

КУРС «КОРПУСНА ЛІНГВІСТИКА» ДЛЯ СТУДЕНТІВ ФІЛОЛОГІЧНИХ СПЕЦІАЛЬНОСТЕЙ

Поспехова А.Л., Шведова М.О.

***Національний технічний університет
«Харківський політехнічний інститут», м. Харків***

Корпуси в наш час є необхідним інструментом лінгвістичних досліджень. На курсі корпусної лінгвістики студенти різних філологічних спеціальностей вчаться користуватися корпусами, доступними в інтернеті для пошуку, а також створювати власні для конкретних дослідницьких потреб з використанням наявних інструментів, передусім засобів для опрацювання української мови.

Наявні посібники з корпусної лінгвістики розкривають передусім важливі теоретичні поняття і методологію, але не описують багатьох конкретних інструментів, які можуть бути корисні для лінгвіста на практиці під час роботи з текстами. Головна причина цього та, що такі інструменти стрімко розвиваються і їх набір оновлюється швидше, ніж друковані посібники. Для інструментів, призначених для української мови цей процес став особливо інтенсивним з 2022 року у зв'язку з війною, міжнародною увагою до України і значною міграцією українців. Це стимулювало розвиток нових українських лінгвістичних ресурсів, укладання корпусів, вдосконалення інструментів автоматичного перекладу для української мови.

Наш курс корпусної лінгвістики має на меті систематизувати нові ресурси для української мови і показати можливості їх використання на практичних задачах, щоб студенти могли користуватися ними у власних дослідженнях. Курс був розроблений в НТУ ХПІ і модифікований для слухачів курсів з цифрової гуманітаристики, організованих Єнським університетом за підтримки фонду DAAD для студентів з різних університетів України.

У межах курсу ми розглядаємо такі теми:

1. Корпуси української мови, доступні онлайн.
2. Мова корпусних запитів CQL.
3. Розпізнавання текстів (OCR) і коригування.
4. Метарозмітка текстів корпусу.
5. Морфологічна розмітка текстів корпусу.
6. Вирівнювання паралельних текстів.
7. Компіляція корпусу для пошуку і публікації онлайн.
8. Інструменти для аналізу і візуалізації корпусних даних.

Курс має допомогти студентам філологічних спеціальностей зорієнтуватися в галузі, обрати найкращий готовий корпус для свого дослідження або створити власний. Знайомство з запропонованими інструментами розширює діапазон можливостей для роботи з текстами; інструменти прості у використанні, доступні студентам без навичок програмування.