LINGUODIDACTIC POTENTIAL OF CORPUS TECHNOLOGY IN THE DEVELOPMENT OF SPECIALIZED DICTIONARIES AND TEXTS

Shayusupova Nargiza Bakhtiyorovna
Lecturer, The Branch of Astrakhan State Technical
in Tashkent region, Republic of Uzbekistan
E-mail: nargiza.shayusupova@gmail.com

Abstract: Corpus technology has emerged as a transformative tool in linguistic research and language teaching. This article explores the linguodidactic potential of corpus technology, particularly in the creation and application of special dictionaries and texts for educational purposes. It outlines how corpora can facilitate the development of specialized language resources that cater to specific domains and highlights their role in improving language learning, especially, in fields such as professional and academic communication. The paper concludes by addressing challenges and suggesting practical applications of corpus-based tools in language teaching.

Key words: linguistic pedagogy, corpus, corpora, corpus technology, advantages and challenges, approaches to study

INTRODUCTION

The increasing integration of digital technologies into language learning has significantly transformed linguistic pedagogy. Among the modern technologies, corpus technology offers unprecedented access to authentic language use across various contexts. A corpus is a large, structured collection of written or spoken texts that can be analysed computationally to uncover patterns in language usage [1]. Corpus linguistics has been effectively employed in language teaching, research activity, and lexicography, particularly in the development of specialized dictionaries and texts. Corpus-based dictionaries and educational texts allow learners to encounter language in its natural form, enhancing their understanding of word usage, collocations, and context-specific meanings.

METHODS.

Corpus Technology: A Brief Overview of its Functions.

Corpus technology relies on the systematic collection and analysis of real-world language data. The technology involves software tools that process large datasets of texts (corpora) to identify patterns, such as frequency of word use, collocations, and grammatical structures. The application of corpora and corpus technology in language education has been a focus for researchers, lexicographers, and educators for over twenty years. Since the 1980s, insights gained from corpus-based language analysis have shaped and impacted teaching resources and reference materials, especially in the context of English as a Foreign Language (EFL) [2].

Modern corpora, such as the British National Corpus (BNC) or the Corpus of Contemporary American English (COCA), have been crucial in understanding both general and specialized language use.[3] In linguodidactic applications, corpus technology performs two core functions:

1. Data-Driven Learning (DDL): Data-driven learning is based on the principle of "cutting out the middle man," a reference to learning language directly from language rather than from mediated resources such as textbooks, grammar, dictionaries, and teachers [4]. Learners can directly interact with corpora, exploring authentic texts and discovering language rules through observation.

2. Development of Learning Resources: Corpus analysis supports the creation of specialized dictionaries, glossaries, and educational texts by providing evidence-based examples of how language is used in particular fields, such as medicine, engineering, or legal contexts.

RESULTS

Advantages of Corpus technology.

In the context of language education, especially for specific purposes (ESP), learners often require access to domain-specific vocabulary and usage patterns. Specialized dictionaries and texts tailored to fields such as medicine, law, or

business provide focused linguistic resources that reflect the terminology and communication styles used within those fields. Traditional language learning materials may not provide the precision or contextual accuracy needed in these areas, making corpus-based resources invaluable [5].

Corpus technology enhances the value of these resources within the following criteria:

- Ensuring Authenticity: Corpus data ensures that the vocabulary and phrases included in dictionaries or texts reflect how native speakers use language in real-world contexts.

- Promoting Relevance: By analysing domain-specific corpora, lexicographers can develop dictionaries that include the most relevant and frequently used terms within a specific field.

- Improving Contextual Understanding: Corpora provide examples of how words and phrases function within sentences, aiding learners in understanding not just definitions but also collocations, register, and pragmatic usage.

Linguodidactic Applications of Corpus Technology.

1. Development of Specialized Dictionaries.

Corpus-based dictionaries are grounded in real-world language usage. By analyzing domain-specific corpora, lexicographers can compile dictionaries that feature terminology commonly used in professional or academic fields. For instance, medical corpora allow the development of medical dictionaries that contain accurate, context-relevant terms. In language education, learners benefit from exposure to these dictionaries as they encounter language that is relevant to their academic or career goals. Specialized dictionaries also provide insight into word frequency, meaning, and variation across registers. A dictionary for business English, for example, might highlight the difference in terminology used in financial reports versus casual workplace communication, giving learners a well-rounded understanding of professional language

2. Creation of Authentic Educational Texts.

Corpus technology enables the design of authentic learning materials by analysing how language is used in specific genres or professional fields. For example, a corpus of legal texts can help language educators create educational materials for learners of legal English, exposing them to contract language, courtroom communication, or legal documentation. These texts not only introduce learners to technical vocabulary but also teach them to recognize common structures and rhetorical patterns in professional discourse.

3. Data-Driven Learning (DDL)

Corpus technology supports the idea of Data-Driven Learning (DDL), where students are encouraged to analyse authentic language data themselves. This method shifts the learning process from a teacher-centered approach to a learner-centered one, empowering students to observe patterns in language use directly. By interacting with a corpus, learners can explore how specific terms or grammatical structures function across various contexts, enhancing their linguistic intuition and critical thinking skills. For example, learners studying business English might use a business-specific corpus to analyse how certain phrases, such as "market share" or "profit margins," are used in annual reports, advertisements, or business correspondence. By identifying patterns in context, learners can gain a more nuanced understanding of language usage and apply this knowledge in real-world communication.

1. Challenges in Corpus-Based Language Education.

Despite its potential, the use of corpus technology in developing specialized dictionaries and texts faces certain challenges:

- Accessibility: While corpora are proving increasingly available, access to high-quality, domain-specific corpora can still be limited, particularly for less commonly studied fields.

- Complexity: Navigating and analysing corpora requires a certain level of technical expertise, which may be difficult for novice language learners or educators unfamiliar with the technology.

- Customization: Tailoring corpus-based resources to meet specific learner needs can be time-consuming, as corpora must often be filtered or adapted to suit educational goals.

CONCLUSION

Corpus technology offers significant linguodidactic potential, particularly in the development and use of specialized dictionaries and texts for language learners. By providing authentic, contextually relevant examples of language use, corpus-based resources help learners engage more deeply with the specific vocabulary and linguistic structures needed in professional and academic fields. As this technology continues to evolve, overcoming challenges related to accessibility and complexity will be crucial in fully realizing its educational benefits. Future research should explore ways to further integrate corpus tools into everyday language learning and investigate their impact on learner outcomes.

REFERENCES

1. McEnery T., Hardie, A. Corpus linguistics: Method, theory and practice. Cambridge University Press. United Kingdom, 2012.
2. Braun S., Kohn K., Mukherjee J. Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods. //English Corpus Linguistics. Vol 3, 2006. - Pp. 1-4.
3. Biber D., Conrad, S., Reppen R. Corpus linguistics: Investigating language structure and use. Cambridge University Press., United Kingdom, 1998.
4. Thomas, J. Deriving extended collocations from full text for student analysis and synthesis. In: A. Leńko-Szymańska and A. Boulton (Eds.) Multiple Affordances of Language Corpora for Data-driven Learning. Amsterdam; Philadelphia, PA: John Benjamins, 2015. Pp 85–108.

5. Römer, U. Corpora and language teaching. In A. Lüdeling & M. Kytö (Eds.). Corpus linguistics: An international handbook. Vol. 1. Mouton de Gruyter, 2008. Pp. 112-130.