириличну аббревіатуру.

НФ: Root locus plot → Графік кореневого локусу.

ПС: Постійний буквальный переклад при наявності українського відповідника.



- ПР: Замінити «кореневий локус» на «кореневий годограф» якщо є root, locus  
НФ: so the response is not oscillatory → тому відгук не є осциляторним.  
ПС: Постійний буквальний переклад при наявності українського відповідника.  
ПР: Замінити «осциляторний» на «коливальний» якщо є oscillatory, response.  
НФ: if the plant culture requires PI level control → якщо культура рослин вимагає ПІ-регулятор рівня.  
ПС: Англійське слово “plant” – це “рослина”, “технологічний об’єкт керування”, “технологічна установка” і “завод”. Однак первинний сенс в тематиці автоматизації зустрічатись буде дуже зрідка.  
ПР: Замінити “культура рослин» на «культура виробництва» якщо є plant culture.  
НФ: Extractive distillation → Екстрактивна дистиляція.  
ПС: DL іноді слово “distillation” перекладає як «ректифікація», а іноді як «дистиляція».  
Правильний переклад – «ректифікація», оскільки майже завжди мається на увазі апарат з флегмою.  
ПР: Замінити «дистиляція» на «ректифікація» якщо в distillation.  
НФ: Ternary Map (Mole Basis) → Тернарна карта (кротячий базис).  
ПС1: Одне зі значень англійського слова «mole» має значення «кріт».  
ПР1: Замінити «кротячий базис» на «мольні одиниці» якщо є mole basis.  
ПС2: Буквальний переклад з некоректним застосуванням слова «карта».  
ПР2: Замінити «тернарна карта» на «трикутна діаграма» якщо є ternary map.  
НФ: to handle large disturbances → для обробки великих збурень.  
ПС: DL для «large disturbances» вибирає далеко не перший сенс дієслів. «Долати» - це другий сенс дієслова handle, однак DL чомусь обрав «обробки» замінити «для обробки великих збурень» на «для подолання великих збурень», якщо в оригіналі є слово “disturbance”.  
ПР: Замінити «для обробки великих збурень» на «для подолання великих збурень», якщо є “disturbance”.  
НФ: is used to set the number of stages in each column and feed locations → використовується для встановлення кількості стадій у кожній колоні та місць подачі.  
ПС: В англійській мові «тарілка» ректифікаційної колони може бути перекладена як “stage”, “tray” і навіть “plate”. Однак “stage” – найбільш частотний варіант. Буквальний переклад його не вірний.  
ПР: Замінити «стадія» на «тарілка» якщо є distillation, column.  
НФ: A 1 min deadtime is included in the temperature loops -> До температурних контурів включено час очікування 1 хв.  
ПС: Українським відповідником «мертвого часу» є час запізнення.  
ПР: Замінити «час очікування» на «час запізнення» якщо є dead ?time.  
НФ: 1 atm case with economizer; decreases in feed flowrate → Випадок з економайзером при тиску 1 атм; зменшення швидкості потоку корму.  
ПС1: Буквальний переклад “flow rate” при українському відповіднику «витрата».  
ПР1: Замінити «швидкість потоку» на «витрату потоку» якщо в АТ є flow ?rate.  
ПС2: DL бере перше значення іменника feed й отримує “корм”, що є зовсім за контекстом.  
ПР2: Замінити «корм» на «живлення» якщо є feed.  
НФ: An underdamped second-order system has a characteristic equation → Недодемпфована система другого порядку має характеристичне рівняння.  
ПС: Буквальний переклад слова “underdamped” замість українського відповідника.  
ПР: Замінити «недодемпфована» на «коливальна» якщо в АТ є underdamped, system.  
Сформовані правила для додавання в базу необхідно віднести до певної категорії та занести в спеціальному вигляді в відповідний до цієї категорії файл.

В першу чергу відпрацьовуються наступні пари токенів:

- amod (adjectival modifier) – прикметник + іменник
- nmod (nominal modifier) – іменник + іменник
- conj (conjunction) – сполучені слова.
- obj (object) – дієслово + об’єкт
- csubj (clausal subject) – дієслово + суб’єкт

Далі відпрацьовується заміна окремих слів на окремі слова.

Наступний крок - заміна послідовностей з нефіксованою кількістю слів (в порядку починаючи з найбільшої кількості). Останній крок - заміна послідовностей, що вимагає програмного коду (на поточному етапі це виявлення змінних для того, щоб скорегувати невірний переклад “manipulated variable” як керованої змінної. Звіти програми для двох статей після досягнення рівня задовільного перекладу показані на рис. 1. і 2.

79	To illustrate this feature let us consider a single tank with the process, load, transmitter and valve transfer functions discussed above with the block diagram shown in Fig. 2. There are two closed-loop transfer functions. The “servo” transfer function relates the response of the controlled process variable to a change in the controller setpoint. Using Aspen Dynamics labeling, the signals into the controller are the process variable signal (PV) from the transmitter and, the setpoint signal (SP).	Щоб проілюструвати цю особливість, розглянемо один резервуар з описаними вище передавальними функціями процесу, навантаження, датчика і клапана, блок-схему якого показано на рис. 2. Існує дві замкнені передавальні функції. Передавальна функція “сервоприводу” пов’язує реакцію контрольованої змінної процесу на зміну уставки регулятора. Використовуючи маркування Aspen Dynamics, сигналами, що надходять до контролера, є сигнал змінної процесу (PV) від датчика та сигнал уставки (SP).	Щоб проілюструвати цю особливість, розглянемо один резервуар з описаними вище <b>передатними</b> <b>передаточними</b> функціями процесу, навантаження, <b>датчик</b> <b>давача</b> і клапана, блок-схему якого показано на рис. 2. Існує дві замкнені <b>передавальні</b> <b>передаточні</b> функції. <b>Передавальна</b> <b>Передаточна</b> функція “сервоприводу” пов’язує реакцію <b>контрольованої</b> <b>керованої</b> змінної процесу на зміну <b>уставки</b> <b>завдання</b> регулятора. Використовуючи маркування Aspen Dynamics, сигналами, що надходять до контролера, є сигнал змінної процесу (PV) від <b>датчик</b> <b>давача</b> та сигнал <b>уставки</b> <b>завдання</b> (SP).
80	The controller output signal (OP) goes to the valve in the exit line.	Вихідний сигнал контролера (OP) надходить на клапан у вихідній лінії.	Вихідний сигнал контролера (OP) надходить на клапан у вихідній лінії.
81	Using an integral time of $t_I = 2$ min gives the root locus plot shown in Fig. 9.	Використання інтегрального часу $t_I = 2$ хв дає графік кореневого локусу, показаний на рис. 9.	Використання <b>інтегрального</b> часу <b>ізодрому</b> $t_I = 2$ хв дає графік кореневого <b>локусу</b> , <b>одографа</b> , показаний на рис. 9.

Рис. 1. Фрагмент звіту програми щодо результатів корегування перекладу статті [13]

51	###3.1. Basic control structure for 10 atm case	###3.1. Базова структура управління для випадку 10 атм	3.1. Базова структура <b>управління</b> <b>САК</b> для випадку 10 атм
52	The well-established control structure proposed by Grassi (1992) is the basis for the design, which was used in the previous work for the high solvent flowrate case at 1 atm. Fig. 6 shows the structure in which the solvent is ratioed to the feed and both reflux ratios are controlled. Conventional PI controller are used except for level loops, which are proportional with $K_c = 2$ . The basic control structure consists of the following loops.	Добре відома структура управління, запропонована Grassi (1992), є основою для проектування, яка була використана в попередній роботі для випадку з високою швидкістю потоку розчинника при 1 атм. На рис. 6 показано структуру, в якій співвідношення розчинника і подачі контролюється, а також обидва співвідношення рефлюксу. Використовується звичайний ПІ-регулятор, за винятком контурів рівня, які є пропорційними з $K_c = 2$ . Базова структура управління складається з наступних контурів.	Добре відома структура <b>управління</b> , <b>САК</b> , запропонована Grassi (1992), є основою для <b>проектування</b> , <b>синтезу</b> , яка була використана в попередній роботі для випадку з високою <b>швидкістю</b> <b>витратою</b> потоку розчинника при 1 атм. На рис. 6 показано структуру, в якій співвідношення розчинника і подачі <b>контролюється</b> , <b>регулюється</b> , а також обидва <b>співвідношення</b> <b>рефлюксу</b> , <b>флегмових</b> <b>числа</b> . Використовується звичайний ПІ-регулятор, за винятком контурів рівня, які є пропорційними з $K_c = 2$ . Базова структура <b>управління</b> <b>САК</b> складається з наступних контурів.
53	1. Feed is flow controlled. This is the throughput manipulator.	1. Подача регулюється за витратою. Це маніпулятор пропускної здатності.	1. Подача регулюється за витратою. Це маніпулятор пропускної здатності.
54	2. Solvent flow is ratioed to feed flow.	2. Потік розчинника співвідноситься з потоком подачі.	2. Потік розчинника співвідноситься з потоком подачі.
55	3. Reflux drum levels are controlled by distillate flows.	3. Рівні в зворотному барабані контролюються потоками дистилату.	3. Рівні в <b>зворотному</b> <b>барабані</b> <b>флегмовому</b> <b>баці</b> контролюються потоками дистилату.
56	Fig. 6. Base-case control structure using S/F, RR1 and RR2.	Рис. 6. Базова структура управління з використанням S/F, RR1 і RR2.	Рис. 6. Базова структура <b>управління</b> <b>САК</b> з використанням S/F, RR1 і RR2.

Рис. 2. Фрагмент звіту програми щодо результатів корегування перекладу статті [14]

Узагальнені результати по перекладу роботи [13] : з 2337 слів 97 слів було додано й 102 слова видалено. Узагальнені результати по перекладу роботи [14]: з 2300 слів 138 слів було видалено й 147 слів додано. Таким чином, для корекції тексту необхідно змінити приблизно 5% від його слів.

#### **Висновки**

1. Результати автоматичного перекладу текстів наукових статей галузі автоматизації технологічних процесів з англійської на українську з використанням DeepL показують, що в цілому спеціаліста є зрозумілим про що йде мова в оригінальному тексті.
2. Тим не менш, переклад є далеким від літературно прийнятого. Замість прийнятих в українській термінології галузі У цілому помилки, які робить КТ є легко помітними (на кшталт “культура рослин” замість культури виробництва чи “корм колони” замість “живлення колони). Однак, зустрілась й така, яка паплюжить смисл, але є непомітно, коли як вхідна, так і вихідна змінна об’єкту в системі позначень стали “керованими змінними”.
3. У цілому помилки, які робить КТ є легко помітними (на кшталт “культура рослин” замість культури виробництва чи “корм колони” замість “живлення колони). Однак, зустрілась й така, яка паплюжить смисл, але є непомітно, коли як вхідна, так і вихідна змінна об’єкту в системі позначень стали “керованими змінними”.
4. Помилки та неточності, які припускає КТ мають систематичний характер. Застосування запропонованого підходу до корегування тексту дозволяє істотно компенсувати фактор систематичних помилок, таким чином час очікуваного редагування тексту, який пройшов комп’ютерний перекладач та запропонований коректор буде істотно меншим ніж час редагування тексту, який піддався тільки комп’ютерному перекладу.
5. Статистичний аналіз показує, що в розглянутих перекладах достатньо замінити лише 5% слів для того, щоб істотно покращити літературний рівень тексту.

#### **Список літератури**

1. Нікітін О.К., Зайцев В.М., Толочко Т.О. Приладобудування та автоматизація. Терміни і визначення. Київ : КПІ ім. Ігоря Сікорського, 2019. Ч. 1. 202 с.
2. Michael C., Dragsted B., Elming J. The process of post-editing: A pilot study. *Proceedings of the 8th International NLPCS Workshop*. Denmark, Samfundslitteratur: Copenhagen Business School. P. 131-142.
3. Parra Escartín C., Arcedillo M. A fuzzier approach to machine translation evaluation: A pilot study on post-editing productivity and automated metrics in commercial settings. *Proceedings of the Fourth Workshop on Hybrid Approaches to Translation (HyTra)*, Beijing. Stroudsburg, PA, USA, 2015. P. 40–45. URL: <https://doi.org/10.18653/v1/w15-4107>.
4. Моїсеєва Н., Дзикович О., Штанько А. Машинний переклад: порівняння результатів та аналіз помилок DeepL та Google Translate. *Advanced Linguistics*. 2023. № 11. С.78-82. <https://doi.org/10.20535/2617-5339.2023.11.27759>
5. Ismail A., Hartono R. Errors Made in Google Translate in the Indonesian to English Translations of News Item Texts. *ELT Forum: Journal of English Language Teaching*. 2016. Vol. 5, no. 2. Article 2. URL: <https://journal.unnes.ac.id/sju/index.php/elt/article/view/11228>.
6. Tongpoon-Patanasorn A., Griffith K. Google Translate and Translation Quality: A Case of Translating Academic Abstracts from Thai to English. *PASAA: Journal of Language Teaching and Learning in Thailand*. 2020. No. 60. P. 134–163.
7. Kur M. Method of measuring the effort related to post-editing machine translated outputs produced in the English>Polish language pair by Google, Microsoft and

- DeepL MT engines: A pilot study. *Beyond Philology An International Journal of Linguistics, Literary Studies and English Language Teaching*. 2019. No. 16/4. P. 69–99. URL: <https://doi.org/10.26881/bp.2019.4.03>
8. Ismail A., Hartono R. Errors Made In Google Translate In The Indonesian To English Translations Of News Item Texts. *Journal of English Language Teaching*. 2016. Vol. 5. URL: <https://api.semanticscholar.org/CorpusID:58690342>.
  9. Giraldo Ospina D. L., Naranjo Ruiz M., Romero Ramírez L. C. A methodological proposal integrating the translation process into the study of cognitive effort. *Cadernos de Tradução*. 2022. Vol. 42, no. 1. P. 1–24. URL: <https://doi.org/10.5007/2175-7968.2022.e84845>
  10. Oliver A., Alvarez S., Badia T. PosEdiOn: Post-Editing Assessment in PythOn. *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, Lisboa, 1 November 2020. P. 403–410. URL: <https://aclanthology.org/2020.eamt-1.43.pdf>
  11. Холодна Н., Висоцька В. Технологія виправлення граматичних помилок в україномовному текстовому контенті на основі методів машинного навчання. *Радіоелектроніка, інформатика, управління*. 2023. № 1. С. 114–140. URL: <https://doi.org/10.15588/1607-3274-2023-1-12>
  12. Tmienova N., Sus B. System of Intellectual Ukrainian Language Processing. *Information Technologies and Security* 2019. P. 199–209. URL: <https://ceur-ws.org/Vol-2577/paper16.pdf>.
  13. Luyben W. L. Liquid level control: Simplicity and complexity. *Journal of Process Control*. 2020. Vol. 86. P. 57–64. URL: <https://doi.org/10.1016/j.jprocont.2019.12.008>
  14. Luyben W. L. Control of heat-integrated extractive distillation processes. *Computers & Chemical Engineering*. 2018. Vol. 111. P. 267–277. URL: <https://doi.org/10.1016/j.compchemeng.2017.12.008>

А.О. Стопакевич, А.М. Тігарєв, О.Р. Романюк, О.А. Стопакевич

**AUTOMATIC CORRECTION SYSTEM OF ENGLISH-UKRAINIAN  
COMPUTER TRANSLATION FOR TECHNICAL TEXTS IN THE FIELD OF  
AUTOMATION OF TECHNOLOGICAL PROCESSES**

A.O. Stopakevych<sup>1</sup>, A.M. Tigarev<sup>1</sup>, O.R. Romanyuk<sup>1</sup>, O.A. Stopakevych<sup>2</sup>

<sup>1</sup>State University of Intellectual Technologies and Telecommunications

1, Kuznechna, Odesa, 65029, Ukraine

<sup>2</sup>National Odesa Polytechnic University

1, Shevchenko Ave., Odesa, 65044, Ukraine

email: stopakevich@gmail.com

The purpose of the work is to develop a system of automatic correction of computer-translated texts with specific terminology, which is inherent in scientific and technical texts in the field of automation of technological processes. An analysis of the reasons why computer translators cannot achieve high-quality English-Ukrainian translation of technical texts in the specified field is given. It was concluded that within the limits of the approach used by modern computer translators, the quality of such translations cannot be improved. An analysis of the experience of translation correction by professional translators, available metrics for evaluating the process and results of correction of computer translations, and available software solutions for working with texts written in the Ukrainian language was carried out. It was concluded that for computer English-Ukrainian translation, the only practically significant approach to assessing its quality is measuring the amount of work that must be performed by a professional translator in order for the text to meet literary standards. The analysis of scientific texts that were translated by DeepL showed that the amount of such work can be significantly reduced, since the errors made by this translator are systematic. Thus, by analyzing the mistakes made by a computer translator, it is possible to form universal correction rules for all texts translated by a certain translator of the industry, which can be performed automatically by the software application. The effectiveness of the approach is demonstrated on the example of the development of rules resulting from the analysis of the results of the translation of two scientific articles. It is shown that replacing approximately 5% of words in computer translation significantly increases its quality.

**Keywords:** correction, computer, translation, automation, technological, processes, NLP, program, python, terminology, deepl.