



УДК 354:328.185

[https://doi.org/10.52058/2786-6165-2024-9\(27\)-500-514](https://doi.org/10.52058/2786-6165-2024-9(27)-500-514)

Хоптяр Алла Олександрівна кандидат філологічних наук, доцент, декан факультету іноземної філології, Кам'янець-Подільський національний університет імені Івана Огієнка, вул. Огієнка, 61, м. Кам'янець-Подільський, 32301, <https://orcid.org/0000-0003-4277-3856>

Катуніна Ольга Сергіївна кандидат економічних наук, доцент, доцент кафедри математичного моделювання та статистики, Навчально-науковий інститут "Інститут інформаційних технологій в економіці", Київський національний економічний університет імені Вадима Гетьмана, просп. Берестейський, 54/1, м. Київ, 03057, <https://orcid.org/0000-0001-7584-0037>

Калініченко Тетяна Миколаївна кандидат педагогічних наук, доцент кафедри загального мовознавства і романо-германської філології, факультет іноземної філології, Харківський національний педагогічний університет ім. Г.С. Сковороди, вул. Алчевських, 29, м. Харків, 61002, <https://orcid.org/0000-0001-7672-5858>

ВИКОРИСТАННЯ МОВНИХ КОРПУСІВ У ДОСЛІДЖЕННІ УСНОГО ТА ПИСЬМОВОГО ПЕРЕКЛАДУ: АНАЛІЗ ВЕЛИКОМАСШТАБНИХ МОВНИХ ДАНИХ

Анотація. Використання мовних корпусів у дослідженні усного та письмового перекладу стало важливим інструментом у сучасній лінгвістиці та перекладознавстві. Корпуси дозволяють аналізувати великомасштабні мовні дані, що сприяє глибшому розумінню мовних структур, закономірностей та особливостей перекладацької діяльності. Дане дослідження присвячене вивченню методів і підходів до використання мовних корпусів для аналізу усного та письмового перекладу. Метою дослідження є визначення основних переваг та обмежень використання корпусних даних у перекладознавстві, а також виявлення ключових аспектів, які впливають на якість перекладу. У статті розглядаються різні типи мовних корпусів, включаючи паралельні, порівняльні та спеціалізовані корпуси, і їх застосування для дослідження перекладу. Особлива увага приділяється автоматизованим методам обробки мовних даних, таким як статистичні та машинні підходи, які



дозволяють проводити точний аналіз великих обсягів текстів. Результати дослідження показують, що використання мовних корпусів значно покращує якість аналізу та інтерпретації перекладених текстів, дозволяє виявляти частотність вживання певних лексичних одиниць, граматичних конструкцій та стилістичних прийомів. Наприклад, за допомогою паралельних корпусів можна детально аналізувати, як різні перекладачі передають одні й ті самі фрагменти тексту, що допомагає ідентифікувати успішні перекладацькі стратегії. Порівняльні корпуси, у свою чергу, дозволяють зіставляти тексти оригіналу та перекладу для виявлення тенденцій і відмінностей у мовленні різними мовами. Корпусний аналіз сприяє виявленню помилок та недоліків у перекладацькій діяльності, що в свою чергу допомагає покращити навчальні програми для підготовки професійних перекладачів. Наприклад, систематичний аналіз помилок, зроблених студентами-перекладачами, дозволяє визначити найбільш проблемні зони та розробити відповідні навчальні матеріали. В подальшому дослідженні планується розширити аналіз, включивши більше мовних пар та різних жанрів текстів, щоб отримати більш детальну картину перекладацьких процесів і вдосконалити методологію перекладу. Це сприятиме підвищенню якості перекладу як у професійній, так і в освітній сферах, забезпечуючи краще розуміння перекладацьких технік і стратегій. Одним із ключових напрямків майбутніх досліджень є вивчення впливу культурних і контекстуальних факторів на переклад. Мовні корпуси можуть бути корисними для аналізу, як різні культурні контексти впливають на вибір лексики та стилю перекладу.

Ключові слова: мовні корпуси, усний переклад, письмовий переклад, великомасштабні дані, автоматизований аналіз, перекладознавство, лінгвістичні структури, Natural Language Processing, Big Data, мовні моделі.

Khoptiar Alla Oleksandrivna PhD in Philological Sciences, Associate Professor, Dean of the Faculty of Foreign Philology, Kamianets-Podilskyi Ivan Ohienko National University, 61 Ohienka St., Kamianets-Podilskyi, 32301, <https://orcid.org/0000-0003-4277-3856>

Katunina Olha Serhiivna PhD in Economical Sciences, Associate Professor, Department of Mathematical Modeling and Statistics, Institute of Information Technologies in Economy, Vadym Hetman Kyiv National Economic University, 54/1 Beresteyskyi Ave., Kyiv, 03057, <https://orcid.org/0000-0001-7584-0037>



Kalinichenko Tetiana Mykolayivna PhD in Pedagogical Sciences, Associate Professor of the Department of General Linguistics and Romance and German Philology, Faculty of Foreign Philology, H.S. Skovoroda Kharkiv National Pedagogical University, 29 Alchevskyh St., Kharkiv, 61002, <https://orcid.org/0000-0001-7672-5858>

**THE USE OF LANGUAGE CORPSES IN THE RESEARCH OF
INTERPRETATION AND WRITTEN TRANSLATION:
ANALYSIS OF LARGE-SCALE LANGUAGE DATA**

Abstract. The use of language corpora in the study of oral and written translation has become an important tool in modern linguistics and translation studies. Corpora allow for the analysis of large-scale language data, which contributes to a deeper understanding of language structures, regularities and peculiarities of translation activity. This study is devoted to the study of methods and approaches to the use of language corpora for the analysis of oral and written translation. The purpose of the study is to determine the main advantages and limitations of using corpus data in translation studies, as well as to identify key aspects that affect the quality of translation. The paper examines different types of language corpora, including parallel, comparative and specialized corpora, and their application to translation research. Special attention is paid to automated methods of language data processing, such as statistical and machine approaches, which allow accurate analysis of large volumes of texts. The results of the study show that the use of language corpora significantly improves the quality of analysis and interpretation of translated texts, allows to identify the frequency of use of certain lexical units, grammatical constructions and stylistic techniques. For example, using parallel corpora, it is possible to analyze in detail how different translators render the same pieces of text, which helps to identify successful translation strategies. Comparative corpora, in turn, allow comparing original and translated texts to identify trends and differences in speech in different languages. Corpus analysis helps to identify errors and shortcomings in translation activities, which in turn helps to improve educational programs for training professional translators. For example, a systematic analysis of mistakes made by student translators allows to identify the most problematic areas and develop appropriate training materials. In further research, it is planned to expand the analysis by including more language pairs and different genres of texts in order to obtain a more detailed picture of translation processes and to improve translation methodology. This will contribute to the improvement of the quality of translation in both professional and educational spheres, providing a better



understanding of translation techniques and strategies. One of the key directions for future research is the study of the influence of cultural and contextual factors on translation. Linguistic corpora can be useful for analyzing how different cultural contexts influence the choice of vocabulary and translation style.

Keywords: language corpora, oral translation, written translation, large-scale data, automated analysis, translation studies, linguistic structures, Natural Language Processing, Big Data, language models.

Постановка проблеми. Мовні корпуси відіграють важливу роль у дослідженні перекладу, як усного, так і письмового. Вони представляють собою великі обсяги текстових даних, які можна аналізувати для вивчення лінгвістичних явищ, виявлення закономірностей та розробки нових методів перекладу. Використання великих мовних даних дозволяє отримувати більш точні результати і розширювати межі перекладацьких досліджень.

Основні проблеми, пов'язані з використанням мовних корпусів у перекладацьких дослідженнях, включають: обмеженість доступних корпусів, де всі мови мають достатньо великий та якісний корпус для проведення досліджень, якість даних, коли всі тексти у корпусах є однаково якісними, що може вплинути на результати досліджень, обробка та аналіз великих даних: необхідність використання спеціальних інструментів та методів для обробки великих обсягів даних, відмінності між усним та письмовим перекладом: усний переклад має свої особливості, які важко врахувати при аналізі письмових текстів. Для подолання цих проблем застосовуються різні методи:

1. Розробка нових корпусів для недостатньо представлених мов, а також розширення існуючих.
2. Використання інструментів для автоматичної анотації, аналізу та перекладу текстів.
3. Порівняння усних та письмових текстів для виявлення специфічних рис кожного типу перекладу.

Використання мовних корпусів є невід'ємною частиною сучасних перекладацьких досліджень. Незважаючи на існуючі проблеми, методи обробки великих даних дозволяють отримувати цінну інформацію та підвищувати якість перекладу. Подальший розвиток технологій та розширення доступних корпусів відкривають нові можливості для дослідження усного та письмового перекладу.

Аналіз останніх досліджень і публікацій. В останні роки використання мовних корпусів у дослідженнях перекладу значно зросло завдяки розвитку технологій та зростанню доступності великих масивів даних.

Одним із найважливіших напрямків є використання корпусів для вдосконалення алгоритмів автоматичного перекладу. Сучасні системи, такі як Google Translate та DeepL, активно використовують великі паралельні корпуси для тренування моделей. Останні дослідження зосереджені на підвищенні якості перекладу за допомогою глибокого навчання та нейронних мереж.

Дослідники використовують корпуси для вивчення різних лінгвістичних явищ, які мають важливе значення для перекладу, таких як фразеологізми, ідіоми та культурні реалії. Наприклад, робота Гізберта Грінеберга (2022) показала, як великі корпуси можуть допомогти виявити тонкощі перекладу культурно специфічних висловів. Корпуси дозволяють вивчати відмінності та спільні риси між різними мовами та культурами. Це особливо важливо для перекладачів, які працюють з літературними та технічними текстами. Наприклад, дослідження Люсі Джонс (2023) зосереджене на аналізі перекладів художніх творів з англійської на китайську мову.

Gisbert Grineberg (2022). "Cultural Nuances in Translation: Using Corpora to Decipher Idiomatic Expressions." – у цьому дослідженні автор розглядає, як великі корпуси допомагають перекладачам справлятися з культурними особливостями мов.

Lucy Jones (2023). "Cross-Cultural Literature Translation: A Corpus-Based Study of English to Chinese Translations." дослідження присвячене аналізу перекладів художніх текстів з англійської на китайську, зокрема вивченню адаптації культурних реалій. Mark Baker (2021). "Machine Translation and Corpus Linguistics: Advances in Neural Networks." Публікація, що розглядає новітні досягнення в області машинного перекладу та роль корпусів у тренуванні нейронних мереж.

Використання мовних корпусів у дослідженні перекладу має значний потенціал для вдосконалення якості перекладу, розуміння культурних відмінностей та розвитку нових методів у комп'ютерній лінгвістиці. Останні дослідження демонструють важливість інтеграції корпусних даних у практику перекладу та відкривають нові горизонти для подальших наукових розвідок.

Мета статті це дослідження ролі мовних корпусів у галузі перекладу, зокрема, в усному та письмовому перекладі, через аналіз великомасштабних мовних даних. Стаття має на меті визначити важливість мовних корпусів у сучасних перекладацьких дослідженнях та практиці, оцінити основні проблеми та виклики, з якими стикаються дослідники під час використання мовних корпусів, проаналізувати методи та інструменти, які використовуються для обробки та аналізу великомасштабних



мовних даних, розглянути останні дослідження та публікації у цій сфері, щоб виділити ключові тенденції та досягнення, а також запропонувати рекомендації щодо подальшого розвитку та вдосконалення використання мовних корпусів у перекладі. Стаття прагне показати, як мовні корпуси можуть підвищити якість перекладу, покращити розуміння культурних та лінгвістичних особливостей, а також сприяти розробці нових методів та технологій у сфері перекладу.

Виклад основного матеріалу. Думки дослідників щодо поняття перекладу є важливою основою для застосування методології, заснованої на інструментарії корпусної лінгвістики, для підготовки перекладачів, оскільки порушують проблему вибору об'єкта дослідження, тобто типу перекладу. переклад, коли, ким, для Існує основа для створення шаблону корпусу та питання репрезентативності, тобто рішення про те, які конкретні тексти можуть бути включені до корпусу, який використовуватиметься для вивчення цього конкретного типу перекладу. [1, с. 126]. Це має вирішальне значення для аналізу та інтерпретації даних, і його потрібно розуміти.

Найбільше хвилює перекладачів питання про те, як об'єкт співвідноситься з нормою, оскільки увага зазвичай зосереджується не тільки на об'єкті дослідження, але й на суб'єкті. Лінгвістичні особливості перекладних і неперекладених текстів можуть довести чи спростувати ефективність методів перекладу. Використання порівнянних корпусів розглядається як спосіб дослідження аспектів використання мови перекладачем, які не є результатом навмисних, контрольованих процесів. [2, с. 36]. Продукт перекладу може надавати непрямі докази когнітивної обробки, яка йде на переклад.

У дослідженні когнітивної складності та граматичної ясності англійської мови Гюнтер Роденбург висунув гіпотезу. Доказом принципу складності є формальні контрасти, пов'язані з видаленням або додаванням і заміною елементів закритого класу. У разі більш очевидних варіантів перевагу буде надано більш очевидним з когнітивної точки зору. [3, с. 151]. Більшість повідомлень англійською мовою можна розглядати як частину загальної моделі явності, і це можна пояснити когнітивною складністю завдання перекладу. Низку вимог до дескриптивної моделі використання мови сформулював у своєму дослідженні Ян Артс. [4, с. 6]: модель повинна поєднувати кількісні та якісні описи даних, встановлювати зв'язок між зовнішніми явищами та внутрішніми явищами щодо мовної системи та дозволяти описувати повний спектр різновидів. [5, с. 7].

Існують вимоги до опису мовного вживання, які відображені в описі інших вчених. Одним із джерел доказів для покращення опису структури

та використання є Грем Кеннеді. Характер лінгвістичних досліджень тексту не змінився завдяки прогресу комп'ютерних технологій. [6, с. 2]. Теоретичні можливості, які пропонує мовна система, протиставляються кількісному та якісному аналізу, який описує її під час використання. Кеннеді припускає, що інтерес до використання мови полягає не лише в тому, що відбувається, а в тому, що може статися. [7, с. 15]. Він підкреслює, що хоча теорії можуть бути отримані з корпусних досліджень, корпусна лінгвістика не вважається лінгвістичною теорією. Корпусні дані можна інтегрувати з іншими джерелами лінгвістичних доказів і застосовувати в різних рамках лінгвістичного опису, зосереджуючись на різних аспектах використання мови.

Корпусна лінгвістика є відносно новою галуззю дослідження, і гіпотеза про її потенційне застосування в практиці викладання перекладу все ще знаходиться на ранніх стадіях. Після ретельного вивчення досліджень, проведених лінгвістами та педагогами, ми виявили, що хоча концепція використання інструментів корпусної лінгвістики є відносно новою, вона активно розвивається та досягла позитивних результатів у певній групі лінгвістів-методистів.

Підхід до викладання корпусної лінгвістики може бути застосований практично в ширшій теоретичній системі перекладу. Оскільки він пропонує спосіб аналізу використання мови під час процесу перекладу. На відміну від більшості досліджень у сфері корпусної лінгвістики, корпусне перекладознавство не розглядає переклад переважно у зв'язку з вихідним текстом, а натомість підкреслює його як форму створення тексту та засіб комунікації. Зокрема Мона Бейкер пропонує використовувати [8, с. 234] під час викладання порівнянного корпусу (корпусу перекладів мовою та порівнянного корпусу оригінальних текстів тією самою мовою), який має охоплювати шаблони, які або обмежені текстом перекладу, або які зустрічаються з більшою чи меншою частотою в тексті перекладу [9, с. 238].

Вона зазначає, що ці лінгвістичні особливості можуть бути пов'язані з певною мовою, але ми також можемо отримати уявлення про загальні характеристики перекладеного тексту та самого процесу перекладу. Це призводить до концепції «універсальних особливостей перекладу», які часто називають «універсальними пекарями». Вона визначила кілька аспектів перекладу, які можна досліджувати за допомогою подібних корпусів [10, с. 250]. Наприклад, переклади можуть бути більш явними на багатьох рівнях порівняно з неперекладеними текстами, і вони можуть певним чином спрощувати, нормалізувати або стандартизувати.

Щоб перевірити цю гіпотезу, ми використали інструменти корпусної лінгвістики, зокрема паралельні корпуси, як метод навчання перекла-



дачів під час практики перекладу з англійської мови. На основі створеної нами колекції офіційних документів було створено паралельний корпус документів англійською та українською мовами. Оцінивши завдання студентів, було виявлено, що паралельний корпус є цінним ресурсом для перекладачів. Зокрема, було помічено, що коли зустрічалися незнайомі терміни, запити робилися під час процесу перекладу шляхом створення списків узгодженості на основі наданого нами паралельного корпусу [11, с. 43].

Зокрема, найбільше запитань викликав переклад таких словосполучень, як: «цим» – 13 запитів, «фізичний відображення» – 7 запитів, «діє до» – 5, «фінансова компенсація» – 5 запитів. У результаті ми твердо переконані, що включення паралельних корпусів у викладання практики перекладу є дуже ефективним та інноваційним підходом до навчання. Загальноприйнятим поглядом на дослідження перекладу на основі корпусу є те, що це методологія, яка дозволяє поєднувати або інтегрувати лінгвістичні [12, с. 50] Вивчення перекладу та різноманітних підходів до нього, включаючи лінгвістичні та культурні перспективи, детально досліджується в цьому дослідженні [12, с. 47]. Ці цілі тісно пов'язані з моніторингом зв'язку між текстом і контекстом, а також моделей дій у реальному житті [13, с. 265].

О. Мейсон визнає потенційну користь дослідження на основі корпусів у галузі перекладу. Він також застерігає від нехтування переконливими намірами, що стоять за створенням мови. Він підкреслює важливість урахування впливу жанру, дискурсу та мети тексту на перекладацькі рішення, а також інших факторів, таких як комунікативні цілі автора та сам процес перекладу. Корпуси надають певні типи даних, які повинні бути інтегровані у відповідну теоретичну структуру та поєднані з іншими даними з інших джерел, якщо дослідження має вийти за рамки тривіального використання комп'ютерних засобів при аналізі англомовних текстів демонструє тенденцію до використання в роботі перекладача.

Майбутні перекладачі змогли знайти їх типові риси завдяки функціям WordSmith Tools: Concord, Wordlist і lusters [14, с. 20-35]. У документах спостерігається тенденція до використання простих речень з прямим порядком слів, відсутність прямої мови, велика кількість скорочень, називних конструкцій. Особливості, виділені студентами, порівнювалися з порівняльним корпусом на веб-сайті. Функція Key Words демонструє тенденцію до широкого використання в документах термінологічних одиниць, абревіатур і усталених виразів. За допомогою цього виду аналізу можна виявити деякі риси та закономірності, пов'язані з офіційно-діловим стилем [14].

Щоб вийти за межі вивчення певних шаблонів чи встановлених норм, дослідження мають бути доповнені інформацією про процес перекладу, а також аналізом на додаткових лінгвістичних і синтаксичних рівнях. Мовні особливості та вибір перекладачів слід розглядати в контексті жанру, мети тексту та структури дискурсу. У цих випадках існують причинно-наслідкові зв'язки. Цей тип дослідження може надати нам гіпотези для подальшої перевірки. Розумно припустити, що вихідні мови для перекладу документів мають широкий спектр лінгвістичних елементів, які можуть бути використані для передачі типів значень, які передаються універсальними термінологічними одиницями.

Роль перекладача полягає в тому, щоб вибрати найбільш підходящий еквівалент, який максимально точно передаватиме те саме значення. Те, що ці лексичні одиниці вживаються в різних контекстах, може мати низку причин, зокрема: відсутність такого поняття в мові перекладу, що змушує перекладача використовувати перекладацькі трансформації, щоб компенсувати необхідну семантичну одиницю, універсальність мовних одиниць, що спонукає перекладача використовувати прийом конкретизації, щоб уникнути спотворення змісту.

Зокрема, для аналізу великомасштабних мовних даних перекладу необхідно використовувати технології NLP (Natural Language Processing) та Big Data. Вони є ключовими інструментами для автоматизованого перекладу, особливо завдяки використанню мовних моделей, що навчаються на великих корпусах текстових даних. Технології NLP дозволяють комп'ютерам розуміти, інтерпретувати та генерувати людську мову, а Big Data забезпечує доступ до величезних масивів текстової інформації для навчання цих систем. Завдяки взаємозв'язку цих двох технологій перекладацькі моделі можуть ефективно навчатися, підвищуючи точність та контекстуальну релевантність перекладів.

Окрім цього, Big Data [15] допомагає покращувати якість перекладів, надавати доступ до спеціалізованих термінологій, технічної документації та прикладів використання мов у різних контекстах. Це дає змогу створювати моделі, що адаптуються до різних галузей знань. У результаті NLP та Big Data сприяють створенню інструментів, які забезпечують не лише швидкий, але й високоякісний переклад, зменшуючи потребу в людській участі.

Роль NLP і Big Data в процесі перекладу є взаємодоповнюючою, проте кожна з цих технологій має свої унікальні переваги. NLP забезпечує структуроване розуміння мови, аналізуючи граматику, синтаксис і семантику, що особливо важливо для письмового перекладу. Вона дозволяє виявляти найточніші відповідники в іншій мові, зберігаючи



граматичні та стилістичні особливості оригінального тексту. Для усного перекладу NLP також корисне, оскільки допомагає в режимі реального часу аналізувати й генерувати мовні відповіді, забезпечуючи плавність та природність перекладу [16].

Big Data дозволяє працювати з великими обсягами тексту, накопичуючи та обробляючи інформацію з численних джерел. Це особливо корисно для перекладу рідкісних термінів, технічних текстів або діалектів, оскільки доступ до великої кількості даних допомагає забезпечити точність і різноманітність перекладів. Завдяки Big Data, мовні моделі можуть навчатися на широкому спектрі реальних прикладів, що підвищує їхню здатність адекватно перекладати як письмові, так і усні тексти в різних контекстах. Технології NLP забезпечує розуміння структури і змісту мови, а Big Data надає необхідний контекст і приклади для підвищення точності перекладу.

Очевидно, що перевірити гіпотезу за допомогою дослідження перекладів і подібних текстів на основі корпусу неможливо [17, с. 10-40]. Щоб зробити це, необхідно було б провести аналогічні ширші дослідження іншими мовами, як у перекладених, так і в неперекладених текстах. Стосовно наступної гіпотези, використання специфікації в перекладі може співвідноситися з іншими доказами, які перекладачі усувають неоднозначність і роблять явними на багатьох рівнях. У результаті було розглянуто проблему максимізації ефективності роботи викладача практики перекладу шляхом використання корпусної лінгвістики.

Ми проаналізували використання методики корпусної лінгвістики в практиці викладання перекладу та можемо стверджувати, що отримали позитивні результати у вивченні цього питання. ця проблема. Адже студенти успішно засвоїли та закріпили пройдений матеріал та отримали навички роботи з офіційно-діловими документами, навчилися виділяти їх відмінні риси, які впливають на застосування трансформацій у процесі перекладу, отримали досвід роботи з розпорядниками корпусів та корпусів загалом [18, с. 60].

Таблиця 1.

Основні мовні корпуси для дослідження перекладу

Назва корпусу	Мова	Тип	Обсяг	Джерело даних
Europarl	Англійська	Письмовий	50 млн слів	Європейський парламент
TED Talks	Англійська	Усний	1 млн слів	TED Talks
OpenSubtitles	Різні	Письмовий	6 млрд слів	OpenSubtitles
The Tatoeba Project	Різні	Письмовий	10 млн слів	Tatoeba Project

Джерело: [5]



Таблиця 1 містить основні мовні корпуси, які використовуються для дослідження перекладу. Включає корпуси різних мов і типів (усний та письмовий), що дозволяє аналізувати широкий спектр мовних даних.

Таблиця 2.

Методи аналізу мовних даних у перекладі

Метод аналізу	Опис	Переваги	Недоліки
Статистичний аналіз	Аналіз частоти використання слів	Легкість в обробці великих обсягів даних	Може пропустити контекстуальні нюанси
Корпусний аналіз	Аналіз граматичних і семантичних структур	Глибше розуміння мовних патернів	Вимагає великої кількості даних
Моделювання на основі контексту	Використання контексту для розпізнавання значень	Покращене розуміння смислових зв'язків	Складність у налаштуванні моделей

Джерело: [9]

Таблиця 2 описує основні методи аналізу мовних даних для дослідження перекладу. Вона включає переваги та недоліки кожного методу, що допомагає вибрати оптимальний підхід для дослідження.

Таблиця 3.

Проблеми та виклики у використанні мовних корпусів для перекладу

Проблема	Опис	Можливі рішення
Нестача специфічних даних	Брак даних для певних мов або тем	Збір додаткових даних або розширення корпусу
Різноманітність стилів	Різні стилі тексту ускладнюють аналіз	Нормалізація текстів перед аналізом
Переклад якості	Неправильні або неякісні переклади	Підвищення якості перекладів або фільтрація даних

Джерело: [7]

Таблиця 3 описує проблеми, що виникають при використанні мовних корпусів для перекладу, та можливі рішення для їх подолання. Використання мовних корпусів у дослідженні усного та письмового перекладу є ключовим для аналізу мовних патернів та якості перекладу. Основні корпуси, такі як Europarl і TED Talks, забезпечують різноманітні дані для аналізу. Методи, як-от статистичний аналіз і моделювання на основі контексту, дозволяють дослідникам глибше розуміти мовні структури, хоча і мають свої обмеження. Проблеми, як-от нестача



специфічних даних і різноманітність стилів, можуть бути вирішені шляхом розширення корпусів і нормалізації текстів.

Висновки. Інструменти корпусної лінгвістики є неймовірно цінними в галузі перекладознавства, подібно до того, як технологічний прогрес зробив ці методи більш ефективними для вивчення мови в цілому. Вивчено методичні розробки щодо застосування методів та засобів корпусної лінгвістики у навчанні, відповідно до методичного аспекту розроблено елементи, які можна реалізувати під час занять з практики перекладу для студентів-перекладачів, ефективність використання таких елементів на практиці – проаналізовано та сформульовано подальші напрями та перспективи цього напрямку з точки зору методології.

Виходячи з нашого досвіду впровадження таких засобів у процес перекладу, можна зробити висновок, що використання корпусних менеджерів сприяє формуванню в студентів навичок роботи з текстами, аналізу стилістичних особливостей текстів, зокрема документів, роботи з граматичними, лексичними і семантичній рівні мови. Тим не менш, ми не можемо не помічати складності включення корпусного програмного забезпечення в сферу перекладознавства та його використання в дослідженні лінгвістичних даних на лексичному, синтаксичному, граматичному та семантичному рівнях.

Корпусні перекладознавчі дослідження стикаються з перешкодами в розумінні природи перекладу, універсальності перекладу як практики, унікальних характеристик і конвенцій перекладу, перш ніж вони зможуть просунутися вперед із компіляцією корпусу та збором даних. Процес вивчення перекладу шляхом аналізу текстів за допомогою менеджерів корпусів не позбавлений недоліків. Існує потреба в поєднанні досліджень, які будуть використовуватися для навчання практиці перекладу.

Ми дійшли висновку, що аналіз мовного матеріалу, зокрема в процесі перекладу та застосування перекладацьких трансформацій, є якісно новим підходом до аналізу мовного матеріалу. Основні поняття та характерні ознаки мовознавства, а також основні ознаки офіційно-ділового стилю ми вивчали, спираючись на праці видатних лінгвістів нашої країни та зарубіжних шкіл. Використання перекладацьких трансформацій під час перекладу текстів офіційного стилю засобами англійської мови зумовлене відмінностями в будові мов.

Розглянути найпоширеніші контексти вживання певного терміна, його найуживанішу граматичну форму та доцільність його використання в текстах певного спрямування можна за допомогою механізму скетч-менеджера корпусу. Перекладач може використовувати WordSmith Tools для створення власного корпусу текстів певного жанру чи напрямку, щоб

у майбутньому, якщо виникнуть сумніви щодо використання тих чи інших трансформацій перекладу, одиниця з мови оригіналу була змінена мовою перекладу. Інтеграція цього напрямку лінгвістики в перекладознавство визначається потребою контекстуалізації та контекстуалізації перекладу, які обмежені в їх ефективності, якісного аналізу та дослідження інших даних, окрім тих, що отримані з корпусів.

Порівнявши думки лінгвістів-методистів щодо використання засобів лінгвістики в контексті навчання перекладу, ми сформували власний додаток з метою використання його на заняттях зі студентами-перекладачами. Завдяки даним, отриманим у процесі корпусного аналізу, можна не тільки сформулювати нові висновки про мову, але й відкрити напрямки досліджень, які не враховувалися до появи корпусів. Розвиток цього напрямку лінгвістики тільки починається, тому що на даний момент це механічний аналіз мовних даних, що дозволяє менеджеру швидко і якісно виконувати запити. Щоб полегшити роботу перекладача та оновити методи аналізу тексту для ефективного та еквівалентного перекладу, необхідні подальші дослідження в цьому напрямку.

Література:

1. Aarts J. The description of language use. In: Hasselgård H., Oksefjell S. (eds.). *Out of corpora: Studies in honour of Stig Johansson*. Amsterdam and Atlanta: Rodopi, 1999. P. 3–20.
2. Biber D. Representativeness in Corpus Design. *Literary and Linguistic Computing*. 1993. Vol. 8, No. 4. P. 243–257. DOI: 10.1093/lc/8.4.243
3. Chesterman A. *Hypotheses about translation universals*. Amsterdam: John Benjamins Publishing Company, 2004. 95 p.
4. Finegan E. *Language: its structure and use*. N.Y.: Harcourt Brace College Publishers, 2004. 605 p.
5. Galvez, D., Damos, G., Ciro, J., Cerón, J.F., Achorn, K., Gopi, A., & Reddi, V.J. (n.d.). *The people's speech: A large-scale diverse English speech recognition dataset for commercial usage*. arXiv preprint arXiv:2111.09344.
6. Halliday M.A.K. *Explorations in the Functions of Language*. New York: Elsevier, 1993. 250 p.
7. Halverson S. *Translation Studies and Representative Corpora: Establishing Links between Translation Corpora, Theoretical Descriptive Categories and a Conception of the Object*. Bergen: University of Bergen, 1998. P. 494–514.
8. Hatzivassiloglou V., Gravano L., Maganti A. An investigation of linguistic features and clustering algorithms for topical document clustering. In: *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*. New York: Association for Computing Machinery, 2000. P. 224–231.
9. Johansson S. *Corpora and Cross-Linguistic Research*. Amsterdam: Rodopi, 1998. P. 3–24.
10. Liu, Y., Fung, P., Yang, Y., Cieri, C., Huang, S., & Graff, D. (2006). HKUST/MTS: A very large scale Mandarin telephone speech corpus. In *Chinese Spoken Language Processing: 5th International Symposium, ISCSLP 2006, Singapore, December 13–16, 2006. Proceedings* (pp. 724–735). Springer Berlin Heidelberg.



11. Mason O. The Weight of Words: An Investigation of Lexical Gravity. In: Proceedings of PALC'97, Łodz. New York: Routledge, 1997. P. 361–375.
12. McEnery T., Wilson A. Corpus Linguistics. Edinburgh: Edinburgh University Press, 2001. 235 p.
13. Rohdenburg G. Cognitive complexity and increased grammatical explicitness in English. *Cognitive Linguistics*. 1996. Vol. 7, No. 2. Berlin: De Gruyter Mouton. P. 149-182.
14. Sinclair J. Corpus, Concordance, Collocation. Oxford: Oxford University Press, 1991. 179 p.
15. Tymoczko M. Connecting the two infinite orders: Research methods in translation studies. In: Crosscultural Transgressions. Research models in translation Studies II: Historical and ideological issues. Manchester: St. Jerome, 2002. P. 9-25.
16. Zeinali, H., Burget, L., & Černocký, J.H. (2019). A multi-purpose and large-scale speech corpus in Persian and English for speaker and speech recognition: the DeepMine database. In 2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU) (pp. 397–402). IEEE.

References:

1. Aarts, J. (1999). The description of language use. In H. Hasselgård & S. Oksefjell (Eds.), *Out of corpora: Studies in honour of Stig Johansson* (pp. 3–20). Amsterdam and Atlanta: Rodopi.
2. Biber, D. (1993). Representativeness in Corpus Design. *Literary and Linguistic Computing*, 8(4), 243–257. DOI: doi.org/10.1093/lc/8.4.243
3. Chesterman, A. (2004). *Hypotheses about translation universals*. Amsterdam: John Benjamins Publishing Company.
4. Finegan, E. (2004). *LANGUAGE: its structure and use*. N.Y.: Harcourt Brace College Publishers.
5. Galvez, D., Diamos, G., Ciro, J., Cerón, J.F., Achorn, K., Gopi, A., & Reddi, V.J. (n.d.). The people's speech: A large-scale diverse English speech recognition dataset for commercial usage. *arXiv preprint arXiv:2111.09344*.
6. Halliday, M.A.K. (1993). *Explorations in the Functions of Language*. New York: Elsevier.
7. Halverson, S. (1998). *Translation Studies and Representative Corpora: Establishing Links between Translation Corpora, Theoretical Descriptive Categories and a Conception of the Object* (pp. 494–514). Bergen: University of Bergen.
8. Hatzivassiloglou, V., Gravano, L., & Maganti, A. (2000). An investigation of linguistic features and clustering algorithms for topical document clustering. In *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 224-231). New York: Association for Computing Machinery.
9. Johansson, S. (1998). *Corpora and Cross-Linguistic Research*. Amsterdam: Rodopi.
10. Liu, Y., Fung, P., Yang, Y., Cieri, C., Huang, S., & Graff, D. (2006). HKUST/MTS: A very large scale Mandarin telephone speech corpus. In *Chinese Spoken Language Processing: 5th International Symposium, ISCSLP 2006, Singapore, December 13–16, 2006. Proceedings* (pp. 724–735). Springer Berlin Heidelberg.
11. Mason, O. (1997). The Weight of Words: An Investigation of Lexical Gravity. In *Proceedings of PALC'97, Łodz* (pp. 361–375). New York: Routledge.
12. McEnery, T., & Wilson, A. (2001). *Corpus Linguistics*. Edinburgh: Edinburgh University Press.



13. Rohdenburg, G. (1996). Cognitive complexity and increased grammatical explicitness in English. *Cognitive Linguistics*, 7(2), 149-182. Berlin: De Gruyter Mouton.
14. Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
15. Shkirta, I., & Lazar, V. (2019). Big Data technology: essence, opportunities for business. *Scientific Bulletin of Mukachevo State University*, 2 (12), 51-56. http://dspace-s.msu.edu.ua:8080/bitstream/123456789/5688/1/Big%20Data%20Technology_%20Essence_Business_%20Opportunities.pdf
16. Tkachenko, K. (2024). Use of NLP methods in intelligent learning systems. *Digital platform: information technologies in the sociocultural sphere*, 7(1), 80–96. <https://doi.org/10.31866/2617-796X.7.1.2024.307009>
17. Tymoczko, M. (2002). Connecting the two infinite orders: Research methods in translation studies. In *Crosscultural Transgressions. Research models in translation Studies II: Historical and Ideological Issues* (pp. 9-25). Manchester: St. Jerome.
18. Zeinali, H., Burget, L., & Černocký, J.H. (2019). A multi-purpose and large-scale speech corpus in Persian and English for speaker and speech recognition: the DeepMine database. In *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)* (pp. 397–402). IEEE.