

УДК 371.14:811.111

DOI <https://doi.org/10.24919/2308-4863/72-4-20>**Анастасія ШЕВЧУК,***orcid.org/0000-0002-5005-3800**кандидат філологічних наук, доцент,  
доцент кафедри іноземної та української філології  
Луцького національного технічного університету  
(Луцьк, Україна) a.shevchuk@lutsk-ntu.com.ua***Юлія ЛІТКОВИЧ,***orcid.org/0000-0003-4962-0617**кандидат філологічних наук, доцент,  
доцент кафедри іноземної та української філології  
Луцького національного технічного університету  
(Луцьк, Україна) litkovych@ukr.net*

## КОРПУСО-БАЗОВАНЕ НАВЧАННЯ: ТРУДНОЩІ ІМПЛЕМЕНТАЦІЇ ПІДХОДУ

Стаття присвячена аналізу ключових викликів, пов'язаних із інтеграцією мовних корпусів на заняттях з філологічних дисциплін, оскільки питання впровадження інформаційних технологій в навчальний процес динамічно набирає обертів. Аналіз лінгвістичних даних, мовних помилок, вивчення академічного мовлення та адаптація під потреби користувачів – це лише деякі аспекти, що роблять ці корпуси невід'ємною частиною навчання.

Впровадження корпусних технологій надає значний потенціал для оптимізації робочих процесів та підвищення ефективності діяльності. Однак, необхідно враховувати можливість виникнення труднощів, пов'язаних із великим обсягом даних, необхідністю їх структурування та забезпечення адекватного аналізу інформації. Завдяки правильній стратегії і управлінню цими викликами, користувачі зможуть максимально використовувати переваги корпусних технологій. Очевидними постають труднощі як зі сторони викладача, так і – здобувача, які найчастіше належать до царини доступу, застосовності та технічних можливостей реалізації.

Для одних важливо мати адекватний доступ до розширених корпусів даних для якісного викладання та досліджень. Інші ж повинні мати достатні ресурси для отримання доступу до відповідних корпусів для навчання та самостійного вивчення. Забезпечення широкого та рівного доступу є ключовим елементом успішної імплементації корпусних технологій у навчальний процес. Технічні виклики в реалізації корпусних технологій можуть включати в себе потребу в інфраструктурі з високою обчислювальною потужністю та спеціалізованих програм, що вимагає від учасників освітнього процесу додаткових технічних навичок та ресурсів. Питання застосовності передбачає важливість того, щоб корпусні дані були релевантними для навчальних цілей, а також для досліджень та професійного розвитку.

Завдяки всебічному аналізу стаття має на меті надати уявлення про ефективні стратегії подолання цих викликів та максимізації переваг вивчення мови на основі корпусів в контексті цифровізації освітнього процесу.

**Ключові слова:** корпус, мова, дані, корпусо-базоване навчання, прикладна лінгвістика, технології.

**Anastasiia SHEVCHUK,***orcid.org/0000-0002-5005-3800**Candidate of Philological Sciences,  
Associate Professor at the Department of Foreign and Ukrainian Philology  
Lutsk National Technical University  
(Lutsk, Ukraine) a.shevchuk@lutsk-ntu.com.ua***Yuliia LITKOVYCH,***orcid.org/0000-0003-4962-0617**Candidate of Philological Sciences,  
Associate Professor at the Department of Foreign and Ukrainian Philology  
Lutsk National Technical University  
(Lutsk, Ukraine) litkovych@ukr.net*

## CORPUS-BASED LEARNING: CHALLENGES OF IMPLEMENTATION

The article aims to analyse the main challenges related to integrating language corpora in philological disciplines. The issue of introducing information technologies in the educational process is gaining momentum. Linguistic data

*analysis, language error identification, academic language study, and adapting it to user needs are some of the aspects that make corpora an integral part of learning process.*

*The use of corpus-based technologies presents a significant opportunity to streamline workflows and improve efficiency. However, it is crucial to acknowledge the potential challenges that come with handling large volumes of data, structuring it, and conducting thorough analysis. By implementing effective strategies and managing these challenges appropriately, users can fully leverage the advantages of corpus technology. Difficulties arise for both teachers and students, often related to access, applicability, and technical feasibility.*

*Ensuring broad and equitable access to extended corpora is crucial for qualified teaching and research, as well as for self-study purposes. The use of corpus technologies in education can only be successful if everyone has sufficient resources to access relevant corpora. Technical challenges in implementing corpus-based technologies may include the need for high power computing infrastructure and specialised software, requiring additional technical skills and resources from those involved in the educational process. The issue of applicability highlights the importance of making corpus data relevant for educational purposes, research, and professional development.*

*The article aims to provide insights into effective strategies for maximising the benefits of corpus-based language learning in the context of digitalisation of the educational process. It also discusses ways to overcome the challenges associated with this type of learning and outlines the prospects for further research.*

**Key words:** *corpus, language, data, corpus-based learning, applied linguistics, technologies.*

**Постановка проблеми.** Сучасна освіта переживає значущі трансформації завдяки активному впровадженню інформаційних технологій та популяризації абсолютних цифровізаційних тенденцій. Комп'ютеризація навчального середовища стає неминучою, оскільки розвиток інформаційної компетентності – одне з ключових умінь новочасного фахівця. Враховуючи динамічність розвитку підходів до навчання, останніми роками спостерігаємо значний сплеск інтересу до корпусної лінгвістики та її глибокого застосування з навчальною метою, зокрема під час мовної підготовки здобувачів філологічних напрямків. Термін «корпусна лінгвістика» виник та розвивався завдяки спільним зусиллям багатьох дослідників. Однак, піонерами апробування корпусних методів в лінгвістиці з середини 20 століття вважаються американські лінгвісти Чарльз Філлмор та Едвард Джонс. Вони вперше застосували поняття корпусу для аналізу лінгвістичних явищ, започатковуючи тим самим розвиток корпусної лінгвістики як самостійної галузі.

У класичному трактуванні корпусна лінгвістика як вивчення мови на основі прикладів «реального» ужитку (McEneaney, Wilson, 2001), згодом у новому розширеному «компіляцією та аналізом корпусів» (Cheng, 2012), набула популярності завдяки здатності пропонувати емпіричне розуміння використання мови. Це у свою чергу дало поштовх до багатовекторного застосування корпусних даних, наділяючи галузь низкою тісних міждисциплінарних зв'язків. Зважаючи на доволі активне і широке залучення корпусних ресурсів у навчальний процес, особливе значення такі технології мають для філологів. Утім, на противагу значному потенціалу корпусної лінгвістики існує низка труднощів, які не отримали досі належної уваги в наукових колах. Тому надзвичайно важли-

вим та актуальним постає питання виокремлення проблемних аспектів, а також представлення певних стратегій щодо їхнього подолання.

**Метою** пропонованої статті є систематизований огляд ключових труднощів використання корпусо-базованого підходу до навчання дисциплін філологічного спрямування та пошук ефективних стратегій їх подолання.

**Аналіз останніх досліджень.** Досягнення корпусної лінгвістики значно розширюють можливості вивчення мовних явищ, адже чимало вчених і дослідників працювали над розвитком цієї галузі протягом років. Джон Сінклер, засновник лінгвістичного проєкту COBUILD, підкреслював важливість повторюваних словосполучень і готових речень у педагогічному контексті через їхній широкий потенціал у розвитку вільного володіння мовою, точності та ідіоматичності. Марк Дейвіс відіграв важливу роль у розробці доступних корпусів, таких як Corpus of Contemporary American English (COCA) – величезними за обсягом сучасними лінгвістичними ресурсами, що сприяє викладанню мови та проведення досліджень на основі корпусу. Мартін Вайсер, Вікторія Жуковська, Соломія Бук розвивають напрямок використання корпусів у мовній освіті. Дуглас Бібер, Крістіан Джонс та Деніел Уоллер працюють із вивченням граматики англійської мови на основі корпусу. Надя Нессельхауф та Сільвіан Грейнджер аналізують спеціалізовані навчальні корпуси. Методологією корпусної лінгвістики цікавився Тоні МакЕнері, що згодом вплинуло на інноваційні підходи до викладання лексики, граматики та дискурсу на заняттях іноземної мови як другої. Кожен із дослідників здійснив вагомий внесок у вивчення та вдосконалення можливостей застосування корпусів для аналізу мовних даних. Відтак колективні зусилля цих впливових вчених

підкреслюють вагомість корпусної лінгвістики в освітньо-науковому середовищі, що не лише розширює сферу наукових пошуків, а й збагачує педагогічну практику новими актуальними підходами до навчання.

**Виклад основного матеріалу.** Беззаперечним є факт багатоспектрового застосування корпусних технологій з огляду на світовий досвід розвитку цієї лінгвістичної галузі. Використання текстових корпусів – значних за обсягом систематизованих колекцій текстів або транскрипцій аудіозаписів, зібраних в електронній формі для відбиття певної мови, мовного варіанту або іншої лінгвістичної сфери (Kübler, Zinsmeister, 2015), відкриває далекоюсяжні обрії вивчення найрізноманітніших аспектів мови в аутентичних контекстах. Все ж попри зростаюче зацікавлення перспективами та активні дослідження у сфері корпусної лінгвістики, імплементація корпусних ресурсів не є легкою справою. Важливо звернути увагу на основні труднощі, які постають перед поціновувачами корпусно-базованого навчання.

Однією з центральних та досить помітних проблем є робота з **технічними аспектами** корпусної лінгвістики, що може обмежувати здатність максимально використовувати потенціал корпусів.

Навігація та використання спеціалізованих програм для роботи з корпусами може стати викликом, особливо для тих, хто не володіє достатньою комп'ютерною грамотністю. Наприклад, для доступу до British National Corpus (BNC) потрібне спеціальне програмне забезпечення, BNCweb, яке надається самим Британським національним корпусом, а Corpus of Contemporary American English (COCA) доступний через інтерфейс в Інтернеті, але також існує опція встановлення спеціального програмного забезпечення для користування окремими додатковими функціями. Задля подолання цього бар'єру надзвичайно важливо забезпечити належну підготовку з володіння технічними інструментами аналізу корпусних даних та організацію підтримки. На початку роботи, важливо звертатися до інструментарію з інтуїтивно зрозумілими та дружельюбними інтерфейсами, забезпечити базові знання та навички, а також розробити педагогічні матеріали з інструкціями та порадами з використання корпусів у вивченні мови.

Недостатнє володіння потрібними навичками ефективної інтеграції корпусно-базованих методів навчання також дещо загострено піднімає питання педагогічної майстерності. У такому разі постійний професійний розвиток сприятиме вправному орієнтуванню та використанню технічних параметрів.

Іншим серйозним викликом може стати **обмежений доступ** до усього спектру функціональних можливостей як загальних, так і спеціалізованих корпусів. Часто розробники дозволяють користуватися своїм продуктом у повному обсязі (the Open American National Corpus (OANC)), або протягом пробного періоду, або згідно з умовами передплати (BNC), яка теж може диференціюватися залежно від кількості зареєстрованих осіб, параметрів пошуку, часу тощо. Наприклад, доступ до The Oxford English Corpus (OEC) зазвичай надається дослідникам, лінгвістам та установам за спеціальними угодами або ліцензіями з видавництвом Оксфордського університету; громадськість не має дозволу користуватися матеріалами ресурсу.

У такому разі допомагає встановлення партнерських зв'язків з лінгвістичними організаціями для отримання доступу до їхніх корпусних баз, заохочення освітніх установ розробляти та удосконалювати власні спеціалізовані корпуси або ж залучення фінансової підтримки для розширення доступу до корпусів. Такі заходи спрямовані на забезпечення широкого доступу до різноманітних корпусів, що підвищить ефективність їхнього використання під час навчання та проведення досліджень.

Ще одним критичним моментом є **невідповідність навчальним потребам** та рівню користувачів. Інколи текстові корпуси можуть бути складними для адаптації до конкретних навчальних потреб (Corpus of Middle English Prose and Verse (CMP)) або рівню мовних навичок, що може ускладнювати їхнє ефективне використання в освітньому процесі. Вирішення питання доцільно вбачати у створенні або виборі корпусу, який відповідатиме конкретному рівню та потребам користувачів, розробці персоналізованих завдань та вправ, які дозволятимуть здобувачам працювати з корпусом згідно індивідуальних можливостей, організації тренувань та курсів з корпусної лінгвістики, використанні корпусів для створення мовленнєвих ситуацій на основі конкретних завдань та згідно з поставленими навчальними цілями. Ці підходи дозволяють адаптувати використання текстових корпусів до різних рівнів навчання та індивідуальних потреб користувачів.

Необхідно поступово вводити в навчальний план діяльність, засновану на корпусах, дозволяючи учасникам звикнути до нового підходу. Розпочинати слід із простіших завдань із поступовим переходом до складнішого лінгвістичного аналізу. Обов'язковим складником організації роботи у цьому контексті мають стати чіткі вказівки та

інструкції щодо інтерпретації та використання даних корпусу з ціллю допомогти ефективно взаємодіяти з корпусами, не відчуваючи переваження.

Багато існуючих корпусів можуть не відповідати різним лінгвістичним фонам і рівням володіння мовою, що викликає потребу розробки спеціальних ресурсів для задоволення цих унікальних вимог, зокрема створення спеціалізованих колекцій з навчальною метою. Наприклад, Cambridge Learner Corpus та Learner's Corpus of English (LCE). Ці корпуси створені для вивчення англійської мови як іноземної. Такий підхід швидше гарантуватиме звикання до структурованої роботи з корпусними даними, розвиваючи згодом впевненість і компетентність.

Цілком зрозуміло, що обсяги адаптованих ресурсів значно менші, а відповідно постає питання балансу між автентичністю та спрощенням, тоді як справжній надійний корпус повинен відповідати вимогам «репрезентативності, автентичності, відібраності, збалансованості, машиночитаності» (Жуковська, 2013). Дисбаланс у використанні автентичних (реальних) ресурсів та спрощених (адапованих) матеріалів становить загрозу адекватного розуміння та сприйняття результатів мовного аналізу. З одного боку, автентичність корпусних даних може вмщати лінгвістичні особливості, які виходять за межі безпосереднього сприйняття користувачем, що потребує ретельного аналізу та вибору додаткових ресурсів. А з іншого боку, текстові корпуси, доступні для навчальних цілей, можуть бути обмеженими за обсягом та не повністю відображати різноманіття мовного матеріалу чи теми. Тому важливо ретельно підбирати корпуси, щоб надати здобувачам більше специфічного та релевантного матеріалу, за потреби використовувати їх у поєднанні з іншими ресурсами, такими як підручники, статті, відеоматеріали, що допоможе подолати обме-

ження та забезпечити релевантне та ефективне використання інформації.

Визнання та вирішення цих невідкладних аспектів має ключове значення для успішного впровадження корпусної лінгвістики в навчальний процес, а зосередження на практичному застосуванні вказує на безпосереднє значення корпусів для вдосконалення мовних навичок, що загалом націлене на забезпечення консистентності та стабільності у впровадженні корпусо-базованого підходу до навчання.

**Висновки та перспективи подальших розвідок.** Інформаційно-комп'ютерні технології прискорюють процеси отримання, обробки та зберігання лінгвістичного матеріалу. Корпуси – цінний навчальний інструмент, використання якого потребує врахування багатьох аспектів, зокрема й тих, що описані у статті, щоб підвищити мотивацію до навчання за оновленими підходами, сприяючи формуванню та вдосконаленню професійної компетенції в сучасних умовах. Кожна перешкода потребує окремого комплексного підходу до елімінації з урахуванням цілої низки ситуативних моментів.

Очевидно, що питання інтеграції технологій потребує докладного аналізу того, як можна поліпшити доступність та спростити їх використання. Одним з найперспективніших способів врегулювання питання інтеграції корпусної лінгвістики в навчання вважаємо створення та впровадження певних стандартів на національному рівні, які будуть визначати основні принципи та методи використання корпусної лінгвістики з навчальною метою. Допоміжним елементом може стати міжнародна співпраця з університетами та мовними асоціаціями з ціллю обміну досвідом, створенням рекомендацій, які допомагатимуть навчальним закладам впроваджувати корпусні методи на різних етапах навчального процесу збагачуючи можливості професійної підготовки.

#### СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Жуковська В. Вступ до корпусної лінгвістики: навчальний посібник. Житомир: ЖДУ ім. І. Франка. 2013. 142 с.
2. Cheng Winnie. Exploring Corpus Linguistics: Language in Action. New York/London: Routledge. 2012. 241 p.
3. Kübler Sandra, Zinsmeister Heike. Corpus Linguistics and Linguistically Annotated Corpora. London: Bloomsbury Publishing. 2015. 312 p.
4. McEnery Tony, Wilson Andrew. Corpus linguistics: an introduction. Edinburgh: Edinburgh University Press. 2001. 235 p.

#### REFERENCES

1. Zhukovska V. (2013) Vstup do korpusnoi linhvistyky [Introduction to Corpus Linguistics]: navchalnyi posibnuk. – Classbook. Zhytomyr: ZDU im. I. Franka. 142 [in Ukrainian].
2. Cheng W. (2012) Exploring Corpus Linguistics: Language in Action. New York/London: Routledge. 241.
3. Kübler S., Zinsmeister H. (2015) Corpus Linguistics and Linguistically Annotated Corpora. London: Bloomsbury Publishing. 312.
4. McEnery T., Wilson A. (2001) Corpus linguistics: an introduction. Edinburgh: Edinburgh University Press. 235.