

Вікторія Калашник
Національний технічний університет
«Харківський політехнічний інститут»

СТВОРЕННЯ ПОЛІВАРІАНТНОГО ПАРАЛЕЛЬНОГО КОРПУСУ УКРАЇНСЬКИХ ПЕРЕКЛАДІВ ПОВІСТІ ДЖОРДЖА ОРВЕЛЛА «ANIMAL FARM» ТА ЙОГО ВИКОРИСТАННЯ ДЛЯ ДОСЛІДЖЕННЯ ВАРІАНТНОСТІ УКРАЇНСЬКОЇ МОВИ

У цій роботі розглядається створення поліваріантного паралельного корпусу українських перекладів творів Джорджа Орвелла як інструменту для дослідження варіантності української мови та підтримки її ресурсності. Для реалізації майбутнього корпусу здійснено вирівнювання оригінального тексту повісті «Animal farm» з сімома українськими перекладами за допомогою програми LF Aligner. Матеріали підготовлено для подальшого корпусного опрацювання. Робота виконується в межах ініціативи ParaRook, яка спрямована на розвиток україномовних корпусних ресурсів.

Особливу увагу приділено варіативності перекладацьких рішень та можливостям майбутнього застосування корпусу в перекладознавстві й обробленні природної мови. У тезах наводяться приклади мовних розходжень між перекладами, що демонструють стилістичне, лексичне й ідеологічне розмаїття українських версій одного й того ж тексту. Створена база даних має потенціал стати повноцінним корпусом і використовуватись для лінгвістичних досліджень, навчання моделей машинного перекладу та автоматичного перекладу.

Ключові слова: поліваріантний переклад, паралельний корпус, Джордж Орвелл, LF Aligner, обробка природної мови.

This paper explores the development of a multi-variant parallel corpus of Ukrainian translations of George Orwell's works as a tool for studying

language variation and enhancing the linguistic resourcefulness of the Ukrainian language. To facilitate the creation of the future corpus, the original text of the novella “Animal Farm” was aligned with seven Ukrainian translations using the LF Aligner tool. The materials have been prepared for further corpus processing. The project is carried out within the framework of the ParaRook initiative, which aims to develop Ukrainian-language corpus resources.

Particular attention is paid to the variability of translation choices and the potential future applications of the corpus in translation studies and natural language processing. The thesis presents examples of linguistic differences among the translations, which demonstrate stylistic, lexical, and ideological diversity in Ukrainian renditions of the same text. The resulting dataset has the potential to become a fully developed corpus and to be used for linguistic research, machine translation training, and automatic paraphrasing.

Key words: multi-variant translation, parallel corpus, George Orwell, LF Aligner, natural language processing.

У мовознавстві та перекладознавстві зростає значення цифрових лінгвістичних ресурсів, які зберігають мовну спадщину та сприяють аналізу мовних явищ. З огляду на зростання ролі української мови в публічному й академічному просторі, актуальним є розвиток україномовних корпусів для наукових і прикладних цілей, зокрема в галузі оброблення природної мови (NLP) [1, с. 105].

Окрему наукову цінність у межах корпусної лінгвістики становлять поліваріантні паралельні корпуси — такі, що включають кілька перекладених версій одного й того ж оригінального тексту, як однією, так і різними мовами. Прикладами таких ресурсів є багатомовні корпуси InterCorp [8] та ParaSol [9], які містять і українські тексти.

Дослідження варіантності перекладів є важливим напрямом перекладознавчих досліджень. Зокрема, Богдан Стасюк досліджує проблему перевидання та редакції перекладів оповідання

Джека Лондона «Біла тиша» з використанням методики зіставлення окремих фрагментів текстів вручну, що можна розглядати як прототип поліваріантного корпусу [10]. Створення поліваріантного паралельного корпусу із можливістю пошуку дає технічні можливості для продовження таких досліджень на більшому матеріалі, а також використання в різноманітних лінгвістичних, культурологічних дослідженнях, для підготовки перекладачів, у завданнях з оброблення природної мови.

1. Паралельні корпуси у проекті ParaRook

Одним із важливих напрямів розвитку українських корпусних ресурсів є ініціатива ParaRook. У її межах вже сформовано декілька корпусів зіставлення перекладів з англійської, іспанської, японської, китайської німецької та французької мов на українську. Особливістю проекту є напівручний спосіб вирівнювання текстів, що забезпечує вищу якість анотації порівняно з автоматично згенерованими корпусами [3, с. 14]. Саме цей підхід ліг в основу мого дослідження, присвяченого створенню поліваріантного корпусу українських перекладів повісті *Animal Farm*.

2. Різні переклади одного тексту в ParaRook

Попереднім етапом у дослідженні варіативності перекладів стало створення двох паралельних корпусів для ParaRook на основі повісті Джека Лондона *The Seed of McCooy*, вирівняних за допомогою програми InterText [6]. У межах курсової роботи було поетапно зіставлено один оригінальний текст і два українські переклади, що дозволило виявити стилістичні та граматичні розбіжності, а також оцінити функціональні можливості доступних інструментів вирівнювання.

3. Поліваріантний корпус перекладів *Animal Farm*

Наступним етапом дослідження стало створення поліваріантного корпусу на основі *Animal Farm* Джорджа Орвелла, що

включає один оригінал і сім українських перекладів, виконаних у різні періоди (табл. 1).

Таблиця 1

Українські переклади «Animal Farm»

NAME	DATE	TRANSLATOR
Ферма тварин	1947	Іван Чернятинський
Хутір тварин	1984	Ірина Дибко
Рай для тварин	1991	Юрій Шевчук
Скотоферма	1991	Олексій Дроздовський
Скотохутір [пер. з рос.]	1992	Наталя Околітенко
Колгосп тварин	2020	Богдана Носенок
Колгосп тварин	2021	В'ячеслав Стельмах

Інструментом для вирівнювання слугувала програма LF Aligner у версії 4.25 – безкоштовний застосунок для створення поліваріантних паралельних корпусів, доступний на платформі SourceForge [11]. Це дало змогу сформуванню високоточної паралельну таблицю відповідностей між фрагментами оригіналу (обсягом понад 1300 речень) й усіма наявними перекладами.

Для ілюстрації варіативності перекладів можна розглянути приклад із першого розділу:

Man serves the interests of no creature except himself. (Дж. Орвелл)
 Людина не служить інтересам жодного створіння крім себе.

(І. Чернятинський)

Людина ніколи не ділилася й не ділиться з нами своїми успіхами.

(І. Дибко)

Людина дбає тільки про власні інтереси. (Б. Носенок)

Людина не діє ні в чийх інтересах, окрім власних. (В. Стельмах)

Людина знає свою і тільки свою користь. (Ю. Шевчук)

У людей є лиш одні інтереси – їхні власні! (О. Дроздовський)

Людей ніщо не обходить, окрім них самих. (Н. Околітенко)

Крім лексичних та стилістичних розбіжностей, під час вирівнювання виявлено також помилки перекладу, зокрема щодо роду персонажів. Деякі переклади некоректно інтерпретують

стать тварини при першій її появі у тексті, навіть коли стать чітко визначена в оригіналі займенником *she, herself* або випливає з імені: *First came the three dogs, Bluebell, Jessie, and Pincher...* У перекладах Івана Чернятинського (1947) та Ірини Дибко (1984) це було подано так:

- Спершу прийшло трое собак: Мурко, Лисько, Бровко... (1947);
- Першими були пси Белов, Джесов і Пінков... (1984).

Це призвело до курйозної ситуації, коли дві з трьох собак, які від початку були позначені як персонажі чоловічої статі, у третьому розділі народили (It happened that Jessie and Bluebell had both whelped soon after the hay harvest, giving birth between them to nine sturdy puppies). Також у цьому випадку можна прослідкувати, які трансформації або обхідні рішення обирали перекладачі, аби приховати неузгодженість:

- Трапилось якраз, що незабаром після сінокошу **суки** пощенилися і привели на світ дев'ять гарних цуценят (1947);
- Після жнив Белова і Джесова стали матерями, обидві привели аж по дев'ять щенят (1984).

Таким чином, аналіз поліваріантного корпусу, побудованого на семи перекладах *Animal Farm* дозволяє не лише зафіксувати розбіжності в лексичному та стилістичному оформленні тексту, а й виявити глибші проблеми, пов'язані з перекладацькою інтерпретацією, мовними нормами, ідеологічним контекстом тощо.

4. Перспективи використання поліваріантного корпусу

Поліваріантні паралельні корпуси відкривають нові можливості як для теоретичних, так і прикладних досліджень у перекладознавстві. Вони дають змогу аналізувати, як різні перекладачі передають ті самі одиниці оригіналу, виявляти перекладацькі стратегії, ступінь інтерпретації, адаптації, редагування та нормування. Такий матеріал може бути використаний як для

оцінювання якості перекладу, так і для створення навчальних ресурсів із прикладами реальної варіантності [7, с. 242].

З технічного боку, поліваріантні корпуси мають потенціал для задач NLP – зокрема, як референтні дані для оцінювання якості машинного перекладу [4, с. 114]. Наявність кількох перекладених версій одного тексту дозволяє не лише порівнювати результат, згенерований програмою, з перекладом людини, а й наочно показує, що існує кілька можливих варіантів перекладу.

Поліваріантні корпуси мають потенціал для міждисциплінарних досліджень – від мовних змін до культурних впливів на переклад. Їх реалізація на українському матеріалі підсилює ресурсність мови та інтегрує українську лінгвістику у світовий науковий простір.

Список використаних джерел:

1. Жуковська В. В. Вступ до корпусної лінгвістики. Житомир : Вид-во ЖДУ ім. І. Франка, 2013. 141 с.
2. Shvedova M. The General Regionally Annotated Corpus of Ukrainian (GRAC, uacorporus.org): Architecture and Functionality / Kyiv National Linguistic University. 2020. URL: <https://ceur-ws.org/Vol-2604/paper36.pdf>.
3. Shvedova M., Lukashevskiy A. *Creating Parallel Corpora for Ukrainian: a German-Ukrainian Parallel Corpus (ParaRook||DE-UK)*. The Third Ukrainian Natural Language Processing Workshop (UNLP). 2024. Pp. 14–22. URL: <https://aclanthology.org/2024.unlp-1.3/> (дата звернення: 28.05.2025).
4. Жуковська В. В. Лінгвістичний корпус як новітній інформаційно-дослідницький інструментарій сучасного мовознавства / *Вчені записки ТНУ імені В. І. Вернадського. Серія: Філологія. Соціальні комунікації*. 2020. Т. 31 (70). № 3. Ч. 1. С. 113–119.
5. Cheesman T., Flanagan K., Thiel S., Rybicki J., Robert S. Laramee, Hope J., Roos A. Multi-Retranslation corpora: Visibility, variation, value, and virtue. URL: <https://academic.oup.com/dsh/article/32/4/739/2669776>.
6. Калашник В. Д. Курсова робота з дисципліни «Використання корпусів у лінгвістичних дослідженнях» / Національний технічний університет «Харківський політехнічний інститут». Харків, 2024. 22 с.
7. Baker M. Corpus Linguistics and Translation Studies. Implications and Applications. URL: <https://surl.lu/ushwdx> (дата звернення: 28.05.2025).

8. Rosen A., Vavřín M., Zásina A. J. InterCorp, Release 15 of 11 November 2022. Institute of the Czech National Corpus, Charles University, 2022. URL: <http://www.korpus.cz> (дата звернення: 28.05.2025).
9. Ruprecht von Waldenfels. ParaSol: Introduction to a Slavic Parallel Corpus. 2012. Pp. 293-302. URL: <https://surl.li/ncylo>.
10. Стасюк Б. Конфлікт редакцій старих перекладів і проблема їх перевидання. *Записки Наукового товариства імені Шевченка*. Львів, 2019. Том 272 (CCLXXII). Праці Філологічної секції. С. 496-514.
11. LF Aligner. *Source Forge*. URL: <https://surl.li/lexkoj> (дата звернення: 28.05.2025).