

КОРПУСНІ ДОСЛІДЖЕННЯ ВИКОРИСТАННЯ СЛІВ-ЗАПОЗИЧЕНЬ У ТВОРАХ СУЧАСНИХ УКРАЇНСЬКИХ ПИСЬМЕННИКІВ

Алешко Є.О., Борисова Н.В.

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

На сучасному етапі розвитку корпусної лінгвістики її методи надають широкі можливості для проведення різноманітних досліджень в межах лінгвістичних, соціологічних та історичних дисциплін. Основою будь-якого корпусного дослідження є лінгвістичний корпус – зібрання текстів певною мовою, представлених в електронній формі, та спеціальним чином розмічене, щоб дослідник міг швидко і в повному обсязі знайти в корпусі потрібну йому інформацію. Однією з основних функцій лінгвістичного корпусу є можливість отримання лінгвостатистичних даних про частоту використання тих або інших словоформ або словосполучень, тобто частотність вживання. Ця функція дозволяє, наприклад, у соціолінгвістичному дослідженні визначити вживання мовної одиниці певною соціальною, віковою, гендерною групою; в історичному аспекті – простежити етимологію та зміну семантики мовної одиниці. Таким чином, в дослідженні із застосуванням корпусів використовуються лише точні емпіричні дані. Для вирішення задачі дослідження використання слів-запозичень у творах сучасних українських письменників було створено власний корпус та розроблено програму автоматичного аналізу корпусу. Корпус складається з 10 підкорпусів. Кожний підкорпус містить по 4-5 творів одного з десяти українських письменників: Леся Подерев'янського, Ліни Костенко, Лубка Дереша, Луко Дашвара, Марії Матіос, Оксани Забужко, Олександра Мехеда, Сергія Жадана, Тараса Прохаско, Юрія Андруховича. Кожний твір містить приблизно від 1000 до 3000 знаків. Для роботи програми автоматичного аналізу створеного корпусу також використовувалась лінгвістична база даних слів-запозичень української мови, що містить більше 500000 слів. База даних формувалась автоматично за допомогою відповідної програми. Вибір слів-запозичень у базу даних здійснювався з існуючих електронних орфографічних словників української мови за визначеними правилами. Правила формувалися за результатами аналізу літературних джерел з цієї проблематики. Відбір здійснювався за 50 правилами, що відповідають різним мовам, з яких відбулося запозичення. В результаті роботи програми автоматичного аналізу корпусу за кожним підкорпусом отримуємо список слів-запозичень, які було знайдено у творах, що входять до підкорпусу, із зазначенням частоти їх вживання. Подальший аналіз отриманих результатів дозволив зробити висновок, що сучасні українські письменники майже не використовують слова-запозичення у своїх творах: частота їх вживання в основному дорівнює 1. Звичайно, це стосується лише творів, що увійшли до корпусу. Розширення існуючого корпусу дозволить підтвердити або спростувати це твердження. Крім того перспективою даного дослідження може бути отримання інших статистичних даних та закономірностей використання слів-запозичень.