УДК 37:004:81'25
Yemelianova O., Yehorova O.
Sumy State University, Sumy, Ukraine
ORCID ID: 0000-0002-3277-1227
ORCID ID: 0000-0002-3225-5580

CORPUS TOOLS IN LANGUAGE EDUCATION: TRENDS, PRACTICES, AND PEDAGOGICAL INSIGHTS

DOI 10.14308/ite000818

Abstract. *The article is devoted to the study of the corpus approach effectiveness in the process of teaching foreign languages and translation. The aim of the article is to investigate the functionality of the corpus approach in translation-oriented teaching of foreign languages by assessing its impact on students' vocabulary acquisition, written competence, and students' engagement; to identify key conditions for the productive inclusion of the corpus approach in modern practices of teaching foreign languages and translation. The authors emphasize the need to analyze the potential of corpus tools to ensure linguistic authenticity, improve learning through the use of corpus data, personalize and contextualize learning, and promote the development of student autonomy. The authors demonstrate the effectiveness of the corpus approach in teaching foreign languages and translation, especially in vocabulary acquisition, which is very important for future translators, and suggest practical examples of working with COCA to study word combinations and word formation. The students' feedback reveals their enthusiasm and positive attitude towards the use of corpora in training. The application of RhymeZone for the development of creative skills and NetSpeak for the study of lexical-grammatical models is also considered. The authors analyze the advantages of the corpus approach, including scientific validity, objectivity, development of autonomy, increased motivation, differentiated and personalized learning, as well as preparation for real speech interaction. The article highlights certain limitations of corpus-based language pedagogy, which include technological barriers, methodological complexity and certain challenges for teachers, ethical and legal concerns, as well as issues related to data interpretation. Prospects for further research lie in a thorough study of the potential of the corpus approach in teaching translation and linguistics students the genre specificity of the analyzed texts, their lexical and grammatical features, and methods of accurate translation from the original language to the target language.*

Keywords: *corpus-oriented language pedagogy, corpus approach, differentiated and personalized learning.*

The problem statement. One of the critical issues of modern language education is the obvious divergence between corpus-based approaches potential and practical implementation of corpus tools in foreign language acquisition and translation training. It is necessary to analyse how the mighty capacity of corpus tools can be meaningfully utilized to guarantee linguistic authenticity in the process of teaching and learning, enhance data-driven learning, and promote learner autonomy.

Analysis of recent research and publications. In the era of rapid development of information technology and digitalization of all spheres of human activity – and that's where we are now – language as an object of study is undergoing profound transformations. Linguistics is apt in attuning to these changes by actively introducing advanced methods of language analysis based on large arrays of real text data. One of the most promising areas in this field is corpus linguistics, which in recent years has merged thoroughly with language pedagogy, becoming an integral part of theoretical research and applied disciplines. Corpus-based language pedagogy (CBLP) is the language teaching approach that uses linguistic corpora, which are the systematized collections of texts or spoken language that have been digitized and annotated for computer processing, as the major source of teaching material. This approach allows us to go beyond the unnatural patterns that

© The Author(s) 2026

are quite often typical of traditional textbooks, especially written by non-native authors, and provide quality learning based on genuine, authentic language data. The role of corpus instruments in language teaching has undergone significant changes recently. The growing number of scientific research and publications in this domain witnesses these transformations. We observe the shift from the simple use of corpora as a data storage to more integrated and dynamic approaches where corpus tools become a vital part of foreign language teaching and learning processes. One of the key areas of active exploration is the development and use of specialized educational corpora. In contrast to general corpora, which may be too extensive and complicated for novice students, educational corpora are created with due regard for learners' specific needs. Studies show the efficiency of corpus use that is focused on certain lexical and grammatical phenomena, that help students stay concentrated on relevant material [1], [2]. Integration of corpus tools into the curricula and teaching methods are also actively researched; that is done in parallel with corpus development. We observe a noticeable shift from occasional use to systematic implementation of corpus approaches in teaching foreign languages. M. Callies, for example, provides practical recommendations for pre- and in-service teachers on how to implement corpus linguistics into the day-to-day teaching practice [3]. The study focuses on "corpus-browsing activities with higher learner autonomy…in which the pupils retrieve and analyze corpus data themselves to find out typical collocations of adjectives and adverbs" [3, p. 260]. That promotes language awareness and facilitates the development of autonomous learning skills. The scholars D. Li et al., while performing a systematic review of corpus-based instruction in EFL classroom learning, note that "uneven focus on specific proficiency levels restricts the broader application of CBI (Corpus-based instruction)" [4, p. 11]. Another important direction of research is the application of corpus tools in assessing language competence. Very often, traditional methods of assessment can't comprise all the nuances of language usage. On the contrary, corpus approaches give the opportunity to analyze a wide range of linguistic phenomena in an authentic context. The study of M. V. Kusumaningrum and P. Ardi proves that corpus metrics can be applied to assess the complexity and diversity of vocabulary in students' written work; it demonstrates that "lexical sophistication is an important component of academic writing proficiency" and is "one of the keys to measure academic success" [5, p. 54–55]. Technological innovations play a crucial role in corpus tools development. Corpus technologies became more accessible to a wide range of teachers and students who don't have profound knowledge of programming due to the availability of cloud computing and the development of user-friendly interfaces. In this context, works devoted to the development of intuitive platforms and applications, such as the platform CorpusMate introduced by P. Crosthwaite and V. Baisa [6], are of exceptional importance. We observe a growing interest in interdisciplinary approaches, where corpus linguistics is combined with other areas of research such as artificial intelligence, machine learning, and psycholinguistics. Researchers suggest using genetic algorithms, the "approach has the potential to be implemented in e-learning environments, intelligent virtual tutoring systems, and adaptive testing platforms, where the continuous improvement of teaching methods and student engagement are essential" [7]. While corpus linguistics as a scientific field is thought to have emerged in the mid-20th century, the ideas of systematically collecting and analyzing texts have much older roots. The computer era made it possible to process vast volumes of text information automatically. The last quarter of the 20th century witnessed a true corpus boom. As scholars point out, "Since the 1980s, corpora have been widely used in all areas of language studies and have greatly advanced our understanding of language use and patterns" [8]. At this period appeared such large-scale projects as the British National Corpus (BNC) [9], Corpus of Contemporary American English (COCA) [10], as well as multilingual and parallel corpora. Further advancement of digital media and Internet technologies have led to an exponential growth of accessible textual data, supporting the development of cutting-edge corpora that are dynamic, regularly updated, and thematically specialized. This trend has brought corpus-based language teaching to the forefront. Data-Driven Learning (DDL) gained exceptional recognition and support among educators who promote the idea that students should "discover" linguistic patterns themselves through work with concordances (i.e. examples of word usage in context), rather than

© The Author(s) 2026

merely memorize rules from a textbook [11], [12]. Identification of previously unresolved aspects of the general problem. Despite the number of works in corpus-based language pedagogy (CBLP), a thorough analysis of the corpus-based approach to teaching translation and philology students has not been carried out yet. That constitutes the relevance of the research. Formulation of the article's goals. The purpose of the article is to investigate the effectiveness of the corpus-based approach in translation-focused language instruction by estimating its influence on students' vocabulary acquisition, writing competency, and learners' engagement; to identify major conditions for its productive blending within modern language teaching frameworks. Presentation of the main research material. Corpus-based language pedagogy relies on linguistic corpora that are electronic collections of texts (or recordings of spoken language) assembled according to certain principles for linguistic analysis. Modern corpora can encompass millions, and sometimes billions, of words. Corpora can be classified according to different criteria, including the following: 1) language: corpora may be monolingual (e.g., the British National Corpus (BNC) [9], multilingual (e.g., UNITEX [13], or parallel (e.g., Open Parallel Corpora [14]); 2) type of speech: corpora may consist of written texts (e.g., literary texts, newspapers), spoken data (e.g., interviews, lectures, conversations), or be a combination of both (mixed corpora); 3) topic: corpora may be general-purpose or specialized (e.g., medical, legal, academic corpora); 4) time of creation: corpora can be classified as historical, modern, or dynamic (i.e., regularly updated). The quality of any corpus is assessed by its representativity, i.e. by its capacity to reflect authentic language in a varied social, functional, and stylistic contexts. To make a corpus suitable for a linguistic analysis it typically undergoes an annotation process, during which some metadata and linguistic tagging are applied. The most widely used types of annotation comprise: 1) morphological or part-of-speech (POS) tagging – assigning grammatical categories to each word; 2) syntactic parsing – identifying the structure of the sentence and breaking it down into its constituent parts; 3) semantic annotation – labeling words or phrases with semantic categories, ontologies, and semantic roles; 4) discourse annotation – indicating rhetorical structures and connections between discourse units. Corpus linguistics operates a range of specialized analytical tools. The central one to corpus-based analysis is the concordance which is a list of occurrences of a certain word or phrase within the context. As a rule, it is 5 words to the right, and 5 words to the left. Concordance lines allow researchers and language learners to explore authentic usage, lexico-grammatical combinability, synonyms, collocations, and stylistic coloring. The concordance underpins several key corpus-based concepts: 1) collocations – stable, frequently co-occurring word combinations. Collocation analysis helps future translators and linguists master their foreign language use by adopting natural, idiomatic constructions; 2) N-grams – sequences of N words (e.g., bigrams, trigrams) that occur together with high frequency. For instance, to sum up is a common trigram in academic English; 3) frequency – the number of occurrences of a word or construction in a corpus. Curriculum design relies heavily on frequency lists. For example, the 2,000 most frequent words of the specialized corpus may form the foundation of vocabulary lists to be mastered by students. Such reputable corpus management software programs as AntConc [15], Sketch Engine [16], WordSmith Tools [17], and LancsBox [18] provide user-friendly interfaces for performing search, filtering, and visualizing corpus data. The Corpus-based Language Pedagogy (CBLP) rests on a set of principles that differentiate it from traditional approaches to language teaching. 1) Authenticity of the language material is the major advantage of CBLP. Students work with the texts produced by the native speakers in real-life settings. It is crucial for exploring stylistic variations of the language, idiomatic expressions, jargon, etc.. As G. Gilquin notes, "Corpora of native or expert language can help decide what to focus on or not in teaching, for example by showing what is frequent in language, what is distinctive of a certain genre, etc. They can also provide detailed and nuanced descriptions of any aspect of language or of specific registers or genres (e.g. academic writing)" [19]. 2) Data-Driven Learning (DDL) encourages students to analyse corpus data and discover linguistic patterns and infer rules autonomously. As Schmidt explains, DDL "is a method for using corpora by language learners via direct classroom instruction" [20]. For instance, a student is supposed to analyse a certain amount of examples with the definite article the and the indefinite article a/an in the concordance lines and

identify patterns by themselves, instead of just mechanically learn the use of articles according to the information in the course book. This approach develops metalinguistic awareness, critical thinking and autonomy in learning. 3) Focus on usage rather than norm. CBLP demonstrates what people say (i.e., how language is actually used), but not how they should say (i.e., how it ought to be used according to normative grammar). It is of vital importance in the context of contemporary language variation, where norms are shaped by practice and discourse rather than by academic standards. 4) Personalization and contextuality are exceptionally important for CBLP. Corpora allow educators to adapt learning to particular students' needs, that is, to provide specialized professional collection of text for thorough examination and analysis (e.g., academic writing, legal texts, medical communication). Furthermore, learner corpora, i.e. collections of students' own written or spoken output, can be used to track their learning progress, identify recurring mistakes, and provide students with personalized feedback. Corpus-based approach in teaching foreign languages and translation proves to be highly efficient. One of the most successful areas of corpora use is vocabulary learning [21, p. 344–347]. The application of the corpus approach vividly demonstrates that word meaning is determined by its context and collocations – in contrast to traditional vocabulary acquisition methods, where students learn isolated words. The observation of practical implementation of the Corpus of Contemporary American English [10] (COCA) during 2023–2025 academic years for teaching translation and linguistic students word families and collocations revealed significant learning gains. Students were thoroughly instructed on how to use each feature offered by COCA. The feature COLLOCATES helps students find natural word pairings and by doing so get a better understanding of natural word usage. (see Figure 1)

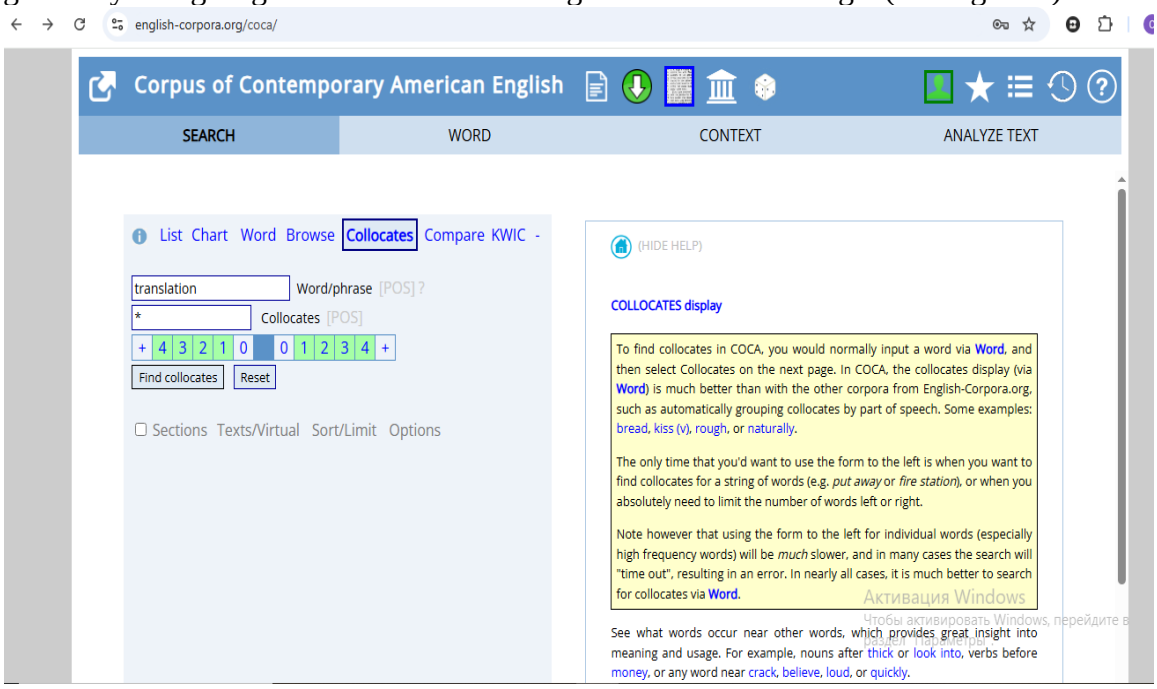

Fig. 1. COCA COLLOCATES feature.

By interpreting the result, students learn various types of collocates, such as noun + noun (N+N), noun + adjective (N+Adj), noun + verb (N+V), and noun + adverb (N+Adv.) with a desired word. The system provides examples of the authentic context where a certain lexeme or collocations are used. Additionally, COCA includes links to reliable external resources that encourage students to look deeper into linguistic phenomena. This fosters critical thinking and promotes creative approaches to language learning. (See Figure 2.)

Fig. 2. COCA COLLOCATES feature extended.

We, as language instructors, consistently highlight the idea that words should be learned in word families ensuring a wide range of students' vocabulary.The BROWSE feature offered by COCA makes the process of word families formation engaging and fun. (See Figure 3.)

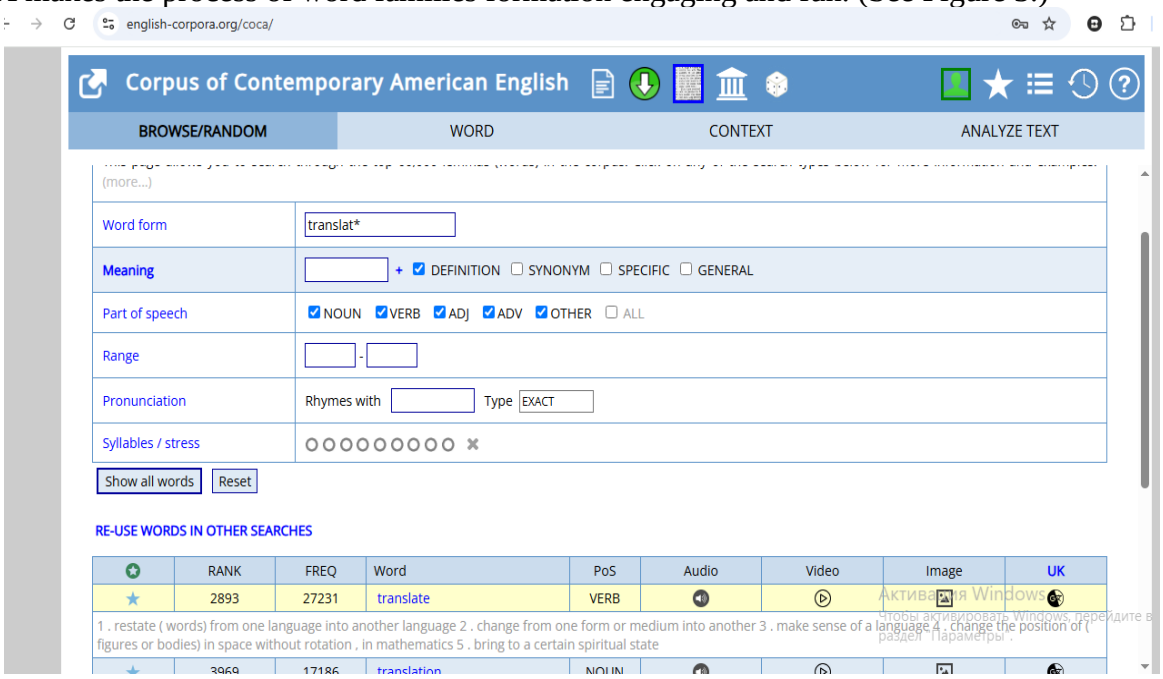

Fig. 3. COCA BROWSE feature

Students are taught to use the feature BROWSE to find the words with different prefixes and suffixes that belong to the same word family. Along with enhancing vocabulary skills, this activity improves their word formation skills, enriching their knowledge and preparing for the future job as a translator. Feedback from students indicates a highly positive reception of corpus-based learning. The vast majority of students (97%) felt positive about using corpora for learning foreign languages and translation. Many reported that it was an absolutely new experience for them and that before that course, they had never used corpora as a language learning tool. Their sincere responses reflect genuine enthusiasm about CBLP. One student noted: "I am very pleased with the fact that during the university classes we were taught not only theoretical knowledge, but also practical tools that

help us learn languages or work with translations. One such tool is the English-Corpora. It's amazing! The English-Corpora broadens the way how I can thoroughly work on the language, making me think and learn" (Dariia, student). The experimental teaching at the translation department during 2023–2025 as to enlarging students' vocabulary via corpora application and an engaging approach demonstrates sustainable results. Students responded well to creative yet intellectually challenging tasks. Recognizing that rhymed words are generally easier to memorize, we incorporated the use of the RhymeZone corpus [22] in our teaching practice to stimulate students' active involvement into the process of knowledge acquisition and enjoyment from the learning process. One illustrative practice involved a warming-up activity at the beginning of the lesson, where students were asked to compose short poems on a specific topic or using a keyword. Having a limited time slot (usually 5–7 minutes), students worked individually or in groups to produce poems. The session concluded with a secret ballot as to whose poem takes the 1 st , 2d, 3d place, giving the activity a particular competitive streak. For example, here are some poems the groups of first-year translation students came up with (using the corpus RhymeZone) at the open demonstrative practical class titled "Translator's Trade" conducted as part of the course on the discipline "The Practical Course of the English Language" on April 10, 2025. Between two worlds that translator stands, He changes words with his own steady hands. He waves the thoughts from tongue to tongue – A bridge of meaning softly spun. (Group 1 poem) The translator's skill – a need for all, Unites the worlds, standing tall. From heart to heart, and word to word, Translation's a wonder, truly heard. (Group 2 poem) Practice shows that it is most effective to introduce creative rhyming activities in group settings first, as some students who have never tried rhyming before, may experience discomfort or hesitation. However, collaborative group work often fosters a supportive environment, so many students – quite often and to their own surprise – discover an ardent inner poet, who was just waiting for the right moment to be called and recognized. Corpora are of exceptional assistance in helping students identify lexical-grammatical patterns, improving grammatical competence, and exploring controversial cases of a certain lexeme or structure usage. Unlike theoretical textbook explanations, corpus data give a well-grounded insights based on actual usage. Among translation and linguistics students at __ University, NetSpeak [23] corpus has gained particular recognition. It offers clear and reliable examples in both English and German, which is especially relevant for students enrolled in the educational program 035/B12 "Germanic Languages and Literatures (Translation included)"; the tool is user-friendly and easy to navigate. When invited to provide feedback, the students expressed strong enthusiasm. For example, some of them noted: "Netspeak is an app that is useful and definitely a must-have for language learners. It is especially useful for translators, who need to know 100% how to use words and in what order to put them. We are all human and can get confused, so this site will be a lifesaver in such situations" (Valeriia, student); "I really liked this resource and would happily recommend it to others. It is exactly what you need to learn a language in everyday use. With the help of Netspeak, you don't need to be afraid of learning less familiar phrases and you can be sure that they actually exist and are used" (Margaryta, student). Corpora can help develop and enhance students' writing skills [24]. While writing essays, scientific articles, or engaging in everyday or business correspondence, students make many mistakes. Corpora allow students to compare their own texts with authentic ones for reference. For instance, Michigan Corpus of Upper-Level Student Papers (MICUSP) [25] provides access to a wide range of academic writing samples, including argumentative essays, creative writing, critiques/evaluations, proposals, reports, research papers, and response papers. It is possible to choose disciplines, student levels, nativeness, textual features. The Academic Word Suggestion Machine (AWSUM) [26] is an indispensible tool for teaching students authentic language structures characteristic of different parts of a research paper, such as: abstract, introduction, method, results, discussion, conclusion. Given that translation students are encouraged to engage in scientific and research work from their first year, familiarizing them with these tools is essential for building advanced writing competence and academic literacy. The analysis shows that CBLP offers a range of significant benefits for foreign language learners, particularly in higher education contexts. Among its key advantages are: 1) scientific validity and objectivity, as the

© The Author(s) 2026

corpus-based approach is empirically foregrounded. Students are taught to rely on the analysis of authentic language data rather than on subjective preferences of teachers or instructors. Students stop questioning the educational material offered for analysis, which they often did while studying at the secondary school (as they mention that themselves); that, in turn, reduces the risk of transmitting outdated or erroneous ideas; 2) development of learner autonomy and metacognition. Working with corpora promotes independent learning, where students formulate questions, test hypotheses, and interpret results based on corpus evidence. This helps prepare autonomous learners ready for lifelong education. Moreover, language through data analysis fosters metacognition, i.e. the awareness of one's cognitive processes, so that students become more conscious of how they learn, not just what they learn; 3) increased learner motivation. The interactive nature of corpus tools offers, elements of discovery and play, opportunities to "decipher" the language as a riddle, and guarantees that the learning process becomes more interesting and exciting; 4) differentiated and personalized learning, since corpus tools allow instructors to tailor content to students' specific needs, proficiency levels, and academic disciplines; 5) enhanced learner readiness for real-world language use. Practice shows that corpus-trained students are better at navigating authentic texts, more adept at recognizing idioms, set combinations, collocations, and stylistic devices. They are less prone to use outdated expressions and give preference to natural, contextually appropriate constructions. Despite obvious advantages, CBLP faces several serious limitations, among which are the following: 1) technological barriers. Effective corpus use often depends on stable Internet and adequate hardware resources. Additionally, some tools like Sketch Engine [16] or AntConc [15] require a certain level of digital literacy. Moreover, some interfaces can be quite challenging for learners, so some students may require careful scaffolding and step-by-step instruction on how to use features offered by a certain tool effectively and with confidence. It is very true of the English Corpora [27], which offers an extensive range of analytical tools. Without instruction, students may feel overwhelmed; however, with proper guidance on the principles how the tool operates and what amazing tricks it can perform for the English language learner, students unanimously claim that this is the right place to go to hone their linguistic skills; 2) methodological complexity for teachers. Implementing corpora in teaching practice requires proper training in corpus linguistics. Many teachers express concerns regarding the time spent on designing corpus-based tasks; furthermore, they can experience fear of losing control over the teaching process, when unpredictable student inquiries arise from the corpus data; 3) ethical and legal issues connected with copyright, data privacy, and data bias (some corpora may overrepresent certain dialects, sociolects, or demographic groups, which distorts the overall language perception; 4) challenges of linguistic data interpretation. For students without prior experience, analyzing concordances or frequencies may be rather confusing and discouraging. Without proper guidance, false generalizations and misinterpretations may occur; 5) limitations of spoken corpora. The largest and most widely used corpora are written with spoken corpora very limited in scope. Oral speech is still poorly represented, whereas tools for automatic transcription and annotation are still developing. This restricts corpus-based approaches in teaching pronunciation, intonation, and conversational techniques. Given the rapid advancement of information technologies nowadays, the future of CBLP looks promising. The integration of corpora with artificial intelligence is already underway. As noted by Mark Davies, "English-Corpora.org will soon offer something entirely new: the ability to combine the depth and reliability of corpus data with the analytic power of Large Language Models (LLMs) like GPT, Gemini, Claude, Perplexity, Llama, Mistral, and DeepSeek" [28]. Artificial intelligence and machine learning open up new horizons for automatic generation of exercises based on corpus data; personalized recommendations on vocabulary and grammar, and intelligent error analysis with corpus-based correction suggestions. We expect to witness the expansion of corpus types, for example, the emergence of real-time dynamic corpora based on platforms like Twitter, Telegram, and YouTube or multimodal corpora that blend text, audio, video, and body language data for comprehensive discourse analysis. Another critical prospect is the creation of corpora of small and endangered languages that would support their resilience. Moreover, we are very optimistic about the continued growth of open-access resources like Corpus

of Contemporary American English (COCA) [10] or Sketch Engine For Language Learning (SKELL) [16]. This will reduce barriers for under-resourced regions and promote equity in educational access. It is notable that in some countries (e.g., the UK, Germany, China) corpus methods are already embedded in national language education. In this context, teacher training becomes essential, prompting the development of relevant modules or courses within professional development programs. Conclusions. Corpus-based language pedagogy emerged in response to criticism of traditional methods that often relied on the teacher's intuition or artificially constructed examples and did not always accurately reflect authentic language use. The corpus-based approach offers an alternative – teaching based on factual, objective, and verifiable data. Corpus pedagogy transforms students from passive recipients of knowledge into active researchers, engaging in analysis, comparison, and hypothesis formation. This aligns with the contemporary educational paradigm that emphasizes student-centered problem-oriented, and inquiry-based construction of knowledge. Corpus-based language pedagogy is a powerful approach that, with the proper introduction and instruction, can qualitatively change the process of language learning and professional training. Its strength relies on scientific validity, material authenticity, student autonomy, and critical thinking development. However, it is important to remember that corpora are educational assistants, and do not substitute a teacher in the classroom (both traditional and virtual). Educators should keep abreast with the data at hand, latest technological innovations, and consider linguistic intuition, human factor influence, and the correlation between lexemes frequency use and their cultural value. Prospects for further research. Further research should focus on exploring the potential of CBLP in teaching future linguists and translators to work with genre-specific texts, their lexical, grammatical, and pragmatic dimensions. Particular attention should be given to developing methodologies that support accurate and adequate translation from the source language into the target one, and curricula redesign to effectively integrate CBLP into existing linguist and translator training programs. Acknowledgement. The paper and the practices described in it were inspired by the participation of the authors in the Online Corpus-Based Language Pedagogy Workshop Series 2025 "Advancing Language Education: Integrating Corpus Technology and AI" and AI-Empowered CBLP Lesson Design & Teaching Competition hosted by The Education University of Hong Kong (EdUHK), Linguistics and Modern Languages Studies by the courtesy of MA, Qing (Angel) (PhD), Associate Dean (Research & Postgraduate Studies), Associate Professor, Faculty of Humanities.

REFERENCES

1. Hejazi, H. (2022). A corpus-based investigation of lexical bundles and keyness in B1, B2 and C1 ESL learners' academic writing (Master's thesis, Department of Linguistics, University of Liverpool).
2. Hou, H. (2014). Teaching specialized vocabulary by integrating a corpus-based approach: Implications for ESP course design at the university level. English Language Teaching, 7(5), 26–35. https://doi.org/10.5539/elt.v7n5p26
3. Callies, M. (2019). Integrating corpus literacy into language teacher education: The case of learner corpora. In M. Paquot & S. Granger (Eds.), Learner corpora and language teaching (pp. 246–264). John Benjamins. https://doi.org/10.1075/slcs.201.12cal
4. Li, D., Noordin, N., Ismail, L., & Cao, D. (2025). A systematic review of corpus-based instruction in EFL classroom. Heliyon, 11, e42016. https://doi.org/10.1016/j.heliyon.2025.e42016
5. Kusumaningrum, M. V., & Ardi, P. (2020). A corpus analysis of lexical sophistication in LLT Journal: A journal on language and language teaching. ELTR Journal, 4(1), 53–75. https://doi.org/10.37147/eltr.v4i1.39
6. Crosthwaite, P., & Baisa, V. (2024). A user-friendly corpus tool for disciplinary data-driven learning: Introducing CorpusMate. International Journal of Corpus Linguistics, 29(4), 595–610. https://doi.org/10.1075/ijcl.23056.cro

© The Author(s) 2026

7. Popescu, D. A., Bold, N., & Stefanidakis, M. (2025). A systematic model of an adaptive teaching, learning and assessment environment designed using genetic algorithms. Applied Sciences, 15(7), Article 4039. https://doi.org/10.3390/app15074039
8. Ma, Q., Yuan, R. (Eric), Cheung, L. M. E., & Yang, J. (2022). Teacher paths for developing corpus-based language pedagogy: A case study. Computer Assisted Language Learning, 37(3), 461–492. https://doi.org/10.1080/09588221.2022.2040537
9. British National Corpus. (n.d.). British National Corpus. http://www.natcorp.ox.ac.uk/
10. Davies, M. (2008–). Corpus of Contemporary American English (COCA). English-Corpora.org. https://www.english-corpora.org/coca/
11. Men, H. (2020). Data-driven learning in enhancing learners' language idiomaticity. International Journal of Emerging Technologies in Learning, 15(23), 27–41. https://doi.org/10.3991/ijet.v15i23.19023
12. Giampieri, P. (2020). Data-driven learning in English for academic purposes class. Languages and Linguistics in Higher Education, 10(1), 217–233. https://doi.org/10.1515/cercles-2020-2006
13. Unitex/GramLab. (2025). Unitex/GramLab: Lexicon-based corpus processing suite. https://unitexgramlab.org/
14. Tiedemann, J. (2012). Parallel data on the web and in OPUS. In Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC). European Language Resources Association. https://opus.nlpl.eu/
15. Anthony, L. (2024). AntConc (Version 4.3.1) [Computer software]. Waseda University. https://www.laurenceanthony.net/software/antconc/
16. Kilgarriff, A., Rychlý, P., Smrz, P., & Tugwell, D. (2004). Sketch Engine [Computer software]. https://www.sketchengine.eu/
17. Scott, M. (2023). WordSmith Tools (Version 8) [Computer software]. Lexically.net. https://www.lexically.net/wordsmith/
18. Brezina, V., & Platt, W. (2025). #LancsBox X (Version 5.5.1) [Computer software]. Lancaster University. http://lancsbox.lancs.ac.uk
19. Gilquin, G. (2024). From second language acquisition research to foreign language teaching through the prism of corpora. Ampersand, 13, 100204. https://doi.org/10.1016/j.amper.2024.100204
20. Schmidt, N. (2023). Unpacking second language writing teacher knowledge through corpus-based pedagogy training. ReCALL, 35(1), 40–57. https://doi.org/10.1017/S0958344022000106
21. Yemelianova, O. V. (2024). English corpora for teaching translation students. In I. V. Ushchapovska (Ed.), Best EU and Ukrainian practices in education for sustainable development: Monograph (pp. 338–361). Sumy State University.
22. RhymeZone. (n.d.). RhymeZone: The ultimate rhyming dictionary. https://www.rhymezone.com/r/rhyme.cgi?Word=translator&typeofrhyme=perfect&org1=syl&org2=l&org3=y
23. NetSpeak. (n.d.). NetSpeak: A new way to search for information in English. https://netspeak.org/
24. Durrant, P. (2025). What can a corpus tell us about school writing? Findings, challenges, and future directions. Applied Corpus Linguistics, 5(2), 100134. https://doi.org/10.1016/j.acorp.2025.100134
25. Michigan Corpus of Upper-Level Student Papers. (n.d.). MICUSP: Michigan Corpus of Upper-Level Student Papers. https://micusp.elicorpora.info/main
26. Academic Word Suggestion Machine (AWSUM). (n.d.). AWSUM: Academic Word Suggestion Machine. https://langtest.jp/awsum/
27. Davies, M. (2008–). English Corpora.org. https://www.english-corpora.org/corpora.asp
28. Davies, M. (2025, August). Integrating AI / LLMs into English-Corpora.org. https://www.english-corpora.org/ai-llms

© The Author(s) 2026

Ємельянова Олена, Єгорова Олеся
Сумський державний університет, Суми, Україна
ORCID ID: 0000-0002-3277-1227
ORCID ID: 0000-0002-3225-5580

ІНСТРУМЕНТИ КОРПУСНОГО АНАЛІЗУ В МОВНІЙ ОСВІТІ: ТРЕНДИ, ПРАКТИКИ ТА МЕТОДИЧНЕ РОЗУМІННЯ

Анотація. *Стаття присвячена вивченню ефективності застосування корпусного підходу в процесі навчання іноземних мов та перекладу. Метою статті є дослідити ефективність корпусного підходу в перекладоорієнтованому навчанні іноземних мов шляхом оцінки його впливу на засвоєння словникового запасу студентами, письмову компетентність та залученість учнів; визначити ключові умови для продуктивного включення корпусного підходу до сучасних практик викладання іноземних мов та перекладу. Автори наголошують на необхідності аналізу можливостей корпусних інструментів задля забезпечення лінгвістичної автентичності, покращення навчання через використання корпусних даних, персоналізації та контекстуальності навчання, сприяння розвитку автономії студентів. Автори демонструють ефективність корпусного підходу у навчанні іноземних мов та перекладу, особливо в засвоєнні лексики, що є вельми важливим для майбутніх перекладачів, використовуючи практичні приклади роботи з COCA для вивчення словосполучень та словотвору. Наводяться відгуки студентів, які свідчать про їхній ентузіазм та позитивне ставлення до використання корпусів. Також розглядається застосування RhymeZone для розвитку творчих навичок та NetSpeak для вивчення лексико-граматичних моделей. Автори аналізують переваги корпусного підходу, включаючи наукову обґрунтованість, об'єктивність, розвиток автономії, підвищення мотивації, диференційоване та персоналізоване навчання, а також підготовку до реальної мовленнєвої взаємодії. Стаття висвітлює певні обмеження корпусно-орієнтованої мовної педагогіки, до яких належать: технологічні бар'єри, методична складність та певні виклики для викладачів, етичні та юридичні питання, а також проблеми інтерпретації даних. Перспективи подальших досліджень полягають в ґрунтовному вивченні потенціалу корпусного підходу у навчанні студентів-перекладачів жанровій специфіці аналізованих текстів, їх лексичним та граматичним особливостям та способам точного перекладу з мови оригіналу на мову перекладу.*

Ключові слова: *корпусно-орієнтована мовна педагогіка, корпусний підхід, диференційоване та персоналізоване навчання.*

Licensed under CC BY-NC-SA 4.0
https://creativecommons.org/licenses/by-nc-sa/4.0/

Стаття надійшла до редакції 18.12.2025
The article was received 18 December 2025