

## СУЧАСНІ ГОНИ УКРАЇНСЬКОЇ КОРПУСНОЇ ЛІНГВІСТИКИ

ДЕМСЬКА О. Текстовий корпус: ідея іншої форми / О.Демська. –

К.: «Нац. ун-т «Києво-Могилянська академія», 2011. – 282 с.

«Чи готове сучасне українське мовознавство до появи у ньому електронного текстового корпусу та корпусної лінгвістики? Якщо так, то яким чином це виявляється? Якщо ж ні, то чому?» – так починає розмову з читачем автор монографії «Текстовий корпус: ідея іншої форми» Оріся Демська. І справді. Чи готове?

Сучасний етап розвитку мовознавства спрямований на розв'язання передовсім прикладних завдань. Інтенсивний поступ такого нового напрямку в сучасній науці про мову, як корпусна лінгвістика, уможлиблює вивчення, аналіз, інтерпретацію тощо мовних реалій дещо під іншим кутом зору, ніж у так званій класичній, традиційній лінгвістиці. А погляд на мовні реалії, з іншого боку, дає змогу отримувати відповідно нові результати та робити нові висновки.

Нині корпусна лінгвістика пропонує власні дослідницькі методи для вивчення різних аспектів мови: її морфологічної будови, синтаксичного устрою, лексичної та фраземної семантики тощо. Проблематика корпусного напрямку доволі розгалужена і передбачає студії, по-перше, загальної теорії корпусної лінгвістики, по-друге, кореляції корпусної лінгвістики та інших лінгвістичних дисциплін, по-третє, типології текстових корпусів і методики інтерпретації корпусних даних, по-четверте, засад створення текстових корпусів природних мов тощо.

«Корпусне відображення мови може охоплювати форми її побутування, рівні, функціональну стратифікацію, історичні варіанти мови, а також інші аспекти, що й безпосередньо детермінує появу того чи того текстового корпусу конкретної мови. Існування тих чи тих текстових корпусів української мови пов'язане із завданням якомога ширшого корпусного відображення української мови: її структурних рівнів, функціональних страт, форм побутування, історичних варіантів тощо», – зазначає у висновках авторка монографії.

О.Демська говорить про стрімке входження корпусної лінгвістики у славистику. Сьогодні вже йдеться про чеську, сербську, болгарську, польську, словенську, словацьку, російську корпусні лінгвістики. Кожна з них презентована передусім корпусами національного типу власної мови.



Зрозуміло, що постала потреба створити національний корпус і української мови. *Корпус сучасної української мови*, за визначенням дослідниці, – це перетворена на електронну форму, стандартно організована, програмно опрацьована, репрезентативна вибірка текстів сучасної української мови, призначена передусім для наукового її вивчення. І монографія О.Демської «Текстовий корпус: ідея іншої форми» є, фактично, першою спробою теоретичного обґрунтування цього масштабного задуму – і за розмахом роботи, і за значенням для престижу української лінгвістики.

Як слушно зазначає авторка рецензованої праці, у сучасному мовознавстві вже давно сформувалася тенденція обґрунтовувати дослідження мовної та мовленнєвої дійсності, використовуючи репрезентативну множину текстів національної мови, або інакше корпус. І ця множина, цей корпус, мусить бути репрезентативним, щоб результати дослідження відповідали критеріям істинності. Не можна не погодитися з твердженням дослідниці, що будь-який корпус текстів природної мови, щоб бути корпусом, а не довільним зібранням комп'ютерних текстів, має відповідати щонайменше п'яти корпусним ознакам. Це, по-перше, *відібраність*, що передбачає застосування чітких правил відбору текстів до корпусу; по-друге, *репрезентативність*, яка полягає у здатності корпусу якнайповніше та якнайкоректніше відображати мову чи якийсь її рівень чи форму побутування; по-третє, *збалансованість*, що досягається введенням до корпусу пропорційної кількості текстових уривків чи текстів, з одного боку, ху-

дожньої літератури, наукових, публіцистичних, офіційно-ділових, теологічних та ін., а з іншого – літературних, діалектних та ін.; по-четверте, *стандартність*, яка забезпечує узаконену, однозначну, мовнонезалежну обробку даних довільної природної мови. Категорія стандартності в дослідній парадигмі корпусної лінгвістики функціонує паралельно з категорією *нормативності*, але, на відміну від останньої, має чіткий технічний характер, що уможлиблює аналогічні дослідження на багатьох корпусах, еволюцію самого корпусу і як безпосереднє синхронне, так і діахронне використання корпусного ядра даних.

Загалом монографія О.Демської «Текстовий корпус: ідея іншої форми» узагальнює прогресивний досвід світових корпусних студій і пропонує найоптимальніший для української мови формат створення корпусу сучасної української мови. Так, європейські корпусні ідеї здобувають на українському ґрунті оригінальне національне втілення. Книжка вміщує велику кількість таблиць, тегсетів, позначок – усього того, що зветься «корпусною розміткою». Визначено основні принципи корпусної лінгвістики, описано методику здійснення корпусного дослідження – фактично, читачам запропоновано інструкцію до укладання корпусу.

Крім очевидного науково-теоретичного внеску, книжку О.Демської характеризує стилістична легкість викладу непростих лінгвістичних проблем, зокрема єдності мовознавства з іншими науками, насамперед логікою, математикою, психологією, кібернетикою, що сприяло появі корпусної лінгвістики з текстовим корпусом у центрі. О.Демська, увібравши найкращі традиції наукового стилю української мови, додає граційної невимушеності викладу таких складних для сприйняття традиційного філолога понять, як електронний текстовий корпус національної мови, його концепції, призначення, мета створення, структура, а також місце у лінгвістичних студіях. Крім цього, у монографії також розкрито аспект становлення текстового корпусу, викладено засади планування й укладення корпусу сучасної української мови.

**Дмитро БУЧКО,**  
доктор філологічних наук, професор  
Тернопільського НПУ ім. В.Гнатюка