

Мовний світ Франкової прози крізь призму лінгвостатистики

Рецензія на монографію

Бук С. Н. Велика проза Івана Франка: електронний корпус, частотні словники та інші міждисциплінарні контексти: монографія. Львів: ЛНУ імені Івана Франка, 2021. 424 с.

Buk S. N. Long Prose Fiction by Ivan Franko: Electronic Corpus, Frequency Dictionaries, and Other Intedisciplinary Contexts. Lviv: Lviv university, 2021. 424 p.

Мовознавче франкознавство знає Соломію Бук як авторку частотних словників «Перехресних стежок» (2007), «Основ суспільності» (2012), «Воа constrictor» (2013), онлайн конкордансу «Перехресних стежок», а також численних статей (близько 50!) про електронний корпус І. Франка та методологію і стратегію комплексного системного лінгвостатистичного опису ідіолекту автора. Нещодавно у Львівському національному університеті імені Івана Франка (за рекомендацією вченої ради) вийшла нова монографія дослідниці «Велика проза Івана Франка: електронний корпус, частотні словники та інші міждисциплінарні контексти», в якій підсумовано дотеперішні її напрацювання, поглиблено методику і створено лінгвістично анотований корпус текстів І. Франка як ґрунтовний мультифункційний ресурс для вивчення лексики, морфології, семантики, структури та інших аспектів великої прози письменника. Водночас праця С. Бук являє першу спробу комплексного квантитативного аналізу великоформатної прозової спадщини Франка і є дуже важливою для дескрипції авторського ідіостилю, мовної картини його художнього світу через встановлення синтагматичних зв'язків того чи того слова, зрештою, уможливаючи в перспективі лінгвостилістичний аналіз усієї Франкової прози.

У сучасному мовознавстві на основі наукової літератури авторка слушно виокремлює головні загальні тенденції дослідження мови письменника: лексикографічну параметризацію, увагу до кількісної складової частини опису ідіолекту, застосування комп'ютерних технологій, зокрема, створення корпусу текстів.

Праця має серйозне історико-методологічне підґрунтя (розділ 1), яке ставить творчість І. Франка у широкий контекст світових тенденцій дослідження мови письменників. Адже, словник, побудований на творах певного письменника, сягає

своїм корінням ще Стародавньої Греції, де до творів Гомера уклали глосарії слів, що вийшли з активного вжитку. Перші конкорданси створювали вручну і лише для особливо важливих текстів, насамперед для Біблії, Корану, а згодом і для творів деяких видатних авторів, як-от В. Шекспіра, Й. В. Гете, М. Павича, а в Україні – Т. Шевченка, Лесі Українки, І. Котляревського, Г. Квітки-Основ'яненка та ін.

З'явилася нова галузь мовознавчої науки — статистична лексикографія, яка, зокрема, займається теорією та практикою укладання частотних словників, а на стику авторської та статистичної лексикографії виник частотний словник мови письменника, і це закономірно, оскільки для опису ідіолекту письменника необхідні не лише якісні, а й кількісні характеристики лексики його творів. Саме тому для найвидатніших майстрів пера різних часів та народів уклали частотні словники, де наводили кількість уживань слова в обстежених текстах певного обсягу. Так, існують частотні словники Евріпіда, Софокла, Арістотеля, В. Шекспіра, В. Гюго, Дж. Джойса, К. Чапека, Ф. Достоєвського тощо. В українській лексикографії інформацію про частоту вживання слова містять вже словники мови Т. Шевченка та Г. Квітки-Основ'яненка, словопоказчики до творів І. Котляревського, Лесі Українки, Ю. Федьковича, А. Тесленка, словники мови Франкових збірок «Зів'яле листя» та «Мій ізмарагд» (обидві авторства З. Терлака) та інших письменників.

Із появою комп'ютерних технологій опрацювання мовного матеріалу виникло поняття електронного корпусу текстів письменника. В монографії С. Букцій проблемі присвячено другий розділ, в якому описано світову й українську практику побудови таких текстових корпусів. Доступні натеper онлайн корпуси текстів В. Шекспіра, М. Достоєвського, В. Даля, а в україністиці відомі такі проєкти на основі поезії Г. Сковороди і Т. Шевченка (Канада, Інститут українських студій). Фокусом дослідження Н. Дарчук стали лінгвістичні засади створення електронних корпусів поетичних творів Т. Шевченка та Лесі Українки (Київський університет, www.mova.info), існують дослідницькі корпуси текстів В. Шевчука, Ю. Шевельова (Шереха), В. Стуса, М. Лепкого, які становлять значний поступ в сучасному мовознавстві.

Джерелами корпусу текстів великої прози І. Франка, на основі якого виконано лінгвостатистичні обчислення у монографії Соломії Бук, стали останні прижиттєві видання таких творів: «Борислав сміється» (1880–1881), «Захар

Беркут» (1883), «Не спитавши броду» (1885–1886), «Для домашнього огнища» (1892), «Основи суспільності» (1894–1895), «Перехресні стежки» (1900), «Воа constrictor» (1905–1907), «Великий шум» (1907), «Петрії і Добошуки» (1909–1912). Як наслідок, загальний обсяг текстів отриманого КТ ВПФ становить близько 500 000 слововживань, що майже співпадає із загальним обсягом частотного словника сучасної української художньої прози (1981), а це дало авторці змогу зіставляти кількісні параметри ідіостилю І. Франка з українською прозою загалом. Ця процедура виявила як універсальні загальномовні лінгвостатистичні характеристики, так й оригінальні параметри вокабулярію І. Франка. Важливою частиною роботи став польсько-український та українсько-польський паралельний корпус автоперекладів І. Франка.

Отже, укладення частотних словників усіх дев'яти Франкових великоформатних творів уможливило їх детальну лінгвостатистичну характеристику, здійснену в третьому розділі монографії. Особливо імponує різнобічність аспектів лінгвального аналізу, адже для кожного твору визначено обсяг тексту і словника, багатство словника, середню повторюваність слова в тексті, кількість слів з частотою 1 (hapaх legomena), індекс винятковості тексту, індекс винятковості словника, кількість високочастотної лексики та індекс концентрації у тексті та словнику. У четвертому розділі також здійснено детальну морфолого-статистичну характеристику текстів, а саме виявлено індекс епітетизації (відношення суми вживань іменників до суми вживань прикметників), індекс дієслівних означень (відношення суми вживань прислівників до суми вживань дієслів), ступінь номінальності (відношення суми вживань іменників до суми вживань дієслів). Сам І. Франко ретельно ставився до добору в тексті тієї чи іншої частини мови. У його статті про переклад «Каменярів» знаходимо власні поетові міркування з цього приводу: «Коли слова перших трьох категорій [іменники, дієслова, прикметники] творять головну основу словесного твору, надаючи йому зміст і акцію, то слова другої категорії [займенники, прислівники, злучники] — се наче тінювання в малярстві, що оживляє і упластичнює картину» [Франко 1976–1986. Т. 39: 7–20].

У роботі Соломії Бук надзвичайно важливо, що авторка запропонувала цікавий статистичний параметр ідіостилю письменника – індекс діалогічності (кількісне зіставлення прямого та авторського мовлення) І. Франка. Такий підхід демонструє суттєво більшу частку прямого мовлення стосовно авторського (40,6

: 59,4 %) порівняно з українською художньою прозою ХХ ст. (28,2 : 71,8 %). Причому дослідниця зазначає, що пропорція неоднакова у творах письменника і залежить від тематики твору: історичний роман потребує ширшого авторського наративу, а твір на сучасні для автора теми має більше діалогічної інтеракції персонажів. На матеріалі Франкових творів поряд із художньою прозою ХХ ст. також цікаво спостережено дві тенденції: 1) зі збільшенням обсягу тексту зменшується його лексичне багатство; 2) для текстів із однаковим обсягом індекс різноманітності вищий для авторської розповіді порівняно з мовленнєвою взаємодією персонажів. На відміну від авторського мовлення, орієнтованого, за винятком деяких оповідних жанрів, на писемну форму, пряме мовлення Франкових персонажів, що відображає особливості усного розмовного мовлення, в перспективі дасть змогу реконструювати розмовний стиль західного варіанта української літературної мови зламу ХІХ–ХХ ст.

П'ятий розділ монографії присвячено множинним дослідницьким підходам на стику мовознавства, літературознавства, математики і фізики, засвідчуючи загальну тенденцію міждисциплінарності у гуманітарній сфері. Тут показано розподіли мовних одиниць на прикладі широко відомого закону Ципфа; пораховано закон Менцерата–Альтмана на різних рівнях одиниць, таких як літера і фонема–слово, склад–слово, клауза–речення. Це значно розширює методологічний діапазон запропонованого дослідження.

У монографії Соломії Бук вперше випрацьовано методику і принципи створення лінгвістично анотованого корпусу текстів І. Франка, його будови, структури, враховано варіантність творів, різні їх видання і т.д., здійснено лематизацію. Концепція, принципи, методологія і стратегія корпусно-лексикографічної та лінгвостатистичної параметризації великої прози І. Франка з використанням новітніх комп'ютерних технологій може бути застосованою для дескрипції ідіолектів інших письменників.

Книга щедро інкрустована таблицями та рисунками, ґрунтується на поважному списку літератури (746 позицій), значна частина яких – іноземними мовами. Слід зазначити, що видання має п'ять рецензентів – високоповажаних фахівців у сфері прикладного та українського мовознавства: В. А. Широков (академік НАНУ, директор Українського мовно-інформаційного фонду НАН України), Н. П. Дарчук (КНУ ім. Т. Шевченка), І. М. Кочан (ЛНУ ім. І. Франка), О. П. Левченко (НУ «ЛП»), Я. Мачутек (Словацька академія наук).

Результати проекту, представленого у книзі, крім самостійної теоретичної й практичної ваги, можуть слугувати одним з етапів роботи над словником мови письменника, а також над створенням повного корпусу текстів І. Франка. Виявлені у праці дані — цінні для укладання словника мови письменника тлумачного типу, зокрема, конкорданс унаочнює сполучуваність та семантику слів; зведений частотний словник подає повний список лем, морфологічно анотований корпус забезпечує інформацією про усю різноманітність словоформ кожної лексеми та їх частиномовну належність. Ці результати також важливі для створення словника західноукраїнського варіанта української мови зламу ХІХ–ХХ ст.

Інтердисциплінарний підхід, застосований у роботі С. Бук, актуалізує множинні дослідницькі підходи до лінгвістичної експлікації Франкової прозової спадщини. До того ж, засягати цієї праці дають змогу підійти до аналізу мовотворчості видатного письменника за різними лінгвістичними напрямками, зокрема з погляду лексики, семантики, психосемантики, граматики, тексту і наративу, стилістики, прагматики та інших аспектів. Праця львівської мовознавиці легко вливається у мейстрім світової лінгвістики з помітними тенденціями до лексикографічної пераметризації в дослідженні мовотворчості письменників. Дослідниця запропонувала новий спосіб представлення статистичних особливостей функціонування мови творів Франка із урахуванням усієї варіативності текстів письменника, що уможлиблює комплексне окреслення лінгвостатистичної моделі авторського ідіолекту. Монографія С. Бук буде цікава для науковців і практиків у сфері франкознавства, лінгвістики (зокрема прикладної), літературознавства, перекладознавства, математики, статистики, філософії науки.

Швець Алла Іванівна – доктор філологічних наук, заступник директора з наукової роботи Інституту Івана Франка НАН України, alla_shvec@ukr.net