

*Пампурак Анжела Миколаївна,  
Національний університет «Острозька академія»*

## **ДО ПИТАННЯ ПРО СТВОРЕННЯ МІЖНАРОДНОГО АНГЛОМОВНОГО КОРПУСУ ЄСЕ СТУДЕНТІВ (ICLE)**

Використання комп'ютерів стало звичним і невід'ємним явищем у всіх сферах знань і діяльності у наш час. Глобальна інформатизація життя суспільства призвела до нового розуміння ролі, яку відіграє лінгвістичне забезпечення баз даних і баз знань, як основ різноманітних систем штучного інтелекту та експертних систем, які обслуговують комунікацію між людиною і комп'ютером. Зокрема комп'ютеризація лексикографічної діяльності дала поштовх до стрімкого розвитку корпусної лінгвістики, науки предметом дослідження якої є процес створення, обробки та використання корпусів. Доцільність створення корпусів текстів часто недооцінюється і піддається жорсткій критиці, тому висвітлення їх переваг та опис основних етапів роботи над створенням корпусу забезпечують актуальність статті.

Корпус – це не просто електронна бібліотека, зібрання текстів чи словників певної мови, це інформаційна система, що надає можливість детального вивчення й аналізу спеціально відібраних та підготовлених текстів, які входять до складу корпусу. Для полегшення аналізу текстів їх необхідно опрацювати та закодувати, тобто внести додаткову інформацію (яку називають анотацією або розміткою) у вигляді відповідних позначок. Отже, зібрання текстів або їх фрагментів в електронному вигляді, що супроводжується розміткою, називається корпусом.

Створення та використання корпусів надає такі переваги і можливості:

- забезпечує накопичення, надійне зберігання та можливість багаторазового опрацювання зібраних матеріалів різних мов для проведення лінгвістичних досліджень та порівняльних аналізів;
- дозволяє розширити та автоматизувати аналіз мовного матеріалу, на якому ґрунтується кожне лінгвістичне дослідження;
- надає можливість проведення швидкого та якісного аналізу великого об'єму текстів, що підвищує достовірність висновків дослідження;
- забезпечує можливість формування списків активної лексики студентів, частотності вживання слів тощо, що дозволяє вирішити низку окремих завдань вивчення та викладання мов;

■ є одним з важливих компонентів, що застосовуються для розробки систем автоматичного аналізу тексту, забезпечення роботи інформаційно-пошукових систем та здійснення машинного перекладу.

На базі лексикографічної лабораторії *LAXILAB* НУ «Острозька академія» відбувається робота над створенням української частини міжнародного навчального англомовного корпусу есе студентів, відомого під назвою *International Corpus of Learner English (ICLE)*. Створення корпусу організоване *The Louvain Centre for English Corpus Linguistics* і по завершенню проекту він матиме вигляд 19 підкорпусів різних мов, кожен з яких складатиметься з 200,000 слів.

Першим кроком до створення підкорпусу є збір необхідної інформації, тобто есе студентів, що мають рівень володіння англійською мовою «*advanced*». За вимогами проекту кожен студент може надати лише одне есе обсягом від 500 слів. Есе можуть бути двох типів: *argumentative essay writing* (есе, що виконуються протягом тривалого періоду з використанням допоміжних засобів) та *literature examination papers* (есе, що виконуються як письмове завдання на парі). Жодне з відібраних есе не повине містити інформацію переписану з інших джерел.

Другим етапом роботи є перетворення звичайного зібрання електронних текстів у корпус, що здійснюється шляхом їх кодування з використанням відповідної розмітки. Для створення повної бази даних необхідні не лише есе, але й певна інформація про автора, умови та вимоги до їх написання, тому кожен з учасників проекту заповнює розроблену анкету. Запитання анкети спрямовані на отримання інформації, яка в майбутньому надасть дослідникам можливість здійснення різноманітних лінгвістичних та соціолінгвістичних досліджень. Наприклад, аналізу есе студентів, що володіють англійською мовою на рівні «*advanced*», порівняння есе написаних учасниками чоловічої та жіночої статі або порівняння окремих підкорпусів.

Участь у проекті та створення українського підкорпусу відкриє доступ до решти навчальних корпусів, що надасть багато можливостей для розробки нових засобів навчання англійської мови, участі в національних та міжнародних дослідженнях.

### Список літератури:

1. Демська-Кульчицька О. М. Основи національного корпусу української мови [Текст]. — К.: Ін-т укр. мови, 2005. — 219 с.
2. Черницький В.Б. Комп'ютерна лексикографія: Навчальний посібник [Текст]. — М.: НУК, 2004. — 84 с.
3. The Louvain Centre for English Corpus Linguistics [Електронний ресурс] // <http://www.uclouvain.be>