

Довгопол Галина Олександрівна

доцентка кафедри філології та перекладу

Київський національний університет

технологій та дизайну (м. Київ)

## КОРПУСНИЙ ПІДХІД У ВИВЧЕННІ ПРИСЛІВНИКІВ

Корпусна лінгвістика дає матеріал для різного роду досліджень мови і її варіантів, і визначає основний метод аналізу текстів на базі корпусів (Corpus-Based Approach). Корпусний підхід, або метод лінгвістичного дослідження, заснований на корпусах текстів, орієнтований на прикладне вивчення мови, його функціонування в реальних середовищах і текстах. Наприклад, словниковий аналіз на базі корпусів допомагає розкрити контекстне вживання тих чи інших слів, особливо синонімічних (наприклад, *a lot of, lots of, a great deal of, large amounts of*), частотну сполучуваність їх з іншими словами, регулярність в тих чи інших стилях, і чітко визначити їх семантику. (2)

За допомогою корпусних методів, зокрема виявлення будь-яких текстових прецедентів і з'ясування кількісних даних їх вживання можна класифікувати наявні прислівники у певному тексті. Корпуси текстів служать неоціненним ресурсом реальних прикладів вживання тих чи інших морфологічних одиниць, за умови, що користувачі мають доступ до корпусів і знайомі з технікою роботи з корпусом.

Незважаючи на наявність безлічі типів розмітки, більшість реально існуючих корпусів відноситься до корпусів морфологічного або синтаксичного типу (останні в англійській літературі називають *treebanks*). При цьому слід підкреслити, що корпус з синтаксичною розміткою явно або неявно включає в себе і морфологічні характеристики лексичних одиниць.

Основною одиницею корпусу текстів є слововживання (або токен, англ. *tokens*, часто їх називають словами, *words*), але також корпусна лінгвістика

оперує поняттями словоформа (type), основна форма, лема (lemma), зворот, стійке словосполучення (Multiword unit), речення (sentence). Обсяг створюваного корпусу текстів в прийнятих одиницях залежить від цілей створення. Він може бути невеликим при вивченні частоти вживання букв, буквосполучень, звуків, звукосполучень. Набагато більшим він повинен бути при вивченні морфологічних явищ.

Серед лінгвістичних типів розмітки виділяються: морфологічна, синтаксична, семантична, анафорична, просодична, дискурсна та ін. (1)

Стосовно морфологічної розмітки слід зауважити, що в іноземній термінології використовують термін *part-of-speech tagging* (POS-tagging), дослівно – частиномовна розмітка. Насправді морфологічні мітки включають не тільки ознаку частини мови, а й ознаки граматичних категорій, властивих даній частині мови. Це основний тип розмітки: по-перше, більшість великих корпусів є якраз морфологічно розміченими корпусами, по-друге, морфологічний аналіз розглядається як основа для подальших форм аналізу – синтаксичного і семантичного, і, по-третє, успіхи в комп'ютерній морфології дозволяють автоматично з великим ступенем правильності розмічати корпуси великих розмірів.

Основні характеристики методу, що визначають його надійність і достовірність, такі:

- є емпіричним і аналізує реальні слововживання в природному мовному середовищі,
- використовує досить велику, репрезентативну вибірку текстів,
- активно використовує комп'ютери і спеціальні програми-конкорданси для аналізу в автоматичному та інтерактивному режимах роботи,
- базується на методах статистичного і якісного аналізу тексту,
- є цільовим, тобто повинен бути орієнтований на реальне додаток і результати.

Однією з важливих особливостей методу аналізу на базі корпусів текстів є дослідження не лише суто лінгвістичних явищ (граматичних або лексичних функцій слів, їх зв'язків з іншими лексемами), але і таких явищ, як, наприклад, частотності лексем або граматичних конструкцій в тих чи інших жанрах, діалектах.

Звернення до даних корпусів обумовлено декількома причинами: по-перше, в них міститься достатньо великий масив прикладів з текстів різних жанрів як британського, так і американського варіантів англійської мови; по-друге, BNC і COCA розташовані на одній платформі (URL: <http://corpus.byu.edu>) і побудовані за одним принципом: володіють однаковим інтерфейсом і функціоналом і, найголовніше, ідентичними критеріями якісної і кількісної (статистичної) обробки представленого в них мовного матеріалу.

Таким чином, використання корпусної лінгвістики допоможе успішно вирішувати як наукові, так і педагогічні завдання. Методи корпусної лінгвістики сьогодні входять в стандартний репертуар багатьох досліджень частин мови.

#### СПИСОК ЛІТЕРАТУРИ

1. Демська-Кульчицька О. (2010) Корпусна рецепція тексту. *Наукові записки*. Т. 111. Сер. Філологічні науки, 3-6.
2. Жуковська В.В. (2013) Ресурси корпусної лінгвістики у дослідженні історичної динаміки мови. *Матеріали міжнародної наукової конференції «Слово і речення: синтактика, семантика, прагматика»*, Київ. унт ім. Б.Грінченка, 151-156.