

Т. В. Бобкова,

Київський національний лінгвістичний університет, м. Київ

ДО ВИЗНАЧЕННЯ КОРПУСНОЇ ЛІНГВІСТИКИ В СУЧАСНОМУ МОВОЗНАВСТВІ

До визначення корпусної лінгвістики в сучасному мовознавстві. Т.В. Бобкова. У статті досліджуються основні проблеми й підходи до визначення корпусної лінгвістики, її місце в науковій парадигмі.

Ключові слова: корпусна лінгвістика, корпус, галузь, методологія, теорія.

К определению корпусной лингвистики в современном языкознании. Т.В. Бобкова. В статье рассматриваются основные проблемы и подходы к определению корпусной лингвистики, ее место в научной парадигме.

Ключевые слова: корпусная лингвистика, корпус, отрасль, методология, теория.

To Corpus Linguistics Defining in the Contemporary Linguistics. T.V. Bobkova. The article deals with the main problems and approaches of corpus linguistics defining; the corpus linguistics place in science paradigm is analyzed.

Keywords: corpus linguistics, corpus, branch, methodology, theory.

Актуальність визначення корпусної лінгвістики пов'язана з нагальною потребою розвитку цього напрямку в сучасному українському мовознавстві. Відомо, що на сьогодні Україна не має доступного репрезентативного національного корпусу текстів, що відображає функціонування української мови.

Мета статті – встановити проблеми й підходи до визначення корпусної лінгвістики в сучасному мовознавстві. Досягнення поставленої мети передбачає виконання таких **завдань**: 1) проаналізувати наявні визначення корпусної лінгвістики в сучасному мовознавстві й намітити основні підходи; 2) окреслити місце корпусної лінгвістики в науковій парадигмі.

Проблеми визначення статусу корпусної лінгвістики пов'язані, насамперед, з емпіричним характером нового напрямку. Початок розвитку комп'ютерної корпусної лінгвістики, як і введення в науковий обіг терміна «корпус текстів», пов'язують із розробкою в 1960-80-х рр. електронних корпусів першого покоління: Браунівського (У. Френсис, Г. Куцера, 1963-64 рр.), Ланкастерського (Дж. Ліч, 1970 р.) та Лондонсько-Лундського корпусу текстів (Р. Кверк, Я. Свартвік, 1980 р.). Принципи розробки й відповідні детермінативні ознаки корпусів визначались їх прикладним застосуванням, або дослідним призначенням. З самого початку, середини 1960 рр., засадничі принципи нової парадигми були представлені як емпіричні, про що свідчить перша ґрунтовна праця з корпусної лінгвістики *English Collocation Studies* (J. McH. Sinclair, S. Jones, R. Daley) [17, с. 136-37].

Відомо, що становлення нової лінгвістичної парадигми відбувалося в умовах поширювання впливу теорій Н. Хомського й несприйняття корпусних досліджень лінгвістами-теоретиками [6, с. 272-74]. Як послідовник раціоналізму, Н. Хомський вважав, що дослідження емпіричних даних абсолютно не має сенсу, оскільки суть лінгвістики полягає саме у вивченні мовної компетенції, а не її відображення – мовної діяльності. Погляди Н. Хомського щодо адекватності інтуїції та неадекватності корпусів стали підґрунтям ортодоксальної позиції лінгвістів-теоретиків наступних поколінь [6, с. 273; 12, с. 190].

Саме тому по цей час у мовознавстві немає одностайної думки щодо визначення корпусної лінгвістики: складно визначити статус «недисциплінованої дисципліни» [16]. У сучасному мовознавстві корпусна лінгвістика традиційно визначається як лінгвістична галузь [12, с. 2; 16; 3, с. 6-8; 4, с. 7], методологія [11, с. 106; 13, с. хііі; 19, с. 1-2; 8, с. 2, 34]; теорія [19, с. 1-2] або як загальний філософський підхід у лінгвістиці [19, с. 48; 7, с. 12-16; 15, с. 49].

Виокремленню корпусної лінгвістики як самостійної **галузі** мовознавства сприяло визначення власного предмету, об'єкту й мети дослідження, формування термінологічного апарату, теоретичного й практичного підґрунтя й власних дослідних прийомів [3, с. 8]. На пострадянському просторі, де початок власне корпусних досліджень припав на перше десятиріччя ХХІ ст., корпусна лінгвістика традиційно визначається як галузь [4, с. 7], самостійна дисципліна в межах комп'ютерної [5, с. 74] або – прикладної лінгвістики [3, с. 8].

Однак, граничний характер та особливості розвитку дисципліни унеможливають однозначне визначення статусу корпусної лінгвістики в західній лінгвістичній традиції. Так, на вирішальне запитання інтерв'ю про статус корпусної лінгвістики (як галузі мовознавства, методу, методології, дисципліни, підходу або інше) думки опитуваних лінгвістів розійшлися [15]. Зокрема, більша частина опитуваних вважають за краще не навішувати один ярлик через особливий дослідний характер галузі: Г. Астон [15, с. 2-3], П. Бейкер [15, с. 17], Т. Б. Сардіна [15, с. 31-30] визначають корпусну лінгвістику одночасно як **галузь науки** й **методологію**.

Однією з причин відсутності одноголосності у визначенні статусу нової парадигми, крім її емпіричного характеру, є широке розуміння [16] загального терміна «корпусна лінгвістика», що об'єднує значну галузь лінгвістичних досліджень [15, с. 157]: від розробки теоретичних засад планування корпусів до практичного їх створення й фактичного аналізу даних.

Саме різноманіття дослідницьких завдань корпусної лінгвістики й застосовуваних для їх розв'язання особливих методів доводить, що визначення корпусної лінгвістики як **методології** здається безпомилковим [15,

с. 49]. Такої думки, за Ч. Меєром [13, с. xi], дотримується більшість лінгвістів (С. Конрад [15, с. 49], М. Девіс [15, с. 65], К. Хайленд [15, с. 101], С. Лавіуса [15, с. 137], Дж. Ліч [15, с. 157-58]): корпусна лінгвістика – це більше ніж просто галузь лінгвістики, це методологічна основа лінгвістичних досліджень. Б. Лоу аргументує таке визначення хронологією подій у мовознавстві: корпусна лінгвістика спочатку народилася як методика, а вже потім як наука [15, с. 180].

Принциповим аргументом на користь визначення корпусної лінгвістики як методології також є ґрунтування досліджень (теоретичних або прикладних) не на надуманих, а на реальних мовних даних. Так, наприклад, у прикладній лінгвістиці розробка парсингів і теггерів будується на вигаданих моделях речень, а має розроблятися на базі великої колекції текстів. У цьому розумінні корпусна лінгвістика є не окремою парадигмою в мовознавстві, а об'єднуючою методологією лінгвістичних досліджень [13, с. xiii].

Ще одне свідчення того, що корпусна лінгвістика являє собою методологію, можна знайти в створенні залежно від мети різних типів корпусів текстів та оформленні різних напрямків лінгвістичних досліджень: психолінгвістики (вивчення дитячого мовлення – Child Language Exchange System (CHILDES)), діахронічної лінгвістики (Diachronic Corpus of Present-Day Spoken English (DCPSE), діалектології і корпусної діалектометрії (на базі корпусів діалектних варіантів – The Freiburg English Dialect Corpus, Estonian Dialect Corpus) і соціолінгвістики (дослідження мовлення різних вікових груп: The Bergen Corpus Of London Teenage Language (COLT)).

Корпусна методологія приваблює лінгвістів різних переконань [13, с. xiу], про що свідчить зростання популярності використання емпіричних даних корпусів у лінгвістичних дослідженнях. Однак, твердження, що лінгвістика корпусів являє собою методологію, є вірним лише «на тривіальному рівні» [16], і не є синонімом «вільної від теорії» дисципліни: створення корпусу та його використання диктується теоретичним підґрунтям дослідження, вираженим або прихованим [18, с. 35].

Безумовно, корпусна лінгвістика, що виникла як суто практична дисципліна, має також і теоретичний характер. Зокрема, поділ між корпусною лінгвістикою й генеративною граматикою в 1960-80 рр. базувався на помилковому твердженні про дескриптивний, емпіричний характер корпусної лінгвістики й теоретичний – генеративної граматики [13, с. 1]. Сьогодні корпусна лінгвістика – це не тільки сукупність нових методів і комп'ютерних інструментів для отримання даних дискурсу, а й теоретизування мовних даних, що призводить до якісно нового розуміння мови [19, с. 1; 17, с. 138].

Результати корпусного аналізу є більш придатними для функціонального аналізу мови, оскільки фокусуються не на формальному описі, а на використанні мови як комунікативного засобу [13, с. 5]. Завдяки збалансованості й стильовому різноманіттю використовуваних корпусних ресурсів результати корпусних досліджень оцінюються не тільки як фактичні й достовірні, а й теоретично орієнтовані [13, с. xii].

Питання щодо теоретичного підґрунтя корпусної лінгвістики у сучасній лінгвістиці є дискусійним. Так, у північноамериканській лінгвістичній традиції теоретичним підґрунтям корпусної лінгвістики вважається структуралізм [3, с. 16-17], а в західноєвропейській лінгвістиці – Фьорзіанська традиція з виокремленням лінгвістичного опису мови та теорії, базованої на систематичному дослідженні мовної варіативності, а не на інтуїції дослідника [15, с. 49; 12, с. 242, 247].

Визначення корпусної лінгвістики як **лінгвістичної теорії** базується, по-перше, на визнанні тексту головним об'єктом теоретичної рефлексії у мовознавстві [7, с.14]. По-друге, на послідовному висхідному аналізі, який починається з дослідження емпіричних даних і власне відрізняє корпусну лінгвістику від інших фундаментальних лінгвістичних теорій [17, с. 138-39]. Однак, як лінгвістична теорія сучасна корпусна лінгвістика отримала остаточне визначення завдяки чіткому розмежуванню корпусно-базованої (corpus-based) і корпусно-керованої методології (corpus-driven) досліджень [19, с. 17].

Корпусно-базований підхід передбачає використання корпусу текстів у якості ілюстративного матеріалу на підтвердження заздалегідь висунутої за інтуїцією дослідника гіпотези. При цьому дослідження корпусу здійснюється за кількісними та якісними параметрами на основі теоретичних припущень, базованих не на корпусних даних [14, с. 115; 12, с. 241]. Натомість корпусно-керований підхід, або нео-Фьорзіанський [12, с. 242], передбачає побудову лінгвістичного опису безпосередньо на корпусних даних. У цьому випадку основні докази ґрунтуються на встановлених в корпусі повторюваних моделях і розподілі частот текстових одиниць [14, с. 115].

Таким чином, корпусно-керована методологія дослідження передбачає застосування суто емпіричного підходу й неможливість редагування отриманих результатів на користь заздалегідь сформульованих теоретичних припущень. Корпусно-кероване дослідження здійснюється через ітеративний аналіз корпусу відповідно до його внутрішніх лінгвістичних особливостей, які не можуть бути встановлені апіорно, а мають бути виведені безпосередньо з корпусу [14, с. 115]. Остаточним результатом застосування корпусно-керованого підходу є повне генерування з корпусу моделі мови, тобто побудова нової лінгвістичної теорії. І хоча, на думку В. Тойберта, на сьогодні сформовано лише контури нової теорії, саме завдяки власній методології досліджень корпусна лінгвістика може претендувати на науковість [17, с. 139-140].

Визначення корпусної лінгвістики як теорії або методології, за Е. Тоніні-Бонеллі, визначається надзвичайною складністю, оскільки корпусна лінгвістика не є галуззю дослідження, а методологією, яка значно перевершує власне методологічну роль [19, с. 1]. Саме тому значна група дослідників визначає корпусну лінгвістику не тільки як дослідницьку методологію вивчення мови, а фактично як новий філософський підхід, що характеризує корпусну лінгвістику як принципово нову парадигму [11, с. 106; 19, с. 1; 9, с. 2; 15, с. 49; 7, с. 14-16]. Таке визначення корпусної лінгвістики аргументується наявністю спільних характеристик корпусних досліджень і певною дослідницькою філософією [15, с. 49-50; 7, с. 9-10].

Зокрема, всі корпусні дослідження характеризуються зосередженням на: 1. Емпіричному аналізі усних або письмових текстів, створених у природних умовах [11, с. 107; 19, с. 12; 15, с. 49; 7, с. 9]. 2. Використанні корпусів текстів. 3. Застосуванні прийомів комп'ютерного аналізу [19, с. 12; 15, с. 50; 7, с. 9]. 4. Поєднанні кількісного аналізу з якісним – функціональною інтерпретацією мовлення [11, с. 107; 19, с. 12; 15, с. 50; 7, с. 9]. 5. Оперуванні в рамках контекстуальної і функціональної теорії значення [19, с. 12; 7, с. 10].

Як зазначено вище, корпусні дослідження передбачають обов'язкове використання комп'ютера, тому достатньо розповсюдженим є визначення корпусної лінгвістики як технології, і частково це так [15, с. 49, с. 214]. Однак такий підхід лише підкреслює штучні характеристики корпусної лінгвістики, але приховує її філософське підґрунтя, що передує застосуванню комп'ютерних технологій. Ґрунтовному аналізу власне філософії корпусної лінгвістики й впровадженню логічного підходу в корпусні дослідження присвячено роботу Т. Лагера [10, с. 18-43].

Таким чином, орієнтація лінгвістичних досліджень на емпіричні дані сприяла формуванню якісно нового філософського підходу до вивчення мови, відкрила нові шляхи її осмислення. За Е. Тоніні-Бонеллі, навіть дивно уявити, що лише збільшення обсягу мовних даних і точні обчислення їх характеристик змогли призвести до корінних філософських змін у сучасній лінгвістичній теорії [19, с. 48].

Визначення певної галузі дослідження передбачає окреслення її місця в науковій парадигмі. Традиційно корпусну лінгвістику відносять до гуманітарних наук, власне до лінгвістичних дисциплін [16; 3, с. 6-7; 2, с. 46]. Оскільки за своєю суттю корпусна лінгвістика є гуманітарною дисципліною, то, на думку А. Реноф, не слід плутати її з комп'ютерною лінгвістикою, що базується на математичному підґрунті, адже дисципліни відрізняються на філософському рівні [16].

Історично поява корпусної лінгвістики була відповіддю на необхідність вивчення англійської мови як іноземної [17, с. 137-138]. Вона не була реакцією на генеративну граматику Н. Хомського або когнітивну лінгвістику, а мала чітко визначене місце в британському контексті прикладної лінгвістики, з його акцентом на викладання мови та укладання словників. Однак, на думку Е. Тоніні-Бонеллі, серед інших дисциплін прикладної лінгвістики завдяки власній методології корпусна лінгвістика визначається статусом «надприкладної дисципліни» [19, с. 2].

Граничний характер дисципліни пояснює визначення корпусної лінгвістики як міждисциплінарної галузі дослідження, що поєднує інформаційні, комп'ютерні технології і власне лінгвістику [1, с. 9], або займає місце між точними й гуманітарними науками [17, с. 140-141]. На думку В. Тойберта, корпусна лінгвістика вирішує стандартні завдання гуманітарних наук, тобто, встановлення значення. Однак, при цьому методологія корпусної лінгвістики спирається на статистичні інструменти й програмне забезпечення, що забезпечує надійні й достовірні результати лінгвістичних досліджень. Отже, особливість визначення корпусної лінгвістики в науковій парадигмі полягає в тому, що за об'єктом дослідження вона належить до гуманітарних наук [16], а за методологією – до точних [17, с. 140-141].

Здійснене дослідження дозволяє дійти таких висновків: 1. У сучасному мовознавстві корпусна лінгвістика визначається як галузь, методологія дослідження, теорія або як загальний філософський підхід у лінгвістиці. 2. Як самостійна галузь мовознавства корпусна лінгвістика характеризується власним предметом, об'єктом, метою дослідження, термінологічним апаратом, теоретичним, практичним підґрунтям і власною методологією. 3. Як методологія дослідження корпусна лінгвістика визначається застосуванням статистичних методів, комп'ютерних технологій і програмного забезпечення. 4. Визначення корпусної лінгвістики як лінгвістичної теорії базується на визнанні тексту в якості головного об'єкту лінгвістичного дослідження й науковому характері власної методології дослідження. 5. Усі корпусні дослідження ґрунтуються на певній дослідницькій філософії. 6. У науковій парадигмі за об'єктом дослідження корпусна лінгвістика належить до гуманітарних наук, а за методологією – до точних. 7. Основними проблемами визначення корпусної лінгвістики є емпірична спрямованість, міждисциплінарний характер дисципліни й широке розуміння терміну корпусна лінгвістика.

Література:

1. Грудева Е. В. Корпусная лингвистика : учеб. пособие / Е. В. Грудева. – М. : ФЛИНТА, 2012. – 165 с.
2. Дарчук Н. П. Дослідницький корпус української мови: основні засади і перспективи / Н. П. Дарчук // Вісник Київського національного університету імені Тараса Шевченка. Серія: Літературознавство. Мовознавство. Фольклористика. – К. : Видавничо-поліграфічний центр «Київський університет», 2010. – № 21. – С. 45-49.
3. Демська-Кульчицька О. Основи національного корпусу української мови : [монографія] / Оріся Демська-Кульчицька; [відп. ред. О. Л. Перевозчикова]. – К. : Інститут української мови НАНУ, 2005. – 219 с.
4. Захаров В.П., Богданова С.Ю. Корпусная лингвистика: учебник / В. П. Захаров, С. Ю. Богданова. – Иркутск : Изд. Иркутского государственного лингвистического университета, 2011. – 161 с.
5. Карпіловська Є. А. Вступ до комп'ютерної лінгвістики: комп'ютерна лінгвістика: підручник / Євгенія Анатолівна Карпіловська. – Донецьк : ТОВ «Юго-Восток, ЛТД», 2006. – 188 с.
6. Лендау Сидні І. Словники: мистецтво та ремесло лексикографії / Сидні І. Лендау; [пер. з англ. 2 вид.]. – К. : К. І. С., 2012. – 480 с.
7. Плуґян В. А. Корпус как инструмент и как идеология: о некоторых уроках современной корпусной лингвистики / В. А. Плуґян // Русский язык в научном освещении. – М., 2008. – № 2 (16). – С. 7-20.
8. An International Handbook / Ed. by Anke Lüdelling, Merja Kytц. – Walter de Gruyter – Berlin – New York : Corpus Linguistics, 2008. – Volume 1. – 776 p.
9. Bennett G. R. Using CORPORA in the Language Learning Classroom: Corpus Linguistics for Teachers. – University of Michigan Press: Ann Arbor, 2010. – 144 p.

10. Lager T. A logical approach to computational corpus linguistics / Torbjörn Lager // Doctoral Dissertation for the degree of Doctor of Philosophy. – Göteborg : Göteborg University, Department of Linguistics, 1995. – 326 p.
11. Leech G. Corpora and theories of linguistic performance / In Directions in corpus linguistics: proceedings of Nobel Symposium 82 / ed. Svartvik J. – Berlin : Mouton de Gruyter, 1992. – P. 105-122.
12. McEnery T., Wilson A. Corpus Linguistics / T. McEnery, A. Wilson. – Edinburgh : Edinburgh University Press, 2001. – 235 p.
13. Meyer Ch. F. English Corpus Linguistics. An Introduction. Cambridge : Cambridge University Press, 2002. – 168 p.
14. Mukherjee J. The state of the art in corpus linguistics: three book-length perspectives / J. Mukherjee // English Language and Linguistics. – Vol. 8.1. – Cambridge University Press, 2003. – P. 103-119.
15. Perspectives on Corpus Linguistics. Studies in Corpus Linguistics / Eds. V. Vander, S. Zyngier, G. Barnbrook. – Vol. 48. – Amsterdam : John Benjamins Publishing, 2011. – 256 p.
16. Renouf An. Corpus linguistics: past and present / in Corpora in Use: In honour of Professor Y. Huizhong // eds. Wei Naixing, Wenzhong, Li, Pu Jianzhong. – 2005. – Режим доступу : www.rdues.bcu.ac.uk/publ/Past_and_Present.pdf
17. Teubert W. Linguistique de corpus: un alternative / Semen. Critical Discourse Analysis I. Les notions de contexte et d'acteurs sociaux / par Adèle Petitclerc, Philippe Schepens. – Vol. 27. – Presses Universitaires de Franche Comté. – Collection Annales Littéraires, 2009. – P. 130-152.
18. The handbook of English linguistics / ed. Bas Aarts, April McMahon. – Malden MA : Blackwell Publishing, 2006. – 824 p.
19. Tognini-Bonelli E. Corpus Linguistics at Work. Studies in Corpus Linguistics / Elena Tognini-Bonelli. – Amsterdam : John Benjamins Publishing, 2001. – 244 p.