

**УДК 371.143:811.111**

***В.В. Жуковська***

*Житомирський державний університет*

*імені Івана Франка*

*канд. філол. наук, доцент*

**Застосування корпусних технологій у навчанні та вивченні  
іноземної мови**

Цифрові технології стали невід'ємною частиною життя кожної людини і міцно проникли у всі галузі суспільного життя. В останні роки відзначається активне використання інформаційних та комунікаційних технологій і в педагогічних цілях, адже сучасні учні – це цифрові аборигени (digital natives) [23], які виростили з цифровими технологіями і не уявляють свого життя без різноманітних гаджетів. У зв'язку з цим у процесі розвитку іншомовної комунікативної компетенції особливої ваги набуває розвиток інформаційної компетенції суб'єктів навчального процесу, що передбачає володіння новими інформаційними технологіями, розуміння сфер їх застосування та уміння користуватися інтернет-ресурсами. Особливе місце у цьому аспекті відводиться технологіям корпусної лінгвістики, адже цей напрям прикладної лінгвістики "може привнести новий зміст у вивчення і навчання іноземній мові та має великий потенціал для того, щоб удосконалити сучасні педагогічні практики" [24: 84]. Лінгводидактичне використання можливостей текстових корпусів вже доволі давно обговорюється у зарубіжній і вітчизняній лінгвістичній та педагогічній літературі (Бублик І.Ф. [1], Бук С.Н. [2], Дідук-Ступ'як Г.І. [4], Колкунова В.В. [7], Нагель О.В. [8], Саєнко Н.С. [9], Сисоєв П.В. [10], О'Кіфі, МакКарті [21] та ін.). Проте, незважаючи на

доволі активне і широке залучення корпусних ресурсів у навчальний процес за кордоном, у нашій країні технології корпусної лінгвістики ще не отримали належної уваги від освітян. **Метою** запропонованої статті є надати систематизований огляд шляхів застосування технологій корпусної лінгвістики у навчанні та вивченні (іноземної) мови.

Корпусна лінгвістика є одним із найбільш перспективних напрямків сучасного теоретичного і прикладного мовознавства. Ця відносно нова лінгвістична галузь розпочала своє активне становлення у 60-х роках ХХ століття у зв'язку із інтенсивним розвитком комп'ютерних технологій. Застосування комп'ютерів та спеціального програмного забезпечення суттєво змінило спосіб дослідження мови та значно полегшило роботу по збору лінгвістичних даних. Без зусиль, лише за декілька секунд, стало можливим здійснювати пошук у багатомільйонних текстових масивах (лінгвістичних корпусах), будувати конкорданс для будь-якого слова, одержувати дані про частоту словоформ, лексем, граматичних категорій, синтаксичних конструкцій, відстежувати зміни у частоті і контексті мовної одиниці у різні хронологічні періоди, одержувати дані про лексичну (колокацію) та граматичну (колігацію) сполучуваність і т. ін. [5: 8-9]. Центральним для корпусної лінгвістики є поняття "*корпус текстів*", під яким розуміється значний за обсягом, представлений в електронному вигляді, уніфікований, структурований, розмічений, філологічно компетентний масив мовних даних, створений для вирішення конкретних лінгвістичних завдань [6: 3]. Мовні корпуси базуються на текстах, відібраних з "реальних" автентичних джерел: книг, газет, журналів, інтерв'ю, промов і т. ін., таким чином подаючи мову у "живому, природньому" середовищі її функціонування.

Сучасне корпусне мовознавство має у своєму дослідницькому арсеналі значну кількість корпусів різних типів і розмірів на різних мовах. Багато з них наявні в мережі Інтернет у вільному доступі. Так, наприклад, найбільш відомими корпусами *англійської мови* є Британський національний корпус (British National Corpus) [26], Лінгвістичний корпус англійської мови (The Bank of English) [33], Американський національний корпус (American National Corpus) [25], Мічиганський корпус академічної англійської мови (Michigan Corpus of Academic Spoken English) [31]; *німецької мови* – Корпус Берлінської Брандербузької академії наук (DWDS-Corpus) [29], Німецька лексика онлайн (Projekt Deutscher Wortschatz) [32], Корпуси інституту Німецької мови (LIMAS-Corpus) [30], Корпус німецької мови (IDS-Corpora) [30]; *французької мови* – Корпус французької мови (Corpus de Référence du Français parlé) [27], Корпус розмовної французької мови (Un corpus d'entretiens spontanés) [34].

Одним із найбільш популярних корпусних ресурсів є [corpus.byu.edu](http://corpus.byu.edu), головним розробником якого є професор Марк Дейвіс (Brigham Young University in Provo, Utah, USA.). Зазначений ресурс містить низку новітніх корпусів на різних мовах та матеріалів, розроблених на їх основі (див. [28]).

На сьогоднішньому етапі свого існування корпуси масштабно застосовуються у навчанні та вивченні іноземної мови. Як зазначає Дж. Флауердью [15: 329], "жоден словник чи граматика не здатні повністю описати всю мову", тому текстовий корпус покликаний заповнити цю прогалину, надаючи у розпорядження користувачів дані, які б вони не змогли отримати з традиційних підручників і довідникових матеріалів. Отож, викладачі та студенти все частіше звертаються до корпусу для здійснення швидкого різноманітного пошуку та аналізу найрізноманітнішої лінгвальної інформації, а

конкорданс визнається зручним інструментом для ефективного опанування мовою.

Виокремлюють три способи використання корпусів у навчанні і вивченні (іноземної) мови: 1) навчальні матеріали на основі даних корпусів, 2) ілюстративні приклади із корпусів, 2) вправи та завдання із використанням даних корпусу (т.б. навчання на основі даних – *data-driven learning*) [11]. Розглянемо детальніше зазначені способи.

Корпуси як джерела емпіричних даних відіграють вагомую роль у підготовці якісних *навчальних матеріалів*. Визнано, що традиційні шкільні та університетські граматики та підручники досить часто проілюстровані штучно створеними чи відредакованими прикладами вживання певної лінгвістичної одиниці чи явища. Такі приклади є малоефективними для успішного навчання учнів використанню мови як засобу комунікації, оскільки вивчене ними може кардинально різнитися від реального функціонування мови. У цьому відношенні дослідження корпусів природньої мови надає більш точний та деталізований опис мови, що у свою чергу відображається у змісті педагогічних матеріалів, граматик та словників, а також дизайні навчальних програм та підручників.

У галузі *навчальної лексикографії* корпусна лінгвістика не лише впровадила нові методи аналізу мовного матеріалу, але і значно розширила межі наукового дослідження. Одним із останніх нововведень у лексикографічній практиці стало розміщення у словниковій статті крім традиційної інформації про значення та семантичні відношення одиниці також й інформації про її сполучуваність та дистрибуцію. Ініціатором першого суто корпусобазованого словника англійської мови COBUILD (1987) був Джон Синклер. Сьогодні ж переважна кількість словників укладається на основі даних лінгвістичних корпусів. Наприклад, практично всі

відомі словники англійської мови (Collins, Webster, Macmillan і т.д.) створюються на матеріалі власних, закритих для загального доступу корпусів, котрі дозволяють зробити словник репрезентативним та відобразити в ньому останні тенденції в мові. Адже вивчення корпусів дозволяє одержувати точні дані про лексичний склад мови, певної підмови, жанру чи індивідуального стилю письменника, виявити останні зміни в лексичному складі мов, різні його варіації (наприклад, поява й зникнення неологізмів) та включити ці дані до словника. Саме сучасність цих словників, побудованих на прикладах із «живих» реальних текстів, сприяє їхньому комерційному успіху.

Вже звичною практикою стало використання корпусів для укладання *довідкової літератури* – граматик, навчальних словників та довідників як для носіїв мови, так і для тих, хто її вивчає. Граматики все більше звертають увагу на лексику, а словники (особливо навчальні) включають у статті словника й граматичну інформацію. На сьогодні вже існує декілька корпусобазованих граматик англійської мови, наприклад, загальних (Collins COBUILD English Grammar (1990); Longman Grammar of Spoken and Written English (1999)) та таких, що акцентують увагу на певних аспектах (Collins COBUILD Grammar Patterns 1: Verbs (1996)).

Таким чином, хоча це і не швидкий процес, проте все більше навчальних матеріалів на сьогоднішній день стають "корпусоінформованими". Зростає тенденція до вкладання коштів у розробку значних за обсягом корпусів такими великими видавцями і розробниками навчальних матеріалів, як Cambridge University Press, Oxford University Press, Pearson-Longman, Collins-COBUILD та Macmillan.

*Навчання на основі даних (data-driven learning)*, при якому корпус стає головним джерелом та способом отримання лінгвістичних знань

й умінь, отримує все більше прихильників серед сучасних студентів та вчителів. Термін "data-driven learning" (DDL) був вперше введений Тімом Джонсом у 1990 році для позначення навчання мові на основі даних текстових корпусів [17]. Цей підхід передбачає використання на заняттях комп'ютерно-генерованих конкордансерів для дослідження закономірностей вживання певних лінгвістичних моделей чи одиниць у реальних текстових прикладах, а також для розробки різноманітних вправ та завдань основі корпусних даних [18: iii]. При цьому підході базоване на корпусі програмне забезпечення використовується для підтримки інтерактивної навчальної діяльності, а залучення "живого" мовлення з корпусу сприяє пробудженню у студентів інтересу до процесу опанування мовою та заохоченню їх до самостійного вивчення автентичного мовного матеріалу. За допомогою різних вправ студенти вчаться спостерігати за даними лінгвістичного корпусу, висувати гіпотези та самостійно формулювати правила вживання тих чи інших одиниць (індуктивний підхід) або перевіряти на даних корпусу валідність правил з навчальних граматики чи підручників (дедуктивний підхід). Через застосування корпусних технологій студенти набувають (або принаймні удосконалюють) такі важливі вміння як вміння прогнозувати, спостерігати, помічати, думати, міркувати, аналізувати, інтерпретувати, рефлексувати, досліджувати, робити висновки (індуктивно чи дедуктивно), фокусуватися, здогадуватися, порівнювати, диференціювати, теоретизувати, висувати гіпотези та перевіряти їх [22: 277].

Виконуючи дослідника мови, студенти стають більш втягнутими у процес вивчення, більш активними і в цілому більш автономними, а вчитель при цьому виступає лише координатором, який опосередковує взаємодію учнів з корпусом, направляє та сприяє навчанню. Ця активна роль суб'єкта навчальної діяльності знайшла

відображення у ключових метафорах напряду, зокрема у спеціалізовані літературі з навчання на основі даних учнів називають "мандрівниками" [12: 22], "дослідниками" [16: 101], а зі слогану Т. Джонса "Кожен студент – це Шерлок Холмс" – "мовними детективами" [16: 101]. А. Боултон [13: 37] зазначає, що DDL ставить студентів у центр навчального процесу, мотивуючи їх приймати підвищену відповідальність за власне опанування матеріалом, а не просто бути пасивними сприймачами правил, що надаються викладачем чи підручником. Зробивши власне лінгвістичне дослідження та прийшовши до самостійних висновків, студент стає більш впевненим та свідомим [20].

У навчанні на основі даних розрізняють два підходи у використанні корпусів: *непрямий* (т.б. вивчення певних мовних явищ на основі прикладів, відібраних з корпусу викладачем чи розробником навчальних матеріалів) та *прямий* (т.б. навчання відбувається із використанням студентами відповідного комп'ютерного забезпечення для самостійного аналізу даних корпусу) [19]. Обидва підходи – це протилежні вісі континууму, між якими розміщуються навчальні ситуації з різним ступенем керованості викладачем. На одному кінці вісі розташовуються процеси з жорстким контролем, коли викладач використовує корпус, щоб визначити, які мовні одиниці чи явища необхідно вивчити на занятті, а потім пропонує для аналізу ретельно перевірені приклади з корпусу у друкованому вигляді, супроводжуючи цю інформацію питаннями, що спрямовують студентів до формулювання потрібних висновків. На протилежному кінці знаходяться ситуації, коли більш підготовлені студенти самостійно працюють з корпусом, щоб знайти відповіді на питання, що їх цікавлять [11].

У своєму дослідженні П. Грабовий надає систематизований виклад можливостей використання корпусних ресурсів у лінгводидактичних цілях [3: 52-53]:

1. *Навчання граматики*: введення граматичного матеріалу, ілюстрація особливостей використання граматичних конструкцій безпосередньо в мовленні; розроблення завдань на корегування помилок, тестових завдань з граматики; вивчення специфіки використання граматичних конструкцій з огляду на соціокультурні характеристики мовців, вікові параметри, регіон проживання тощо; створення практичних завдань та вправ на використання граматичних конструкцій.

2. *Вивчення лексики*: створення студентами власних словників, які ґрунтуються на результатах тематичного пошуку по корпусу зразків як усного, так і писемного мовлення; використання конкордансів слів відповідно до частотності їхнього вживання; виконання ІНДЗ та творчих проектів, які стосуються особливостей функціонування тих чи інших лексичних одиниць; створення вправ націлених на розширення словникового запасу студентів тощо.

3. *Навчання читання та розуміння автентичних текстів*: використання тестів, розміщених в корпусі, для розвитку навичок загального розуміння тесту; визначення реалізації контекстуального значення слова; визначення тематики тексту та ін.

4. *Навчання письма*: використання корпусів та конкордансів як довідкового матеріалу при створенні власного письмового виловлювання; засвоєння шаблонів писемного мовлення; розвиток навичок використання граматичних конструкцій та лексичних одиниць у різних типах та жанрах письма тощо.

У цілому, набір вправ та завдань з використанням даних корпусу є доволі великим й обмежується лише уявою розробника [14: 162].



Зауважимо, що навчання на основі даних має низку переваг та недоліків. Зокрема, найбільшою перевагою цього методу є аналіз "живого" мовлення, що функціонує у "природних" реальних ситуаціях. Користувачі мають змогу звернутися до набагато більшого масиву автентичних прикладів вживання конкретної лінгвістичної одиниці, ніж пропонується у класі, та виявити моделі її граматичної та лексичної сполучуваності, особливості вживання у текстах різних жанрів/регістрів, словотвору і т. ін. Такий спосіб вивчення лінгвістичного матеріалу сприяє розширенню словникового запасу студентів та підвищує рівень опанування ними різними мовленнєвими моделями.

До недоліків прямого застосування корпусу студентами відносять труднощі роботи з корпусним інтерфейсом, відсутність навичок правильного запису пошукових запитів, отримання невідповідних прикладів, неправильна інтерпретація отриманих прикладів. Проте до головних недоліків навчання за допомогою даних слід все-таки віднести відсутність необхідних навичок роботи з комп'ютером та Інтернетом як у викладачів, так і в студентів; неналежні технічні умови у навчальних закладах; низький рівень мовної підготовки студентів; значні затрати часу та зусиль, що витрачаються на розробку та підготовку завдань.

Разом з тим, у підсумку відзначимо, що беззаперечні переваги та очевидна користь навчання на основі даних суттєво переважають вказані недоліки та обмеження у застосуванні корпусів у навчальному процесі з (іноземної) мови. На наше глибоке переконання, сучасні вчителі іноземних мов повинні вміти ефективно поєднувати у своїй професійній діяльності останні здобутки лінгвістичної та педагогічної науки. Саме з цією метою до освітньої програми підготовки магістрів спеціальності 014.02 Середня освіта. Мова і література (англійська) у

навчально-науковому інституті іноземної філології ЖДУ імені Івана Франка кафедрою міжкультурної комунікації та прикладної лінгвістики включено окрему дисципліну "Використання корпусних технологій у навчанні та вивченні іноземних мов". Ця дисципліна покликана сформувати більш цілісне, поглиблене бачення професійної діяльності майбутнього вчителя іноземних мов та розвинути навички використання корпусних технологій на заняттях з іноземної мови.

### **СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ ТА ЛІТЕРАТУРИ**

1. Бублик І.Ф. Застосування даних корпусної лінгвістики до навчання письмовому перекладу / І.Ф.Бублик // Сучасні суспільство в контексті гуманітарного знання. Вчені записки ХГУ "НУА" – 2013. – С. 513-518.

2. Бук С. Н. Учнівські корпуси в методиці викладання іноземної мови / С. Н. Бук // Теорія і практика викладання української мови як іноземної. – Львів, 2007. – Вип. 2. – С. 19-24.

3. Грабовий П. Використання ресурсів корпусної лінгвістики в процесі підготовки фахівця з англійської мови / П. Грабовий // Збірник наукових праць – Частина 1. – 2013. – С.48-54

4. Дідук-Ступ'як Г. І. Лінгводидактичні можливості корпусної лінгвістики / Г. І. Дідук-Ступ'як // Наукові записки ТНПУ ім. В. Гнатюка. – 2010. – № 1. – С. 105-109.

5. Жуковська В.В. Вступ до корпусної лінгвістики: навчальний посібник / В. В. Жуковська – Житомир: Вид-во ЖДУ ім. І. Франка, 2013. – 142 с.

6. Захаров В.П. Корпусная лингвистика: Учебно–метод. пособие. / В.П. Захаров – СПб., 2005. – 48 с.

7. Колкунова В.В. Формування іншомовних навичок студентів-філологів у контексті корпусної лінгвістики / В.В. Колкунова //

Філологічні науки. Наукові записки НДУ ім. М.Гоголя – Книга 2 – 2016. – С. 138-141.

8. Нагель О. В. Корпусная лингвистика и ее использование в компьютеризированном языковом обучении / О. В. Нагель // Язык и культура. – 2008. – № 4. – С. 53-59.

9. Саєнко Н.С. Корпусний підхід у навчанні іноземних мов у технічному університеті / Н.С. Саєнко // Педагогічні науки: теорія, історія, інноваційні технології – №1 (55). – 2016. – С. 142-151.

10. Сысоев П. В. Лингвистический корпус в методике обучения иностранным языкам / П. В. Сысоев // Язык и культура. – №1 (9). – 2010. – С. 99-111.

11. Bennett G. R. Using corpora in the language learning classroom: Corpus linguistics for teachers / G. R. Bennett – Ann Arbor: University of Michigan Press, 2010. – 144 p.

12. Bernardini S. Systematising serendipity: Proposals for concordancing large corpora with language learners / S. Bernardini // Rethinking language pedagogy from a corpus perspective – Frankfurt am Main: Peter Lang, 2001. – P. 225-234.

13. Boulton A. testing the limits of data-driven learning: Language proficiency and training / A. Boulton // ReCALL – Vol. 21(1) – 2009. – P. 37-51.

14. Breyer Y. My Concordancer : tailor-made software for language teachers and learners / Y. Breyer // Corpus Technology and Language Pedagogy – Frankfurt am Main: Peter Lang, 2006 – P. 157-176.

15. Flowerdew J. Corpora in language teaching / J. Flowerdew // The Handbook of Language Teaching – Oxford, UK: Wiley-Blackwell, 2009. – P. 327-350.

16. Johns T. Contexts: The background, development, and trailing of a concordance based CALL program / T. Johns // Teaching and Language corpora – New York: Longman, 1997. – P. 100-115.
17. Johns T. From printout to handout: grammar and vocabulary teaching in the context of data-driven learning / T. Johns T. // CALL Austria – Issue 10. – 1990.– P. 14-34.
18. Johns T., King P. English Language Research Journal Vol.4: Classroom Concordancing / T. Johns, P. King // Birmingham: The University of Birmingham, 1991. – P. 27-45.
19. Leech G. Teaching and language corpora: A convergence / G. Leech // Teaching and language corpora – New York: Addison Wesley Longman, 1997. – P.1-23.
20. Mair C. Empowering non-native speakers: the hidden surplus value of corpora in continental English departments / C. Mair // Teaching and Learning by Doing Corpus Analysis. – Amsterdam & New York: Rodopi, 2002. – P. 119-130.
21. O’Keeffe A. From corpus to classroom: language use and language teaching / A. O’Keeffe, M. McCarthy, R. Carter. – Cambridge: Cambridge University Press, 2007. – 315 p.
22. O’Sullivan I. Enhancing a process-oriented approach to literacy and language learning: the role of corpus consultation literacy / I. O’Sullivan // ReCALL 19(3) – 2007. – P. 269-286.
23. Prensky M. Digital Natives, Digital Immigrants / M. Prensky // On the Horizon — Vol. 9, № 5. – NCB University Press. – 2001. – Режим доступа: <http://www.marcprensky.com/writing/Prensky%20-%20Digital%20Natives,%20Digital%20Immigrants%20-%20Part1.pdf>.
24. Romer U. Corpus research and practice: What help do teachers need and what can we offer? / U. Romer // Corpora and Language Teaching. – Philadelphia: John Benjamins, 2009. – P.83-98.

## СПИСОК КОРПУСІВ

25. American National Corpus – Режим доступу:  
<http://americannationalcorpus.org/>
26. British National Corpus – Режим доступу:  
<http://www.natcorp.ox.ac.uk/>
27. Corpus de Référence du Français parlé – Режим доступу:  
<http://sites.univprovence.fr/delic/corpus/index.htm>
28. corpus.byu.edu – Режим доступу: <https://corpus.byu.edu/>
29. DWDS-Corpus – Режим доступу:  
[http://www.dwds.de/pages/pages\\_textba/dwds\\_textba.htm](http://www.dwds.de/pages/pages_textba/dwds_textba.htm)
30. LIMAS-Korpus – Режим доступу:  
<http://www.korpora.org/Limas/> IDS-Korpora – Режим доступу:  
<http://www.ids-mannheim.de/kt/korpora.html>)
31. Michigan Corpus of Academic Spoken English– Режим доступу: <http://quod.lib.umich.edu/m/micase/>
32. Projekt Deutscher Wortschatz – Режим доступу:  
<http://wortschatz.uni-leipzig.de/>
33. The Bank of English – Режим доступу:  
<http://www.collins.co.uk/Corpus/CorpusSearch.aspx>)
34. Un corpus d'entretiens spontanés – Режим доступу:  
<http://www.llas.ac.uk/resources/mb/80>)