

УДК 378.016:[811.161.2+81'25

Михайло ЛУКАЩУК,

кандидат філологічних наук, асистент кафедри англійської мови Чернівецького національного університету імені Юрія Федьковича, м. Чернівці;

ORCID: <https://orcid.org/0000-0002-2071-9325>

ВИКОРИСТАННЯ ПАРАЛЕЛЬНИХ КОРПУСІВ У НАВЧАННІ УКРАЇНСЬКОЇ МОВИ ТА ПЕРЕКЛАДІ

Браунівський корпус (*The Brown University Standard Corpus of Present-Day American English*) – електронна колекція зразків текстів американського варіанта англійської мови різних жанрів – був першим в історії корпусом, який містив 500 зразків різних жанрів розміром 2000 та більше слів сучасної англійської мови, на основі якого вивчали дистрибуцію слів у щоденному вжитку. Корпус уклали Генрі Кучера та Нельсон Франсіс у Браунівському університеті (США). На той час він містив 1014312 слів. 1967 року вони опублікували розвідку, у якій надали статистичні дані власне Браунівського корпусу.

З появою комп'ютерів на початку 1990-х років було розроблено перший, найбільший на той час Британський національний корпус (*British National Corpus, BNC*), що містив 100467090 орфографічних слів. Це був збалансований корпус, який залучав лексику з більшості сфер суспільного життя, що складало приблизно 90 млн. слів, зафіксованих у галузі прикладних, гуманітарних та природничих наук, художній літературі, мистецтві, релігії та філософії, комерції та фінансів, розваг та політиці. Додатково йшлося також про розмовний підкорпус, який містив близько 10 млн. слів.

Після 2000-х років з подальшим розвитком комп'ютерної техніки з'явилися корпуси розміром понад мільярди слів. Лідерами серед англомовних країн стали британські лінгвісти, хоча нині американські науковці поступово наздоганяють досягнення британців.

Сьогодні корпусну лінгвістику використовують у різних галузях мовознавства: лексикографії (під час розробки словників), у когнітивній і прикладній лінгвістиках, соціолінгвістиці, у перекладознавстві та дидактиці (у процесі навчання англійської мови). Великі корпуси полегшують роботу під час обробки текстів та надають достатньо прикладів, щоб провадити роботу у згаданих вище напрямках.

Паскуаль Перез-Паредез подає таке визначення корпусних досліджень: «корпусна лінгвістика вивчає мову, визначаючи, як репрезентативні тексти конкретного жанру відображають дискурсивні практики справжніх мовних користувачів» [3, с. 1]. Корпусна лінгвістика досліджує мову емпірично, тобто акцентує увагу на тому, як лексичні одиниці в реальному мовленні, а також як тексти різних жанрів відображають мовлення окремих користувачів. У Великій Британії під керівництвом Рандольфа Квірка було запроваджено програму *Survey of English Usage* (огляд вживання англійської мови), яка, починаючи з 1959 року, мала на меті реєструвати реальне вживання англійської мови на сучасному етапі, що й стало початком корпусної лінгвістики у цій країні.

Однією із перших мовознавчих галузей для застосування корпусної лінгвістики у світовій практиці була лексикографія. Сьогодні, працюючи над словниками англійської мови, їхні укладачі обов'язково використовують і створюють корпуси текстів різних стилів та жанрів, а вже тоді результати цієї роботи втілюють у своїх лексикографічних проєктах. Серед популярних словників,

видавництва яких використовують корпуси для їх створення, є *Longman Dictionary of Contemporary English*, *The New Oxford Dictionary of English*, *Macmillan English Dictionary*, нове видання 20-томного *The Oxford English Dictionary*. Зараз також на основі корпусів перебувають на стадії розробки та будуть втілені тільки у електронному форматі *Oxford Learner's Dictionary*, *Oxford Collocation Dictionary*, *Collins COBUILD Advanced Learner's English Dictionary*, *Collins English Thesaurus*, *Oxford Thesaurus of English*, *The Chamber's Dictionary & Thesaurus* та багато інших. Раніше лексикографи самі подавали ілюстрації до значень різних слів. З появою корпусної лінгвістики вони послуговуються прикладами речень із щоденного вжитку різних стилів та жанрів.

Одним із мільярдних корпусів є утворений із вебсайтів англomовних країн різних жанрів iWeb, який складається з 14 мільярдів слів. Серед українських корпусів можемо згадати такі: корпус ГРАК, який є зібранням текстів, починаючи з 1816 року, що складає 1,78 мільярда слів, засвідчених у різних стилях та жанрах. Іншим корпусом української мови є Лейпцизький корпус української мови, що складається із 1,5 мільярда слів з новинних сайтів, вебсторінок та текстів у загальному доступі, та корпус української мови Київського національного університету ім. Т. Шевченка розміром 100 мільйонів слів, що складається із публіцистичних текстів, художньої прози, поетичної мови, законодавчих, наукових та фольклорних текстів.

Корпуси текстів можуть бути використані для навчання мови, зокрема української та англійської. Вони містять тексти різних стилів і жанрів, що може допомогти студентам опанувати конкретні теми чи галузі знань за потреби. Здобувачі освіти засвоюють граматику за допомогою зразків або моделей. Як зазначає Ксяофей Лу, «процес вивчення конструкцій ґрунтується на прикладах та

включає «вивчення асоціацій в межах та між конструкціями». Цей процес розглядається як раціональний у тому сенсі, що ментальна модель, яка побудована студентами на будь-якому відрізку процесу вивчення мови, відображає вивчені можливості різних асоціацій форми та значення, які студенти акумулювали із свого сумарного досвіду вивчення мови» [1, с. 47].

Важливим чинником у навчанні іноземної мови для студентів є частотність вживання слів. Частотність є одним із принципів використання мови. Отже, вчитель може вибирати потрібні зразки текстів для того, щоб студенти опанували конкретні лексичні одиниці чи граматичні конструкції. Дж. Байбі підкреслює, що «...повторення дій веде до утворення конструкцій; отже, у мові також ми бачимо, що повторення є необхідним компонентом формування граматики... Причина, через яку частотність чи повтор відіграють роль у формуванні граматики полягає в тому, що розум чутливий до повторів. Це є загальний принцип домену; тобто він застосовується не тільки до мови, але й до інших когнітивних доменів» (Bybee, 2007: 8) *цит. за* [3, с. 11]).

Отож, коли раніше за підручником студенти мали можливість вивчати слова та конструкції в одному контексті, то з використанням корпусної лінгвістики вчитель може відібрати необмежену кількість повторів потрібної конструкції чи лексичної одиниці в реальному автентичному використанні носіями цієї мови. І це дає змогу суттєво розширити набір моделей та лексичних одиниць опановуваних студентами мов.

Ще одна сфера, у якій корпусна лінгвістика може бути використана дуже ефективно, це переклад із однієї мови на іншу. У цій галузі доречно використовувати паралельні корпуси. Паралельний корпус – це анотований корпус, що містить тексти мови-оригіналу та їх точних перекладів

мовою перекладу. У ЄС всі законодавчі акти та інші офіційні документи перекладають всіма мовами членів Євросоюзу. Стефанія Мейсі та Мішель Зала про паралельні корпуси зазначають таке: «...паралельні та порівняльні корпуси прогресивно зростають з різними цілями. За останні три десятиліття вони виявилися основним інструментом для машинного перекладу та багатомовної обробки природної мови та обов'язковим елементом у контрастивній лінгвістиці, у перекладі та навчанні іноземних мов » [2, с. 38].

Студенти можуть ефективно засвоювати граматичні конструкції та лексичні одиниці з конкретної сфери за допомогою паралельних корпусів. На жаль, поки що немає українсько-англійського паралельного корпусу, проте зі вступом України до Євроспільноти паралельні корпуси будуть створені, оскільки виникне необхідність всі законодавчі акти та інший документообіг перекладати також і українською мовою.

Отже, корпуси можна ефективно використовувати з різною метою, зокрема для вивчення української та англійської мов, у галузі перекладу та різноманітних лінгвістичних досліджень.

Список літератури

1. Lu, Xiaofei. (2023). *Corpus Linguistics and Second Language Acquisition: Perspectives, Issues, and Findings*. New York: Routledge. 156 p.
2. Maci, S. M. & Sala, M. (Eds.). (2022). *Corpus Linguistics and Translation Tools for Digital Humanities: Research Methods and Applications*. Dublin: Bloomsbury Publishing Plc. 227 p.
3. Pérez-Paredes, Pascual. (2021). *Corpus Linguistics for Education: A Guide for Research*. London: Routledge. 175 p.