

## ДО ПРОБЛЕМИ ЗАСТОСУВАННЯ КОРПУСІВ ТЕКСТІВ У СУЧАСНИХ ЛІНГВІСТИЧНИХ ДОСЛІДЖЕННЯХ

**Олександрук Ірина Володимирівна,**  
молодший науковий співробітник  
Інституту мовознавства ім. О.О. Потебні НАН України

У сучасному мовознавстві активно застосовуються електронні (цифрові) корпуси текстів для проведення різноманітних досліджень, зокрема семантичних. Найповніше в українській науковій літературі інформація з цього напрямку представлена в працях співробітників Українського мовно-інформаційного фонду НАН України, у першу чергу в монографіях [1, 2]. Питання розроблення та застосування корпусів текстів у сучасних дослідженнях представлені також у працях С. Н. Бук, О. І. Ванівської, Н. П. Дарчук, О. М. Демської-Кульчицької, В. В. Жуковської, Є. А. Карпіловської, А. А. Лучик, М. О. Шведової та інших.

Визначень поняття «лінгвістичний корпус» у сучасній науці є дуже багато, але якщо зібрати найсуттєвіші риси та ознаки цього об'єкта, то їх можна узагальнити в такій дефініції: великий за обсягом, уніфікований, структурований, розмічений і філологічно компетентний цифровий масив текстів природною мовою, споряджений спеціальними програмними засобами для введення, корекції, пошуку, аналізу, структурування та виконання інших операцій з лінгвістичною інформацією.

Лінгвістичний корпус виконує такі основні функції: 1) репрезентації (представлення) лінгвістичних даних; 2) маркування (тегування, анотування; розмічання) текстів; 3) експлікації даних [2, с. 89].

Оскільки корпусне опрацювання мови дає змогу проводити дослідження не лише на окремих прикладах, а на репрезентативних масивах мовного матеріалу, у нашому дослідженні, метою якого є розроблення мовно-інформаційного інструментарію для багатопараметричного опису лексичної семантики відносних прикметників української мови, інструментом та матеріалом дослідження обрано Український національний лінгвістичний корпус (УНЛК) [1], створений в Українському мовно-інформаційному фонді під керівництвом В. А. Широкова. УНЛК слугує базою ілюстративного матеріалу для укладання словників, а також для проведення різноманітних мовознавчих досліджень.

На основі даних корпусу розроблено семантичний класифікатор іменників, синтаксично зв'язаних із відносними прикметниками як інструмент ознаки семантичної валентності для опису лексичного значення відносних прикметників із тлумаченнями відсильного типу (так звані формули квазісемантики). Наприклад: «Прикм. до...» та «Стос. до...» (**АВАНГАРДНИЙ**, а, е. 1. Прикм. до **авангáрд** 1; **ЖУРНАЛЬНИЙ**, а, е. Стос. до журналу) [3]. Зазначений результат призначений для розв'язання проблеми експліцитного представлення

інформації в електронних тлумачних словниках, що сприятиме впровадженню комп'ютерних технологій у дослідження з лексичної семантики, орієнтованих на створення систем опрацювання природної мови.

УНЛК, обсягом більше 200 мільйонів слововживань, є частиною Інтегрованої лексикографічної системи (ІЛС), тому вибір контекстів здійснюється автоматично із залученням програми автоматичного морфологічного аналізу тексту та програми лематизації. Інструментом автоматичного морфологічного аналізу і лематизації в ІЛС слугує електронний граматичний словник української мови. Сервіси УНЛК дають змогу сформувати текстові вибірки різних стилів, зокрема наукового, публіцистичного, художнього тощо, виокремити тексти конкретного автора, або ж за певний часовий період, а також отримати статистичні дані на основі великих масивів мовної інформації, що забезпечує достовірність результатів дослідження.

Лінгвістичний корпус доцільно і можливо застосовувати для вирішення широкого кола мовознавчих проблем з використанням як традиційних методів дослідження, так і методів комп'ютерних технологій, спрямованих на автоматизоване опрацювання мовної інформації.

### **Список літератури**

1. Корпусна лінгвістика : монографія / В.А. Широков та ін. Український мовно-інформаційний фонд НАН України. Київ : Довіра, 2005. 472 с.
2. Лінгвістично-інформаційні студії : праці Українського мовно-інформаційного фонду НАН України : у 5 т. / В.А. Широков та ін. Т. 4 : Корпусна та когнітивна лінгвістика. Київ. Український мовно-інформаційний фонд НАН України. 2018. 246 с.
3. Словник української мови online. Томи 1-11 (А-ОБМІЛЬ). URL: <https://services.ulif.org.ua/expl/Entry/index?wordid=25129&page=835>\_(дата звернення 17.10.2022).