CORPUS-BASED VS GENAI-SUPPORTED LEARNING: THE POTENTIAL IN EFL TEACHING

Yuliia Klymovych
PhD, Senior Lecturer,
Zhytomyr Ivan Franko State University, Zhytomyr, Ukraine

Corpus-based learning and artificial intelligence (AI) have become extremely popular over the past decade for their unique potential to personalize and enrich English as a Foreign Language (EFL) learning experience. The approaches are deeply intertwined with technology, relying on computer-based, data-driven learning to improve foreign language acquisition.

Corpora have long provided learners with authentic examples of language in context, revealing patterns of grammar, vocabulary, and usage (Al-Gamal & Mohammed Ali, 2019). However, with the rise of generative AI, language learning has entered a new era. AI-powered tools offer a more interactive and personalized experience, immediate feedback and conversational opportunities that can engage learners in ways traditional methods cannot. But despite these promising advancements, current generative AI tools lack the ability to directly analyze large, structured language corpora, such as the Corpus of Contemporary American English (COCA) or the British National Corpus (BNC). This limitation highlights the need to balance the strengths of both methods and explore how they can complement one another in EFL teaching. In our research we want to investigate whether AI can fully substitute corpora-based approach in language learning.

To clarify, the distinction in our analysis lies not simply between corpora and generative AI but rather between concordancers and generative AI systems. While both approaches utilize language corpora, they do so in fundamentally different ways.

Concordancers, tools traditionally used in corpus-based learning, rely on curated and structured corpora and allow users to search, analyze, and interpret language patterns based on these large collections of authentic texts. Concordancers provide educators and learners direct access to a body of verified linguistic data, facilitating a precise analysis of vocabulary, collocations, syntax, and context-specific language use.

Generative AI systems, on the other hand, like OpenAI's GPT models, also draw upon extensive corpora. However, the sources used to train such models are typically broader, encompassing vast, publicly available data sources, and are less curated for academic or linguistic purposes (Grothaus, 2023). The training data is not organized or searchable in the same way as traditional language corpora, and the AI's function is not to provide authentic instances of language but to generate coherent, contextually relevant responses based on learned language patterns.

Crosthwaite & Baisa (2023) in their research consider some of the advantages that corpus tools hold over chatbot-like GenAI:

- **knowing the data** (*concordancers provide clear insights into the domain of texts from which the corpus data is derived*);
- **authenticity** (*corpus data represents real and actually produced by humans language*);
- **replicability** (*concordancers allow for consistent retrieval of language patterns, offering "hard evidence" for linguistic analysis, whereas generative AI produces varied responses to identical prompts*);
- **multimodality** (*concordance tools support various forms of data representation, including colored concordances, statistical tables, visual charts, and relationship maps of lexical and grammatical units*);
- **safety** (*most corpus tools typically require minimal user information, thus reducing privacy concerns*);
- **hallucinations** (*generative AI's outputs may sometimes lack precision, with "hallucinations" affecting response reliability*);
- **active learning** (*working with corpora and concordancers requires significant inductive learning processes*).

Conversely, generative AI offers distinct advantages over corpus-based learning, particularly in terms of **accessibility** and **personalization**. Generative AI is highly user-friendly, requiring minimal technical expertise or prior training for effective interaction, which allows users of all backgrounds to engage with language tasks immediately and confidently. In contrast, working with traditional corpora and concordancers often requires a foundational understanding of corpus analysis techniques and tools, as well as familiarity with linguistic terminology and search strategies. Another notable advantage of generative AI is **cost-effectiveness**. While many language corpora and concordancers require paid subscriptions or licenses, most generative AI platforms offer free access.

Both concordancers and generative AI tools offer unique advantages and limitations for use in EFL classroom settings. Concordancers and traditional corpus-based tools remain unmatched in terms of reliability and validity of language data. They allow educators and learners to explore authentic, evidence-based language patterns, providing structured data that can be easily verified and analyzed for accuracy (Meunier, 2024). For EFL activities that require precise linguistic analysis, such as examining collocations, identifying genre-specific vocabulary, or studying grammatical structures, concordancers are invaluable. For example, teachers can guide students to analyze common collocations in COCA or BNC to deepen their understanding of word associations in academic writing.

Generative AI tools, with their ease of use, accessibility, and often free availability, provide a flexible option for language practice and basic concept learning without the need for extensive technical training. Generative AI can be effectively used for activities such as conversational practice, grammar drills, or vocabulary expansion, where students can receive immediate, interactive feedback (Hartwell & Aull, 2023). For instance, students might use an AI chatbot (for writing or speaking) to simulate real-life dialogues, helping them practice conversational English in various contexts or ask the AI to suggest synonyms and paraphrase sentences.

Based on the comparison of corpus-based learning and AI-powered learning in EFL teaching, it's clear that each approach offers distinct advantages and limitations. If an EFL classroom's goals include data accuracy, replicability, and deep language analysis, corpus-based tools are the preferred choice. However, for accessible, interactive, and user-friendly platforms that enhance language practice without requiring deep linguistic expertise, generative AI tools can be an effective addition. Each tool has its place in the EFL classroom, and the combination of AI and corpus-based resources has the potential to create a more efficient and holistic language learning experience by pairing real-world language data with responsive, adaptive technology.

References

Al-Gamal, A. A. M., & Mohammed Ali, E. A. M. (2019). Corpus-based method in language learning and teaching. *International Journal of Research and Analytical Reviews*, 6(2), 473–476.

Crosthwaite, P., & Baisa, V. (2023). Generative AI and the end of corpus-assisted data-driven learning? Not so fast!, *Applied Corpus Linguistics*, 3(3). https://doi.org/10.1016/j.acorp.2023.100066.

Grothaus, M. (2023). What is a "corpus"? And why is everyone in AI suddenly talking about it? Here's what you need to know. *Fast Company Middle East.* Retrieved November 6, 2024, from https://fastcompanyme.com/technology/what-is-a-corpus-and-why-is-everyone-in-ai-suddenly-talking-about-it-heres-what-you-need-to-know/

Hartwell, K., & Aull, L. (2023). Editorial Introduction – AI, corpora, and future directions for writing assessment, *Assessing Writing*, 57. https://doi.org/10.1016/j.asw.2023.100769.

Meunier, F. (2024, March 28). *GenAI-supported vs corpus-aided language teaching* [Conference presentation]. Empowering Language Education: When Corpus-based Language Pedagogy Meets with AI, The Education University of Hong Kong, Hong Kong, China. https://www.researchgate.net/publication/379483507_GenAI-supported_vs_corpus-aided_language_teaching