

Жуковська В. В.

Київський національний лінгвістичний університет

ЛІНГВІСТИЧНИЙ КОРПУС ЯК НОВІТНІЙ ІНФОРМАЦІЙНО-ДОСЛІДНИЦЬКИЙ ІНСТРУМЕНТАРІЙ СУЧАСНОГО МОВОЗНАВСТВА

Стаття презентує систематизований огляд корпусної лінгвістики як інноваційного інформаційно-семіотичного напрямку емпіричного мовознавства: розкрито теоретико-методологічні засади галузі, визначено специфіку її зв'язку з когнітивною і комп'ютерною лінгвістикою, означено сучасні напрями корпусних досліджень і переваги залучення даних текстового корпусу у мовознавчих дослідженнях. Особлива увага приділена характеристиці лінгвістичного корпусу як багатофункціональної мовно-інформаційної системи, призначеної для вирішення широкого спектра дослідницьких питань; окреслено прототипові ознаки сучасного корпусу, визначено типи кількісних даних, що надаються корпусом, обґрунтовано необхідність застосування відповідних статистичних методів і спеціалізованих комп'ютерних програм обробки мовних даних для верифікації результатів корпусного аналізу. Розмежовано два магістральні напрями сучасних корпусних досліджень (корпусо-базований vs корпусо-керований) та аргументовано специфіку зазначених підходів до аналізу даних корпусу. У статті деталізовано пропонувану автором інтегровану методика проведення квантитативно-корпусного дослідження, яка інкорпорує принципи корпусо-базованого і корпусо-керovanого напрямів. Висловлено думку, що активне проникнення корпусних технологій у всі галузі мовознавства та суміжні науки зміцнить дослідницькі позиції корпусної лінгвістики у майбутньому та сприятиме появі нових міждисциплінарних і міждисциплінарних досліджень такого багатогранного феномену, як людська мова. З огляду на високу інформаційну насиченість і технологічність нинішніх корпусів запропоновано включати навчальні курси з корпусних технологій у професійну підготовку магістрів і докторів філософії на філологічних факультетах університетів України.

Ключові слова: корпусна лінгвістика, лінгвістичний корпус, типи корпусних даних, корпусо-базоване дослідження, корпусо-кероване дослідження.

“<...> every corpus that I've had a chance to examine <...> has taught me facts that I couldn't imagine finding out about any other way”.

Fillmore Ch. J. “Corpus linguistics” or “Computer-aided armchair linguistics” (1992)

Постановка проблеми. Останні десятиліття розвитку мовознавства позначені експонентним ростом кількості прикладних досліджень, що значною мірою зумовлюється дедалі більшою «потребою у використанні природномовних механізмів в інформаційно-комп'ютерних системах» [3, с. 262]. Новітньою галуззю прикладних мовознавчих студій, котра демонструє надзвичайно швидкі темпи розвитку, є корпусна лінгвістика [35, с. 427], а найбільш популярним мовно-інформаційним ресурсом, «альфою й омегою лінгвістики» [17, с. 8], визнається текстовий корпус. На думку Дж. Егберта та П. Бейкера, причина сьогоденішнього масового захоплення

корпусами й корпусною методологією є надзвичайно простою: корпуси надають об'єктивні свідчення того, як «реальні люди використовують реальну мову» [18, с. 4]. Для великої когорти сучасних дослідників лінгвістично та інформаційно анотований корпус із мультифункціональним програмним забезпеченням виступає надійним джерелом мовних даних та ефективним інструментом отримання об'єктивних і статистично верифікованих висновків. Застосування даних різноманітних корпусів постулюється як невіддільна ознака мовознавчого дослідження, що претендує на надійність і вірогідність отриманих результатів [6, с. 108].

Аналіз останніх досліджень і публікацій. Корпусна лінгвістика та різнопланові дослідження на основі корпусів перебувають у центрі уваги багатьох дослідників. Науковий доробок сучасного мовознавства щороку поповнюється новими посібниками [8; 10; 16; 37; 43], колективними монографіями [33; 15; 13], лекційними матеріалами [22] та різноманітними оглядами останніх досягнень у галузі [23; 34]. Змістова наповненість цих публікацій і глибина порушених питань варіюється залежно від наукових інтересів авторів, проте традиційно розглядається історія становлення корпусної лінгвістики, окреслюються її теоретико-методологічні настанови, презентуються різні класифікації лінгвістичних корпусів, специфікуються технологічні можливості програмного забезпечення для роботи з корпусними даними, обговорюються результати різноаспектних корпусних досліджень. Водночас питання теоретичного статусу корпусної лінгвістики, окреслення прототипових ознак сучасного корпусу, роль статистичних методів у корпусних дослідженнях і напрями корпусного аналізу ще потребують подальших обговорень.

Постановка завдання. Запропоноване дослідження має на меті окреслити сучасну корпусну лінгвістику як інноваційну галузь емпіричного мовознавства та розкрити переваги залучення даних лінгвістичного корпусу у сучасних мовознавчих розвідках. Звичайно, поставлена мета є доволі амбітною, і її досягнення потребуватиме наукової праці значно більшої, ніж журнальна публікація. Сподіваємося, що запропонований систематизований огляд статусу корпусної лінгвістики як інформаційно-семіотичного напрямку та головних понять корпусного дослідження – таких як лінгвістичний корпус, прототипові ознаки корпусу, типи корпусних даних, корпусо-базоване vs корпусо-кероване дослідження, – здійснений на основі аналізу спеціалізованої літератури в галузі за останні роки, допоможе емпірично налаштованим лінгвістам зорієнтуватися в останніх тенденціях корпусного мовознавства.

Виклад основного матеріалу. За доволі коротку історію існування (з кінця ХХ ст.) корпусна лінгвістика утвердилася як основна емпірична парадигма науки про мову [2, с. 239], що ґрунтується на квантитативних засадах і покликана зробити мовознавство «експериментальною наукою у сучасному значенні цього слова» [3, с. 263]. До компетенції галузі входять 1) аналіз (зазвичай) великих колекцій природномовних текстів у електронному форматі з використанням

спеціалізованого комп'ютерного програмного забезпечення [8, с. 93] та 2) дослідження прикладних мовознавчих проблем у комунікативних процесах, у фокусі уваги перебуває не мова як система і не проблема її формалізації, а процес змістовної комунікації мовою [2, с. 240]. Як новітній інформаційно-семіотичний напрям мовознавства корпусна лінгвістика виявляє тісні зв'язки з комп'ютерною і когнітивною лінгвістикою. Від першої вона отримує інструменти обробки мовного матеріалу – комп'ютерне програмне забезпечення, що дозволяє дуже швидко й ефективно опрацьовувати великі обсяги мовного матеріалу, шукати відповідні одиниці, сортувати результати пошуку й анутовати тексти [19, с. 4], а з другою має спільні базові передумови дослідження: «мовну стихію, що представлена у нескінченній кількості текстів і є тією реальністю, яку й необхідно вивчати, не зводячи її до обмеженого набору структурних схем, інваріантів та ідеальних парадигм» [4, с. 28]. Тож корпуси постають потужними мовно-інформаційними системами, що активно використовуються для вирішення широкого спектра дослідницьких питань практично в усіх галузях мовознавства: лексикографії, граматиці, лексикології та семасіології, стилістиці, перекладацьких студіях, прагматиці, соціолінгвістиці, психолінгвістиці, дискурсолгії, когнітивній лінгвістиці, лінгвістичній варіантології, а також у навчанні та вивченні (рідної та / або іноземної) мови, літературознавчих дослідженнях і т. ін. [див.: 10; 12; 13; 20; 24; 26; 27; 37; 40; 44].

Головна мета лінгвістики корпусів полягає у кращому розумінні правил, які керують мовою загалом або певним аспектом мови (регістром, діалектом і т. ін.) [17, с. 7], та охопленні досліджуваної предметної галузі (мови) в усіх її проявах і повноті явищ [7, с. 7]. Беззаперечною перевагою корпусних досліджень є можливість відтворити лінгвістичний експеримент і незалежно перевірити отримані результати, а застосовані комп'ютеризовані процедури аналізу сприяють усуненню когнітивної, соціальної та політичної упередженості [18, с. 2], мінімізуючи суб'єктивізм й особисті вподобання дослідника.

Об'єктом і предметом дослідження галузі виступає лінгвістичний корпус, що характеризується низкою дистинктивних ознак (базованістю на вибірці, збалансованістю, репрезентативністю, порівнюваністю і природномовністю) [16, с. 6; 17, с. 3], які чітко відрізняють його від електронних бібліотек, збірок електронних текстів та архівів. Результатом бурхливого розвитку корпусної

лінгвістики в останні два десятиліття є зростаюча у геометричній прогресії кількість різноманітних корпусів [див. огляди: 5; 20, с. 81–114; 21, с. 9–11; 38, с. 10–17; 43 с. 15–25; і каталог корпусів: 14], укладених із різною дослідницькою метою. Не всі з них повною мірою втілюють вказані ознаки та відповідають канонам таких еталонних корпусів, як, наприклад, Браунівський корпус чи Британський національний корпус. Тож, як зазначає С. Грайс, у сучасному розумінні лінгвістичний корпус виступає прототиповою категорією [22, с. 4] і дефініюється як укладене з дослідницькою метою зібрання продукованих у природних комунікативних умовах матеріалів (текстових (усних / писемних), графічних, аудіо, відео), що є репрезентативним і збалансованим відповідно до певного варіанту / регістру / жанру, розміщене на електронному носії й оснащено відповідним програмним забезпеченням [17, с. 3; 21, с. 7; 22, с. 4; 43, с. 13].

Корпус як величезний масив мовних даних надає у розпорядження дослідника один тип свідчень (*evidence*) – це кількісні дані про функціонування мовної одиниці будь-якого рівня трьох видів: **фактичні дані** (*factual evidence*) (одиниця x зафіксована у корпусі ‘існує (x)’), **дані про частотність** (x вживається ‘ $f(x)$ ’ разів у корпусі) і **дані про взаємодію** (x співживається з іншою одиницею у більш / менш часто, ніж це було б випадково) [42, с. 61] (див. рис. 1).

Якщо перші корпусні дослідження базувалися на кількісних підрахунках мовних одиниць і переважно фокусувалися на аналізі співживання (колокації / колігації) слів, то сьогодні застосування сирих (*raw*) даних про частоту вже не є достатнім. Сучасні корпусні дослідження обов’язково передбачають застосування відповідних статистичних процедур (наприклад, методів колострукційного аналізу: колексемного аналізу, розмежувального колексемного аналізу, коваріаційного колексемного аналізу) та спеціалізова-

них комп’ютерних програм опрацювання великих обсягів складних даних (як-от, вільно поширюваний програмний комплекс RStudio), формуючи новий напрям – *квантитативна корпусна лінгвістика* [11; 21; 22; 29]. Водночас трактувати лінгвістику корпусів як суто квантитативний напрям неправомірно, оскільки кожен етап роботи з корпусними даними потребує всебічної інтерпретації та пояснень дослідника зі зверненням до ретельного прочитання текстів корпусу та залучення відповідних корпусних інструментів.

Корпусна лінгвістика є вповні сформованою галуззю сучасних мовознавчих досліджень, проте питання щодо її теоретичного статусу ще й досі однозначно не вирішене. Наприклад, систематизуючи підходи до трактування корпусної лінгвістики, Ш. Тейлор наводить такі визначення: корпусна лінгвістика – це інструмент (*a tool*), метод (*a method*), методологія (*a methodology*), методологічний підхід (*a methodological approach*), дисципліна (*a discipline*), теорія (*a theory*), теоретичний напрям (*a theoretical approach*) або парадигма (*a paradigm*) (теоретична чи методологічна) [39, с. 180]. Проте, за великим рахунком, усі перераховані підходи можна поділити на ті, які визнають корпусну лінгвістику як *окремий мовознавчий напрям*, що має потенціал удосконалити та чіткіше окреслити низку теорій мови [30, с. 1], і на ті, які ставляться до лінгвістики корпусів як до *методології дослідження*, покликаної вдосконалити аналіз вживання та варіативності мови [10; 28, с. 105–106]. Підставою для такого поділу є специфічне розуміння вихідної точки та кінцевої мети корпусного аналізу.

Традиційно виділяють два напрями корпусних досліджень, що відбивають дихотомію раціонального й емпіричного, дедуктивного й індуктивного у науковому дослідженні: від теорії до корпусу – корпусо-базовані (*corpus-based*) та від корпусу до теорії – корпусо-керовані (*corpus-driven*) дослідження [1; 41].

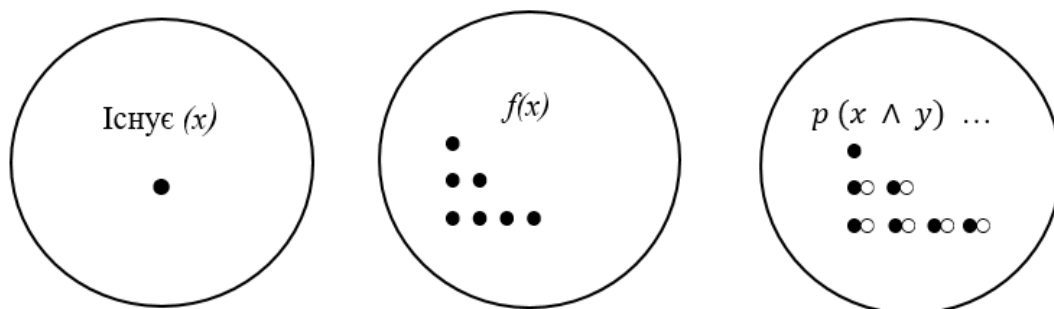


Рис. 1. Типи корпусних даних (адаптовано з [42, с. 62])

Корпусо-базовані дослідження мають на меті «описати та пояснити лінгвістичні моделі варіативності та вживання» [9, с. 163] на основі репрезентативної вибірки реальних текстів (тобто даних корпусу). Прихильники підходу вважають, що призначення корпусного аналізу – перевірити або уточнити наявні у лінгвістичних джерелах гіпотези або лінгвістичні теорії (моделі) [8, с. 95]. Базоване на корпусі дослідження є значною мірою дедуктивним, оскільки починається з уже наявних описів, категорій і класифікацій, якими оперує лінгвіст до початку роботи з корпусом, що потім тестуються на даних корпусу. У цьому разі використання корпусу є методологією дослідження, а базований на корпусі підхід є *підтверджувальним (confirmatory)* [25, с. 59] за своєю суттю.

Корпусо-керований напрям є радикальним продовженням корпусо-базованого напрямку. У цьому разі корпус не є частиною методу, а слугує єдиним засобом доступу до мовної компетентності [17, с. 9]. Головна мета керованого корпусом дослідження полягає в ідентифікації раніше не встановлених лінгвістичних категорій та одиниць [9, с. 168] і виявленні нових лінгвістичних конструкцій за допомогою індуктивного аналізу корпусів [9, с. 169]. Прибічники підходу використовують неанотовані корпусні дані для формулювання гіпотез про мову без звернення до вже наявних лінгвістичних теорій, а тому вибудовують теорію виключно на даних, отриманих з аналізу корпусу, та вважають, що пряме спостереження корпусних даних є вихідною точкою лінгвістичного аналізу. Перевагою корпусо-керованих досліджень є репрезентативність аналізованих даних, їхня надійність і відтворюваність. Такий підхід до аналізу корпусу дозволяє виявити властивості, непомітні за традиційного дедуктивного підходу, та має *дослідницький (exploratory)* характер [25, с. 59].

Отже, у корпусо-базованих розвідках корпуси аналізуються у квантитативному і квалітативному аспектах на основі вже наявних теоретичних засад, а подекуди – із залученням даних не лише корпусу, а у корпусо-керованих дослідженнях розглядається повністю корпусо-генерована модель мови [41, с. 65–100]. Попри вказані відмінності, обидва підходи мають низку спільних дослідницьких настанов: 1) опору на частотність, використання корпусів і їх опрацювання у якісному та кількісному аспектах; 2) зміщення дослідницької стратегії з вивчення норми на вивчення узусу; 3) автоматичне отримання інформації за допомогою пошукових запитів,

яке може призводити до отримання великого обсягу матеріалу, що не завжди є релевантним; 4) поширення формально-морфологічного підходу, за якого пошук прикладів базується на морфологічній або на буквеній формі; 5) використання квантитативних методів, що дозволяють враховувати частотні характеристики досліджуваних одиниць, і заміна інтроспективної оцінки матеріалу точними кількісними даними про вживання; 6) опора на автоматичне анотування, яке може мати певні неточності; 7) увага до контексту у широкому розумінні (дослідження колокацій, ключових слів, конструкцій передбачає врахування оточення досліджуваної одиниці) [1, с. 42].

Насправді протиставлення корпусо-базованих і корпусо-керованих досліджень є дещо перебільшеним. Ці два підходи слід розглядати як крайні точки [32, с. 8] континууму корпусного аналізу. Відмінності між ними мають градуальний характер і свідчать лише про «ступінь корпусності» [1, с. 42; 31] конкретного дослідження. Наприклад, у нашому аналізі відмінностей між синонімічними граматичними конструкціями [*with*[Subj_{COMMON NOUN}][V_{PARTICIPLE I}]] та [*with-less*[Subj_{COMMON NOUN}][V_{PARTICIPLE I}]] була апробована інтегрована методика квантитативно-корпусного дослідження, що інкорпорує принципи обох напрямів [див.: 45] і включає такі положення: 1) поєднання підтверджувального та дослідницького аспектів: з одного боку перевіряються робочі гіпотези, сформульовані з огляду на теоретико-методологічні засади застосованого у роботі лінгвістичного напрямку, а з іншого – аналізовані корпусні дані самі слугують джерелом нових гіпотез і положень; 2) анотований корпус є легітимним джерелом мовних даних; 3) дані з анотованого корпусу відбираються за допомогою автоматизованого пошуку, який здійснюється вбудованим корпусним менеджером; 4) повне врахування усіх випадків вживання досліджуваної одиниці, жоден приклад не виключається лише через те, що якийсь критерій неможливо до нього застосувати; 5) отримані дані, типово достатньо великого обсягу, піддаються *систематичному та вичерпному* кількісному (статистичному) аналізу, що має на меті перевірку теоретичних гіпотез (підтверджувальний аспект) або формулювання нових (дослідницький аспект); 6) залучення до аналізу корпусних метаданих (інформація про мовців / авторів, жанри, реєстри й т. ін.), що розширює масштаб корпусного дослідження і перетворює його на «кероване даними» (*data-driven*) (у термінах [35]) дослідженням.

Висновки і пропозиції. Підсумовуючи, наголосимо, що лінгвістичний корпус – це не просто джерело ілюстративних прикладів, а потужний інструмент, багатofункціональна лінгво-інформаційна система для проведення різнопланових мовознавчих досліджень. Завдяки ефективному поєднанню технологічного та людського компонентів у лінгвістичному аналізі корпусна лінгвістика відкриває нові перспективи для мовознавчої науки. У найближчому майбутньому корпусна методологія лише зміцнить свої дослідницькі позиції внаслідок активного проникнення її технологій у всі мовознавчі галузі та суміжні

науки, що сприятиме міжпарадигмальним і міждисциплінарним дослідженням такого багатогранного феномену, як людська мова. Водночас висока інформаційна насиченість і технологічність нинішніх корпусів вимагає від лінгвіста не лише ґрунтовних знань мовознавчого профілю, але й обізнаності з основами (лінгво)статистики, комп'ютерного програмування та моделювання. Навчальні курси з корпусних технологій повинні стати нормативною компонентою силабусів професійної підготовки магістрів і докторів філософії на філологічних факультетах у вітчизняних університетах.

Список літератури:

1. Голубкова Е. Е. Использование лингвистических корпусов при решении семантических проблем. *Методы когнитивного анализа семантики слова: компьютерно-корпусный подход* / под общ. ред. В. И. Заботкиной. Москва, 2015. С. 39–80.
2. Заячковская О. О. Концептуальный анализ семантики эмоционального лексикона с опорой на корпусные данные. *Методы когнитивного анализа семантики слова: компьютерно-корпусный подход* / под общ. ред. В. И. Заботкиной. Москва, 2015. С. 243–268.
3. Коннова М. Н. Когнитивные сдвиги в семантике темпоральных лексем: корпусный аспект. *Методы когнитивного анализа семантики слова: компьютерно-корпусный подход* / под общ. ред. В. И. Заботкиной. Москва, 2015. С. 269–310.
4. Копотев М. В. Принципы синтаксической идиоматизации. Helsinki : Helsinki University Press, 2008. 52 с.
5. Коцюк Л. М., Коцюк Ю. А. Класифікаційна парадигма корпусу текстів за особливостями його дизайну, структури та способами використання, а також способом фіксації та індексації текстових даних. *Наукові записки Національного університету «Острозька академія». Серія «Філологія»*. 2020. Вип. 9 (77). С. 106–110.
6. Кузнецова Ю. Л., Велейшикова Ю. Л. Современные корпусные исследования языка: новые подходы. *Вопросы языкознания*. 2010. № 6. С. 108–124.
7. Лінгвістично-інформаційні студії : праці Українського мовно-інформаційного фонду НАН України : у 5 т. / В. А. Широков та ін. Т. 4 : Корпусна та когнітивна лінгвістика. Київ, 2018. 246 с.
8. Baker P. Corpus methods in linguistics. *Research methods in linguistics* / Ed. L. Litosseliti. London, New York, 2010. P. 93–113.
9. Biber D. Corpus-based and corpus-driven analyses of language variation and use. *The Oxford Handbook of Linguistic Analysis* / eds. B. Heine, H. Narrog. Oxford, 2010. P. 159–191.
10. Biber D., Reppen R. Introduction. *The Cambridge Handbook of English Corpus Linguistics* / eds. D. Biber, R. Reppen. Cambridge, 2015. P. 1–8.
11. Brezina V. Statistics in corpus linguistics. A Practical Guide. Cambridge : Cambridge University Press, 2018. 296 p.
12. Contemporary corpus linguistics / ed. P. Baker. London : Continuum International Publishing Group, 2012. 356 p.
13. Corpus linguistics on the move. Exploring and understanding English through corpora / eds. M. J. López-Couso, B. Méndez-Naya, P. Núñez-Pertejo, I. M. Palacios-Martínez. Leiden, Boston : Brill Rodopi, 2016. 367 p.
14. Corpus Resource Database (CoRD) : веб-сайт. URL: <http://www.helsinki.fi/varieng/CoRD/> (дата звернення: 08.09.2020).
15. Corpus-based approaches to grammar, media and health discourses. Systemic Functional and other perspectives / Eds. D. Yang, W. Li. Singapore: Springer, 2020. 396 p.
16. Crawford W. J., Csomay E. Doing corpus linguistics. New York, London : Routledge, Taylor & Francis Group, 2016. 163 p.
17. Desagulier G. Corpus linguistics and statistics with R. Introduction to quantitative methods in linguistics. Cham: Springer International Publishing, 2017. 353 p.
18. Egbert J., Baker P. Introduction. *Triangulating Methodological Approaches in Corpus-Linguistic Research* / eds. P. Baker, J. Egbert. New York, London, 2016. P. 1–19.

19. Egbert J., Baker P. Introduction. *Using Corpus Methods to Triangulate Linguistic Analysis* / eds. J. Egbert, P. Baker. New York, London, 2020. P. 1–21
20. Friginal E. *Corpus Linguistics for English Teachers. New Tools, Online Resources, and Classroom Activities*. New York, London : Routledge, Taylor & Francis Group, 2018. 349 p.
21. Gries St. Th. *Quantitative Corpus Linguistics with R. A practical introduction*. New York, London : Routledge, Taylor & Francis Group, 2017. 272 p.
22. Gries St. Th. *Ten Lectures on Corpus Linguistics with R. Application for Usage-Based and Psycholinguistic Research*. Leiden, Boston : Brill, 2020. 298 p.
23. Grisot C. Cohesion, Coherence and Temporal Reference from an Excremental Corpus Pragmatic Perspective. *Yearbook of Corpus Linguistics and Pragmatics*. Springer, Springer International Publishing, 2018. 319 p. URL: <https://doi.org/10.1007/978-3-319-96752-3> (дата звернення 07.09.2020).
24. Hansen B. *Corpus linguistics and sociolinguistics: a study of variation and change in the modal systems of world Englishes*. Leiden; Boston : Brill, 2018. 413 p.
25. Jensen G. B., McGillivray B. *Quantitative historical linguistics. A corpus framework*. Oxford: Oxford University press, 2017. 229 p.
26. Jones Ch., Waller D. *Corpus Linguistics for Grammar: A Guide for Research*. New York, London : Routledge, Taylor & Francis Group, 2015. 201 p.
27. Lange C., Leuckert S. *Corpus Linguistics for World Englishes: A Guide for Research*. New York, London : Routledge, Taylor & Francis Group, 2020. 220 p.
28. Leech, G. *Corpora and theories of linguistic performance*. *Directions in corpus linguistics* / ed. J. Startvik. Berlin, 1992. P. 105–122.
29. Levshina N. *How to do Linguistics with R. Data Exploration and statistical analysis*. Amsterdam / Philadelphia : John Benjamins Publishing Company, 2015. 443 p.
30. McEnery T., Hardie A. *Corpus Linguistics: Method, Theory and Practice*. Cambridge : Cambridge University Press, 2012. 292 p.
31. McEnery T., Gabrielatos C. English corpus linguistics. *The Handbook of English Linguistics* / eds. B. Aarts, A. McMahon. Oxford, 2006. P. 33–71.
32. McEnery T., Xiao R., Tono Y. *Corpus-Based Language Studies: An Advanced Resource Book*. London, New York, 2006. 408 p.
33. *New directions in corpus-based translation studies* / eds. C. Fantinuoli, F. Zanettin – Berlin : Language Science Press, 2015. 164 p.
34. Park H., Nam D. Corpus linguistics research trends from 1997 to 2016: A co-citation analysis. *Linguistic Research*. 2017. № 34 (3). P. 427–457.
35. Rayson P. From key words to key semantic domains. *International Journal of Corpus Linguistics*. 2008. № 13. P. 519–549.
36. Rühlemann Ch. *Corpus Linguistics for Pragmatics: A Guide for Research* – New York, London : Routledge, Taylor & Francis Group, 2019. 206 p.
37. Stefanowitsch A. *Corpus Linguistics. A Guide to the methodology*. Berlin : Language Science Press, 2020. 490 p.
38. Szudarski P. *Corpus Linguistics for Vocabulary. A Guide for Research*. New York, London : Routledge, Taylor & Francis Group, 2018. 224 p.
39. Taylor Ch. What is corpus linguistics? What data says. *ICAME Journal*. 2008. № 32. P. 179–200.
40. Timmis I. *Corpus Linguistics for ELT: Research and practice* New York, London : Routledge, Taylor & Francis Group, 2015. 212 p.
41. Tognini-Bonelli E. *Corpus Linguistics at Work*. Amsterdam : John Benjamins, 2001. 219 p.
42. Wallis S. Grammar and corpus methodology. *The Oxford Handbook of English Grammar* / eds. B. Aarts, J. Bowie, G. Popova. Oxford, 2020. P. 59–86.
43. Weisser M. *Practical Corpus Linguistics: An Introduction to Corpus-Based Language Analysis*. Wiley, Blackwell, 2016. 287 p.
44. *Yearbook of Corpus Linguistics and Pragmatics 2015. Current approaches to discourse and translation studies* / ed. J. Romero-Trill. Springer : Springer International Publishing, 2015. 328 p.
45. Zhukovska V. A Quantitative Corpus-Driven Approach to Disambiguation of Synonymous Grammatical Constructions. *Proceedings of the 4th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2020)*. Volume I: Main Conference, Lviv, Ukraine, April 23–24, 2020. CEUR Workshop Proceedings 2604, CEUR-WS.org 2020. P. 507–522. URL: <https://dblp.uni-trier.de/pers/hd/z/Zhukovska:Viktoriiia> (дата звернення 08.09.2020)

Zhukovska V. V. LINGUISTIC CORPUS AS A STATE-OF-THE-ART INFORMATION AND RESEARCH TOOL IN CONTEMPORARY LANGUAGE STUDY

This study presents an overview of corpus linguistics as an innovative information-semiotic approach in empirical linguistics. The paper substantiates the theoretical and methodological underpinnings of the field, identifies its connection with cognitive and computational linguistics, discusses modern areas of corpus research and demonstrates the feasibility of utilising text corpus data in modern linguistic research. Particular attention is devoted to the characteristics of the linguistic corpus as a powerful multifunctional language information system used to address a broad range of research issues. The article defines prototypic features of the modern corpus, specifies types of quantitative data provided by the corpus, justifies the need for utilizing appropriate statistical methods and specialized computer data processing programs to verify the results of corpus analysis. Two dominant approaches in corpus research (corpus-based vs corpus-driven) are distinguished, the specifics of these approaches to a corpus analysis are outlined. In this paper, the author introduces the integrated methodology of conducting a quantitative-corpus research, incorporating the principles of corpus-based and corpus-driven approaches. It is stated that the active penetration of corpus technologies in all fields of linguistics and related sciences will strengthen the research position of corpus linguistics in the future and contribute to cross-paradigmatic and cross-disciplinary research of such a multifaceted phenomenon as human language. Given the high information and technical specifications of contemporary corpora, courses on corpus technologies should be included into the university training of Masters and Doctors of Philosophy in Linguistics.

Key words: *corpus linguistics, linguistic corpus, types of corpus data, corpus-based research, corpus-driven research.*