

Київський національний університет імені Тараса Шевченка
Навчально-науковий інститут філології
Кафедра мов і літератур Далекого Сходу та Південно-Східної Азії

**Особливості синтаксичних трансформацій в китайсько-
українському публіцистичному перекладі у фокусі корпусного аналізу**

Кваліфікаційна робота
освітнього ступеня «бакалавр»
Студентки IV курсу ОС «Бакалавр»
спеціальності 035 «Філологія» / 035.065
«східні мови та літератури (переклад включно)»
Освітньої програми «Китайська
мова і література та переклад, англійська мова»

Бурди Анастасії Юріївни

Науковий керівник:
канд.філол.н., асист. Козоріз Олександр Павлович

Рецензент:
канд.філол.н., асист. Нестеренко Ольга Олександрівна

«Допущено до захисту»

Протокол засідання кафедри

Мов і літератур Далекого Сходу

та Південно-Східної Азії

Протокол № 14 від 5 червня 2024 р.

Завідувачка кафедри _____ доц. Ісаєва Н.С.

КИЇВ-2024

ЗМІСТ

СПИСОК СКОРОЧЕНЬ	4
ВСТУП.....	5
РОЗДІЛ 1. ТЕОРЕТИЧНІ ЗАСАДИ ВИВЧЕННЯ СИНТАКСИЧНИХ ТРАНСФОРМАЦІЙ В КОНТЕКСТІ КИТАЙСЬКО-УКРАЇНСЬКОГО ПУБЛІЦИСТИЧНОГО МАШИННОГО ПЕРЕКЛАДУ	8
1.1. Теоретичні засади машинного перекладу в історичному розрізі та на сучасному етапі розвитку	8
1.2. Потенціал НМП у контексті міжкультурного спілкування	12
1.3. Дослідження явища перекладацьких трансформацій та універсалій в контексті НМП	16
1.4. Синтаксичні трансформації у процесі НМП китайсько-українських публіцистичних текстів	19
Висновки до Розділу 1.....	23
РОЗДІЛ 2. КОРПУСНЕ ДОСЛІДЖЕННЯ СИНТАКСИЧНИХ ОСОБЛИВОСТЕЙ МОВ В КОНТЕКСТІ НМП.....	25
2.1. Корпусна лінгвістика як наука: застосування корпусів у НМП.....	25
2.2. Методологія корпусного лінгвістичного аналізу та комп'ютерного опрацювання текстів	28
2.3. Визначення синтаксичних особливостей будови китайської та української мов на основі аналізу паралельного корпусу	34
Висновки до Розділу 2.....	41
РОЗДІЛ 3. ОЦІНЮВАННЯ ЯКОСТІ РОБОТИ СИСТЕМ НМП ЗА СИНТАКСИЧНИМ АСПЕКТОМ У ПРОЦЕСІ КИТАЙСЬКО-УКРАЇНСЬКОГО ПУБЛІЦИСТИЧНОГО ПЕРЕКЛАДУ	42
3.1. Підходи до здійснення оцінювання систем НМП за лінгвістичним аспектом	42

3.2. Методологія здійснення автоматизованого оцінювання систем НМП та критерії якості.....	45
3.3. Гібридне оцінювання ефективності системи НМП DeepL на матеріалі китайсько-українського публіцистичного тексту.....	47
3.3.1. Автоматизоване оцінювання тексту за метрикою BLEU.....	48
3.3.2. Ручне оцінювання та розбір тексту перекладу, визначення перекладацьких трансформацій на основі формальних тенденцій вихідної та цільової мов	51
Висновки до Розділу 3.....	58
ВИСНОВКИ.....	60
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	65
ДОДАТКИ	75
Додаток А. Текст промови «坐轮椅潜行深海» («Плавання морськими глибинами в інвалідному візку») спікерки Сю Остін китайською	75
Додаток Б. НМП тексту промови «坐轮椅潜行深海» («Плавання морськими глибинами в інвалідному візку») спікерки Сю Остін через систему НМП DeepL	78
Додаток В. Ручний переклад тексту промови «坐轮椅潜行深海» («Плавання морськими глибинами в інвалідному візку») спікерки Сю Остін через систему НМП DeepL, виконаний людиною-перекладачем.....	81
Додаток Г. Приклади окремих складених авторських функцій для здійснення корпусного аналізу	84
Додаток Г.1. Функція для визначення загальної кількості складнопідрядних речень в україномовних текстах корпусу	84
Додаток Г.2. Функція для визначення кількості дієслівних фраз в китайськомовних текстах корпусу.....	84
Додаток Г. Приклад авторської функції, складеної для здійснення автоматизованого оцінювання сегменту за метрикою BLEU	85

СПИСОК СКОРОЧЕНЬ

МП – машинний переклад

НМП – нейронний машинний переклад

ШІ – штучний інтелект

ALPAC – Automatic Language Processing Advisory Committee:

Консультативний комітет з автоматичної обробки мови

BLEU – Bilingual Evaluation Understudy

COMET – Crosslingual Optimized Metric for Evaluation of Translation

IDE – Integrated Development Environment

LEPOR – Length penalty (L), Precision of n-grams (E), Precision (P), N-gram position difference penalty (O), Recall (R): метрика оцінювання автоматичного машинного перекладу за настроюваними параметрами

METEOR – Metric for Evaluation of Translation with Explicit ORdering

NIST – National Institute of Standards and Technology

NLP – Natural Language Processing

NMT – Neural Machine Translation

PRIS – Precision (P), Recall (R), Position of n-grams (I), Sentence length penalty (S): метрика оцінювання автоматичного машинного перекладу за настроюваними параметрами

RBMT – Rule-Based Machine Translation (укр. Машинний переклад на основі правил)

SMT – Statistical Machine Translation (укр. Статистичний машинний переклад)

SOV – Subject-Object-Verb [word order]

SVO – Subject-Verb-Object [word order]

TEI – Text Encoding Initiative

TER – Translation Error Rate

XML – Extensible Markup Language

ВСТУП

В основу процесу міжкультурної комунікації закладене мовне різноманіття в світі, що породжує необхідність здійснення перекладу задля трансляції повідомлення. Хоча цей вид діяльності спирається здебільшого на формальні правила певної мови та необхідність підбору відповідників за змістом певного сегменту інформації, справжнє розуміння між представниками відмінних мовних та етнічних груп досягається через додаткову обробку тексту з огляду на конкретну мовленнєву ситуацію. Така складова процесу перекладу становить складнощі для систем автоматизованого машинного перекладу (МП), що є сучасним напрямком розвитку перекладознавства та основою побутових засобів перекладу. Як зазначав Воррен Вівер, один з засновників галузі машинного перекладу, у праці «The Mathematical Theory of Communication» («Математична теорія комунікації»): «Технічна проблема [машинного перекладу]: наскільки точно можна передати символи зв'язку? Семантична проблема: наскільки точно передані символи передають бажане значення? Проблема ефективності: наскільки ефективно отримане значення впливає на реципієнта бажаним чином?» [67, с. 14]

Актуальність дослідження зумовлена зростаючим попитом на залучення систем МП, наразі, зокрема, нейронного машинного перекладу (НМП), в різних сферах людської діяльності та необхідністю підвищення ефективності їхньої роботи через дослідження лінгвістичного аспекту перекладу. Ми розглядаємо синтаксичну будову мов як один з рівнів, який має значний вплив на формування повідомлення. У звичайному житті людини підбір адекватної структури речення, залучення конструкцій в потоці мовлення ґрунтуються на аспектах конкретної ситуації. У китайсько-українській мовній парі, як структурно значно відмінних мов, під час перекладу постає чимало питань щодо доречних синтаксичних трансформацій. Тобто принцип формування повідомлення носієм є багатоплановим, що зумовлює необхідність дослідження синтаксису в контексті перекладу, зокрема в розрізі способів вдосконалення систем НМП.

Метою роботи є дослідження особливостей синтаксичних трансформацій в китайсько-українському публіцистичному перекладі у фокусі корпусного аналізу та з огляду на їхню реалізацію в процесі роботи систем НМП.

Досягнення мети дослідження передбачає виконання таких **завдань**:

1) визначити фундаментальні принципи роботи систем МП в діяхронічному та синхронічному розрізах з огляду на лінгвістичну складову цього процесу;

2) виявити актуальний напрямок досліджень у галузі НМП в рамках китайської та української як пари типологічно відмінних мов;

3) залучити методи корпусної лінгвістики у вивченні будови мов та виявити ефективність відповідного аналізу на основі значних обсягів текстів;

4) теоретично та через практичне дослідження загального корпусу окреслити основні розбіжності синтаксичної будови китайської та української мов в контексті залучення отриманих даних у перекладі;

5) розбити двомовний корпус на сегменти задля визначення динаміки зміни значень показників за окремим параметром синтаксичного аналізу – співвідношення іменникових та дієслівних фраз (NP та VP) у досліджених мовах – та визначити значення отриманих результатів для китайсько-українського перекладу;

6) окреслити перекладацькі трансформації як конкретні механізми китайсько-українського перекладу в контексті вдосконалення систем НМП;

7) здійснити оцінювання роботи системи НМП DeepL;

8) визначити недоліки опрацювання тексту системою НМП DeepL у процесі публіцистичного китайсько-українського перекладу та запропонувати способи їхнього вирішення;

9) навести перспективи лінгвістичних досліджень у галузі НМП для мовної пари китайська-українська та значення отриманих результатів дослідження.

Об'єктом дослідження виступають китайсько-українські публіцистичні тексти, представлені паралельним двомовним корпусом та перекладом окремої промови із загального обсягу за допомогою системи НМП DeepL.

Предмет дослідження – синтаксичні особливості будови китайської та української мов та їхній вплив на процес перекладу, зокрема якість роботи системи НМП.

Новизна роботи полягає в розгляді китайської та української мов як основи для розробки перекладацьких трансформацій та універсалій в контексті підвищення якості роботи систем НМП. Крім того, в центрі нашої уваги постає розробка гнучких способів НМП між структурно відмінними мовами, що здійснюється на основі корпусного аналізу та віднайдення закономірностей в синтаксичній будові мов.

Методи дослідження: під час дослідження було залучено такі підходи: 1) метод аналізу та синтезу, який надав змогу на основі теоретичних праць та надбань науковців світового простору в напрямку перекладознавства визначити основні риси, переваги та недоліки сучасного етапу розвитку цієї галузі; 2) історичний метод, що послугував засобом спостереження за процесом розвитку МП в діяхронічному плані та етапами поступового становлення НМП; 3) критичний метод, який дозволив визначити основні специфічні риси синтаксичної будови китайської та української, зазначити їхній ймовірний вплив на процес публіцистичного перекладу; 4) метод термінологічного аналізу, який дозволив дати визначення окремим поняттям в контексті корпусного аналізу текстів та іменування спостережуваних явищ і тенденцій; 5) порівняльний метод, завдяки якому вдалося побачити збіги та відмінності синтаксису двох розглянутих мов, окреслити коло питань з огляду на НМП та характер необхідних перекладацьких трансформацій у процесі китайсько-українського перекладу публіцистичних текстів; 6) метод корпусного аналізу, що полягає у залученні значного обсягу текстового матеріалу та його опрацювання за допомогою мов програмування задля визначення статистичних показників, які вказують на особливі риси та закономірності в будові досліджуваних природніх мов.

Робота складається зі вступу, трьох розділів і висновків. Загальний обсяг роботи 85 сторінок.

РОЗДІЛ 1. ТЕОРЕТИЧНІ ЗАСАДИ ВИВЧЕННЯ СИНТАКСИЧНИХ ТРАНСФОРМАЦІЙ В КОНТЕКСТІ КИТАЙСЬКО-УКРАЇНСЬКОГО ПУБЛІЦИСТИЧНОГО МАШИННОГО ПЕРЕКЛАДУ

1.1. Теоретичні засади машинного перекладу в історичному розрізі та на сучасному етапі розвитку

Під терміном «машинний переклад», «механічний переклад» або «автоматизований переклад» розуміють комп'ютеризовані системи, що з або без допомоги людини-асистента здійснюють переклад з однієї мови на іншу [34, с. 3]. Це поняття не включає в себе спеціальні інструменти програмного забезпечення, що виступають допоміжними ресурсами для професійних перекладачів (зокрема, що надають доступ до словників, термінологічних баз даних, обладнання для переробки різноформатного повідомлення в текст, редагування тексту тощо). Однак термін «машинного перекладу» (надалі МП) охоплює в собі такі системи, що дозволяють професійним перекладачам та звичайним користувачам контролювати комп'ютери у здійсненні перекладу через підготовку текстів, їхню обробку й виведення результату. Як зазначають дослідники, які активно досліджували питання підвищення ефективності перекладу через застосування останніх технологічних досягнень [34, с. 3], автоматизація цього процесу, в ідеалі, має на меті отримання високоякісного продукту, у створенні якого виключається участь людини-асистента. Попри досить довгий період розробок та вдосконалень, результати оцінювання роботи більшості систем МП на практиці демонструють, що пост-редагування¹ людиною перекладів, отриманих завдяки системам МП, досі є невід'ємною частиною багатоаспектного процесу перекладу [52, с. 8].

Задля визначення векторів розвитку машинного перекладу, зміни ступеня залученості людини-асистента в процес перекладу та основних принципів роботи систем МП на сучасному етапі варто звернутися до історії становлення

¹ *англ.* post-editing

цієї галузі. Виникнення ідей про створення систем МП датується ще 50-ми роками минулого століття, коли було запропоновано використання методів криптографії, статистичного аналізу, логіки, специфічних та універсальних рис властивих системам різних мов, а також теорії Шеннона у здійсненні перекладу в рамках військової сфери [48, с. 18]. Меморандум Воррена Вівера у липні 1949 року послугував для наступного десятиліття поштовхом до активних емпіричних досліджень миттєвої обробки на основі статистики, а також розробок способів перекладу, що ґрунтувалися на теоретичних лінгвістичних знаннях про мови. Цей період в історії МП мав надзвичайно велике довгострокове значення, адже деякі дослідження цього часу, а саме, праці щодо розробки електронних словників та методів синтаксичного аналізу, заклали фундамент для сучасної комп'ютерної лінгвістики та систем штучного інтелекту.

В подальшому, після створення Консультативного комітету з автоматичної обробки мови ²(ALPAC) у 1964 році в Сполучених Штатах і висунення ідеї про нерентабельність машинного перекладу та відсутності необхідності в негайному розвитку цього напрямку Національною академією наук США, спостерігався певний занепад досліджень [52, с. 13]. Упродовж 1960-их років на теренах Західної Європи, Канади та Японії проводилися більш вузько спеціалізовані розробки, зокрема систем на основі таких лінгвістичних понять, як «субмова» (Météo³), «інтерлінгва» та «трансформація» або «трансфер» (Ariane ⁴та SUSY⁵). Здебільшого до розробок залучалися такі мовні пари, як: французько-англійська, англо-італійська, англо-німецька тощо. Останні досягнення в МП та комп'ютерній лінгвістиці, отримані окремими дослідницькими групами, були в подальшому залучені в роботу Комісії Європейських співтовариств⁶, заснованої

² Комітет із семи вчених на чолі з Джоном Р. Пірсом, створений у 1964 році урядом Сполучених Штатів з метою оцінки прогресу комп'ютерної лінгвістики загалом і машинного перекладу зокрема.

³ Система МП для щоденного перекладу звітів про погоду науковцями з Монреалю.

⁴ Трансферна система МП другого покоління, особливістю якої була динамічність і гнучкість у залученні синтаксичної будови мов і граматичних конструкцій; створена науковим угрупованням Grenoble group.

⁵ Мультилінгвальна система, створена Саарбрюккенською групою дослідників з Франції, які залучили експериментальним шляхом різноманітні лінгвістичні прийоми обробки тексту у процесі МП.

⁶ Commission of the European Communities: часто називають просто Європейською комісією (European Commission), є виконавчим органом Європейського Союзу (ЄС), відповідальним за законопроекти, імплементацію рішень, дотримання договорів ЄС і управління повсякденною діяльністю ЄС.

в 1976 році. Регіональні наукові угруповання об'єднувалися в один проект під назвою EUROTRA.

У 1980-их роках у спеціалізованих наукових колах з розробки штучного інтелекту (ШІ) починають закладатися підвалини систем розуміння природної мови. Ідея про вихід машинного перекладу за формальні лінгвістичні межі мови, як-от рівня синтаксису або семантики, у напрямку «розуміння» системами знань про «реальний світ» була висунута в Університеті Карнегі-Меллона, Піттсбург⁷. Цю подію можна вважати першим кроком до розробки підходу в перекладі через проміжні репрезентації на основі позамовних «універсальних» елементів.

Великим поштовхом до відродження досліджень у галузі МП та наступної епохи розвитку послугувала поява комерційних систем перекладу, тобто таких, що були розроблені з огляду на хоча й грубу лінгвістичну якість вихідного результату їхньої роботи, але водночас і їхню високу рентабельність. Основними виробниками стали американські (Weidner and Logos, ALPSystems, Siemens) та японські (Toshiba, Fujits, Mitsubishi) технологічні компанії [34, с. 8]. Поява систем МП на ринку призвела до значного зростання обізнаності громадськості щодо важливості розробки інструментів МП.

Лише на початку 2000-их років достатній рівень технологій надав змогу здійснювати базовий машинний переклад через комп'ютерні системи у їхній формі, відомій нам зараз. Основним методом «навчання» комп'ютерів перекладати стало використання статистичних баз мови із залученням великого об'єму ручної праці [51]. Великим стрибком стала подія 2016 року, коли компанія Google запропонувала модель нейронного машинного навчання на основі попередніх досліджень [26; 37; 56; 66]. Метод навчання комп'ютерів за допомогою ШІ виявився значно ефективнішим і швидшим за статистичний підхід. Зміна вектору у машинному перекладі надала поштовх до розвитку нейронного машинного перекладу (надалі НМП), що ґрунтується на побудові та навчанні нейронних мереж [40, с. 67]. Нейронна мережа – це техніка машинного

⁷ Центр активних розробок у галузі МП на той час, де ініціаторами ідеї про новий вектор досліджень був запропонований у 1988 році дослідниками Ейїтіро Суміта та Цуцумі Ютака (Eiichiro Sumita та Yutaka Tsutsumi).

навчання, при якій приймаються кілька вхідних даних і очікуються конкретні вихідні дані. НМП передбачає обчислення очікуваних результатів нейронної мережі, а також за допомогою різноаспектних поетапних алгоритмів подальше підвищення точності вихідного продукту – тексту перекладу цільовою мовою⁸.

Беручи до уваги історію становлення МП, можемо окреслити наступні його етапи або підходи до процесу перекладу: машинний переклад на основі формальних правил мови (RBMТ⁹), статистичний машинний переклад (SMT¹⁰) нейронний машинний переклад (NMT¹¹). Перший метод спирається на значний за обсягом, заздалегідь визначений набір лінгвістичних правил. Однак цей метод МП наразі майже не використовується на практиці через його громіздкість та низьку ефективність. Другий підхід передбачає створення статистичної моделі зв'язків між словами, фразами та реченнями в заданому тексті й застосовує мовну модель для перетворення цих елементів на перекладений цільовою мовою текст. Третій підхід є сучасною стадією МП і надає змогу «навчати» автоматизовані моделі мовам із перспективою подальшого покращення їхньої роботи.

В основі методу НМП визначальним є принцип того, як працюють нейрони людського мозку і як людина опановує та засвоює матеріал. Функції активації¹² використовуються для моделювання різного стану людських нейронів (збуджений або нейтральний) штучними нейронами, які, з'єднуючись один з одним, утворюють структуру нейронної мережі [40, С. 30–32]. Процес навчання нейронної мережі включає в себе постійне коригування значень мережевих з'єднань. Основний метод навчання називається навчанням з виправленням помилок, тобто відхилення між очікуваним результатом і фактичним результатом нейронної мережі береться як еталон для посилення й зіставлення, а значення змінюються в такий спосіб, аби остаточно зменшити відхилення між отриманим

⁸ *англ.* target language»

⁹ *англ.* Rule-based machine translation

¹⁰ *англ.* Statistical machine translation

¹¹ *англ.* Neural machine translation

¹² Математичні функції, які застосовуються до виходу кожного нейрона та вводять нелінійність в модель, що дозволяє мережі вивчати складні шаблони даних.

перекладом та взірцем. Виходячи з механізму її роботи, нейронна мережа потребує навчання через дані із позначеними наперед відповідями.

З огляду на сучасну тенденцію в підвищенні ефективності роботи автоматичних систем МП, а особливо із залученням систем НМП, все гостріше постає питання про роль людини-асистента або людини-перекладача. Насправді, машинний і ручний переклад не знаходяться у конкурентному відношенні, а доповнюють один одного [36]. Замість виключення людини з процесу перекладу через її заміщення, системи МП за допомогою технологічних засобів створюють нову сферу послуг для майбутніх спеціалістів [25, с. 199]. Роль людини-перекладача набуває нових значень і вимагає ширше коло навичок, однак залишається невід'ємною частиною багатоаспектного процесу перекладу [34, с. 4], що зокрема окреслено викликами роботи систем НМП на сучасному етапі.

Отже, МП пройшов довгий процес розвитку аби набути сучасного свого вигляду – НМП. Історія становлення систем МП налічує неодноразові падіння та підйоми у розвитку галузі, що було зумовлено економічними, соціальними та практично зумовленими чинниками. Однак, наразі ця царина є невід'ємною складовою різних сфер життя людини. Удосконалення систем НМП наразі полягає у машинному навчанні, що здійснюється з безпосередньою участю людини-асистента, тим самим зумовлюючи роль ручної праці з обробки лінгвістичного матеріалу при автоматизованому перекладі.

1.2. Потенціал НМП у контексті міжкультурного спілкування

Нейронний машинний переклад став значним стрибком у розвитку МП для світового простору з огляду на механізми передачі інформації. У порівнянні із попередніми стадіями становлення МП можна виокремити безперечні переваги систем НМП на сучасному етапі формування, а також і виклики, що постають, зокрема, у контексті стилістичної специфіки текстів різними мовами. У нашому дослідженні розглядаємо китайсько-український переклад публіцистичних текстів та його синтаксичний аспект, спираючись на гострі питання перекладознавства та лінгвістичну відмінність між мовною парою.

На нашу думку, дослідники слушно зауважують про такі загальні переваги та особливості систем НМП [68; 72]: значне підвищення якості й адекватності вихідного тексту у порівнянні з попередніми розробками; реалізація «нульового» перекладу¹³ [35, с. 1258], одночасне опрацювання значних текстових масивів, здатність до контекстного перекладу, залучення простіших механізмів, що не вимагають високого ступеня розробки, у порівнянні з системами статистичного перекладу тощо. Такі аспекти вдосконалення процесу МП відкривають нові перспективи у перекладознавстві, однак покривають далеко не всі нюанси процесу перекладу.

Серед викликів, що складають проблеми для розв'язання дослідниками у НМП на сьогодні, виділяють: необхідність пост-редагування [68], недостатній рівень застосовності систем НМП у різних літературних сферах (стилях та жанрах), передача аспекту творчості в тексті та здатність до трансляції художньої образності у перекладі, низька стійкість до помилок і стабільність систем з огляду на окремі явища мови, такі як, зокрема, «полісемія» [20]; існування мов з малим об'ємом доступних баз текстів¹⁴ (корпусів) [51], точність підбору термінології та інші. У контексті машинного китайсько-українського перекладу істотними труднощами є розробка більших за обсягом корпусів для навчання систем та складання кращих за точністю мовних моделей, адаптація систем НМП до стилістичної специфіки текстів відповідними мовами. На нашу думку, задля вирішення цих питань обов'язковим є дослідження методів отримання більш стилістично гнучких систем китайсько-українського НМП, що ґрунтується на вивченні характерних рис систем мов. Ми беремо до уваги цей аспект в контексті підвищення ефективності систем МП як суто лінгвістичний. Для більш вузького фокусу дослідження, розглядаємо синтаксичний рівень китайської та української мов у контексті НМП.

Щодо стилістично зумовлених труднощів в процесі НМП, це дослідження має на меті розгляд публіцистичних текстів. Дослідниками неодноразово було

¹³ *англ.* zero-shot

¹⁴ *англ.* low-resource languages

вивчено спеціальний переклад за допомогою систем МП та підтверджено високу його ефективність [19; 28]. Спеціальний переклад включає в себе переклад текстів, що застосовуються у специфічних сферах, наприклад наукового, офіційно-ділового стилів. Таким текстам властива формальність, шаблонність, насиченість уніфікованою термінологією та часто повторюваною лексикою. Відповідна характеристика текстів, що зменшує варіативність вихідного тексту, разом з наявністю великих об'ємів текстів перекладів – матеріалу для машинного навчання – здійснених людиною-перекладачем, спрощують процес автоматизованого перекладу.

У зіставленні з спеціальним перекладом системами НМП, дещо протилежною постає ситуація з художнім. Тексти цього стилю тяжіють до індивідуального авторського стилю, певних художніх форм, образної передачі змісту, суб'єктивного змалювання дійсності, що вимагає глибинного осмислення семантичного навантаження тексту під час перекладу. Системи НМП, засновані на формальних правилах мов та статистичних даних, нездатні до опрацювання текстів за наведеними аспектами, що робить процес високоякісного художнього перекладу за їхньою допомогою неможливим. Наявні зараз системи НМП, зокрема такі, що побудовані на основі принципів ШІ і нейронних мереж, можуть певним чином здійснювати художній переклад [24]. Механізм полягає в тому, що мережа перебирає надану інформацію по сегментах та генерує наступне слово на основі сусідніх. Згодом вона «дізнається», на яких словах зосередитися, а де створити найкращі контекстуальні зв'язки на основі попередніх прикладів. Цей процес є формою «глибокого навчання»¹⁵ і дозволяє системам перекладу поступово вдосконалюватися з часом. У НМП контекст дешифрування називається «вирівнюванням»¹⁶ і відбувається в механізмі уваги, який розташований між кодером і декодером¹⁷ у системі.

¹⁵ *англ.* deep learning

¹⁶ *англ.* alignment

¹⁷ Кодер – механізм прийняття та розпізнавання тексту, декодер – алгоритм виведення опрацьованого результату цільовою мовою.

Процес розробки нейронних мереж для здійснення художнього перекладу через «розуміння» тексту значно виходить за межі лінгвістики. На нашу думку, більшу увагу варто приділити саме НМП публіцистичних текстів та мовним факторам, що впливають на якість опрацювання текстів комп'ютерними засобами. Публіцистичний стиль відходить від суворой формальності літературної мови, але й не відрізняється високим ступенем художності. Тексти цього стилю тяжіють до розмовного, де залучення частотних лексичних одиниць і граматичних конструкцій реалізує мову щоденного ужитку, що й робить його наразі більш актуальним для вивчення у контексті НМП. Розробка ефективніших механізмів автоматичного перекладу публіцистичних текстів умотивована спрощенням процесу міжкультурного спілкування та залучення здобутків такого дослідження до ширшої сфери використання у повсякденні, наприклад, у веденні побутових розмов, миттєвому перекладі інтерфейсу програм або локалізації бізнесу.

Міжкультурна комунікація, що зокрема реалізується через переклад текстів різними мовами, має безперечну соціальну й культурну цінність. Дослідниками було глибоко досліджено питання перекладу китайською у парі з окремими іншими найбільш поширеними мовами в світі, зокрема англійською [2]. Спеціальний переклад пари китайської та української мов було детально розглянуто з огляду на: синтаксичний аспект у перекладі інформаційно-новинних текстів [3], в перекладі нормативної та технічної документації [17], лексико-граматичні особливості перекладу наукової літератури [1; 7], військової [16], медичної термінології [12]. Питання художнього перекладу було піднято за такими напрямками: переклад авторських творів [8], кінематографу [6] тощо. Крім того, в увазі дослідників опинилися й проблеми перекладу, зумовлені культурою: методи передачі китайських слів засобами української мови [5], переклад номінацій їжі [15]. Ближче вивчення перелічених аспектів мов у процесі перекладу розкривають сутність їхньої будови, відмінностей, схожих рис і засобів точної передачі інформації у міжкультурному спілкуванні.

З огляду на покращення механізмів спілкування, у контексті машинного перекладу існує необхідність в подальших дослідженнях, де до уваги братимуться сучасні технологічні засоби аналізу та актуальні питання перекладу. Публіцистичні тексти є наразі матеріалом, який реалізує якнайширше тематичний спектр міжкультурного спілкування. Згідно із закономірністю Зіпфа про розподіл у жанрах, функціональних стилях та терміносистемах, з першої тисячі найчастотніших слів певної мови близько 62% залучено у текстах публіцистичного стилю [34, с. 85]. Виходячи з цього, вивчення перекладу текстів відповідного стилю не тільки надає можливість здійснення ефективних розробок у МП, але й становить найбільшу цінність в контексті зміцнення позицій української мови у просторі міжнаціонального обміну й розвитку за допомогою перекладу.

Отже, від рівня ефективності роботи систем НМП залежить якість міжкультурного спілкування. Удосконалення їхніх механізмів є актуальним для специфічних сфер життя, а саме професійних галузей, літератури тощо, однак здатність систем НМП до відтворення мовлення щоденного спілкування носіїв є одним з найгостріших питань сьогодення. Публіцистичний репрезентує найбільш частотні одиниці та структури мовлення, а тому тексти цього стилю виступають найбільш вагомим матеріалом у контексті китайсько-українського перекладу як елементу міжкультурної комунікації.

1.3. Дослідження явища перекладацьких трансформацій та універсальї в контексті НМП

Синтаксичний рівень становить одну з найбільших відмінностей між мовами. Коректний підбір конструкцій на рівні речень безпосередньо впливає на якість перекладу, а отже, потребує уваги як один з компонентів процесу МП. Синтаксис є таким пластом розуміння мовлення, який закодує в собі інформацію через структуру й регламентованість форми, зумовлену нормами. Тобто окремі мови з часом формують унікальний спосіб вираження змісту через співвідношення і взаємозв'язок між словами, їхній порядок у реченнях та

менших синтаксичних одиницях. У цьому сенсі синтаксис відіграє таку ж за значенням роль у вираженні сенсу повідомлення, як і добір лексичних одиниць, тобто семантичних структур. За спільними рисами мови поділяються на типи: ізолюючі або кореневі, флективні, аглютинативні. Однак при цьому вони зберігають і такі особливості, що є унікальними і характерні лише окремим з них. Важливість розгляду аспекту синтаксису у НМП полягає в необхідності покращення способів розпізнавання такого «коду» й використання його як надмовного пласту для передачі змісту у процесі перекладу між структурно відмінними мовами.

Типологічна подібність та відмінність будови мов підіймає питання про виокремлення загальних тенденцій у перекладі та розробку конкретних підходів, що можуть мати гнучке застосування до декількох пар мов одночасно на основі спільних закономірностей між ними. Конкретні способи перекладу мають визначення перекладацьких трансформацій, а методи значно ширшого застосування, тобто що можуть бути залучені до більшого числа мовних пар, – перекладацьких універсалій.

Такі особливості та тенденції перекладу китайською на синтаксичному рівні, як перекладацькі трансформації, поставали в увазі багатьох дослідників в Україні [13] та світі [53; 65; 77]. В основі їхніх праць покладено аналіз синтаксичної будови пар мов та визначення розбіжностей між ними. Ми також розглядаємо синтаксичні перекладацькі трансформації для визначення потенційних викликів у перекладі системою НМП, однак актуальність методології полягає у залученні корпусного аналізу саме до китайсько-української мовної пари.

У фокусі нашого дослідження перекладацькі універсалії розглядаються в контексті машинного перекладу як, за визначенням В. Фроулі, «третій код» [31, с. 45] Тобто виокремлення на основі особливостей систем мов певних «правил» або «законів» і частотно повторюваних конструкцій у процесі перекладу підвищує якість вихідного тексту. Серед основних перекладацьких універсалій виділяють спрощення та нормалізацію [49, с. 2]. Спрощення полягає у тенденції

до спрощення засобів мовлення на різних рівнях цільової мови, а нормалізація є тяжінням до шаблонів мови перекладу, що знаходить відображення у використанні типових граматичних структур, пунктуації та сталих виразів або кліше.

Незначний масштаб нашого дослідження та низька комплексність методології не дають змогу про визначення перекладацьких універсалій у китайсько-українському перекладі. Тим не менш, при аналізі окремого випадку НМП китайсько-українського публіцистичного тексту ми маємо можливість ідентифікації вище наведених перекладацьких закономірностей на практиці. Висновки нашої праці можуть слугувати основою для подальших лінгвістичних досліджень у напрямку вивчення китайсько-української мовної пари.

Проведене нами дослідження є малим за обсягом, але, тим не менш, вагомим вкладом у поле досліджень НМП на теренах України. З огляду на останні тенденції розвитку цього напрямку, українська мова має бути ширше вивчена та представлена в контексті пар з найбільш поширеними світовими мовами. У такий спосіб можливо збільшити об'єм матеріалу, який наявний для лінгвістичних розробок міжнародними науковцями, а отже, ще більше залучити українську мову у світову систему розробки інструментарію НМП.

Отже, значна типологічна відмінність китайської та української підіймає питання вивчення синтаксичної будови двох мов через компаративний підхід. Визначення розбіжностей та тенденцій структурного плану є матеріалом для подальших досліджень у напрямку перекладу. З огляду на останні надбання у цій галузі, лінгвістичні праці з аналізу синтаксису китайської та української, закономірностей у перекладі між цією мовною парою та ступеня залучення у цьому процесі універсалій перекладу слугують ключовим інструментом підвищення якості роботи систем НМП.

1.4. Синтаксичні трансформації у процесі НМП китайсько-українських публіцистичних текстів

Для окреслення передбачуваних нами результатів проведення корпусного аналізу текстів китайською та українською варто звернутися до основних розбіжностей синтаксичної будови відповідних мов. На основі значних відмінностей у структурі можливо на практиці прослідкувати перекладацькі трансформації, а також реалізацію системами НМП перекладацьких універсалій [46].

Першим аспектом системи синтаксису мов, який тяжіє більше до статистичних даних, але вартий уваги з лінгвістичної точки зору, це довжина речень. Однією з характеристик мовлення представниками різних культур та народів є обсяг лексичного і граматичного матеріалу, необхідний для висловлення завершеної думки, тобто речення. Основна розбіжність між китайською та українською полягає у формі представлення слів як окремих лексем, що сполучаються для висловлення змісту повідомлення. Тобто у китайській цей процес відбувається за допомогою одного або сполучення ієрогліфів, кожен з яких є сукупністю фонетичного, семантичного і графічного рівнів мови. Українською утворюються формально довші речення через фонетичну характеристику мови – наявність літер, складів та слів як окремих одиниць. Дослідники зазначають, що китайська за формою є більш лаконічною і не потребує багато слів для вираження певної ідеї в потоці мовлення [61]. Можемо зробити припущення про те, що, у зіставленні, середня довжина речення китайською менша, аніж у відповідному україномовному сегменті паралельного корпусу. Крім того, за цим передбаченням можливим стає спостереження універсалій, а саме спрощення чи нормалізації, що може бути побачене через залучення більшої або меншої кількості одиниць мовлення у висловлюванні, формування довших або коротших речень при перекладі з китайською українською.

Іншою рисою будови речення, яка визначальна у процесі перекладу, є порядок слів. Основою речення є структура, що включає в себе суб'єкт, дію й

об'єкт, а також визначає їхнє місце стосовно один одного. В китайській мові визначають дві схеми порядку слів у реченні: типу SVO та SOV, де S¹⁸ – суб'єкт, V¹⁹ – дія, O²⁰ – об'єкт [71]. В системі української мови речення здебільшого тяжіють до порядку слів SVO, хоча через флективність мови, їхнє місце не є настільки чітко визначеним, як в китайській, представником групи ізолюючого або кореневого типу. Наприклад, речення «Він зробив це» можна виразити як «Він це зробив», де сенс не змінюється від перестановки об'єкта та дії відносно один одного. Перший варіант відповідає схемі речення SVO, а другий – SOV, серед яких останній має меншу частоту використання та здебільшого обирається мовцем у певній комунікативній ситуації, наприклад для вираження інтонації та наголосу на окремому члені речення, додавання емоційного забарвлення тощо [14]. В китайській мові аналогічне речення «他做了这件事情。」 не має варіацій з безпосередньою зміною порядку наявних членів речення. Однак може бути залучена синтаксична конструкція з ієрогліфом «把», що перетворює наведене речення схеми SVO на приклад порядку SOV: «他把这件事情做了。」

З огляду на проілюстроване явище, ми припускаємо статистичне відображення розбіжності між мовами за частотою використання тієї чи іншої схеми речення, що маємо змогу простежити в результаті здійснення корпусного аналізу текстів. Тобто, виходячи з описаної характеристики порядку членів речення в двох мовах, передбачаємо, що під час перекладу речення вихідною китайською мовою із порядком слів SVO та SOV трансформуватимуться за принципом SVO у тексті цільовою мовою – українською.

Заглиблюючись у структуру «дерева» речення, а саме наступного рівня його поділу за фразами, дослідниками означено переважання в структурі китайської дієслівних фраз [71] та тяжіння української мови до формування іменникових фраз [4, с. 79]. Наведена розбіжність ілюструє ключові риси зображення явищ дійсності через типологічно зумовлені засоби мовлення

¹⁸ *англ. subject*

¹⁹ *англ. verb*

²⁰ *англ. object*

(відсутність або наявність флексій, словозміни), зокрема, і як особливість мовної картини світу носіїв відповідних мов, яка мала вплив на історичне формування засобів мови. З цього випливає, що в китайській мові дієслова посідають центральне місце у реченні та слугують єдиною ланкою між іншими компонентами. З точки зору світосприйняття, китайці транслюють дійсність через дії, у порівнянні з іншими народами. В українській мові висловлювання формуються навколо іменних частин мов (іменник, прикметник, займенник, числівник) для опису предметів та явищ, а дієслова надають характеристику подій. До прикладу, речення «他 每天 早上 都 去 跑步 。」 в оптимальному перекладі українською звучить як «Кожен ранок він йде на пробіжку.» В першому варіанті бачимо використання двох дієслів поспіль на позначення дії та мети – «去» та «跑步». В еквівалентному другому варіанті вихідною мовою дія «йде» сполучена через прийменник з іменником «пробіжку». Тобто для вираження змісту одного й того самого повідомлення двома мовами залучаються різні частини мови, а саме у варіанті речення китайською простежуємо сполучення двох дієслів, а в перекладі українською – іменник, якому передують дієслово та прийменник.

Варто зауважити, що співвідношення між іменниковими та дієслівними фразами в реченнях обома мовами не є статичним поняттям, адже воно варіюється відносно функціонального стилю. Тексти офіційно-ділового або наукового стилю фокусується на лаконічній подачі інформації, сповненої певних понять, які виражаються через іменні частини мови, а саме іменники. На противагу, художній стиль передбачає більшу кількість описів зображуваного, образного тлумачення понять та специфічного оформлення подій та явищ через залучення прикметників та дієслів. У публіцистичних текстах спостерігаються найбільш частотні шаблони щоденного мовлення, що не зумовлено лише конкретною сферою діяльності людини. За нашим припущенням, тексти саме цього стилю мають відображати загальні закономірності мов, а тому слугують універсальним матеріалом для нашого напрямку дослідження.

Ще однією характеристикою зв'язку компонентів речення, а саме окремих граматичних основ в ньому, є наявність підрядних частин [44]. Дослідники зазначають, що у порівнянні з іншими мовами, такими як англійська, китайська має меншу частотність підрядних частин складного речення. За структурою, вона тяжіє до формування складносурядних речень, або ж безособових складних речень. Синтаксична будова української характеризується широким залученням підрядних різного типу зв'язку. Серед науковців поки що не має одностайного рішення, чи відрізняється китайська за цим аспектом з іншими мовами. Однак нами розглядається це питання задля того, щоб на практиці перевірити наведену теорію, визначити вагомість результату у контексті публіцистичного тексту дослідженими мовами. Крім того, розбіжність за частотністю утворення підрядних частин саме у китайсько-українській мовній парі може бути підґрунтям для дослідження перекладацьких універсалій у напрямку спрощення або ж поширення речення із додаванням другорядних членів речення. Результати такого аналізу можуть бути залучені в удосконаленні систем НМП в контексті як китайської та української, так й інших структурно відмінних мов.

Отже, загальні теоретичні розбіжності у синтаксичній будові китайської та української окреслюють параметри компаративного аналізу двох мов. Такі відмінності потребують конкретного підтвердження при розгляді текстового матеріалу задля подальшого визначення трансформацій та механізмів НМП між китайською та українською.

Висновки до Розділу 1

У Розділі 1 наводимо теоретичні засади МП як окремого відгалуження перекладознавства. Через історичний розріз розвитку цього напрямку окреслюємо зміну механізмів систем МП, які поступово набули форми нейронного машинного перекладу – НМП. Інструменти здійснення МП спочатку засновані на формальних правилах мови були вдосконалені через залучення алгоритмів статистичного аналізу мовного матеріалу та наразі досягли етапу розвитку, який полягає в сполученні різних аспектів перекладу через треновані нейронні мережі.

Закладені в системи НМП принципи роботи хоча і стали значним стрибком у галузі МП завдяки таким перевагам, як: значне підвищення якості й адекватності вихідного тексту у порівнянні з попередніми розробками; реалізація «нульового» перекладу, одночасне опрацювання значних текстових масивів, здатність до контекстного перекладу, залучення простіших механізмів, що не вимагають високого ступеня розробки, у порівнянні з системами статистичного перекладу тощо. Однак сучасний етап розвитку МП все ще не задовольняє всі потреби високоякісного перекладу, а тому напрямок його вдосконалення полягає у таких аспектах: необхідність пост-редагування, недостатній рівень застосовності систем НМП у різних сферах, передача аспекту творчості в тексті та здатність до трансляції художньої образності у перекладі, низька стійкість до помилок і стабільність систем з огляду на окремі явища мови, такі як, зокрема, «полісемія»; існування мов з малим об'ємом доступних баз текстів (корпусів), точність підбору термінології та інших. Ключовим способом визначення недоліків систем НМП та пропонування еталонних текстів перекладу наразі є залучення людини в процес пост-редагування перекладів цільовою мовою.

Вдосконалення систем НМП, зокрема в контексті китайської та української мов, вимагає значний обсяг мовного матеріалу та інформації про формальну характеристику мов, що зумовлює актуальність лінгвістичних досліджень у цій галузі. Системи НМП наразі є активним інструментом міжкультурної комунікації, засобом взаєморозуміння між представниками різних

країн та культур. Китайська та українська є двома типологічно відмінними мовами, що виявляє їхню цінність як об'єкта досліджень за такими аспектами: 1) з точки зору вдосконалення автоматизованого перекладу у цій конкретній парі мов, 2) залучення результатів вивчення формальної складової їхньої будови для використання в розробці універсальних стратегій у багатомовному перекладі.

Синтаксис є не менш важливим планом вираження, ніж лексика, оскільки форма вираження повідомлення також оформлює його цілісний зміст. Закономірними змінами форми в процесі перекладу є перекладацькі трансформації (трансформування синтаксичної структури в нашому випадку) та універсалії (спрощення та нормалізація). Перекладацькі трансформації зокрема можуть бути унікальними для конкретної мовної пари, виходячи з чого ми визначаємо аспекти, за якими, за нашим припущенням, реалізуються зміни синтаксису в контексті китайсько-українського перекладу: довжина речень, порядок слів, частотність утворення дієслівних та іменникових фраз, формування підрядних частин складного речення. Перекладацькі універсалії можемо тоді простежити у загальних закономірностях спрощення чи нормалізації синтаксичної будови в тексті цільовою мовою.

Публіцистичний текст виступає як оптимальний матеріал для вивчення з огляду на його стилістичні особливості. Серед таких рис зазначаємо тяжіння до частотних граматичних конструкцій, фокусування на широкому тематичному спектрі, відсутність зумовленості лише конкретною сферою діяльності, наближеність до побутового мовлення носіїв тощо. Висвітлені нами висновки слугують основою для здійснення практичного дослідження особливостей синтаксичної будови китайської та української, а також простеження їхнього впливу на процес перекладу публіцистичних текстів, зокрема за допомогою систем НМП.

РОЗДІЛ 2. КОРПУСНЕ ДОСЛІДЖЕННЯ СИНТАКСИЧНИХ ОСОБЛИВОСТЕЙ МОВ В КОНТЕКСТІ НМП

2.1. Корпусна лінгвістика як наука: застосування корпусів у НМП

Корпусна лінгвістика — це методологічний підхід, який займає емпіричну позицію щодо до вивчення мови. Він ґрунтується на якісному чи кількісному аналізі масиву письмових текстів або транскрипцій спонтанного чи напів спонтанного мовлення. Корпусні лінгвістичні методи мають потенційно значний вплив на теорію, оскільки за їхньою допомогою можливо підтвердити або спростити теоретичні припущення [18].

Поняття корпусу за своїм визначенням є сукупністю текстів, які скопійовані відповідно до заздалегідь запланованих критеріїв і систематично зберігаються в електронному вигляді. Інакше кажучи, корпус – це набір текстів у машиночитаному форматі, які можуть бути автоматично проаналізовані та оброблені програмними пакетами [30]. Корпуси повинні мати деякі специфікації, щоб повністю відповідати окремому дослідженню. Вирізняють різні типи корпусів відповідно до такої їхньої характеристики, як: кількості залучених мов (одномовний, паралельний, багатомовний, компаративний або контрастивний), розглянутого часового простору (діахронічний та синхронічний), актуальності даних (статичний, моніторинговий або динамічний) та функції (навчальний, помилково-анотований, спеціалізований тощо). В нашому дослідженні для аналізу синтаксичних особливостей китайської та української мов було залучено паралельний двомовний китайсько-український корпус.

Паралельний корпус є найбільш типовим для проведення досліджень у галузі перекладу. Як і випливає з його назви, паралельний корпус визначається як принаймні два тексти, які вирівняні²¹ або сполучені речення за реченням або абзац за абзацом таким чином, що їх можна порівнювати та зіставляти безпосередньо [30, с. 217]. Крім того, паралельні корпуси можуть бути або

²¹ *англ.* aligned

двомовними, або багатомовними. Двомовність передбачає залучення мови А (вихідний текст) і мови В (мови перекладу або тексту цільовою мовою), тоді як у багатомовному паралельному корпусі є одна мова оригіналу та більше ніж одна мова перекладу. Двомовні паралельні корпуси є дієвим орієнтиром для порівняння перекладу певною цільовою мовою; однак багатомовні паралельні корпуси є основою для порівняння більше ніж одного цільового тексту одночасно, тобто є фактично подальшим поєднанням окремих двомовних корпусів в одну систему паралелей.

В основу корпусного підходу в лінгвістичних дослідженнях покладено автентичність мови [39]. Завдяки широкому використанню комп'ютерів та Інтернету письмові корпуси, які часто складаються з літературних і публіцистичних текстів, зараз широко доступні в мережі різними мовами. Наприклад, існують корпуси з кількох мільйонів слів навіть для мов, що налічують невелику кількість носіїв у світі. Джерелами текстів є підручники, художні твори, наукові журнали, телефонні розмови тощо, тобто будь-яка комунікативна побутова або тематично спеціалізована ситуація, в якій відбувається будь-яка мовленнєва діяльність, може утворювати корпус.

Незважаючи на значний прогрес у складанні корпусів, їхній обсяг значної багатьма мовами, що є відображенням активного мовлення носіїв і джерелом для дослідження загальних тенденцій та унікальних особливостей їхнього спілкування, все ще вкрай обмежений. Науковці А. Пранав, С. Чун та М. Вегерс [60, с. 1] зауважують, що 85% мовного матеріалу, який є частиною Консорціуму лінгвістичних даних, представляють лише п'ять мов, тобто: англійську, китайську, арабську, іспанську та японську. З цього випливає, що об'єм текстів іншими мовами, зокрема українською, потребує розширення задля реалізації повноцінної міжкультурної комунікації в контексті перекладу.

Контрастивний корпусний аналіз між типологічно наближеними мовами, як-от для пар мов англійська й іспанська, англійська та французька, були у фокусі багатьох досліджень [38]. На їхній основі можливим є вивчення методів НМП і навчання систем з огляду на глобальне значення цих мов та значну кількість

носіїв у світі. Порівняно з типологічно спорідненими мовами міжмовний контраст таких генетично відмінних мов, як англійська та китайська, китайська та українська, є більш складним, але відкриває нові перспективи для майбутніх наукових розробок в галузі НМП [76, с. 55]. Лінгвістичні теорії, засновані на тісно пов'язаних мовних парах, обмежені у можливості узагальнення результатів. Оскільки висновки, які здаються одностайними при вивченні таких мов, можуть бути спростовані переконливими доказами при розгляді абсолютно різних за типологією мов. Особливо гострим це питання постає у дослідженнях з визначення перекладацьких трансформацій та віднайдення перекладацьких універсалій.

У нашому дослідженні китайсько-українська мовна пара постає матеріалом для вивчення синтаксичного аспекту машинного перекладу. Типологічна відмінність двох мов дає можливість якнайкраще проаналізувати особливості в будові синтаксичних конструкцій кожної з них, дослідити на практиці мовний контраст через корпусний аналіз, визначити ймовірні виклики, що постають перед системою НМП при здійсненні перекладу публіцистичних текстів. Детальне вивчення та опис паралельних корпусів сприяють створенню прийнятних перекладів поточними системами НМП, оскільки такі об'єми текстів зазвичай включають приклади стандартних мовних моделей і методів перекладу. Корпуси пропонують стандартні мовні структури з точки зору лексики, пунктуації та синтаксичних і текстових моделей, які є найбільш оптимальним варіантом для тексту перекладу цільовою мовою [49, с. 20].

Отже, засади корпусної лінгвістики як науки дають змогу проведення ґрунтовних досліджень з контрастивного аналізу мов. Значні обсяги текстів, об'єднаних у корпуси функціонально та тематично, є цінним матеріалом для вивчення процесу перекладу, розгляду лінгвістичних аспектів міжкультурної комунікації, зокрема у мовній парі китайська-українська, здійснення широкоформатного дослідження мовленнєвої діяльності носіїв різних мов у різних сферах життя.

2.2. Методологія корпусного лінгвістичного аналізу та комп'ютерного опрацювання текстів

Корпуси, зокрема мультимовні та паралельні, становлять виклик для подальшого опрацювання текстів, оскільки вони є «сирим» матеріалом і вимагають здійснення додаткової обробки. Такий процес включає автоматичне розпізнавання мовлення, загальний аналіз, машинний переклад, пошук і вилучення інформації та обчислювальна обробка [18, с. 4]. В основі наступних етапів роботи з корпусом лежить токенізація²², тобто сегментація текстів на окремі елементи та визначення їхнього типу, приналежності до частини мови, синтаксичної ролі тощо. У процесі токенізації утворюються токени з анотацією, тобто вони визначаються як: мовні та пунктуаційні знаки, пробіли, символи тощо. Через здійснення відповідного розбору текстів можливо отримати готовий матеріал, «зрозумілий» за формою для комп'ютерних систем. Варто зауважити, що унікальні риси конкретної мови можуть спровокувати певні складнощі обробки. Так, текст українською мовою, у контрасті з китайською, вирізняється наявністю слів з великої літери, що ставить виклик обчислювальним системам через відмінність коду для літер різного регістру. Зважаючи на це, необхідно вирівняти одиниці тексту аби уникнути подальших помилок аналізу.

Актуальний в рамках нашої праці, синтаксичний корпусний аналіз проводиться на основі наступного етапу обробки – парсингу²³. Процес парсингу полягає в аналізі синтаксичної структури текстів через розпізнавання послідовності сегментів, визначених на етапі токенізації, і зіставлення їхнього порядку з нормами відповідної мови, а отже і визначення синтаксичної ролі окремих елементів тексту. Формальні правила мови комп'ютер отримує через завантажену мовну модель²⁴. Результатом парсингу є «синтаксичні дерева», визначена структура речень розглянутої мови, що має зокрема схематичне

²² *англ.* tokenization

²³ *англ.* parsing

²⁴ Мовними моделями є обчислювальні моделі, розроблені для розуміння, генерування та прогнозування тексту чи мовлення певною мовою. Основна функція мовної моделі полягає у визначенні ймовірності послідовності слів або токенів, що робить її фундаментальним компонентом у різноманітних програмах обробки природної мови.

зображення. Це поняття стосується деревоподібних діаграм або «маркерів фраз», «парсів», які використовуються під час аналізу [58]. Наприклад, речення «Дівчинка сиділа на стільці» може бути представлено таким візуальним зображенням (див. рис. 1):

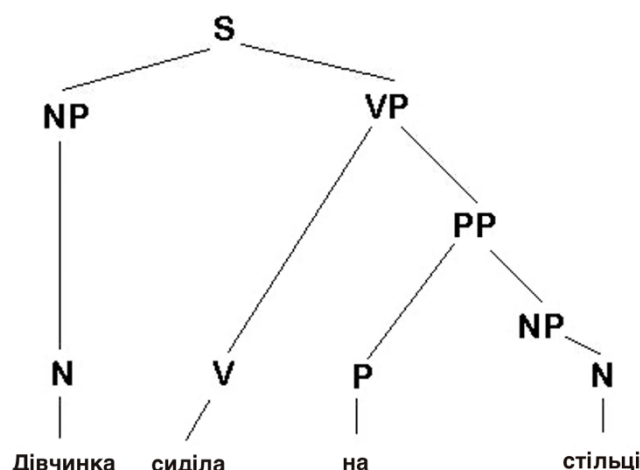


Рисунок 1. Схематичне зображення «синтаксичного дерева» речення

Джерело: розроблено автором на основі [45, с. 10].

Для простоти й зручності зображення «дерева» речення використовуються загальноприйняті скорочення на позначення його компонентів, назви яких залучають з термінології англійською мовою ²⁵ (див. список скорочень). На розглянутому рисунку 1 маємо такі елементи: S – речення, NP – іменникова фраза, VP – дієслівна фраза, PP – прийменникова фраза, N – іменник, V – дієслово, P – прийменник [45, с. 10]. Окремими латинськими літерами зазначаються сегменти речення за їхньою приналежністю до певної частини мови (іменник, дієслово, прийменник), а сполучення цих літер демонструють зв'язки між ними (іменникова, дієслівна та прийменникова фрази). Визначення компонентів речення у такий спосіб та їхня ідентифікація та здійснення парсингу необхідні для подальшого автоматизованого корпусного аналізу [32, с. 63].

Складання корпусів та їхня попередня обробка є клопітким та трудомістким ручним видом діяльності. Для зручності використання і

²⁵ Задля єдності термінології світового наукового простору в галузі лінгвістики застосовують терміни англійською: sentence, noun, verb, preposition, noun phrase, verbal phrase, prepositional phrase тощо.

поширення матеріалу досліджень, залучаються бази корпусів різними мовами з вільним веб-доступом, які є вже анотованими попередньо, а також містять метадані про тексти. Метадані є інформацією про функціональний стиль, джерела, об'єм текстів, соціолінгвістичні аспекти мовців, місце і час збору матеріалу тощо [18, с. 5]. Така характеристика текстів може бути використана у дослідженнях як урахування екзогенних, тобто позамовних чинників, що як впливають на мовленнєву діяльність окремих носіїв, так і породжують масштабні зміни природніх мов, викликають нові тенденції. Крім того, єдність формату токенизації всесвітньо створених корпусів є досить важливим з огляду на потребу їхнього швидкого поширення і легкого використання при аналізі значних обсягів даних різними мовами [11]. Загальноприйнятими стандартами анотації корпусів, за рекомендацією Worldwide Web Consortium (W3C) є Extensible Markup Language (XML) та Text Encoding Initiative (TEI) [18, с. 4]. У нашому дослідженні було використано паралельний китайсько-український корпус публіцистичних текстів TED20202 попередньо сегментований за стандартами XML, доступний на платформі Opus [70].

Корпус TED20202, залучений до нашого дослідження, складений дослідниками Н. Раймерс та І. Гуревич в рамках наукової праці «Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation» при Корнельському університеті [63]. Загальний багатомовний корпус складається з близько 4000 окремих промов TED, що доступні понад 100 мовами. Переклад скриптів був здійснений волонтерами з усього світу, котрі як безпосередні носії національних мов зробили свій вклад в створення вільного доступу до текстів різними мовами. Факт залучення людей до процесу перекладу, а не автоматичних систем, гарантує наявність у текстах перекладу актуальних одиниць мовлення для публіцистичного стилю.

Зручність використання корпусу TED20202 полягає в можливості обрати перекладені тексти, вирівняні між певною парою мов, у нашому випадку – китайською та українською. Обсяг залученого до нашого дослідження двомовного корпусу складає 12,925 речень, у яких наявні 33,508 токенів

китайською та 170,224 українською. Вихідною мовою корпусу є китайська, а українська виступає цільовою. Попри відносно незначний об'єм корпусу, з огляду на потреби дослідження широких тенденцій на рівні національних мов, за допомогою відповідного матеріалу ми маємо змогу визначити ключові розбіжності синтаксичної будови мов, оцінити якість корпусного дослідження в контексті компаративної лінгвістики між китайською та українською мовами.

Для опрацювання корпусу, а також проведення цілісного аналізу текстів використовується різноманітний інструментарій. Серед спеціалізованих засобів є веб-інструменти, програмне забезпечення, адаптоване для різних операційних систем, а також окремі бібліотеки ²⁶ для використання при написанні власних програм і певних функцій штучною мовою Python ²⁷ (див. додаток Г). Такі платформи, як Corpus-analysis.com, пропонують вибір інструментів комп'ютерних обробки відповідно до конкретних цілей і завдань [74]. У ході дослідження нами було залучено одну з найбільш функціональних і часто використовуваних бібліотек мови Python, яка спеціалізована на опрацюванні природніх мов ²⁸, – spaCy. Оскільки китайська, як мова ізолюючого типу, є більш складною для комп'ютерних систем з точки визначення частин мови і зв'язку між ними, то для опрацювання відповідного корпусу застосовано бібліотеку jieba. Інтегрованим середовищем для розробки ²⁹ власних функцій для аналізу корпусу нами було обрано Visual Studio Code.

Задля цілісного аналізу синтаксичної структури речень китайською та українською, а також виокремлення ключових рис будови мов на рівні синтаксису, було окреслено певні етапи роботи. Вони ґрунтуються на конкретних завданнях окремих функцій, за допомогою яких було здійснено аналіз паралельного корпусу:

- 1) визначити загальну довжину речень за кількістю мовних знаків та слів;

²⁶ Набори готових функцій та методів, що застосовують при написанні коду мовами програмування.

²⁷ Мова програмування, що зокрема широко застосовується для написання програм з опрацювання природніх мов комп'ютерними системами та складання мовних моделей.

²⁸ *англ.* natural language processing / NLP

²⁹ *англ.* integrated development environment / IDE

2) розбити речення на іменникові та дієслівні фрази (NP та VP), визначити їхнє співвідношення у досліджених мовах;

3) встановити найбільш частотний порядок слів у реченнях відповідно до структур SVO, SOV, окреслити їхнє статистичне співвідношення у кожній з двох мов;

4) віднайти частотність утворення складнопідрядних речень в китайській та українській як властивість будови мови.

Методологія здійснення конкретних завдань аналізу полягає в певних етапах роботи з корпусом:

1. Задля отримання загальних статистичних даних про корпус (середню довжину речень, кількість речень загалом та складних зокрема) залучаємо бібліотеки Python: `re` (для зчитування тексту), `sraCy` (для опрацювання обох мов) та `jieba` (спеціально для визначення структури китайського речення). Функції аналізу за наведеними параметрами ґрунтуються та таких елементах тексту, як пунктуаційні знаки та пробіли, що дозволяє визначити кінець кожного речення. Кількість складних речень підраховуємо через зчитування граматичних основ речень за означеними токенами корпусу. Наприклад, у реченні «我开头讲到臀部我最尾想讲：大家要行路同开会», яке в корпусі представлено без більшості знаків пунктуації, визначається як складне через наявність основ «我开头讲到», «我想讲» та «要行路同开会». Тобто ключовим фактором визначення типу речення є не пунктуація, а приналежність мовних одиниць до частин мови, зв'язок між ними.

2. Написання програм для опрацювання зв'язків компонентів речення між собою вимагає залучення мовних моделей (див. додаток Г). У нашому дослідженні застосовуємо моделі запроваджені бібліотекою `sraCy`: «zh_core_web_sm» для китайської та «uk_core_news_sm» для української. На їхній основі ми складаємо програми для підрахунку за такими параметрами: 1) випадки утворення порядку речення за схемами SVO/SOV; 2) кількість іменникових і дієслівних фраз (NP та VP); 3) кількість підрядних частин

складного речення. Програма завдяки наявним у бібліотеках модулів та заздалегідь визначеним токенам у корпусі визначає приналежність компонентів речення до частин мови, а через мовну модель зіставляє із синтаксичними структурами та їхнім визначенням. Частотність формування синтаксичних одиниць в текстах за відповідними параметрами аналізу підраховуємо вручну за отриманими даними програм:

1) частотність порядку слів SVO/SOV за формулою: кількість речень структури SVO/SOV поділено на загальну кількість речень, де наявні S, V, O, та помножуємо на 100%;

2) співвідношення NP та VP: ділимо значення NP на VP;

3) частотність утворення складних речень: кількість складних речень ділимо на загальну суму речень у корпусі;

4) частотність утворення підрядних частин: ділимо кількість підрядних частин у корпусі на загальну кількість складних речень.

3. Визначення динаміки зміни значень, а саме показників співвідношення в текстах NP та VP, полягає в розгляді окремих сегментів. За допомогою програми для розбиття двомовного корпусу на фрагменти отримуємо 15 файлів, де кожний містить 10 речень. Порожній рядок після кожного тексту прийнятий як аспект визначення кількості текстів. Для кожного сегменту тексту китайською та українською застосовуємо раніше прописану програму визначення кількості NP та VP (див. додаток Г.2.). Після цього залучаємо відповідний алгоритм визначення співвідношення NP : VP для кожного сегмента і фіксуємо значення для обидвох частин корпусу у хронологічному порядку. Однак через єдність формули підрахунку (NP поділено на VP) але постійне переважання першого показника в текстах українською, а другого – в реченнях китайською, отримані значення неможливо порівняти безпосередньо. Коефіцієнт NP : VP в корпусі українською набуває значень від 10 до 20, а китайською – від 0 до 1. Для паралельного зіставлення показників отримані значення коефіцієнта співвідношення нормалізовано, тобто поділено на найбільше значення, отримане

для корпусу кожної з мов. У результаті маємо значення коефіцієнта співвідношення від 0 до 1, які подаємо у вигляді графіка.

Отже, спираючись на методологію опрацювання двомовного корпусу, а також на основі текстового матеріалу TED20202 та зазначених завдань щодо розгляду синтаксичної структури речень китайської та української, ми визначаємо конкретні етапи здійснення корпусного аналізу, формуємо детальні алгоритми дій на кожному з них та обираємо показники та форму для подачі результатів дослідження. Зазначені інструменти та послідовні методи є основою для опрацювання текстів китайською та українською, визначення особливостей синтаксичної будови, зіставлення результатів та проведення їхнього контрастивного аналізу з огляду на застосування у НМП.

2.3. Визначення синтаксичних особливостей будови китайської та української мов на основі аналізу паралельного корпусу

Кожний з параметрів, визначених завданнями корпусного аналізу синтаксичної будови китайської та української мов, ми опрацьовуємо та обчислюємо на основі окремих функцій та їхніх сполучень мовою Python. Процес складання програм, а також детальні коментарі поетапних дій, доступні у вільному доступі через віддалений публічний репозиторій на платформі GitHub³⁰ [69]. Точні обчислені результати аналізу наводимо у таблиці нижче. Для зручності відображення обчислення значень за параметрами, приймаємо умовні позначення для певних показників:

n – загальна кількість речень у корпусі,

$K1$ – середня довжина речень за кількістю мовних знаків,

$K2$ – середня довжина речень за кількістю слів,

m – загальна кількість речень, де наявні S, V, O;

$P1$ – кількість речень структури SVO,

$P2$ – кількість речень структури SOV,

³⁰ Місце, де зберігаються дані у вигляді файли, доступ до яких наданий усім користувачам мережі.

NP – кількість іменникових фраз,

VP – кількість дієслівних фраз,

CS – загальна кількість складних речень у корпусі,

Sb – кількість підрядних частин у CS.

Розглянемо значення кожного з параметрів у контексті компаративного аналізу (див. табл. 1). Висновки про відмінність синтаксичної структури досліджених мов та їхній вплив на процес перекладу мають підтверджувати наші попередні припущення у Розділі 1, або ж спростовувати їх.

Таблиця 1

**Результати корпусного аналізу текстів китайською та українською
за синтаксичними параметрами**

Параметр аналізу	Отримані показники	
	Корпус китайською (TED2020.uk-zh.zh)	Корпус українською (TED2020.uk-zh.uk)
1	2	3
Загальна кількість речень у корпусі (n)	3514	13861
Середня довжина речень за кількістю мовних знаків (K1)	18.58	84.32
Середня довжина речень за кількістю слів (K2)	11.334	11.829
Загальна кількість речень, де наявні S, V, O (m)	2309	2313
Кількість речень структури SVO (P1)	2058	2312
Частотність структури SVO (P1/m * 100%)	89, 13%	99,96%

1	2	3
Кількість речень структури SOV (P2)	251	1
Частотність структури SOV ($P2/m * 100\%$)	10,87%	0,04%
Кількість іменникових фраз (NP)	38282	6386
Частотність формування NP на одне речення (NP / n)	10,89	0,46
Кількість дієслівних фраз (VP)	54091	498
Частотність формування VP на одне речення (VP / n)	15,39	0,036
Співвідношення випадків утворення NP до VP (NP : VP)	304:431	103:8
Загальна кількість складних речень у корпусі (CS)	1957	4987
Частотність утворення CS на загальну кількість речень у корпусі ($CS/n * 100\%$)	55,69%	35,98%
Підрядні частини CS (Sb)	1892	3750
Частотність утворення підрядних частин на одне CS (Sb/CS)	0,97	0,75

Джерело: розроблено автором на основі [70].

Першим і загальним статистичним параметром аналізу визначаємо кількість речень у кожній з половин паралельного корпусу. Результат може здатися на перший погляд нерелевантним за розбіжністю в показниках, де обидва мовами розглянуто однакову кількість текстів ідентичних за обсягом і змістом. Однак такі значення демонструють особливість будови кожної з мов. Китайська тяжіє до формування складносурядних речень, що представляють собою прості речення із завершеною думкою, які поєднуються між собою через кому. На противагу, в українській речення частіше розбиваються на прості, відокремлені крапкою або іншими знаками пунктуації на позначення кінця

речення. Наведена риса синтаксичної будови у процесі мовлення ініціює додаткові формальні зміни при перекладі. Окреслена закономірність формулювання речення проявляється в процесі перекладу через синтаксичну трансформацію – спрощення структури речення при українсько-китайському перекладі та розширення й ускладнення при зворотному. Таку зміну будови речення, зокрема, можна розглядати як вид перекладацької універсалії – спрощення (формування простих речень замість складних) та нормалізацію (відповідність нормам української мови) [59, с. 75].

Другим параметром ми окреслюємо середню довжину речень за підрахунком мовних знаків і слів. Кількість мовних знаків, тобто літер в українській мові та ієрогліфів у китайській, вказує на характер речення з точки зору обсягу залучених мовних одиниць для висловлення завершеної думки. Насправді, поняття слова в обидвох мовах є різним, оскільки в українській одне слово охоплює в середньому більше мовних знаків, аніж в китайській, де мовний знак поєднує фонетичне, семантичне і граматичне значення, а також є їхнім графічним відображенням [23, с. 5]. Як бачимо з результатів, довжина речень за першим аспектом значно відрізняється, однак розбіжність значення за другим є досить малою – приблизно 0,5 слова. Все ж навіть за таким мінімальним числом можна спостерігати властивість китайської як більш лаконічної, адже в її структурі в реченні залучається менша кількість як знаків, так і повнозначних слів [23, с. 10]. У контексті двостороннього публіцистичного перекладу, ефективним є залучення універсалії спрощення та розширення плану вираження задля охоплення повного сенсу висловлювань в потоці мовлення.

Наступним аспектом розгляду стала структура речення за порядком розміщення суб'єкта, дії та об'єкта в реченні (позначені як S, V, O відповідно), тобто така, що відповідає схемам SVO та SOV [71, с. 3]. Варто зауважити, що аналізу підлягали речення, де присутні кожен з елементів структур, тобто де граматична основа представлена підметом і присудком, що вказує на об'єкт дії. За наведеними значеннями у таблиці, у китайській та українській однаково переважає перший з двох типів порядку членів речення – SVO. Однак у корпусі

китайською трапляється і другий, коли в текстах українською мовою майже одноставно залучається лише перший. Отримані результати підтверджують нашу теорію про панування структури SVO в синтаксисі української мови, а також ілюструють аналогічну тенденцію у текстах китайською. Виходячи з практичного дослідження, можемо зауважити, що формування речень за другою зі схем властиве майже виключно для китайської. Цей аспект синтаксису й призводить до перекладацьких трансформацій у процесі двостороннього перекладу, які полягають у перестановці членів речення, особливо з китайської на українську, аналогічно до висновків, зроблених дослідниками Г. Чжао та Ц. Цао у контексті китайської та англійської [33, с. 396].

Відходячи від загальних зв'язків основних компонентів речення, ми беремо до уваги менші «гілки» синтаксичного «дерева» – іменникові та дієслівні фрази (позначені NP та VP відповідно), – які були обчислені за кількістю, частотою використання і співвідношенням. Значна розбіжність отриманих показників щодо загальної кількості випадків NP та VP вказують на ключові риси двох мов – лаконічність в китайській, а з іншого боку, та залучення більше залежних слів, поєднання їх у фрази в українській. Наше припущення щодо кількісної характеристики наявності іменникових та дієслівних фраз у досліджених мовах ґрунтується на висновках роботи Т. Г. Ліна. Науковець детально зазначає відмінності синтаксичної будови китайського та англійського речень [45, с. 13], що дозволяє нам окреслити гіпотезу щодо можливості прослідкувати розбіжності правил будови речень у китайській та українській.

За отриманими даними аналізу бачимо, що підтверджується наше припущення стосовно переважання NP у реченнях публіцистичних текстів українською мовою та VP – китайською. Це означає, що при перекладі відбуваються трансформації з заміни головних і залежних слів у реченні за їхньою приналежністю до частин мови. Наприклад, процеси можуть бути висловлені українською через іменники, а китайською – через дієслова. Описана закономірність може породжувати граматичні перекладацькі трансформації, що призводять до змін і на рівні синтаксису.

Задля виявлення більш чіткої динаміки значення співвідношення NP : VP у корпусах, нами було побудовано графічне відображення відповідного показника в окремих сегментах мовного матеріалу, а саме фрагментах по 10 речень (див. рис. 2). Показник NP : VP був нормалізований: коефіцієнт набуває значення від 0 до 1, що дає змогу компаративного аналізу корпусів. З такої точки зору через зіставлення окремих сегментів нами виявлено певну тенденцію: прослідковується збільшення кількості іменникових фраз NP в конкретній частині текстів однієї мови водночас з підвищенням значення VP в реченнях іншою і навпаки. Така закономірність висуває питання про явище взаємної компенсації NP та VP у контексті китайсько-українського перекладу. Виявлена нами особливість синтаксичної будови речень дослідженими мовами вимагає більш детального подальшого вивчення на більшому обсязі матеріалу, однак становить інтерес у визначенні чітких методів перекладу та перекладацьких трансформацій.

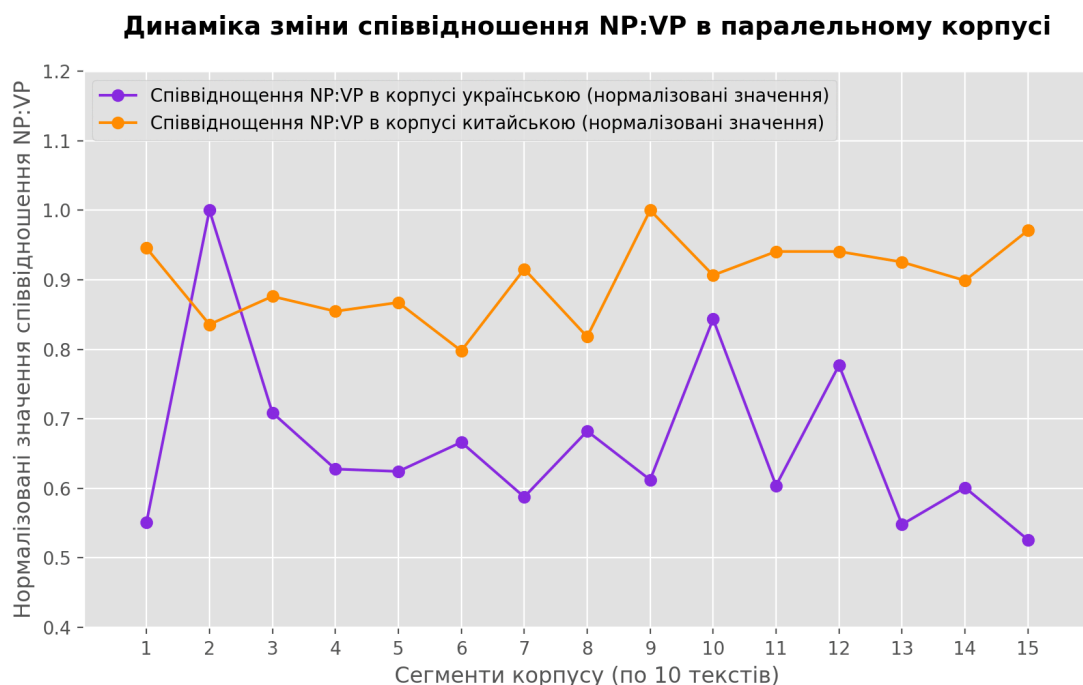


Рисунок 2. Графік співвідношення NP : VP у текстах китайською та українською в паралельному корпусі.

Джерело: розроблено автором на основі [70].

Іншим параметром аналізу є наявність підрядних частин (Sb) у складних реченнях, а саме їхнє обчислення за кількістю і частотою вживання відносно загального обсягу залучення складних речень у текстах корпусу. Згідно з висновками китайського дослідника Вей Цзясюе, складнопідрядні речення становлять виклик у процесі перекладу, наприклад між китайською та англійською, що вимагає розробку конкретних підходів задля збереження оригінального стилю і змісту через доречне додавання, віднімання синтаксичних конструкцій та інші методи [75, с. 289]. Зауважимо, що як було зазначено у першому пункті аналізу, в китайській прості речення тяжіють до об'єднання в складносурядні, а тому і частка складних речень від загальної кількості у всіх текстах є великою. За отриманими результатами, від загальної кількості складних речень (CS) у корпусах дослідженими мовами у текстах китайською частотність утворення підрядної частини на одне речення вища – 0.97, – аніж за аналогічний показник в матеріалі українською – 0.75. Такі значення дають змогу зробити припущення про синтаксичні перекладацькі трансформації, при яких підрядні частини китайського речення замінюються на сурядні або певні другорядні члени речення, як-от означення або обставина. Крім того, зменшення кількості підрядних частин речення у перекладеному тексті українською є проявом універсальї спрощення.

Отже, під час фактичного здійснення корпусного аналізу публіцистичних текстів китайською та українською нами було виявлено певні особливості та закономірності у синтаксисі кожної з мов. Отримані статистичні показники є підґрунтям для припущень щодо потенційних викликів та трансформацій у процесі китайсько-українського перекладу: 1) тяжіння речень до складносурядних речень китайською до поділу на окремі у вихідному тексті українською; 2) необхідність доречного додавання членів речення при перекладі на українську задля компенсації оригінального змісту; 3) заміна порядку речення типу SOV на SVO; 4) заміна дієслівних фраз у реченнях китайського тексту на іменникові в українському; 5) перетворення підрядних конструкцій оригінального речення на структурні одиниці іншого типу згідно з нормами

української мови; 6) залучення перекладацьких універсалій спрощення та нормалізації на рівні синтаксису.

Висновки до Розділу 2

У Розділі 2 ми окреслюємо основні засади корпусної лінгвістики та методології корпусного аналізу, а також матеріал для досліджень, які ця галузь пропонує сфері НМП на сучасному етапі його розвитку. У фокусі уваги постає питання використання двомовних корпусів для компаративного аналізу мов, а саме китайської та української, способів обробки текстів за допомогою комп'ютера, що полягає в таких підходах: токенізація, парсинг, складання та залучення функцій і методів мовою програмування Python. Для нашого дослідження ми дібрали інструменти здійснення аналізу (IDE Visual Studio Code, мова Python та бібліотеки re, spaCy й jieba), а також склали послідовний алгоритм його проведення за різними параметрами.

Застосовуючи корпус публіцистичних текстів TED2020, маємо конкретні статистичні дані про окремі особливості синтаксичної будови двох мов, а саме ми визначаємо такі тенденції: 1) тяжіння речень до складносурядних речень китайською до поділу на окремі у вихідному тексті українською; 2) необхідність доречного додавання членів речення при перекладі на українську задля компенсації оригінального змісту; 3) заміна порядку речення типу SOV на SVO; 4) заміна дієслівних фраз у реченнях китайського тексту на іменникові в українському; 5) перетворення підрядних конструкцій оригінального речення на структурні одиниці іншого типу згідно з нормами української мови; 6) залучення перекладацьких універсалій спрощення та нормалізації на рівні синтаксису.

Отримані дані є основою для передбачень нюансів процесу перекладу публіцистичних текстів китайською та українською, перспектив удосконалення механізму НМП. Для чіткого спостереження реалізації особливостей синтаксичної будови двох мов необхідний ручний розбір та аналіз роботи окремої системи НМП, здійснення ручного опрацювання тексту перекладу та його редагування не тільки на формальну синтаксичну будову української мови, але й з огляду на рецепцію тексту носієм, милозвучність та зрозумілість.

РОЗДІЛ 3. ОЦІНЮВАННЯ ЯКОСТІ РОБОТИ СИСТЕМ НМП ЗА СИНТАКСИЧНИМ АСПЕКТОМ У ПРОЦЕСІ КИТАЙСЬКО-УКРАЇНСЬКОГО ПУБЛІЦИСТИЧНОГО ПЕРЕКЛАДУ

3.1. Підходи до здійснення оцінювання систем НМП за лінгвістичним аспектом

Оцінка систем машинного перекладу є важливою галуззю досліджень для оптимізації роботи систем НМТ та підвищення їхньої ефективності. Сучасні системи НМП мають в своїй основі хоча й досить складну систему здійснення перекладу, але таку, що піддається навчанню і вимагає цього процесу для подальшого вдосконалення. З цього випливає, що систематичне проведення оцінки таких систем надає змогу ще більшого підвищення їхньої продуктивності й адаптації до динаміки мовної структури та засобів мовлення у процесі комунікації.

У нашому дослідженні залучаємо систему НМП DeepL, яка є одним з найбільш популярних інструментів НМП на сьогодні. Алгоритми опрацювання тексту, що лежать в основі її роботи, зумовлюють увагу на окремі аспекти процесу перекладу при оцінюванні його якості. По-перше, механізм здійснення перекладу системою DeepL описують як засноване на принципі уваги, або розгляду окремих одиниць та структур мовлення в контексті, їхньому оточенні. Тобто встановлення ймовірності використання того чи іншого відповідника цільовою мовою ґрунтується на частоті його використання у мовленнєвих сполуках з поданими одиницями вихідного тексту. По-друге, зазвичай система НМП використовує архітектуру кодер-декодер і генерує цільовий переклад зліва направо (див. рис. 3). Незважаючи на значні успіхи, моделі НМП страждають від певних недоліків [41]. Однією з найпомітніших є проблема незбалансованих виходів, де префікси перекладу передбачаються краще, ніж суфікси [47] або вибір

системою наступних слів не відповідає якісному адекватному перекладу через неточне передбачення на основі попередніх компонентів тексту [42].

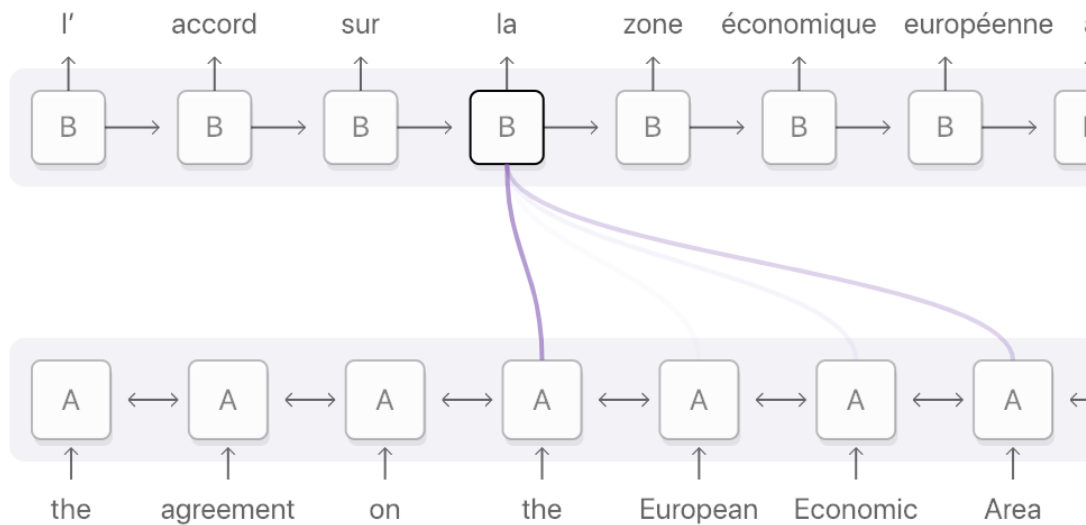


Рисунок 3. Механізм архітектури «уважних мереж кодерів-декодерів», закладений в основу системи НМП DeepL [55].

Існує цілий ряд підходів до оцінки систем НМП, наприклад: оцінка людиною, автоматизоване оцінювання, гібридне, на основі чітко визначених завдань тощо [50]. З точки зору лінгвістики, оцінювання систем НМП ґрунтується на детальному розгляді різних мовних аспектів вихідного тексту та всебічному аналізі якості як за формальною будовою і правилами мови, так і з огляду на сприйняття тексту реципієнтом у потоці мовлення [73]. Тобто робота дослідників-лінгвістів полягає у визначенні багатопланових помилок та недоліків систем НМП задля подальшого покращення їхньої роботи.

В цьому дослідженні приймаємо стратегію гібридного оцінювання якості НМП: здійснення автоматизованого оцінювання та пост-обробки людиною. Підхід оцінювання через залучення окремої людини для опрацювання вихідного тексту надає змогу визначити недоліки роботи НМП з огляду на дотримання формальних мовних правил цільової мови, збереження точного значення загального тексту для адекватного сприйняття реципієнтом, можливість відтворити більш природнє мовлення. Крім того, людина-редактор здатна чітко окреслити перекладацькі трансформації вихідного тексту, зокрема синтаксичні.

Автоматизоване оцінювання якості відбувається через математичні та статистичні розрахунки перекладеного продукту системи НМП у зіставленні з взірцем – довершеним варіантом тексту перекладу цільовою мовою, що створений зазвичай людиною-перекладачем [62]. Задля встановлення ступеня якості обирається певна шкала оцінювання.

На сьогоднішній день існує цілий ряд різних автоматичних систем або метрик оцінювання. Серед них широко використовуються: BLEU, NIST, METEOR, LEPOR, COMET, PRIS, TER (див. список скорочень). Кожна метрика працює за різними алгоритмами і, відповідно, по-різному здійснює процес автоматичного оцінювання. Це означає, що вони мають різні сильні та слабкі сторони, а також відрізняються за типом помилок, за які вони дають вищі або нижчі за значенням пенальті ³¹(штрафи).

З усіх перерахованих вище метрик BLEU є найбільш часто використовуваною. Це одна з перших метрик, яка досягла високого рівня кореляції з людським оцінюванням і породила безліч різних варіацій інших метрик [10]. У дослідженні ми застосовуємо саме цю систему загального оцінювання якості НМП системою DeepL.

Отже, зважаючи на те, що існують різні метрики визначення якості тексту перекладу в результаті НМП, в цьому дослідженні ми обираємо гібридний спосіб здійснення оцінювання роботи системи НМП DeepL як найбільш оптимальний спосіб тестування та визначення якості. Залучаємо систему загального автоматизованого оцінювання BLEU та пост-обробку людиною з фокусом на синтаксичних трансформаціях та особливостях китайсько-українського перекладу публіцистичних текстів.

³¹ *англ.* penalty

3.2. Методологія здійснення автоматизованого оцінювання систем НМП та критерії якості

Метрика оцінювання BLEU є універсальною для визначення якості роботи систем НМП між різними мовами, і яка надає загальний аналіз, не фокусуючись лише на окремих мовних аспектах, тобто без урахування, наприклад, зрозумілості змісту або граматичної правильності [27]. Опрацювання тексту перекладу метрикою BLEU полягає в тому, що окремі речення або фрагменти тексту оцінюються порівняно з набором високоякісних еталонних перекладів [57]. Сегменти, на які розбивається весь вихідний текст називаються n-грамами, що є послідовними сполученнями з трьох або чотирьох слів [50]. Окремі показники оцінювання по сегментах усереднюються, і отримане число є остаточним результатом за шкалою BLEU для даної системи НМП. Таке тестування демонструє, наскільки результат роботи системи машинного перекладу відповідає взірцевому перекладу, виконаному людиною, що є маркером якості. «Чим ближче машинний переклад до професійного перекладу людини, тим він кращий» – це основна ідея BLEU [9, с. 96].

Метрика BLEU опрацьовує текст через заздалегідь складену програму мовою програмування, зазвичай за допомогою Python (див. рис. 4). Закладена в основу функція приймає як вхідні дані список передбачених речень і список списків посилань або взірців, оскільки кожний передбачений сегмент перекладу може мати кілька посилань [54].

```
>>> predictions = ["hello there general kenobi", "foo bar foobar"]
>>> references = [
...     ["hello there general kenobi", "hello there !"],
...     ["foo bar foobar"]
... ]
>>> bleu = evaluate.load("bleu")
>>> results = bleu.compute(predictions=predictions, references=references)
>>> print(results)
{'bleu': 1.0, 'precisions': [1.0, 1.0, 1.0, 1.0], 'brevity_penalty': 1.0, 'length_ratio': 1.1666666666666667, 'translation_length': 7, 'reference_length': 7}
```

Рисунок 4. Приклад програми з опрацювання тексту НМП та обчислення балу за метрикою BLEU у порівнянні з еталонним перекладом [54].

Серед вихідних даних [29]:

1) передбачення перекладу системою НМП (список значень типу string) – сегменти перекладу, що мають бути оцінені;

2) посилання (список списків значень типу string) – взірць для кожного сегменту МП.

Отримані дані:

1) bleu (дробове значення типу float³²) – оцінка за шкалою BLEU,

2) precision (список дробих значень типу float) – середнє геометричне точності відповідності n-грам,

3) brevity_penalty (float) – пенальті або штраф за стислість,

4) length_ratio (float): співвідношення довжин,

5) translation_length (ціле значення типу int³³) – довжина перекладу,

6) reference_length (int) – довжина посилання.

Отримане остаточне значення bleu завжди є числом від 0 до 1. Воно вказує на те, наскільки текст-передбачення схоже на тексти посилань, при чому значення, ближче до 1, демонструє вищий ступінь схожості продукту НМП на еталонні зразки [43, с. 51]. Для зручності зіставлення та аналізу отриманих даних отримане дробове значення bleu переводять у шкалу 0–100. Тоді якість оцінюється за 100-бальною шкалою.

На сьогодні найвищі показники роботи систем НМП за шкалою BLEU на матеріалі текстів публіцистичного та нейтральних стилів складають близько 45–50 балів [22, с. 6]. Вищі показники 70–80 балів були отримані під час дослідження роботи систем НМП на основі спеціалізованих технічних текстів, що зумовлено інваріантністю структур речення та єдністю термінології певної наукової сфери [64, с. 397]. Це означає, що судження щодо результатів проведеного нами оцінювання спиратиметься на діапазон 45–50 балів відносно еталонних, що зумовлено стилістичним спрямуванням публіцистичних текстів, обраних як матеріал дослідження.

³² Тип значення, що є числом з плаваючою точкою, в мові Python.

³³ Тип значення, що є цілим числом, в мові Python.

Залучення ручної пост-обробки людиною є необхідним компонентом нашого дослідження. Під час цього етапу роботи можливо об'єктивно оцінити лінгвістичне підґрунтя якості перекладу системою НМП. Конкретні переваги та недоліки роботи системи НМП постають не на основі статистичних даних та визначених заздалегідь алгоритмів оцінювання, а через ситуативне сприйняття тексту перекладу носієм цільової мови. Якість реалізації синтаксичного плану перекладу безпосередньо впливає на ступінь зрозумілості та релевантності тексту для окремого реципієнта.

Отже, при здійсненні автоматизованого та ручного оцінювання, як двох етапів гібридного способу визначення якості роботи системи НМП, нами обрано різні критерії. Метрика BLEU дає змогу отримати інформацію про загальну ефективність роботи системи DeepL в контексті китайсько-українського публіцистичного перекладу, статистичні дані за синтаксичним аспектом перекладу, а також відповідність тексту перекладу взірцевому продукту праці людини-перекладача. Ручна пост-обробка отриманого тексту системи НМП дозволяє оцінити якість роботи DeepL не тільки за структурною ознакою, а й через призму сприйняття продукту перекладу носієм цільової мови – української. Крім того, під час ручного розбору перекладу, виконаного системою НМП, можливо окреслити конкретні зміни в синтаксичній будові речень, що зумовлені формальною характеристикою кожної з мов, а також визначити, наскільки система НМП DeepL здатна адаптувати синтаксичний рівень перекладу до норм української мови при перекладі публіцистичного тексту з китайської.

3.3. Гібридне оцінювання ефективності системи НМП DeepL на матеріалі китайсько-українського публіцистичного тексту

Задля здійснення оцінювання системи НМП DeepL в контексті китайсько-українського перекладу на конкретному прикладі нами було обрано текст промови TEDTalks від 8 січня 2013 року «坐轮椅潜行深海» («Плавання морськими глибинами в інвалідному візку») спікерки Сью Остін. Цей текст виступає оптимальним матеріалом як текст публіцистичного стилю, який тяжіє

до живої розмовної мови та необтяжений специфічною лексикою або зворотами. Письмовий запис був перекладений китайською з англійської користувачами через вільний доступ на платформі Amara [21]. Текст зазнав правлень багатьох контриб'юторів, що робить його доречним для розгляду з точки зору ступеня об'єктивності, зрозумілості та відповідності нормам китайської мови.

3.3.1. Автоматизоване оцінювання тексту за метрикою BLEU

Перш ніж здійснювати повне гібридне тестування отриманого за допомогою системи DeepL тексту промови «坐轮椅潜行深海» («Плавання морськими глибинами в інвалідному візку») спікерки Сю Остін, визначаємо алгоритм дії під час першої стадії – автоматизованого оцінювання тексту за метрикою BLEU . Цей процес полягає у таких етапах роботи:

- 1) завантаження бібліотеки evaluate мови Python, що містить функції, модулі обробки тексту та алгоритми розрахунку якості в балах за шкалою BLEU;
- 2) підготовка еталонного перекладу обраного тексту людиною-перекладачем, поміщення оригінального тексту, автоматичного перекладу та ручного в окремі файли;
- 3) написання програми, яка приймає у вигляді списку по 4–5 окремих сегментів (речень) тексту машинного перекладу як передбачення системи НМП, а також список списків із відповідними частинами взірцевого ручного перекладу, після чого зіставляє значення і опрацьовує інструментами бібліотеки evaluate, виводить значення: bleu, precision, brevity_penalty, length_ratio, translation_length та reference_length (див. додаток Г);
- 4) поділ загального тексту перекладу на рівні частини по 4–5 речень та здійснення оцінювання кожного сегменту (загалом 7) задля визначення динаміки зміни значення BLEU впродовж всього циклу роботи системи НМП DeepL;
- 5) поєднання коду роботи з окремими сегментами в одній програмі та проведення оцінювання всього вихідного тексту для отримання загальної оцінки роботи DeepL за шкалою BLEU;

б) візуальне відображення результатів оцінювання через побудову графіку за допомогою бібліотеки `matplotlib`, де по осі *x* демонструються сегменти тексту у їхній хронологічній послідовності, а по осі *y* – бали за метрикою BLEU для кожної частини тексту перекладу.

В результаті проведеного загального автоматизованого оцінювання всього тексту промови «坐轮椅潜行深海» («Плавання морськими глибинами в інвалідному візку») спікерки Сюю Остін отримуємо значення:

```
{'bleu': 0.18461040582339483, 'precisions': [0.4794326241134752,
0.2556053811659193, 0.14691943127962084, 0.08375209380234507],
'breivity_penalty': 0.9368348868457684, 'length_ratio': 0.9387483355525965,
'translation_length': 705, 'reference_length': 751}.
```

За наявними показниками прослідковуємо ефективність роботи системи DeepL оцінену в 18.4 балів з 45–50 оптимальних або зі 100 можливих. Як було зазначено раніше, у порівнянні з еталонними значеннями 45–50 балів у сфері досліджень публіцистичного перекладу, якість обраного нами матеріалу можна оцінити як нижче середнього. Це означає, система DeepL здійснює недостатньо якісний китайсько-український переклад за загальним зіставленням із роботою людини-перекладача. Щодо статистичних показників чіткості відповідних *n*-грам та довжини речень, то прослідковуємо високі показники. Такі дані дозволяють судити про здатність системи НМП DeepL добирати коректний об'єм одиниць мовлення.

Варто зауважити, що механізм роботи НМП, який полягає у здійсненні перекладу через передбачення наступного значення зліва направо, зумовив необхідність встановлення динаміки зміни якості вихідного тексту від його початку до кінця. Нами побудовано графік значень BLEU по послідовних сегментах перекладеного матеріалу (див. рис. 5).

Розподілення якості НМП протягом циклу роботи DeepL (за метрикою BLEU)

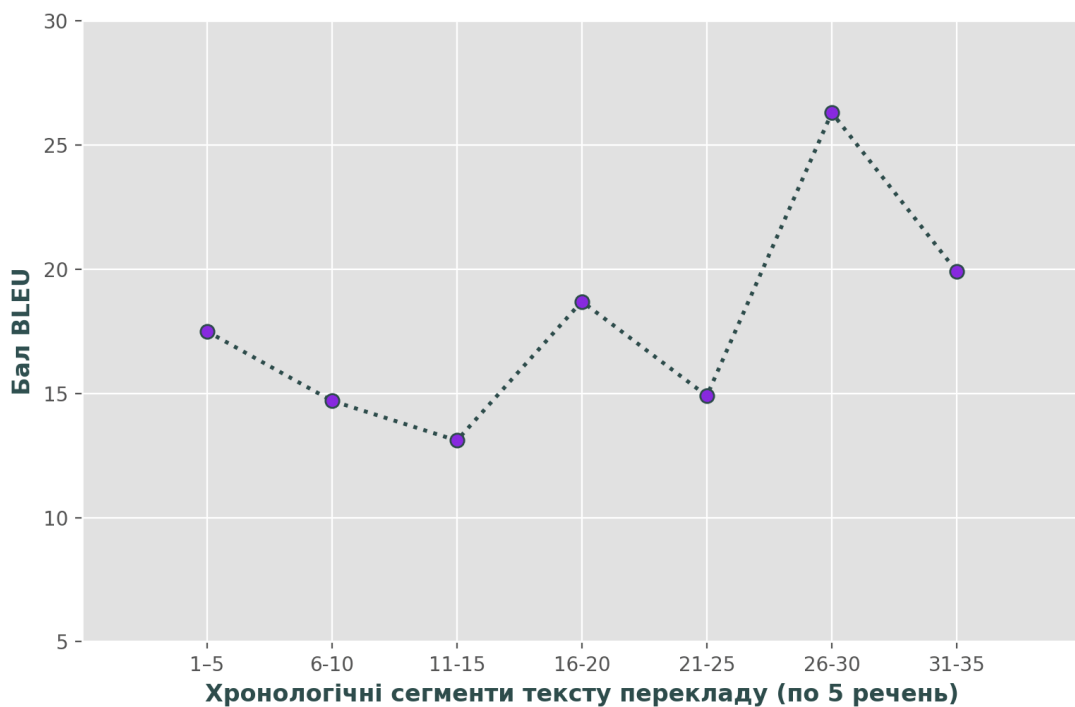


Рисунок 5. Графічне відображення динаміки зміни балу BLEU послідовно впродовж всього циклу роботи DeepL над перекладом розглянутого тексту.

Джерело: розроблено автором на основі [70].

Можемо бачити, що тенденція зміни якості характеризується як спадна до сегменту під номером 3 (речення 11–15) тексту включно, що відповідає ідеї наукової праці дослідників Л. Чжоу, Д. Чжана та Ч. Дзуна [22, с.]. Науковці зауважили зниження показників через неточне передбачення системи НМП та відповідно некоректний підбір компоненту перекладу, що залежить від попереднього вибору системи. Показники якості метрики BLEU сегментів 4–7 опрацьованого нами тексту не відображають точної інформації про зміну значень через малий обсяг дослідження. Фрагмент матеріалу під номером 6 (речення 26–30) відзначаються найвищим значенням bleu – 26.3 балів. Цей показник вказує на те, що якість перекладу системою НМП є досить змінною через динаміку мовлення у публіцистичному тексті, де окремі структури і речення є більш

«зрозумілими» для системи НМП DeepL задля обробки і підбору більш точних відповідників цільовою мовою.

Отже, автоматизоване оцінювання роботи НМП DeepL показало середні показники якості перекладу, а також виявило закономірність часткового зниження показників BLEU у процесі послідовного перекладу. Отримані дані свідчать про задовільний рівень розробки механізмів НМП для китайсько-українського перекладу та виявляють необхідність підвищення ефективності роботи спираючись на формальні характеристики цих мов зокрема.

3.3.2. Ручне оцінювання та розбір тексту перекладу, визначення перекладацьких трансформацій на основі формальних тенденцій вихідної та цільової мов

В результаті наступної стадії гібридного оцінювання системи НМП ми здійснюємо ручне тестування продукту та редагування, створивши еталонний зразок перекладу. Пост-обробка продукту перекладу системою НМП вручну виявляє рівень якості роботи DeepL з публіцистичним текстом між китайською та українською, а також окреслює формальні особливості та тенденції в синтаксичній будові двох мов, що породжують нюанси процесу перекладу. Аналізуючи отриманий текст українською, ми відкидаємо помилки системи НМП на інших мовних рівнях окрім граматичного та синтаксичного, а також звертаємо увагу на вплив особливостей синтаксису двох мов у процесі НМП та їхній вплив на якість отриманого тексту.

Загальну якість роботи DeepL в розглянутому випадку оцінюємо як досить високу. Як бачимо за окремими прикладами перекладу речень, наведених у таблиці 1, системою ефективно здійснено підбір синтаксичних структур без втрати сенсу повідомлення та логічних зв'язків в потоці мовлення. Однак, можемо судити про неспроможність системи НМП адаптувати переклад до норм цільової мови в повній мірі, при чому синтаксична структура речень прямо унаслідується з оригінального матеріалу. Проілюструємо конкретними прикладами результатів автоматичного перекладу (див. додаток Б), зіставляючи

його з вихідним текстом (див. додаток А), а також ручним перекладом (див. додаток В) у вигляді таблиці (див. табл. 2).

Таблиця 2

Порівняльна таблиця автоматизованого НМП та ручного перекладу окремих речень дослідженого публіцистичного тексту

№	Оригінальне речення китайською	Машинний переклад через DeepL	Ручний переклад українською	Необхідні синтаксичні трансформації
1	2	3	4	5
1.	可以话水肺潜水带昇人慨印象系充满刺激同历奇而且同轮椅昇人慨印象完全唔同。	Можу сказати, що враження від підводного плавання сповнені хвилювання і здивування, і вони зовсім не схожі на враження від занурення в інвалідному візку.	Можна сказати, що спорядження для підводного плавання справляє враження, повне хвилювання та пригод. Відрізняється від враження, яке справляють інвалідні візки.	1) заміна частотного сполучника «і» на інші сурядні з аналогічною функцією, 2) відокремлення залежної граматичної основи у вигляді простого речення
2.	因为之前无人见过潜水轮椅系点 所以潜水轮椅面世 象征着新睇法、新认知 而家呢种睇法都响大家心中扎实。	Ніхто ніколи раніше не бачив інвалідного візка для пірнання, тому поява такого візка символізує новий погляд на речі, новий спосіб бачення речей, і цей спосіб бачення речей закріплюється у свідомості кожного з вас.	Оскільки ніхто раніше не бачив нічого спільного між інвалідним візком і оснащенням для пірнання під воду, то запуск інвалідного візка для дайвінгу символізує новий спосіб бачення та нове розуміння. Це нове бачення має сильний вплив на серце кожного.	1) заміна частотного сполучника «і» на інші сурядні з аналогічною функцією, 2) відокремлення граматичної основи складного речення у вигляді простого
3.	我开始做一啲作品 旨在向大家传递 我坐轮椅慨嗰份内心喜悦同自由 我叫佢做能量椅 因为可以同个世界撻过。	Я почала робити деякі роботи, які мали на меті передати внутрішню радість і свободу перебування в інвалідному візку, і я називаю його «енергетичним кріслом», тому що я можу взяти світ за роги.	Я почала проводити різні заходи, аби передати всім почуття внутрішньої радості і свободи, які маєш перебуваючи в інвалідному візку. Я називаю його кріслом сили, тому що можу пересуватися в ньому по всьому по світу.	1) поділ твердження на частини, окремі аргументи у вигляді завершених послідовних речень, 2) урізноманітнення сполучників: висловлення мети як типу підрядного зв'язку
4.	我觉得，因为嗰一刻 渠咁睇紧一件嘢系渠咁前所未见 或者超乎渠咁想像 打破晒渠咁对轮椅慨观感。	Я думаю, що саме момент, коли вони бачать те, чого ніколи раніше не бачили, або те, що виходить за межі їхньої уяви, ламає їхнє сприйняття інвалідних візків і змушує їх поглянути на них знову.	Я думаю, що в той момент вони бачили щось, чого ніколи раніше не бачили або що було за межами їх уяви, і це зруйнувало їх уявлення про інвалідні візки, змушувало поглянути на них новими очима.	1) зменшення кількості підрядних, 2) додавання членів речення для компенсації змісту

1	2	3	4	5
5.	但呢张轮椅就好似一件神奇玩具，令我出返去吹风。	Але цей візок - як чарівна іграшка. Він дозволяє мені гуляти з вітром.	Але ця коляска як чарівна іграшка, бо з нею я можу вибиратися на свіже повітря. Крім того, можливість вийти на вулицю це дуже класно!	1) перехід об'єкта в позицію підмета складнопідрядного речення задля уникнення тавтології, 2) скорочення кількості дієслів, 3) додавання членів речення для компенсації змісту
6.	为咗令你哋明白我讲慨经历我想同大家分享呢件法宝 带畀我慨其中一段体验 同埋渠带咗我去慨世界有几咁精彩。	Щоб ви могли зрозуміти, про що я говорю, я хотів би поділитися з вами одним із вражень, які подарував мені цей скарб, і яким чудовим є світ, в який він мене переніс.	Щоб ви зрозуміли, про що саме я говорю, то хочу поділитися з вами одним із досвідів, які принесла мені ця чарівна зброя, і яким чудовим є світ, у який вона мене привела.	1) уникнення тавтології через опущення членів, 2) зміна підрядної на сурядну частину речення
7.	张轮椅变咗一件我可以又写又画慨嘢。	Візок став чимось, на чому я можу писати і малювати.	Він дав мені змогу писати та малювати.	1) скорочення кількості дієслів, 2) додавання підрядної
8.	当我开始为自己留下快乐同自由慨印记 外界对我慨作品有反应，令我好雀跃。	Коли я почала залишати на собі слід щастя і свободи, я була настільки схвильована, що зовнішній світ відгукнувся на мої роботи.	Коли я почала буквально залишати сліди свого внутрішнього щастя і свободи, то і зовнішній світ відгукнувся на мої роботи. І я була цьому безмежно рада.	1) заміна дієслівної фрази на окреме дієслово, 2) поділ складного речення на простіші, 3) додавання членів речення для компенсації змісту
9.	呢件事话畀我哋听，普通人都睇得到差异慨价值，同箇中慨快乐，当我哋唔系着眼于损伤或者局限，而系用新视野、新角度去睇，我哋就会发现力量同喜悦。	І, скажімо, звичайні люди бачать різницю в цінності та радості, в силі та радості, яку ми знаходимо, коли не дивимося на пошкодження чи обмеження, а дивимося на це з новим баченням, з новою перспективою.	Тож погляньте на цю історію. Звичайні люди можуть побачити різні цінності і відчуті радість, не зосереджуючись на недоліках чи обмеженнях, а дивлячись на це з нових горизонтів і під новими кутами. Тільки тоді ми знаходимо силу та щастя.	1) поділ складного речення на простіші, 2) заміна підрядної на другорядний член речення, 3) додавання членів речення для компенсації змісту

Джерело: розроблено автором на основі [«坐轮椅潜行深海»].

В реченнях № 1 та № 2 остання граматична основа поєднується з попередніми сполучником сурядності «і». Як бачимо за багаторазовим повторенням цього явища в інших частинах тексту, цей сполучник система НМП залучає як універсальний для поєднання тверджень. На нашу думку, доречним у перекладі українською є заміна сполучника на інші сурядні з аналогічною функцією, або ж відокремлення граматичної основи у вигляді простого речення.

Крім того, характерна риса синтаксису китайської мови, яка полягає в об'єднанні окремих одиниць мовлення та простих речень у складнопідрядне зі значною кількістю граматичних основ, породжує перевантаження мовлення українською, наприклад:

У реченні № 3 більш природнім для сприйняття україномовним реципієнта було б розділення твердження на частини. Складне речення українською, особливо в усній формі, не є достатньо простим для розуміння, а тому граматичні основи краще відокремлювати у декілька речень, наприклад: Я почала робити деякі роботи, які мали на меті передати внутрішню радість і свободу перебування в інвалідному візку. Я називаю його «енергетичним кріслом», тому що я можу взяти світ за роги.

Інша особливість синтаксису китайської, що становила складнощі у роботи системи НМП, – частотне використання у мові оригіналу підрядних. У реченні № 4 спостерігаємо, що DeepL дублює загальну структуру речення, через що логічні зв'язки нашаровуються та перешкоджають чіткому й зрозумілому висловленню ідеї. Варто зауважити, що частина речення «змушує їх поглянути на них знову», що в оригінальному тексті висловлено конструкцією 令+ Об'єкт + Дієслово, також відображає калькування синтаксису китайської та нездатність до трансформування тексту відповідно до норм української. Наведена синтаксична структура китайського речення стосується явища переважання дієслівних фраз, що прослідковуємо і на інших прикладах з оригінального тексту, які зіставляємо зі взірцевим перекладом промови «坐轮椅潜行深海» («Плавання морськими глибинами в інвалідному візку») спікерки Сюю Остін (див. додаток В). Так, у реченнях № 4 та № 5 хоча і містять одну й ту саму конструкцію, але під час перекладу українською зазнали різних синтаксичних трансформацій: (5) 但呢张轮椅就好似一件神奇玩具，令我可以出返去吹风。 – Я називаю його кріслом сили, тому що можу пересуватися в ньому по всьому по світу; (6) 为咗令你哋明白我讲慨经历，– Щоб ви зрозуміли, про що саме я говорю, ...

У реченні № 5 – об’єкт «我» у другій частині складносурядного речення переходить у позицію підмета складнопідрядного речення «я», який опускається задля уникнення тавтології з головною граматичною основою. Тобто кількість дієслів загалом скорочується, але додаються інші члени речення для компенсації змісту – обставина «в ньому по всьому по світу». У реченні № 7 аналогічно зменшується частотність дієслів – «ви зрозуміли» замість «змусити вас зрозуміти». Крім того, додається з’ясувальна підрядна.

Прослідкуємо інші прояви розбіжностей синтаксису китайської та української у процесі перекладу в інших фрагментах тексту. На прикладі речення № 7 простежуємо спрощення структури речення, де дві граматичні основи поєднані в одну: 张轮椅变啦一件我可以又写又画慨嘢 – Він дав мені змогу писати та малювати. Тобто українське речення не тяжіє до формування підрядних настільки частотно, як китайське. Можемо також бачити заміну VP на NP: «可以又写又画慨嘢» перетворено на «змогу писати та малювати».

У реченні № 8 бачимо частотне використання VP в китайському реченні: 当我开始为自己留下快乐同自由慨印记 外界对我慨作品有反应, 令我好雀跃 – Коли я почала буквально залишати сліди свого внутрішнього щастя і свободи, то і зовнішній світ відгукнувся на мої роботи. І я була цьому безмежно рада. «有反应» та «令我好雀跃» у перекладі трансформовані в окреме дієслово та двоскладову граматичну основу відповідно. Сурядний зв’язок між другою та третьою частинами початкового речення розірвано у перекладі і утворено просте речення. Додаються і окремі члени речення для формування логічних зв’язків: виникають обставина «буквально» у першій частині речення та додаток «цьому» в останній.

У реченні № 9 під час перекладу складне речення підлягає поділу на простіші. Крім того підрядна обставинна часу оригінального речення замінена на дієприслівниковий зворот в українському. VP «畀我哋听» випадає під час перекладу, як і дієслівне словосполучення з модальним елементом перетворено

на окреме дієслово: «会发现» – «знаходимо». Для підтримання логічних зв'язків до останньої граматичної основи додається «тільки тоді».

Перекладацькі трансформації, виявлені нами на практиці, є нормалізацією тексту виходячи не тільки з формальних ознак цільової мови, а ситуативного добору синтаксичних конструкцій, що робить мовлення необтяженим та простішим для розуміння. Залучення конкретних конструкцій, що спираються на риси мовлення носіїв певної мови, засобів полегшення усного сприйняття повідомлення, а також постають частотними у тексті цільовою мовою в публіцистичному перекладі, можуть бути основою для універсальних підходів у НМП. У контексті розглянутої пари мов, такими часто залученими механізмами перекладу є: 1) спрощення структури складного речення та відокремлення граматичних основ; 2) зменшення кількості підрядних частин через перетворення їх на другорядні члени речення; 3) заміщення VP іншими членами без залежних компонентів речення; 4) спрощення зв'язків «дерева» речення в межах однієї граматичної основи; 5) додавання членів речення для відтворення логічних зв'язків. Перелічені перекладацькі трансформації не були залучені системою НМП DeepL, а отже вказують на явні недоліки механізмів його роботи з огляду на типологічно відмінні мови.

Отже, в рамках синтаксичного аспекту публіцистичного перекладу, гібридне оцінювання вихідного тексту показало як переваги роботи системи НМП DeepL (синтаксична структура речень відповідає нормам української мови, дібрані конструкції повноцінно транслюють зміст повідомлення), так і наразі непереборені виклики цього процесу, а саме: тяжіння до калькування структури китайського речення, утворення складносурядних та складнопідрядних речень з послідовним нашаровуванням залежних граматичних основ, а тому перенавантаження повідомлення мовними одиницями; нездатність до здійснення якісної перестановки та заміни членів речення цільовою мовою, залучення частотних сполучників сурядності без конкретизації зв'язку між граматичними основами тощо. Автоматизоване оцінювання тексту виявило середній ступінь загальної якості роботи та статистичні показники відповідності продукту НМП у

зіставленні зі взірцевим перекладом вручну – 18.4 балів. Однак ручна пост-обробка тексту цільовою мовою продемонструвала коло проблем НМП та механізми вирішення у контексті публіцистичного стилю мовлення. Запропоновані перекладацькі трансформації, як-от: спрощення структури речення українською мовою, зміна порядку й ролі членів речення, перетворення граматичних основ на інші другорядні члени речення – всі вони ґрунтуються як на формальних ознаках двох мов, так і на специфічних випадках мовлення їхніх носії у заданій ситуації.

Висновки до Розділу 3

У Розділі 3 окреслюємо проблему оцінювання систем НМП та лінгвістичну зумовленість цього процесу. Визначаємо види оцінювання НМП, а саме: оцінка людиною, автоматизоване оцінювання, гібридне, на основі чітко визначених завдань тощо. Ми обираємо гібридний підхід до тестування, який є оптимальним для не тільки визначення якості роботи системи НМП DeepL, але й спостереження за проявом синтаксичних особливостей китайської та української в контексті перекладу публіцистичних текстів. Для першої з двох стадій оцінювання (автоматизоване та ручне тестування послідовно) обираємо метрику BLEU.

Розглядаємо методологію проведення першого етапу гібридного оцінювання, визначаємо алгоритм дій підготовки комп'ютерної програми, вихідні та отримані показники функції, розставляємо акцент на параметрах аналізу – окреслення загальної відповідності тексту перекладу еталонному ручному перекладу, динаміки зміни значення балу bleu впродовж всього циклу здійснення НМП системою DeepL. Останній аспект аналізу зумовлений явищем незбалансованих виходів, де префікси перекладу передбачаються краще, ніж наступні сегменти тексту.

Саме у другий етап гібридного оцінювання – ручне тестування – визначаємо основні недоліки роботи системи НМП, наводимо зауваження щодо доречності залучення синтаксичних конструкцій у тексті цільовою мовою (українською) при перекладі публіцистичного тексту з китайської. Зокрема зауважуємо на таких аспектах НМП, що мають бути вдосконалені: 1) тяжіння до калькування структури китайського речення, 2) утворення складносурядних та складнопідрядних речень з послідовним нашаровуванням залежних граматичних основ, а тому перенавантаження повідомлення мовними одиницями; 3) нездатність до здійснення якісної перестановки та заміни членів речення цільовою мовою, 4) залучення частотних сполучників сурядності без конкретизації типу зв'язку між граматичними основами тощо. Пропонуємо перекладацькі трансформації в контексті китайсько-українського

публіцистичного перекладу, що можуть бути закладені в основу вдосконалених механізмів систем НМП: 1) спрощення структури речення українською мовою, 2) зміна порядку й ролі членів речення, 3) перетворення граматичних основ на другорядні члени речення незалежної частини складного речення, 4) перетворення дієслівних фраз, де дія висловлюється сполученням дієслів, на еквівалент українською, який складається з одного або кількох слів на позначення конкретної дії; 5) урізноманітнення структури речення цільовою мовою, адаптуючи ступінь зрозумілості повідомлення під час усного мовлення (адже ми розглядаємо скрипт промови «坐轮椅潜行深海» («Плавання морськими глибинами в інвалідному візку») спікерки Сюю Остін як публіцистичний текст).

Попри закладені складні багатоетапні алгоритми опрацювання тексту, системи НМП, такі як DeepL, все ще зазнають складнощів у здійсненні якісного перекладу, наближеного до живого мовлення носіїв. Синтаксичний рівень відіграє ключову роль в забезпеченні високоефективної комунікації між мовцями, де увага спрямована не тільки на трансляцію точного змісту повідомлення, але й на форму його подачі та структурну простоту для сприйняття. Нами виявлено основні підходи оптимізації китайсько-українського публіцистичного перекладу, що є матеріалом для подальшого вдосконалення систем НМП.

ВИСНОВКИ

Підгрунття якості здійснення процесу перекладу полягає не тільки в формальних теоретичних знаннях про мови і віднайденні еквівалентних лексичних одиниць, але й у різних мовленнєвих аспектах, формі подачі змісту, ситуативно зумовлених чинниках та ступені сприйняття інформації реципієнтом. Оформлення доречної структури повідомлення на рівні синтаксису відіграє важливу роль у трансляції змісту. У ході дослідження синтаксичних особливостей будови китайської та української мов та їхнього впливу на процес публіцистичного перекладу, зокрема якість роботи системи НМП, доходимо висновків:

1) НМП є ефективним методом перекладу на сучасному етапі розвитку МП, широко застосовуваним способом здійснення перекладу в різних сферах життя; на сьогодні системи НМП ґрунтуються на багатофазових алгоритмах обробки даних, як-от нейронні мережі, які в контексті опрацювання природніх мов потребують подальшого вдосконалення за конкретними аспектами (проблема пост-редагування, недостатній рівень застосовності систем НМП у різних сферах, передача аспекту творчості в тексті та здатність до трансляції художньої образності у перекладі, низька стійкість до помилок і стабільність систем тощо), що відкриває широке поле лінгвістичних досліджень;

2) роль людини-перекладача на сучасному етапі розвитку НМП не була заміщена автоматизованими системами перекладу, а піддала впливу цей вид діяльності, трансформувавши її функції: ручний переклад може слугувати основою машинного навчання та вдосконалення механізмів НМП, а ручна пост-обробка перекладу системою НМП необхідна для визначення недоліків продукту й розробки більш досконалих алгоритмів;

3) китайська та українська мови є типологічно значно відмінними мовами (ізолюючого та флективного типу відповідно), що виходить з порівняльного аналізу їхньої синтаксичної будови; цей чинник зумовлює актуальність дослідження такої мовної пари для проведення корпусного компаративного аналізу, віднайдення особливих рис і закономірностей двох мов, розробки

універсальних підходів (перекладацьких трансформацій та універсалій) до здійснення НМП для ширшого кола типологічно відмінних мов;

4) синтаксичний рівень мови є так само визначальним з огляду на оформлення повідомлення, як і лексичний, включає в себе план змісту; коректний вибір синтаксичних структур здійснюється не тільки за формальними нормами мови, але й відносно певної комунікативної ситуації, форми подачі інформації (наприклад письмова або усна), особливостями сприйняття реципієнтом та інших факторів, а тому такий аспект перекладу потребує більшої уваги й дослідження, зокрема, в китайсько-українській мовній парі;

5) публіцистичний текст виступає як оптимальний матеріал для вивчення з огляду на його стилістичні особливості, серед яких зазначаємо: тяжіння до частотних граматичних конструкцій, фокусування на широкому тематичному спектрі, відсутність зумовленості лише конкретною сферою діяльності, наближеність до побутового мовлення носіїв тощо – усі ці риси зумовлюють актуальність залучення корпусів цього стилю до компаративного аналізу та дослідження синтаксичних особливостей китайсько-української мовної пари;

6) корпусна лінгвістика є основою для вивчення особливостей мов на базі значних текстових обсягів (корпусів), опрацьованих заздалегідь через токенизацію та парсинг для можливості їхнього аналізу за допомогою комп'ютера; тож ми використовуємо китайсько-український паралельний корпус публіцистичних текстів TED2020, обсяг якого складає 12,925 речень, серед яких наявні 33,508 токенів китайською та 170,224 українською;

7) залучивши методологію здійснення корпусного аналізу (IDE Visual Studio Code, Python з бібліотеками re, spaCy, jieba) ми склали функції опрацювання двомовного паралельного китайсько-українського корпусу публіцистичних текстів TED2020 та на основі цього матеріалу простежили особливості синтаксичної будови мов; за конкретними встановленими параметрами спостерігаємо такі риси мов:

– довжина речень (речення китайською коротше за відповідник українською),

– порядок слів (схема SVO домінує в обидвох мовах, однак синтаксис китайської включає частотне використання порядку SOV, який майже відсутній в текстах українською),

– частотність утворення дієслівних та іменникових фраз (VP переважають за NP у китайській, а в українській спостерігаємо обернене явище),

– формування підрядних частин складного речення (тяжіння китайської до формування сурядних частин складного речення, а української – до утворення підрядних та поділу довгих висловлювань на окремі аргументи з декількома граматичними основами);

8) додатково, ми простежуємо явище збільшення кількості іменникових фраз NP в конкретній частині текстів однієї мови водночас з підвищенням значення VP в реченнях іншою і навпаки; така закономірність висуває питання про явище взаємної компенсації NP та VP в контексті китайсько-українського перекладу та має потенційне значення для подальших розробок принципів НМП;

9) основою для підвищення якості НМП за аспектом синтаксису є тестування та оцінювання автоматизованих систем; ми обираємо оптимальний метод здійснення оцінювання системи НМП DeepL з перспективою розробки синтаксичних перекладацьких трансформацій та універсалій на основі даних, визначених попередньо корпусним аналізом, – гібридний (дві стадії: автоматизоване оцінювання метрикою BLEU та ручна пост-обробка тексту перекладу);

10) отримані дані автоматизованого оцінювання за метрикою BLEU на матеріалі окремого публіцистичного тексту – скрипті промови «坐轮椅潜行深海» («Плавання морськими глибинами в інвалідному візку») спікерки Сью Остін – виявили середній рівень якості роботи DeepL в контексті китайсько-українського перекладу – 18.4 бали з 45–50 балів, оптимальних для текстів цього стилю; однак отриманий переклад є достатньо задовільним за формою для повноцінного сприйняття повідомлення носієм цільової мови; додатково простежуємо динаміку зменшення якості перекладу DeepL за

метрикою BLEU протягом циклу роботи, що підтверджує теорію дослідників Л. Чжоу, Д. Чжана та Ч. Дзуна;

11) за результатами корпусного аналізу ми зробили припущення про можливі перекладацькі трансформації та універсалії, що мають бути застосовані під час китайсько-українського НМП системою DeepL, більшість з яких підтвердилася; у ході ручного редагування тексту автоматизованого перекладу визначаємо такі необхідні синтаксичні перекладацькі трансформації:

- спрощення структури речення українською мовою,
- зміна порядку й ролі членів речення,
- перетворення граматичних основ на другорядні члени речення незалежної частини складного речення,
- перетворення дієслівних фраз VP, де дія висловлюється сполученням дієслів, на еквівалент українською, який складається з одного або кількох слів на позначення конкретної дії;
- урізноманітнення структури речення цільовою мовою, адаптуючи ступінь зрозумілості повідомлення під час усного мовлення (адже ми розглядаємо скрипт усної промови як публіцистичний текст);

12) перекладацькі універсалії як загальні закономірності синтаксичних змін при перекладі з китайської українською простежуємо у тенденціях спрощення (зменшення кількості членів речення, граматичних основ загалом або поділ складного речення на прості) та нормалізації синтаксичної будови в тексті цільовою мовою (включає узагальнене явище синтаксичних трансформацій та адаптації тексту до норм цільової мови);

13) алгоритми НМП мають ґрунтуватися не тільки на формальних ознаках мов, але також і механізмах адаптації тексту повідомлення і потоці мовлення, зокрема з огляду на його синтаксичну структуру, що грає важливу роль в розробці систем НМП, здатних відтворювати у тексті перекладу мовлення, максимально наближене до такого, що спостерігається у носіїв цільової мови, є більш зрозумілим та милозвучним для реципієнта;

14) пропонуємо певні перспективи лінгвістичних досліджень у напрямку НМП, а саме: фонетичний, лексичний, стилістичний рівні НМП, письмовий та усний переклад, діалекти, розробка систем послідовного та синхронного автоматизованого китайсько-українського перекладу, переклад художніх текстів тощо;

15) практичне значення нашої роботи полягає в можливості залучення отриманих результатів та висновків у дослідженнях з загального мовознавства, перекладознавства, лінгвокультурології, нейронного машинного перекладу та машинного навчання; запропоновані синтаксичні трансформації можуть бути використані при китайсько-українському перекладі публіцистичних текстів у різних сферах діяльності людини; крім того, окреслені нами підходи до зміни синтаксичної структури речення між типологічно відмінними мовами можуть бути закладені в основі більш довершених алгоритмів НМП; висновки роботи можливо залучити при складанні підручників та методичних посібників з перекладу; методологія здійснення корпусного аналізу та гібридного оцінювання системи НМП є основою для подальших лінгвістичних досліджень в контексті китайсько-українського перекладу та вдосконалення засобів автоматизованого перекладу.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Бундуки В. С., Богущ А. М. Відтворення лексико-граматичних особливостей китайськомовної наукової літератури англійською та українською мовами. *Науковий вісник ПНПУ ім. К. Д. Ушинського*. 2018. Вип. 27. С. 42–52. DOI: <https://doi.org/10.24195/2616-5317-2018-27-5> (дата звернення: 17.04.2024).
2. Война М. О., Ракітіна М. І. Паратаксист та гіпотаксист в контексті китайсько-англійського перекладу. *Закарпатські філологічні студії*, 2022. Вип. 22, Т 1. С. 149–153. URL: <http://dx.doi.org/10.32782/tps2663-4880/2022.22.1.25> (дата звернення: 17.04.2024).
3. Война М. О., Тулупова А. І. Китайсько-англійський та китайсько-український переклад інформаційно-новинних текстів (синтаксичний аспект). *Науковий вісник Міжнародного гуманітарного університету. Сер.: Філологія*. 2019 № 39 том 3. С. 55–58. URL: <http://dx.doi.org/10.32841/2409-1154.2019.39.3.13> (дата звернення: 17.04.2024).
4. Вольницька Д. Структурні особливості іменної фрази у сучасній англійській мові. *Studia Philologica*, 2015. № 1. Р. 78–85. DOI: <https://studiap.kubg.edu.ua/index.php/journal/article/view/84> (date of access: 01.05.2024).
5. Гобова Є. В. Проблеми передачі китайських слів засобами української мови. *Китаєзнавчі дослідження*, 2021. №. 1, С. 94–103. URL: <https://chinese-studies.com.ua/index.php/journal/article/view/99> (дата звернення: 17.04.2024).
6. Зайченко С. А. Особливості перекладу сучасного китайського кінематографу. *Китаєзнавчі дослідження*, 2021. № 1-2, С. 126–143. URL: <https://chinese-studies.com.ua/index.php/journal/article/view/173> (дата звернення: 17.04.2024).
7. Зосімова О. Ю., Александрова О. В.. Типологічні характеристики перекладу китайськомовної науково-технічної літератури англійською та українською мовами. *Науковий вісник ПНПУ ім. К. Д. Ушинського*. 2018. Вип. 27.

С. 105–115. DOI: <https://doi.org/10.24195/2616-5317-2018-27-13> (дата звернення: 17.04.2024).

8. Ісаєва Н. С., Акімова А. О. Поезія Лесі Українки в китайськомовному перекладі: психолінгвістичний аспект. *Psycholinguistics. Психолінгвістика*, 2021. Вип. 30(2). Р. 25–45. URI: <http://ep3.nuwm.edu.ua/id/eprint/22261> (дата звернення: 17.04.2024).

9. Козоріз О. Автоматизація оцінки якості перекладу текстів паралельних корпусів. *Актуальні питання гуманітарних наук: Мовознавство. Літературознавство*. 2021. С.94–103 DOI: <https://doi.org/10.24919/2308-4863/43-2-15> (дата звернення: 01.05.2024).

10. Козоріз О. П. Статистичні характеристики мовних одиниць юридичної термінології китайської мови. *Вісник Київського національного університету імені Тараса Шевченка: Східні мови та літератури*, 1(20), 2014. С. 15– 20. URL: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwiYqqCWsLCGAxXRRvEDHZjdBfMQFnoECBMQAQ&url=http%3A%2F%2Ffirbis-nbu.gov.ua%2Fcgi-bin%2Ffirbis_nbu%2Fcgiirbis_64.exe%3FC21COM%3D2%26I21DBN%3DUJRN%26P21DBN%3DUJRN%26IMAGE_FILE_DOWNLOAD%3D1%26Image_file_name%3DPDF%2FVKNU_Sm_2014_1_6.pdf&usg=AOvVaw2Sr5SPT-inemmC25a-CEXt&opi=89978449 (дата звернення: 01.06.2024).

11. Козоріз О. П. Створення власних лінгвістичних корпусів. *Вісник Київського національного університету імені Тараса Шевченка: Східні мови та літератури*, 1(27), 2023. С. 6– 2. DOI: <https://doi.org/10.17721/1728-242X.2021.27.6-12>. (дата звернення: 01.06.2024).

12. Мамчич Д. С. Особливості перекладу китайської професійної лексики у медичній сфері: бакалаврська робота. Київ, 2023. 65 с. URL: <https://ir.library.knu.ua/handle/123456789/6036> (дата звернення: 17.04.2024).

13. Михайленко О. О. Синтаксичні трансформації у перекладі наукової статті. *Мовні і концептуальні картини світу*. 2020. Вип. 1, № 67. С. 68-80. DOI: <https://doi.org/10.17721/2520-6397.2020.1.06> (дата звернення: 01.05.2024).

14. Порядок слів у реченні. *Zno.if.ua*. URL: <https://zno.if.ua/?p=4109> (дата звернення: 01.05.2024).
15. Ребрій О. В., Ребрій О. О. Ми – те, що ми їмо, або труднощі перекладу китайських культурно орієнтованих номінацій їжі з китайської мови українською та англійською. *Китаєзнавчі дослідження*, 2023. №1, С. 68–79. DOI: <https://doi.org/10.51198/chinesest2023.01.068> (дата звернення: 17.04.2024).
16. Самойленко Н. С., Золотухіна Н. А. Особливості перекладу військової термінології з української мови на китайську. *Китаєзнавчі дослідження*, 2023. № 1. С. 104–112. DOI: <https://doi.org/10.51198/chinesest2023.01.104> (дата звернення: 17.04.2024).
17. Тулупова А. І., Вечоринська Т. В. Синтаксичні трансформації у разі українсько-китайського перекладу нормативно-технічної документації. *Китаєзнавчі дослідження*. 2022, №1. С. 83–94. DOI: <https://doi.org/10.51198/chinesest2022.01.083> (дата звернення: 17.04.2024).
18. Adamou E. Corpus linguistic methods: Vol. 1. / ed. by Darquennes J., Salmons J., Vandebussche W. *Language contact: An International Handbook*, De Gruyter. 2019. P. 638–653. DOI: <https://doi.org/10.1515/9783110435351-052> (date of access: 18.04.2024).
19. An Empirical Comparison of Domain Adaptation Methods for Neural Machine Translation/ ed. by C. Chu et al., *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*. Vancouver: Association for Computational Linguistics, vol. 5, 2017. P. 385–391. URL: <https://aclanthology.org/P17-2061> (date of access: 17.04.2024).
20. Araghi S., Palangkaraya A., Webster E. The impact of language translation quality on commerce: The example of patents. *Journal of International Business Policy*, 2023. 23 p. URL: <https://doi.org/10.1057/s42214-023-00157-0> (date of access: 16.04.2024).
21. Austin S. 坐輪椅潛行深海. Amara. URL: <https://amara.org/videos/sMaomPoaEnVN/zh/345178/?team=tcd> (date of access: 18.05.2024).

22. Bawden R., Zhang B., Yankovskaya L. et al. A Study in Improving BLEU Reference Coverage with Diverse Automatic Paraphrasing. Punta Cana: 2020 *Conference on Empirical Methods in Natural Language Processing: Findings*, 2020. URL: <https://hal.science/hal-02981156/document> (date of access: 18.05.2024).
23. Chen H., Liu H. How to Measure Word Length in Spoken and Written Chinese. *Journal of Quantitative Linguistics*. Vol. 23(1). P. 5–29. DOI: <https://doi.org/10.1080/09296174.2015.1071147> (date of access: 18.05.2024).
24. Cullen A. How artificial intelligence «works» in literary translation. Goethe Institut. URL: <https://www.goethe.de/prj/k40/en/lan/lit.html> (date of access: 17.04.2024).
25. Czulo O. Approaching machine translation from translation studies – a perspective on commonalities, potentials, differences. *European Association for Machine Translation Conferences/Workshops*, Dubrovnik, 2014. P. 199–206
26. De Gispert A., Iglesias G., Byrne B. Fast and accurate preordering for smt using neural networks. *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, Denver, Colorado, 2015. P. 1012–1017. URL: <http://www.aclweb.org/anthology/N15-1105> (date of access: 16.04.2024).
27. Demystifying the Blue Score: A Python Implementation for Evaluating Language Quality. Medium, 2023. URL: <https://medium.com/@TheDataScience-ProF/demystifying-the-blue-score-a-python-implementation-for-evaluating-language-quality-c44db88738e0> (date of access: 18.05.2024).
28. Domain Terminology Integration into Machine Translation: Leveraging Large Language Models/ ed. by Y. Moslem et al. *WMT*, 2023. P. 902–911. URL: <https://api.semanticscholar.org/CorpusID:264426476> (date of access: 17.04.2024).
29. Evaluate Library. The Python Package Index (PyPI). URL: <https://pypi.org/project/evaluate/> (date of access: 18.05.2024).
30. Farahani M, Dastjerdi H. Corpora and Translation Studies: Implications and Applications. *Journal of Language and Translation*, 2021, 11(3). P. 215–223.

31. Frawley W. Translation: Literary, Linguistic, and Philosophical Perspectives. Newmark: University of Delaware Press, 1984. 218 p.
32. Gui Xin Pan. A Contrastive Study of the English and Chinese Comparative Constructions. URL: https://space.snu.ac.kr/bitstream/10371/81462/1/05_Gui%20Xin%20Pan.pdf (date of access: 18.05.2024).
33. Guodong Zhao. Contrastive Study of English and Chinese Word Order from the Perspective of Figure-ground Theory—A Case Study of The Moonstone and Its Chinese Version. DOI: <http://dx.doi.org/10.17507/jltr.0702.20> (date of access: 18.05.2024).
34. Hutchins W. J., Somers H. L. An introduction to machine translation. London: Academic Press, 1998. 362 p.
35. Improved Zero-shot Neural Machine Translation via Ignoring Spurious Correlations/ Gu J. et al. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence: Association for Computational Linguistics, 2019. P. 1258–1268. URL: <https://aclanthology.org/P19-1121> (date of access: 17.04.2024).
36. Kaibao H., Yi L. Features of machine translation and its relation with manual translation, *Chinese Translators Journal*. 2016. Vol. 38, no. 5, P. 13–28.
37. Kanouchi S., Sudoh K., Komachi M. 2016. Neural reordering model considering phrase translation and word alignment for phrase-based translation. *Proceedings of the 3rd Workshop on Asian Translation (WAT2016)*. The COLING 2016 Organizing Committee, Japan, Osaka, 2016. P. 94–103. URL: <http://aclweb.org/anthology/W16-4607> (date of access: 16.04.2024).
38. Kaszubski P. Corpora in Applied Linguistics, *ELT Journal*. 2003. № 4, Vol. 57. P. 416–420. DOI: <https://doi.org/10.1093/elt/57.4.416> (date of access: 18.04.2024).
39. Kennedy G. D. An Introduction to Corpus Linguistics: Studies in Language and Linguistics. London: Longman, 1998. 315 p. URL: <https://press.umich.edu/pdf/9780472033850-part1.pdf> (date of access: 18.04.2024).

40. Koehn P. Neural Machine Translation. Cambridge: Cambridge University Press, 2020. 406 p.
41. Koehn P., Knowles R. Six challenges for neural machine translation. *First Workshop on Neural Machine Translation*, 2017. P. 28–39. URL: <https://aclanthology.org/W17-3204> (date of access: 17.04.2024).
42. Konstantinov M. A Guide to Neural Machine Translation With Attention Mechanism. *ELEKS*. URL: <https://eleks.com/research/neural-machine-translation-attention-mechanism/> (date of access: 18.05.2024).
43. Kreutzer J. et al. Quality at a Glance: An Audit of Web-Crawled Multilingual Datasets. *TACL*, 2022. P. 50–72. URL: <https://aclanthology.org/2022.tacl-1.4> (date of access: 18.05.2024).
44. Lin C. C., Hu, H. Linking comprehension and production: Frequency distribution of Chinese relative clauses in the Sinica Treebank. *Text, Speech, and Language Technology Series*. Springer. 2022. P. 22. DOI: https://www.researchgate.net/publication/358039901_Linking_Comprehension_and_Production_Frequency_Distribution_of_Chinese_Relative_Clauses_in_the_Sinica_Treebank (date of access: 01.05.2024).
45. Lin T. H. Comparative Grammar of Chinese and English. National Tsing Hua University: Graduate Institute of Linguistics, 2006. 91 p. URL: https://issuu.com/andygui/docs/comparative_grammar_of_chinese_and_ (date of access: 18.05.2024).
46. Ling L. P., Sepora T. The Differences between English and Chinese Language Sentence Structure and their Impacts to English-Chinese Machine Translation. The 14th International Conference on Translation. 2013. 10 p. URL: https://www.researchgate.net/publication/311933809_The_Differences_between_English_and_Chinese_Language_Sentence_Structure_and_their_Impacts_to_English-Chinese_Machine_Translation (date of access: 18.05.2024).
47. Liu L., Utiyama M., Finch A., et al. Agreement on target- bidirectional neural machine translation. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language*

Technologies. Association for Computational Linguistics. 2016. P. 411–416. URL: <https://aclanthology.org/N16-1046> (date of access: 18.05.2024).

48. Locke W. N., Booth A. D. Machine Translation of Languages – Fourteen Essays. *Translation* / ed. By Weaver W. Cambridge: The Massachusetts Institute of Technology. 1955, P. 15–23.

49. Luo J., Li D. Universals in machine translation?: A corpus-based study of Chinese-English translations by WeChat Translate. *International Journal of Corpus Linguistics*. 2022. Vol. 27, no. 1, P. 31–58. DOI: <https://doi.org/10.1075/ijcl.19127.luo> (date of access: 01.05.2024).

50. Machine Translation Evaluation: The Ultimate Guide. *MachineTranslation.com*. URL: <https://www.machinetranslation.com/blog/machine-translation-evaluation-ultimate-guide> (date of access: 18.05.2024).

51. Machine Translation Explained: Types, Use Cases, and Best Practices. *Phrase*. URL: <https://phrase.com/blog/posts/machine-translation/#the-beginnings-of-machine-translation> (дата звернення: 16.04.2024).

52. Machine translation: An introductory guide/ ed. by D. Arnold et al. Manchester: NCC Blackwell, 1994. 240 p.

53. Mel'cuk I., Wanner L. Syntactic mismatches in machine translation. *Machine Translation*. 2006. Vol. 20. P. 81–138. DOI: <http://dx.doi.org/10.1007/s10590-006-9013-7> (date of access: 01.05.2024).

54. Metric: bleu. *Hugging Face*. URL: <https://huggingface.co/spaces/evaluate-metric/bleu> (date of access: 18.05.2024).

55. Neural Machine Translation With Attention Mechanism: Step-by-step Guide. *Eleks*. URL: https://miro.medium.com/v2/resize:fit:1280/format:webp/1*Mm8sMamk7g4SCdDglAhJNg.gif (date of access: 01.05.2024).

56. Neural network based bilingual language model growing for statistical machine translation / ed. by R. Wang et al. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Doha, Qatar, 2014. P. 189–195. URL: <http://www.aclweb.org/anthology/D14-1023> (date of access: 16.04.2024).

57. Papineni K. et al. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*. Philadelphia: Association for Computational Linguistics. 2002. P. 311–318. URL: <https://aclanthology.org/P02-1040/> (date of access: 18.05.2024).
58. Parsing. Lancaster University. URL: <https://www.lancaster.ac.uk/fss/courses/ling/corpus/Corpus2/2PARSE.HTM#:~:text=Parsing%20involves%20the%20procedure%20of,are%20sometimes%20known%20as%20treebanks> (date of access: 01.05.2024).
59. Pastor G. C. et al. Translation universals: do they exist? A corpus-based NLP study of convergence and simplification. *The 8th Conference of the Association for Machine Translation in the Americas: Research Papers*. 2008. P. 75–81. URL: <https://aclanthology.org/2008.amta-papers.5> (date of access: 18.05.2024).
60. Pranav A., Chung S., Wagers M.. Widening the Net: Challenges for Gathering Linguistic Data in the Digital Age. Submitted to the National Science Foundation SBE 2020 planning activity., 2011. 5 p. URL: <https://people.ucsc.edu/~schung/anandchungwagers.pdf> (date of access: 18.04.2024).
61. Qin T. The Differences between Chinese and English. *Do My Essay.net*. URL: <https://www.domyessay.net/blog/difference-between-english-and-chinese/> (date of access: 01.05.2024).
62. Rei R. et al. COMET: A Neural Framework for MT Evaluation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, 2020. P. 2685–2702. URL: <https://aclanthology.org/2020.emnlp-main.213> (date of access: 18.05.2024).
63. Reimers N., Gurevych I. Making Monolingual Sentence Embeddings Multilingual Using Knowledge Distillation. *ArXiv*. 2020. Vol. abs/2004.09813. URL: <https://api.semanticscholar.org/CorpusID:216036089> (date of access: 01.05.2024).
64. Reiter E. A Structured Review of the Validity of BLEU. *CL*. Cambridge: MIT Press, 2018. Vol. 44, no. 3. P. 393–401. URL: <https://aclanthology.org/J18-3002> (date of access: 18.05.2024).

65. Sang Y., Chen Y., Zhang J. Neural Machine Translation Research on Syntactic Information Fusion Based on the Field of Electrical Engineering. *Applied Sciences*. 2023, Vol. 13, no. 13(23). 14 p. DOI: <https://doi.org/10.3390/app132312905> (date of access: 01.05.2024).
66. Schwenk H. Continuous space translation models for phrase-based statistical machine translation. *Proceedings of COLING 2012: Posters*. The COLING 2012 Organizing Committee, Mumbai, India, 2012. P. 1071–1080 URL: <http://www.aclweb.org/anthology/W12-2702> (date of access: 16.04.2024).
67. Shannon C. E., Weaver W. The Mathematical Theory of Communication. University of Illinois Press, 1949. 125 p.
68. Shevchenko O. M. Neural Machine Translation: Basic principles, strengths and weaknesses. *Monographic series «European Science»*. 2023. Book 18, Part 4. P. 86–92. URL: <https://ela.kpi.ua/handle/123456789/57539> (date of access: 16.04.2024).
69. SweetLink. syntax-corpus-analysis-uk-zh. *GitHub*. URL: <https://github.com/SweetLink/syntax-corpus-analysis-uk-zh> (date of access: 01.06.2024).
70. TED2020. *Opus*. URL: <https://opus.nlpl.eu/TED2020/zh&uk/v1/TED2020> (date of access: 01.05.2024).
71. Tennison S. E. Chinese word order: A look into SVO and and SOV Arguments. *Portland State University: University honors theses*. 2022. 23 p. DOI: <https://doi.org/10.15760/honors.1244> (date of access: 18.05.2024).
72. The Impact and Future of Neural Machine Translation on a Globalized World. *Pollion*. URL: <https://pollion.net/neural-machine-translation/> (date of access: 16.04.2024).
73. Thompson B., Post M. Automatic machine translation evaluation in many languages via zero-shot paraphrasing. *In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2020. P. 90–121. DOI: [10.18653/v1/2020.emnlp-main.8](https://doi.org/10.18653/v1/2020.emnlp-main.8) (date of access: 01.05.2024).
74. Tools for Corpus Linguistics. *Corpus-analysis.com*. URL: <https://corpus-analysis.com> (date of access: 01.05.2024).

75. Wei J. A study on the translation strategies of Chinese run-on sentences from the perspective of functional equivalence theory. URL: <https://www.davidpublisher.com/Public/uploads/Contribute/64d05532ead23.pdf> (date of access: 18.05.2024).

76. Xiao R. Contrastive corpus linguistics: Cross-linguistic contrast of English and Chinese / ed. by B. Zou, S. Smith, M. Hoey. *Corpus Linguistics in Chinese Contexts*. New York: Palgrave Macmillan, 2015. P. 35–62.

77. Xu J., Li J. A syntactic complexity analysis of translational English across genres. *Across Languages and Cultures*. 2021. Vol. 22, no. 2. P. 214–232. DOI: <https://doi.org/10.1556/084.2021.00015> (date of access: 01.05.2024).

ДОДАТКИ

Додаток А. Текст промови «坐轮椅潜行深海» («Плавання морськими глибинами в інвалідному візку») спікерки Сью Остін китайською

好高兴可以响呢度 同大家分享我慨心路历程 同埋讲下呢张轮椅带畀我慨自由

十六年前我开始坐轮椅 一场大病令我唔可以 好似以前咁接触世界

当我开始坐轮椅 渠带畀我全新慨自由

我曾经以为 病后伤残，下半世一定唔好过

但呢张轮椅就好似一件神奇玩具

令我可以出返去吹风

而且，可以出到街真系好爽！

虽然重获自由，我真系好开心 但街外人对我慨态度就完全唔同晒

渠哋好似当我透明 就好似我着咗件隐形斗篷咁

渠哋带住一份成见嚟睇我 觉得我坐轮椅就一定会系点点点

当我问「一提起轮椅 你哋会谗起乜嘢？」 渠哋会话「局限」、「恐慌」
「可怜」、「无自由」

我发现我都受到呢啲成见影响 亦令我从根本对自己慨睇法唔同咗

感觉自己有一部份唔再系自己咁

我再唔系由自己慨观点出发去审视自己 而系受其他人对我慨睇法影响

于是我觉得咁唔掂 我要活出自己慨故事 我要重新演绎，定义自我 [（投影片）寻找自我： 嚟活出自己慨故事时，

我哋学识认真看待 写自己生命的一字一句。—— Davis 2009， TEDx 女性中发表] 我开始做一啲作品 旨在向大家传递 我坐轮椅慨嗰份内心喜悦同自由 我叫佢做能量椅 因为可以同个世界撻过

我将我内心慨感受尽量传递出嚟 透过破格慨摄影 去改变当初我开始坐轮椅时 对自己慨成见

张轮椅变咗一件我可以又写又画慨嘢

当我开始为自己留下快乐同自由慨印记 外界对我慨作品有反应，令我好雀跃

呢啲作品好似打开咗新慨视野 引领艺术转型

艺术创作可以重塑一个人慨自我观感 同时改变固有睇法，改变成见

当我 2005 年开始坐轮椅潜水我发现水肺潜水装置同轮椅都一样畀到你水底慨自由可以话水肺潜水带畀人慨印象系充满刺激同历奇而且同轮椅畀人慨印象完全唔同

我心谗：唔知两件事摆埋一齐会点呢？（笑声、掌声）最终我哋真系做咗部潜水轮椅 而且响过去七年 渠带我体验咗非一般慨水底旅行

为咗令你哋明白我讲慨经历 我想同大家分享呢件法宝 带畀我慨其中一段体验 同埋渠带咗我去慨世界有几咁精彩

（音乐） （掌声） 呢个系我呢世人至今最精彩慨体验 以前种种都无得比

我基本上可以自由咁郁动 360 度通行无阻 体验至高无上慨快乐同自由

至于令我喜出望外，妙不可言慨系 其他人似乎都可以感同身受

渠哋对眼当堂发晒光 又话：「我都想要要部咁慨嘢！」 又有话：「如果你连潜水做得到， 我有嘢唔可以做唔到！」

我觉得，因为嗰一刻 渠哋睇紧一件嘢系渠哋前所未见 或者超乎渠哋想像 打破晒渠哋对轮椅慨观感 以致令渠哋重新看待轮椅

我同时觉得 潜水轮椅想法出现慨嗰一刻 将自由慨呢一份情怀感染到其他人

呢件事话畀我哋听 普通人都睇得到差异慨价值 同箇中慨快乐 当我哋唔系着眼于损伤或者局限 而系用新视野、新角度去睇 我哋就会发现力量同喜悦

轮椅系一部转化器

我而家叫呢部潜水轮椅做「变身器」 系因为渠令我踏入全新慨生命 揭开人生新一页，对自我有新慨觉悟

因为之前无人见过潜水轮椅系点 所以潜水轮椅面世 象征着新睇法、新认知 而家呢种睇法都响大家心中扎实

所以你哋都系作品慨一部份喇！

(掌声)

Додаток Б. НМП тексту промови «坐轮椅潜行深海» («Плавання морськими глибинами в інвалідному візку») спікерки Сью Остін через систему НМП DeepL

Я радий бути тут, щоб поділитися з вами своєю подорожжю і розповісти про свободу, яку подарував мені інвалідний візок.

Шістнадцять років тому я став прикутим до інвалідного візка, і серйозна хвороба заважала мені бачити світ так, як я звик.

Коли я сів у візок, це дало мені новий вид свободи.

Раніше я думав, що бути хворим і калікою - це погано на все життя.

Але цей візок - як чарівна іграшка.

Він дозволяє мені гуляти з вітром.

І це так приємно мати можливість вийти на вулицю!

Я така щаслива, що знову вільна, але люди на вулиці ставляться до мене по-іншому.

Вони ставляться до мене так, ніби я невидима, ніби я ношу плащ-невидимку.

Вони дивляться на мене з упередженням, ніби я в інвалідному візку і я трохи безладний.

Коли я запитую: «Що ви думаєте, коли думаєте про інвалідний візок?». Вони відповідали: «Обмеження», «паніка», «жалість», «відсутність свободи».

Я зрозумів, що перебував під впливом цих упереджень і що я принципово бачив себе інакше.

Я відчув, що частина мене вже не була собою.

Замість того, щоб дивитися на себе з власної точки зору, я піддавався впливу того, як на мене дивилися інші люди.

Тому мені було так погано, що я мусила прожити власну історію, переосмислити і визначити себе [(Слайд) «Знайти себе»: Проживаючи власну історію, ми навчилися серйозно ставитися до власної ідентичності.

Ми навчилися серйозно ставитися до кожного слова нашого життя. -- Девіс 2009, TEDx Women] Я почала робити деякі роботи, які мали на меті

передати внутрішню радість і свободу перебування в інвалідному візку, і я називаю його «енергетичним кріслом», тому що я можу взяти світ за роги.

Я намагаюся передати те, що я відчуваю всередині, змінити те, як я думав про себе, коли вперше почав їздити на візку, через позаконтекстну фотографію.

Візок став чимось, на чому я можу писати і малювати.

Коли я почала залишати на собі слід щастя і свободи, я була настільки схвильована, що зовнішній світ відгукнувся на мої роботи.

Ці роботи, здається, відкривають нові горизонти і ведуть до трансформації мистецтва.

Мистецтво здатне перевернути уявлення про себе, змінити перспективу і стереотипи.

Коли я почав пірнати в інвалідному візку в 2005 році, я зрозумів, що і акваланг, і інвалідний візок дають однакову свободу під водою. Можу сказати, що враження від підводного плавання сповнені хвилювання і здивування, і вони зовсім не схожі на враження від занурення в інвалідному візку.

Я подумав: «Цікаво, що станеться, коли ці дві речі поєднати? (Сміх, оплески) Зрештою, у нас з'явився водолазний візок, і протягом останніх семи років він возив мене у надзвичайні підводні подорожі.

Щоб ви могли зрозуміти, про що я говорю, я хотів би поділитися з вами одним із вражень, які подарував мені цей скарб, і яким чудовим є світ, в який він мене переніс.

Це найкращий досвід, який я мав у своєму житті, і він не схожий ні на що, що я коли-небудь мав раніше.

Я можу практично вільно пересуватися на 360 градусів і відчувати найвищий рівень радості та свободи.

І що зробило мене такою щасливою, такою чудовою, так це те, що всі інші, здавалося, могли зрозуміти мене.

Їхні очі засяяли, і вони сказали: «Я хочу такий!». «Якщо ти можеш пірнати з аквалангом, то немає нічого, чого б я не зміг!»

Я думаю, що саме момент, коли вони бачать те, чого ніколи раніше не бачили, або те, що виходить за межі їхньої уяви, ламає їхнє сприйняття інвалідних візків і змушує їх поглянути на них знову.

Я також думаю, що в той момент, коли з'явилася ідея інвалідного візка для пірнання, вона поширила відчуття свободи на інших.

І, скажімо, звичайні люди бачать різницю в цінності та радості, в силі та радості, яку ми знаходимо, коли не дивимося на пошкодження чи обмеження, а дивимося на це з новим баченням, з новою перспективою.

Візок – це трансформатор.

Я називаю цей візок «трансформером», тому що він дозволяє мені увійти в нове життя, відкрити нову сторінку в моєму житті, по-новому усвідомити себе.

Ніхто ніколи раніше не бачив інвалідного візка для пірнання, тому поява такого візка символізує новий погляд на речі, новий спосіб бачення речей, і цей спосіб бачення речей закріплюється у свідомості кожного з вас.

Тож ви всі є частиною цієї роботи!

(Оплески)

**Додаток В. Ручний переклад тексту промови «坐轮椅潜行深海»
(«Плавання морськими глибинами в інвалідному візку») спікерки Сю
Остін через систему НМП DeerL, виконаний людиною-перекладачем**

Я дуже рада мати змогу поділитися з вами своєю подорожжю та разом з тим розповісти про свободу, яку дав мені інвалідний візок.

Опинилася я в інвалідному візку шістнадцять років тому. Серйозна хвороба забрала в мене можливість торкатися світу так, як раніше.

Однак з інвалідним візком для мене відкрилась абсолютно нова свобода.

Раніше я думала, що з обмеженими можливостями через хворобу друга половина мого життя вже точно не буде легкою.

Але ця коляска як чарівна іграшка, бо з нею я можу вибиратися на свіже повітря.

Крім того, можливість вийти на вулицю це дуже класно!

Хоча я й щаслива знову бути вільною, але ставлення до мене з боку інших тепер зовсім інше.

Здається, ніби я для них прозора, що ніби одягнена у плащ-невидимку.

Вони дивляться косо на мене з упередженням, думаючи, що, якщо сиджу в інвалідному візку, то обов'язково обмежена у всьому.

Коли я запитала: "Що спадає вам на думку, коли ви чуєте про інвалідний візок?", вони сказали: "обмеження", "паніка", "жалість", "обмеженість свободи".

Я усвідомила, наскільки на мене вплинули ці стереотипи і змусили сприймати себе принципово інакше.

Я відчуваю, що частина мене більше не є мною

Я більше не дивлюся на себе з власної точки зору, а радше на мене впливає те, як мене бачать інші. Відчуваю, що частково я не така, якою була раніше.

Через такий дискомфорт мені довелося вдатися до переосмислення та самовизначення [(Слайд Powerpoint) «Знайти себе: живучи своєю власною історією».

«Ми навчилися серйозно ставитися до кожного слова, яке пишемо про своє життя.» —Davis 2009, опубліковано в TEDx Women]

Я почала проводити різні заходи, аби передати всім почуття внутрішньої радості і свободи, які маєш перебуваючи в інвалідному візку.

Я називаю його кріслом сили, тому що можу пересуватися в ньому по всьому по світу.

Я намагалася якнайкраще передати свої внутрішні відчуття через нестандартне зображення, щоб змінити стереотип про себе, коли я опинився в інвалідному візку.

Він дав мені змогу писати та малювати.

Коли я почала буквально залишати сліди свого внутрішнього щастя і свободи, то і зовнішній світ відгукнувся на мої роботи. І я була цьому безмежно рада.

Ці твори ніби відкрили мені нові горизонти й направили в бік мистецтва.

Мистецтво може трансформувати самосприйняття людини, змінити і властиві їй погляди, і стереотипи.

Коли я почала займатися підводним плаванням в інвалідному візку в 2005 році, то виявила, що як з обладнанням для підводного плавання, так і з візком під водою маєш однакову свободу. Можна сказати, що спорядження для підводного плавання справляє враження, повне хвилювання та пригод. відрізняється від враження, яке справляють інвалідні візки.

І це враження від екіпірування для підводного плавання, повне хвилювання та пригод, повністю відрізняється від того, яке справляють інвалідні візки.

Я подумала: чи можна ці дві речі об'єднати? (Сміх, оплески) Зрештою ми справді зробили інвалідний візок для дайвінгу, і за останні сім років я здійснила надзвичайні підводні подорожі.

Щоб ви зрозуміли, про що саме я говорю, то хочу поділитися з вами одним із досвідів, які принесла мені ця чарівна зброя, і яким чудовим є світ, у який вона мене привела. (Музика) (Оплески)

Це найпрекрасніший досвід, який я коли-небудь мала у цьому житті.

Я можу практично вільно пересуватися без перешкод на 360 градусів і відчувати неперевершене щастя та свободу.

Що стосується мого здивування й захоплення, то, здається, ніби інші також відчують те саме.

Вони переглядалися і казали: «Я хочу так само!» Інші казали: «Якщо ти навіть вмієш займатися дайвінгом, немає нічого, чого б я не зміг!»

Я думаю, що в той момент вони бачили щось, чого ніколи раніше не бачили або що було за межами їх уяви, і це зруйнувало їх уявлення про інвалідні візки, змушувало поглянути на них новими очима.

Я також відчуваю, що в той момент, коли виникла ідея інвалідного візка для дайвінгу, то це відчуття свободи заразило інших людей.

Тож погляньте на цю історію. Звичайні люди можуть побачити різні цінності і відчути радість, не зосереджуючись на недоліках чи обмеженнях, а дивлячись на це з нових горизонтів і під новими кутами. Тільки тоді ми знаходимо силу та щастя.

Інвалідний візок це трансформатор, перевтілювач.

Зараз я називаю цей візок для підводного плавання «трансформером», тому що він дозволяє мені зробити крок у нове життя, відкрити нову сторінку в житті та по-новому усвідомити себе.

Оскільки ніхто раніше не бачив нічого спільного між інвалідним візком і оснащенням для пірнання під воду, то запуск інвалідного візка для дайвінгу символізує новий спосіб бачення та нове розуміння. Це нове бачення має сильний вплив на серце кожного.

Отже, ви всі є частиною цієї праці!

(оплески)

Додаток Г. Приклади окремих складених авторських функцій для здійснення корпусного аналізу

Додаток Г.1. Функція для визначення загальної кількості складнопідрядних речень в україномовних текстах корпусу

```
subclauses_uk.py •
analysis-functions > subclause-analysis > subclauses_uk.py > ...
You, 4 weeks ago | 1 author (You)
1 import spacy You, 4 weeks ago • init
2
3 # Завантажуємо модель для української мови
4 nlp = spacy.load("uk_core_news_sm")
5
6 def count_subordinate_clauses(text):
7     doc = nlp(text)
8     subordinate_count = 0
9
10    for sent in doc.sents:
11        # Перевіряємо, чи є поточне речення складнопідрядним
12        if any(token.dep_ == 'advcl' or token.dep_ == 'acl' for token in sent):
13            subordinate_count += 1
14
15    return subordinate_count
16
17 # Зчитуємо текст з файлу
18 with open("uk-zh.txt (1)/TED2020.uk-zh.uk", "r", encoding="utf-8") as file:
19     text = file.read()
20
21 # Розділяємо текст на менші частини (по 500000 символів)
22 text_parts = [text[i:500000] for i in range(0, len(text), 500000)]
23
24 # Ініціалізуємо лічильник підрядних речень
25 total_subordinate_count = 0
26
27 # Обробляємо кожну частину тексту окремо
28 for part in text_parts:
29     subordinate_count = count_subordinate_clauses(part)
30     total_subordinate_count += subordinate_count
31
32 print("Загальна кількість складнопідрядних речень:", total_subordinate_count)
33
34 # Загальна кількість складнопідрядних речень: 3750
```

Джерело: розроблено автором.

Додаток Г.2. Функція для визначення кількості дієслівних фраз в китайськомовних текстах корпусу

```
vp_zh.py ×
analysis-functions > np-vp-analysis > vp_zh.py > ...
You, 4 weeks ago | 1 author (You)
1 import jieba.posseg as pseg
2
3 def count_verb_phrases(text):
4     sentences = text.split('.') # Розділити текст на речення
5
6     total_verb_phrases = 0
7     for sentence in sentences:
8         words = pseg.cut(sentence.strip()) # Токенізація китайського речення
9         verb_phrase_count = 0
10
11        # Підрахунок кількості дієслівних фраз у кожному реченні
12        for word, flag in words:
13            if flag.startswith('v'): # Якщо слово - дієслово
14                verb_phrase_count += 1
15
16        total_verb_phrases += verb_phrase_count
17
18    return total_verb_phrases
19
20 # Завантажити корпус китайських текстів
21 with open('uk-zh.txt (1)/TED2020.uk-zh.zh', 'r', encoding='utf-8') as file:
22     text = file.read()
23
24 verb_phrase_count = count_verb_phrases(text)
25 print("Кількість дієслівних фраз у корпусі: ", verb_phrase_count)
26
27 # Кількість дієслівних фраз у корпусі: 54091 You, 4 weeks ago • init
```

Джерело: розроблено автором.

Додаток Г. Приклад авторської функції, складеної для здійснення автоматизованого оцінювання сегменту за метрикою BLEU

```
mybleu_1.py •
bleu_test > mybleu_separate > mybleu_1.py > ...
You, 3 weeks ago | 1 author (You)
# https://huggingface.co/spaces/evaluate-metric/bleu
1
2
3 import nltk
4 import evaluate
5 from nltk.translate.bleu_score import corpus_bleu
6 predictions = ["Я радий бути тут, щоб поділитися з вами своєю подорожжю і розповісти про свободу, яку подарував мені інвалідний візок.
7 references = [
8     ["Я дуже рада мати змогу поділитися з вами своєю подорожжю та разом з тим розповісти про свободу, яку дав мені інвалідний візок."],
9     ["Опинилася я в інвалідному візку шістнадцять років тому."],
10    ["Серйозна хвороба забрала в мене можливість торкатися світу так, як раніше."],
11    ["Однак з інвалідним візком для мене відкрилась абсолютно нова свобода."],
12    ["Раніше я думала, що з обмеженими можливостями через хворобу друга половина мого життя вже точно не буде легкою."]
13 ]
14 bleu = evaluate.load("bleu")
15 results = bleu.compute(predictions=predictions, references=references)
16 print(results)
17
18 # {'bleu': 0.1756825820356824, 'precisions': [0.40540540540540543, 0.2318840579710145, 0.140625, 0.0847457627118644],
19 # 'brevity_penalty': 0.9602702338479381, 'length_ratio': 0.961038961038961, 'translation_length': 74, 'reference_length': 77} You,
```

Джерело: розроблено автором.