

## **КОРПУСНА ЛІНГВІСТИКА В ЮРИДИЧНОМУ ДИСКУРСІ**

**Давидов М.Д., Бабкова Н.В., Борисова Н.В.**

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

У сучасних умовах постійного зростання інформаційних потоків проблема автоматичного опрацювання природномовної інформації набуває всі більшого значення. Комп'ютерні технології, які прийшли в лінгвістику, запропонували свої технічні можливості для опрацювання, збереження та відбору мовного матеріалу. Це дало поштовх до зародження та розвитку нової галузі мовознавчих досліджень: корпусної лінгвістики. Базуючись на використанні величезних масивів емпіричного матеріалу та залученні комп'ютерних технологій для отримання об'єктивних висновків про функціонування мовних одиниць, лінгвістичні корпуси швидко перетворилися на неоціненний лінгвістичний ресурс. Отримані в результаті корпусного аналізу дані сприяють формулюванню якісно нових висновків про мову та окреслюють такі напрями лінгвістичних досліджень, які до появи корпусів не привертали уваги дослідників. Спираючись головним чином на реальний «живий» мовний матеріал, а не на мовну інтуїцію та інтроспекцію, корпусні дослідження дозволяють абстрагуватися від суб'єктивності дослідника і наблизитися до об'єктивного вивчення мови.

З часу свого становлення до сьогодні корпусна лінгвістика пройшла складний шлях від повного неприйняття у 60-ті роки ХХ століття до магістрального напрямку сучасного мовознавства. Про популярність та інтенсивність розвитку корпусної лінгвістики у світі свідчить як чимале число наукових досліджень у провідних міжнародних наукових виданнях, посібників, підручників, так і самих корпусів різних типів. Вітчизняне корпусне мовознавство ще перебуває на етапі становлення. Відчувається нестача ґрунтовних україномовних навчальних посібників, які б ознайомили майбутніх філологів з базовими принципами та методами проведення корпусних досліджень.

Задача автоматичної класифікації тональності текстів для лінгвістичної експертизи має складну природу й потребує нестандартних підходів до її вирішення. Складність її природи полягає у тому, що вихідними даними є тексти на природній мові. Кожне слово такого тексту несе зміст, а комбінація слів являє собою складну взаємодію змісту кожного слова. На даний час не існує універсального методу моделювання такої взаємодії на мові машини або чисел. Окрему цікавість являють собою комбінування підходів автоматизації, які базуються на лінгвістиці та машинному навчанні.

Незважаючи на складність задачі, вона приваблює велику кількість дослідників у галузі автоматизованої обробки природньої мови та юрислінгвістиці. Пошуки у даній області активно ведуться і вже існують деякі досягнення. Багато розроблених алгоритмів досягають точності більш ніж 85%. Але треба враховувати, що ці результати отримані на тестових даних в умовах експерименту.