

ВИКОРИСТАННЯ ПАРАЛЕЛЬНИХ КОРПУСІВ В ЗАДАЧАХ СТВОРЕННЯ ОНТОЛОГІЙ

Микитюк М.В., Бабкова Н.В.,

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

На сьогоднішній день, в умовах зростання інформаційних потоків проблема автоматичного опрацювання природномовної інформації стає все більш актуальною. Комп'ютерні технології та лінгвістика дали поштовх до зародження та розвитку нової галузі мовознавчих досліджень – корпусної лінгвістики. 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.

Останнім часом все частіше виявляється усвідомлення лінгвістами необхідності фронтального обстеження мови як у феноменальному, так й в концептуальному аспектах. У загальному розумінні лінгвістичний корпус – це представлений у цифровому форматі, великий за обсягом, уніфікований, структурований, розмічений й філологічно компетентний масив текстів природною мовою, який доповнено системою керування — універсальними програмними засобами для пошуку та опрацювання різноманітної лінгвістичної інформації.

В сучасному світі, дані корпусів масштабно використовуються в лексикографії, стилістиці, судовій лінгвістиці, лінгвістичній варіантології, перекладознавстві (особливо паралельні корпуси), соціолінгвістиці, методиці навчання і вивчення іноземної мови та в багатьох інших лінгвістичних дослідженнях.

У прикладній лінгвістиці інтерес до вивчення онтологій існує доволі давно. Онтологічні моделі забезпечують розуміння природної мови комп'ютерними системами, в результаті чого такі моделі знаходять застосування у великій кількості практичних завдань, до яких відноситься семантичне анотування, вилучення інформації з текстів, питально-відповідні системи, машинний переклад, вирішення лексичної багатозначності, тощо. Більшість онтологій розробляється на основі лексики однієї мови, при цьому найменування і описи концептів, що містяться в них, також формулюються однією мовою (часто в обох випадках цією мовою є англійська). Тим не менш, існує термін «багатомовна онтологія», який використовується для позначення двох типів онтологій: 1) призначених для обробки текстів кількома мовами; 2) перекладені різними мовами з метою забезпечення кращого розуміння онтологічних знань користувачами – носіями різних мов.

Таким чином, поєднання двох настільки потужних інструментів – онтологій та паралельних корпусів – в перспективі бачиться як можливість автоматизувати створення багатомовних тезаурусів та тлумачних й синонімічних словників.