

## УЧНІВСЬКІ КОРПУСИ В МЕТОДИЦІ ВИКЛАДАННЯ ІНОЗЕМНОЇ МОВИ

Соломія Бук

*Львівський національний університет імені Івана Франка*

У статті узагальнено можливості використанні корпусів, зокрема, загальномовного, паралельного, учнівського, у методиці викладання іноземної мови. Увагу зосереджено на окремому типі корпусу текстів — учнівському, тобто побудованому на текстах тих, хто вивчає певну мову як іноземну. Він став надійним джерелом інформації для дослідження як особливостей засвоєння іноземної мови (Second Language Acquisition), так і у викладанні іноземної мови (Foreign Language Teaching). Подано його характеристику, зразки проєктів, переваги та дослідницькі можливості.

*Ключові слова:* корпус текстів, учнівський корпус текстів, конкорданс, анотування.

Потреба міжнародного спілкування і, отже, вивчення та викладання іноземної мови (ІМ) – це ті питання, що залишаються актуальними ще протягом довгого часу. Нові перспективи у цій галузі відкрило застосування комп'ютерних технологій, зокрема, корпусів текстів.

Корпус текстів (КТ) – це „машиночитане стандартно організоване зібрання репрезентативних для певної мови, діалекту або іншої підмножини мови писемних або усних текстів, призначених для лінгвістичного аналізу й опису, відібраних і впорядкованих згідно з експліцитними екстра- та інтралінгвальними критеріями” [3: 53]. Надалі такі ознаки корпусу, як існування в електронному (комп'ютерному) вигляді, репрезентативність чи достатньо великий обсяг та анотованість, уважаємо як самі собою зрозумілими, обов'язковими. Отже, інформація, одержана на матеріалі корпусів текстів, – надійні дані для тверджень про мову.

У науковій літературі [1; 4; 5; 17; 20] за різними критеріями виділяють такі типи корпусів: загальномовні та спеціалізовані (наприклад, територіальних чи соціальних діалектів, учнівського мовлення тощо); повних текстів чи їх фрагментів; дослідницькі та ілюстративні; письмового, усного мовлення та мішані; одномовні та багатомовні; синхронічні та діахронічні; статичні та динамічні; збалансовані та моніторингові; паралельні та порівняльні; малі, середні та великі.

Корпуси текстів, обсяг яких становить мільйони слововживань, стали зручним та надійним інструментом дослідження у багатьох (якщо не всіх) галузях мовознавства: лексикографії, семантиці, морфології, статистичній і когнітивній лінгвістиці, перекладознавстві тощо. Не стала тут винятком і методика викладання ІМ. Є декілька можливостей використання КТ у методиці.

1) Використання загальномовного (національного) КТ [8; 10; 12–16; 21], який дасть можливість швидкого різнорівневого пошуку найрізноманітнішої лінгвальної інформації, важливої для методики викладання ІМ: найчастотнішу лексику, сполучуваність слів, уживання прийменників, конкорданси, в яких легко визначити значення слова, дієслівне керування тощо. Конкорданс як спосіб презентації текстової інформації, при якому до кожного реєстрового слова (чи словоформи) подано усі або вибіркові контексти його вживання, визнано чи не найзручнішим інструментом для виявлення лексичного та граматичного оточення лексеми, можливості її сполучуваності та, загалом, повної її синтагматичної характеристики. „Сьогодні видається, що нові словники укладатимуть на основі конкордансів” [19: 169] (пор. [2]).

2) Створення паралельних корпусів (*parallel corpus*). Паралельний корпус — це „збір текстів та їх перекладів іншими мовами” [22], тому складається із як мінімум двох підкорпусів, один з яких містить текст мовою оригіналу, а інший — це детальний його переклад іншою мовою. Як правило, паралельний корпус — двомовний, суттєвою його ознакою є внутрішні зв'язки між сегментами цих різномовних текстів. Записані мовою програмування у відповідному форматі, вони уможливають доступ до безпосередніх перекладних еквівалентів слова, фрази, синтаксичної конструкції. Такий інструмент не тільки полегшує створення перекладних словників, базованих на реальних прикладах перекладацької практики, а й сприяє легшому вивченню ІМ, особливо відповідно до комунікативної стратегії навчання. Застосовуючи двомовні конкорданційні програми, комп'ютерна система висвітлює обидва тексти поряд, і студент легко та швидко перекладає текст, оскільки не змушений шукати кожне невідоме слово у словнику. Також у режимі конкордансу максимально спрощений процес пошуку типу перекладного еквівалента, залежного від контексту.

3) Створення учнівських корпусів (*learner corpus*). Учнівський корпус — це „збір мовних даних, який походить від осіб, які вивчають мову як іноземну” [17]. Тут термін *англ. learner, (пол. uczniowski)* перекладаємо лексемою „учнівський”, похідною від іменника *учень* зі значенням „той, що вчиться, вивчає щось”, тобто під цим прикметником слід розуміти людину безвідносно до віку, яка навчається. Учнівський корпус текстів дає матеріал для глибокого аналізу мовленнєвих даних учня, не лише якісного, а й кількісного; надто частого вживання слова або, навпаки, його уникання.

Тому анотування учнівського корпусу має свою специфіку. Так, залежно від дослідницьких цілей створення, у його зовнішній розмітці, окрім іншого, закодовують дані про учня (вік; стать; рідна мова; іноземні мови, якими володіє; освіта; рівень володіння цією іноземною мовою / рік її вивчення тощо), про умови виконання завдання (наприклад, доступ до словників), про форму відповіді (усна, письмова), довжину, жанр, тему відповіді-розповіді і т. д. [11]. Особливості внутрішнього анотування полягають у тому, що його, на відміну від корпусів літературної мови, виконують вручну, а допущеним помилкам приписують спеціальні індикатори. Останні дають можливість виявляти і класифікувати найчастотніші помилки залежно від рідної мови учня, від його рівня володіння ІМ, від освіти тощо.

Учнівські корпуси текстів широко використовують у дослідженнях як особливостей засвоєння іноземної мови (Second Language Acquisition), так і у викладанні іноземної мови (Foreign Language Teaching). Спільними питаннями дослідження цих лінгвістичних дисциплін можна вважати такі, як різниця у засвоєнні першої (рідної) та другої (іноземної) мов, універсальні процеси у засвоєнні іноземної мови, роль першої мови у вивченні іноземної, мовну нестабільність (variability), індивідуальні особливості, стратегії навчання тощо. Даними для опрацювання вказаного кола проблем слугують мовлення учнів (вільне та провоковане), спостереження, рапорти учнів, інтроспекція, експерименти. Вивчення мовлення учнів має певну традицію, так, наприклад, особливістю аналізу учнівських помилок у часі 1960–70-х років є те, що він базувався на невеликих сукупностях текстів (їхня величина часто не була зазначена), які не мали електронного формату, здебільшого неоднорідних (що часто провокувало неоднозначні, а то й суперечливі висновки), з них виписували тільки певні помилки, а решту матеріалу не брали до розгляду. У порівнянні учнівські корпуси текстів мають низку суттєвих переваг: значно більший обсяг матеріалу, електронний формат, точні критерії збору матеріалу дають можливість не тільки фіксувати помилки, а й сформулювати враження про цілісне мовлення.

На сьогодні найбільшим проектом учнівського корпусу є Міжнародний корпус учнівської англійської – International Corpus of Learner English (ICLE), що нараховує підкорпуси англійського мовлення учнів, які є носіями рідної болгарської, бразильської (варіанта португальської), грецької, данської, китайської, італійської, іспанської, литовської, німецької, норвезької, польської, португальської, російської, турецької, фінської, французької, чеської, шведської мов [9]. Про все більшу актуальність використання цих ресурсів свідчить присвята окремих секцій на міжнародних конференціях „вивченню й викладанню мови та учнівським корпусам текстів” (Language Learning and Teaching and Learner Corpora) [7].

Прикладом поєднання двох різновидів корпусів текстів – паралельного та учнівського – є один із засобів проекту PELCRA ([www.uni.lodz.pl/pelcra/](http://www.uni.lodz.pl/pelcra/); <http://korpus.ia.uni.lodz.pl>). В Інституті англійської філології Лодзького університету з метою підвищити рівень об'єктивності в оцінюванні перекладів, виконаних студентами-перекладазнавцями старших курсів, частину праць зберігають в електронній формі, пов'язаній відповідним чином з оригіналом. Таким чином зібрано кілька десятків перекладів одного і того самого тексту, чиїми авторами є особи, що вивчають цю мову як іноземну. Відповідне опрацювання таких ресурсів дає можливість на підставі великої кількості прикладів виявити різні варіанти і стратегії перекладу різних осіб, а також типові помилки та неточності [17: 47–48].

Перед лінгводидактикою в Україні стоять принаймні два завдання: приєднання до міжнародних проектів учнівських корпусів та створення учнівського корпусу текстів власної мови.

1. Баранов А. Н. Введение в прикладную лингвистику: Учеб. пособие. – М., 2001.
2. Бук С., Ровенчак А. Онлайн конкорданс роману Івана Франка „Перехресні стежки”. – Доступно з: <http://www.ktf.franko.lviv.ua/~andrij/science/Franko/concordance.html>.
3. Демська-Кульчицька О. Основи національного корпусу української мови. – К., 2005.
4. Карпіловська Є. А. Вступ до комп'ютерної лінгвістики. – Донецьк, 2003.
5. Корпусна лінгвістика /В.А. Широков, О.В. Бугаков, Т.О. Грязнухіна та ін. – К., 2005.
6. Национальный корпус русского языка.— Доступен из: <http://www.ruscorpora.ru>.
7. 16th International Conference on Pragmatics and Language Learning, Indiana University – Bloomington, April 14–16, 2005: <http://www.indiana.edu/~pll2005/Workshop.htm>; Seventh International Conference on Teaching and Language Corpora: 1–4 July 2006 <http://tale7.eila.jussieu.fr/welcome.en.shtml>; 12th International CALL Research Conference <http://webh01.ua.ac.be/didascalia/welcome.htm>.
8. British National Corpus. – Available from: <http://www.natcorp.ox.ac.uk>.
9. Centre for English Corpus Linguistics – CECL – Available from: <http://www.fltr.ucl.ac.be/fltr/germ/etan/cecl/Cecl-Projects/Icle/icle.htm>.
10. Český národní korpus. – Available from: <http://ucnk.ff.cuni.cz/>
11. Granger S. Computer learner corpus research: current status and future prospects // Applied Corpus Linguistics: A Multidimensional Perspective / Ed. by U. Connor and T. Upton. – Amsterdam; Atlanta, 2004. – P. 123–145.
12. Hellenic National Corpus (HNC): The ILSP Corpus. – Available from: <http://hnc.ilsp.gr/en/>.
13. Hrvatski nacionalni korpus. – Available from: <http://www.hnk.ffzg.hr>.
14. Hungarian National Corpus. – Available from: [http://corpus.nyud.hu/mnsz/index\\_eng.html](http://corpus.nyud.hu/mnsz/index_eng.html).
15. Korpus Języka Polskiego Wydawnictwa Naukowego PWN. – Warszawa, 2003. – Dostępny z: <http://korpus.pwn.pl>.
16. METU Turkish Corpus. – Available from: <http://www.ii.metu.edu.tr/~corpus/corpus.html>
17. Podstawy językoznawstwa korpusowego / Red. B. Lewandowska-Tomaszczyk.— Łódź, 2005.
18. Polski korpus narodowy. – Dostępny z: <http://www.lodz.pl/pelcra/corpora-pl.htm>.
19. Sinclair J. Corpus. Concordance. Collocation. – Oxford, 1991.
20. Sinclair J. Corpus typology: A Framework for classification // Studies in anglistics. – Stockholm, 1995. – P. 17–33.
21. Slovenský národný korpus/– Available from: <http://korpus.juls.savba.sk>.
22. Teubert W. Comparable or Parallel Corpora? // International Journal of Lexicography. – V. 9. – 1996. – N 3. – P. 245.

## **LEARNER CORPORA IN THE TECHNIQUES OF TEACHING A FOREIGN LANGUAGE**

**Solomija Buk**

*Ivan Franko Lviv National University*

The possibility of corpus usage in the foreign language teaching is generalized in the article. General, parallel, and learner corpora, in particular, are considered for the teaching purposes. A special attention is paid to a specific text corpus type, namely, learner corpus created on the basis of the texts for teaching the

---

language as a foreign one. Such a corpus became a trusted information source for the study both of the Second Language Acquisition and Foreign Language Teaching peculiarities. There are characteristics, project samples, advantages and research possibilities of the learner corpus given here.

*Key words:* text corpus, learner corpus, concordance, annotation.