

ВИМОГИ ДО УКЛАДАННЯ СПЕЦІАЛІЗОВАНИХ МОВНИХ КОРПУСІВ

Маслова Т.Б.

*Національний технічний університет України
«Київський Політехнічний Інститут імені Ігоря Сікорського»*

Завдяки сучасним комп'ютерним технологіям з'явилися нові технічні можливості для обробки, збереження і відбору лінгвістичних даних у вигляді «корпусу», тобто репрезентативного зібрання певним чином упорядкованих текстів, що може розглядатися як зменшена модель мови чи підмови [1, с. 367]. Текстові корпуси знаходять застосування у різних галузях: їх використовують в комп'ютерній лінгвістиці, в корпусній стилістиці, в лексикографії, а також у методиці викладанні іноземних мов [2]. Проте наявність великої кількості текстів, які є доступними для лінгвістичного аналізу, напевно є найбільш корисною для проведення лінгвістичних досліджень, оскільки автоматичне опрацювання реальних мовних явищ допомагає оптимізувати і об'єктивізувати пошук та статистичну обробку мовних даних.

Існує багато так званих національних корпусів, які репрезентують мову в цілому, проте зараз спостерігається зацікавленість у створенні спеціалізованих корпусів з метою відобразити певну підмову і вирішити конкретні лінгвістичні задачі. Ключовими вимогами є репрезентативність і збалансованість корпусів. У випадку відсутності репрезентативності будь-які результати дослідження, отримані в межах корпусу, не можна буде екстраполювати на відповідну ділянку мови, так як буде незрозуміло, як корпус з нею співвідноситься. Збалансованість передбачає пропорційність різних частин корпусу з огляду на такі характеристики як автори, реципієнти і жанри текстів. Забезпечення репрезентативності і збалансованості можливе лише за умови чіткого визначення параметрів сукупності текстів, з якої потрібно зробити вибірку [1; 3]. Параметризація корпусу, тобто оптимальне поєднання різних підходів, методів і критеріїв із врахуванням специфіки конкретного корпусу, зокрема включає: широту (охоплення максимальної кількості текстів), глибину (виокремлення підкатегорій серед обраних текстів), пропорційність (добирання відсутніх текстів у певній пропорції) і часовий вимір.

ЛІТЕРАТУРА

1. Коцюк Л. Параметризаційні характеристики корпусу текстів біатлону сучасної англійської мови / Л. Коцюк, О. Очеретович // Наукові записки Національного університету «Острозька академія». Серія: Філологічна. – 2016. – Вип. 62. – С. 366-368.
2. Маслова Т. Інструменти корпусної лінгвістики у навчанні і дослідженні іноземних мов / Т.Б. Маслова // Smart-освіта: ресурси та перспективи : матеріали II Міжнар. наук.-метод. конф. (Київ, 23 листопада 2016 р.) : тези доповідей. – К.: Київ. нац. торг.- екон. ун-т, 2016. – С. 277-280.
3. Старко В. Параметризація корпусу як спосіб підвищення його репрезентативності та збалансованості / В. Старко, Н. Чейлитко // Українське мовознавство. – 2013. – Вип. 43. – С. 87-94.