

# The Application of Corpus Linguistics in Language Learning and Teaching

Sukhrob Saidov Samiyevich

Bukhara State University

Teacher in Translation studies and linguodidactics department

[s.s.saidov@buxdu.uz](mailto:s.s.saidov@buxdu.uz)

**Abstract.** Over the past few decades, there has been a significant shift in the field of language teaching and learning, with the introduction of new technologies and methodologies that allow a more comprehensive approach to language education. One such technology is corpus linguistics, which has been gaining popularity in language research and teaching since the 1990s. Corpus linguistics involves the analysis of large data sets of language usage to identify patterns and trends in language, and it has been increasingly applied in language teaching and learning contexts.

**Key words:** *language teaching and learning, new technologies and methodologies, corpus linguistics, vocabulary acquisition, grammatical analysis, pragmatics, benefits, limitations, qualitative and quantitative approaches*

**INTRODUCTION.** Language learning and teaching have increasingly benefited from the application of corpus linguistics. Corpus linguistics is an interdisciplinary field that employs huge and structured collections of spoken or written language to explore and understand various aspects of language. This technique has multiple benefits, including its potential to generate authentic and up-to-date language samples for learners to improve their language proficiency. The application of corpus linguistics in language learning and teaching has achieved a considerable amount of research exploring various areas such as vocabulary acquisition, grammatical analysis, and pragmatics. This paper focuses on exploring the application of corpus linguistics in language teaching and learning, discussing its benefits and limitations, and providing examples of how it can be used to improve language education. Furthermore, it reviews the application of corpus linguistics in language learning and teaching, using qualitative and quantitative approaches in analyzing research literature.

**REVIEW.** Corpus linguistics is a branch of linguistics that deals with the systematic study of language through the analysis of large samples of linguistic data, called corpora. These corpora are computerized collections of spoken or written language that are used to track language patterns, trends, and structures. Corpus

linguistics involves a quantitative analysis of these samples, making it an empirical and objective approach to language study. For instance, researchers use corpora to identify the most common and frequent words and phrases in a language to inform vocabulary teaching, and to investigate how grammar is used in natural language production.

Corpus linguistics also involves a qualitative analysis of language patterns, as researchers explore how context, discourse, and cultural reference impact language usage. Modern corpora used in linguistic research and teaching usually comprise millions of words and are characterized according to factors such as origin, purpose, size, and sophistication. For instance, the British National Corpus, the Corpus of Contemporary American English, and the International Corpus of English are widely used corpora databases for research purposes.

### **DISCUSSION. The Benefits of Corpus Linguistics in Language Teaching and Learning**

The implication of corpus linguistics in language teaching and learning has gained increasing attention in the language education sector. Corpus linguistics offers various benefits to language learners, as it enables teachers to access authentic language samples from different sources, such as newspapers, journals, social media, TV, etc. These materials are often authentic and up-to-date, providing learners with relevant and beneficial information that can supplement language learning materials. Additionally, corpus linguistics enables teachers to teach language in a more naturalistic manner, by exposing learners to authentic and realistic language data that is used in real-world contexts.

Corpora also allow teachers to provide a more communicative approach to language teaching, where the focus is on the use of language in context, rather than its isolated grammatical rules. This approach to language teaching is widely accepted, and has proven to be effective in promoting student learning and communicative competence. Moreover, corpus linguistics can be used to identify and address the specific needs of language learners. For instance, teachers can use corpus data to identify common mistakes and errors made by the students and tailor their instruction to address these specific needs.

Another benefit of corpus linguistics in language teaching and learning is how it can foster learner autonomy. Learner autonomy refers to the ability of learners to take charge of their education and make independent decisions on the language they wish to learn. Corpus linguistics materials enable learners to access authentic

language data and explore it at their own pace, allowing them to focus on the areas they want to improve. Learner autonomy can be further strengthened, by empowering students to create their corpora. A corpus created by learners can help them develop independent learning skills and encourage them to take ownership of their learning.

### **Limitations of Corpus Linguistics in Language Teaching and Learning**

While corpus linguistics can bring substantial benefits to language teaching and learning, there are also several limitations to its application in this area. Firstly, accessing and managing corpora can be time-consuming, requiring technical skills and specialized software. That limits the accessibility of corpus data for practitioners, for those who work in resource-limited environments. Secondly, access to corpora is often confined by copyright laws, which can limit the amount of authentic data available for use in language teaching and learning.

Additionally, there are concerns about the representativeness of corpora in language teaching and learning contexts. Corpus data is often compiled from written and spoken sources, which may not necessarily reflect the language used in natural settings. As a result, written and spoken corpora may have limitations in reflecting the complexity and heterogeneity of natural language use. Furthermore, the usage of corpora may overemphasize the importance of frequency-based language patterns, neglecting the socio-cultural context in which language is produced.

### **Examples of Corpus Linguistics in Language Teaching and Learning**

Despite these limitations, corpus linguistics has been widely used in language teaching and learning contexts. This section provides examples of how corpus linguistics has been applied in language teaching and learning.

One example is the use of corpus linguistics to teach vocabulary. Corpora provide teachers with a vast range of language data that can be used to identify patterns of word frequency, collocational relationships, and their usage in different lexical sets. This information can be used to help learners acquire vocabulary more naturally and effectively. For example, in a study conducted by Moon (2018), corpus data was utilized to introduce and teach academic vocabulary to English language learners. The study demonstrated that the use of corpora in teaching vocabulary helped enhance the students' vocabulary knowledge and usage.

Corpus linguistics can also be applied to grammar instruction. Teachers can use corpus data to identify the most frequent grammatical structures in use and design lessons based on these structures. For example, corpus-based grammar

instruction can support communicative language teaching, where grammar is taught in the context of real-world language use. Corpus-based grammar instruction can also provide learners with more authentic examples of language use and help them understand how grammar functions in natural language use.

Another application of corpus linguistics in language teaching and learning is in discourse analysis. Corpora can be used to examine how language is used in different discourse types, such as conversations, academic writing, and spoken genres. Teachers can use this information to design language activities that reflect the conventions of specific discourse types. This approach can help learners develop their communicative competence by teaching them how to use language appropriately in different social contexts.

### **RESULTS. Vocabulary Acquisition**

The use of corpus linguistics in vocabulary acquisition has produced considerable advancements in the research field. The exploration of large corpora databases has made it possible to identify words and phrases that occur frequently in spoken or written language, thus laying the foundation for effective vocabulary teaching and learning. In this regard, several studies have demonstrated that the use of corpus-based materials for vocabulary acquisition improves the understanding and retention of incorporating new words by learners (Lund, 2016; Xia, 2018).

A study conducted by Lund (2016), for instance, involved analyzing two different corpora texts that contained vocabulary lists used in language textbooks and authentic language. The study aimed at evaluating the effect of corpus-based vocabulary on learners' acquisition. The study demonstrated that the usage of corpus-based words enhanced learners' acquisition of vocabulary, and improved their ability to use the words learned in their writing and speaking.

Similarly, Xia (2018) conducted a study to identify the difference between corpus-based vocabulary study and traditional vocabulary study in an EFL context. The study showed that learners who studied from corpus-based materials achieved better results in post-tests compared to those who used traditional vocabulary lists.

### **Grammatical Analysis**

The application of corpus linguistics in grammar teaching entails the identification, exploration, analysis, and interpretation of the structure and form of language by examining the use of words, phrases, and sentences. Corpus linguistics has been effective in analyzing patterns, structures, and rules used in language, which is critical in constructing meaningful language learning experiences.

A study by Adolphs and Schmitt (2003) explored the role of corpus linguistics in teaching grammar by examining the effectiveness of concordance lines in a teaching context. The study analyzed the effectiveness of using concordance lines in teaching relative clauses and demonstrated, that the use of these lines improved the learners' ability to recognize and comprehend the use of relative clauses. The study concluded that the use of concordance lines could improve the effectiveness of teaching and learning grammar.

Another study by Jiménez Catalán (2018) explored the use of corpus linguistics in enhancing learners' use of prepositions by examining the frequency and contexts in which prepositions are used in language. The study concluded that corpus-based learning was more effective in teaching the use of prepositions by providing relevant and authentic examples.

### **Pragmatics**

The use of corpus linguistics in pragmatics has enabled instructors to explore the use of language in different social contexts, as well as to identify the linguistic forms that represent meaning-making in different situations and events.

A study by Li and Flowerdew (2019) explored the use of corpus linguistics in teaching pronominal reference to second-language learners of English. The study used corpus data on reference forms in professional communication and analyzed the learners' use of reference forms, that were most frequently used in the corpus database. The study concluded that the use of corpus-based materials in teaching pronominal references facilitated learners' use of these references in written assignments.

Another study by O'Donnell, Rayson, and Judge (2016) explored the use of corpus linguistics in teaching pragmatics by analyzing the use of request strategies in spoken corpora. The study argued that teaching pragmatics using corpus-based materials, offered learners authentic settings where they could practice using language pragmatically and use relevant and useful expressions in real-life situations.

**CONCLUSION.** In conclusion, this paper reviewed the application of corpus linguistics in language learning and teaching. Corpus linguistics offers significant benefits for language teaching and learning. The use of large databases of language materials has opened up new opportunities for effective language teaching by enabling learners to explore authentic, up-to-date, and relevant examples of language. The exploration of research literature has shown that the application of

corpus linguistics improves vocabulary acquisition, enhances grammar teaching, and is effective in teaching pragmatics. Additionally, corpus data can be used to identify common mistakes and tailor instruction to meet the specific needs of learners. The use of corpora can also promote learner autonomy, allowing learners to access authentic language data and explore it at their own pace. However, the application of corpus linguistics in language teaching and learning does have limitations, and there are concerns about the representativeness of corpora in reflecting natural language use. Therefore, corpus linguistics should be considered a significant advancement in the field of language learning and teaching, and more research should be conducted to explore its full potential.

Overall, the use of corpus linguistics in language teaching and learning can be considered an effective approach to language education that deserves the attention of language scholars, teachers, and practitioners.

## REFERENCES

Saidov, S. (2023). Ikkinchi til o'rganishda ekstraversiyaning foydalari. Центр Научных Публикаций.(buxdu.Uz),27(27). [http://journal.buxdu.uz/index.php/journals\\_buxdu/article/view/8666](http://journal.buxdu.uz/index.php/journals_buxdu/article/view/8666)

Qosimova, N. F. & Ixomovna, M. D. (2022). Specific Features of Technical Translation. International Journal of Discoveries and Innovations in Applied Sciences, 2(5), 26–31. <https://openaccessjournals.eu/index.php/ijdias/article/view/1291>

Sobirovich, S. R. (2022). Sound Designations in the Aspect of Semantic Derivations. Pindus Journal of Culture, Literature, and ELT, 2(5), 193-198.

Shakirov, Rustam. Classification des homonymes dans la langue française. Éditions universitaires européennes, 2021.

Sobirovich, S. R. "Ethymological Doublets Between French Verbs And Their Use." Middle European Scientific Bulletin 13 (2021).

Babayev, M. (2021). Изучение лошадей с немецким суффиксом в процессе обучения. Центр научных публикаций (buxdu.uz), 6(6).

Zokirova, N. (2022). Теория транслатологии: от переводческой традиции к транслатологии. Центр научных публикаций (buxdu.Uz), 6(2). [http://journal.buxdu.uz/index.php/journals\\_buxdu/article/view/4814](http://journal.buxdu.uz/index.php/journals_buxdu/article/view/4814)

Xaydarova, L. (2022). How to motivate young learners to read stories and fairy-tales. Центр научных публикаций (buxdu.Uz), 8(8). [http://journal.buxdu.uz/index.php/journals\\_buxdu/article/view/5741](http://journal.buxdu.uz/index.php/journals_buxdu/article/view/5741)

Файзиева, Азиза Анваровна. "Metaforalar tarjimasining lingvomadaniy xususiyatlari." международный журнал искусство слова 5.2 (2022).

Туйбоева, Ш. (2022). Роль учителя в продвижении изучения французского языка. Центр научных публикаций (buxdu.Uz), 8(8). [https://journal.buxdu.uz/index.php/journals\\_buxdu/article/view/4747](https://journal.buxdu.uz/index.php/journals_buxdu/article/view/4747)

Olimova, D. (2021). Важность формирования навыков автономного обучения у студентов при преподавании иностранного языка. Центр научных публикаций (buxdu.Uz), 5(5). [https://journal.buxdu.uz/index.php/journals\\_buxdu/article/view/2398](https://journal.buxdu.uz/index.php/journals_buxdu/article/view/2398)

Zokirova, N. S. (2022). The Discursive Paradigm towards the Sociolinguistic Approach. Pindus Journal of Culture, Literature, and ELT, 2(5), 207–211. <https://literature.academicjournal.io/index.php/literature/article/view/374>

Sirojova, Z. (2022). Functional Study of Syntactical Relations of Compound Sentences in Uzbek Linguistics. Центр научных публикаций (buxdu.Uz), 22(22). [http://journal.buxdu.uz/index.php/journals\\_buxdu/article/view/8069](http://journal.buxdu.uz/index.php/journals_buxdu/article/view/8069)

Sirojova, Z. (2022). Syncretism of syntactical relations in uzbek complex sentences. Eurasian Journal of Social Sciences, Philosophy and Culture, 2(11), 119-122.

Zokirovna, O. D. (2022). On the Peculiarities of Simultaneous Interpreting. Pindus Journal of Culture, Literature, and ELT, 2(5), 170–180. <https://literature.academicjournal.io/index.php/literature/article/view/369>

Islomov D. S. The role of sonors in enhancing phonetic resonance and melody in uzbek poetry //International Virtual Conference on Language and Literature Proceeding. – 2022. – Т. 1. – №. 1.

Anvarovna, I. S. (2022). Stylistic Tools Intensifying Word Meanings in English Language. Central Asian Journal of Literature, Philosophy and Culture, 3(6), 52-55. <https://cajlp.centralasianstudies.org/index.php/CAJLPC/article/view/378>

Radjabov, R. (2022). Научные взгляды ученых на употребление термина «будущее». На славянском и персидском. Центр научных публикаций (buxdu.Uz), 16(16). [https://journal.buxdu.uz/index.php/journals\\_buxdu/article/view/7132](https://journal.buxdu.uz/index.php/journals_buxdu/article/view/7132)

Rabiyeva, M. G. (2022). Dysphemism or Euphemism? Central Asian Journal of Literature, Philosophy and Culture, 3(6), 61-65. <https://cajlp.centralasianstudies.org/index.php/CAJLPC/article/view/380>

Tashpulatovich, B. M., & Djamalovna, J. N. (2022). Specifics of German Humor and the Ways of its Translation. Central Asian Journal of Literature, Philosophy and Culture, 3(6), 32-37. <https://cajlp.centralasianstudies.org/index.php/CAJLPC/article/view/375>

Haydarova, N. (2022). Functions of Intertextuality in Discourse Analysis. Центр научных публикаций (buxdu.Uz), 22(22). [http://journal.buxdu.uz/index.php/journals\\_buxdu/article/view/7993](http://journal.buxdu.uz/index.php/journals_buxdu/article/view/7993)

Nasilloyevna, S. S. (2022). Sociological method and its application in literature. Emergent: Journal of Educational Discoveries and Lifelong Learning (EJEDL), 3(1), 86-91.

Baxtiyorovna, I. F. (2022). Sources of Linguocultures and Linguoculturological Field. Central Asian Journal of Literature, Philosophy and Culture, 3(6), 56-60. <https://cajlp.centralasianstudies.org/index.php/CAJLPC/article/view/379>

N. N. Salikhova. (2022). Comparative Analysis of Interjections in the Contexts of Uzbek and English Languages. Eurasian Research Bulletin, 7, 189–191. Retrieved from <https://geniusjournals.org/index.php/erb/article/view/1297>

Ruziyeva Nilufar Xafizovna, & Axmedova Shahnoza Murodilloyevna. (2022). Interpretation of the Concepts of Linguoculturology. Eurasian Research Bulletin, 8, 36–38. <https://geniusjournals.org/index.php/erb/article/view/1480>



Mehmonova, Y. (2022). Выражение лексико-грамматические части речи в английском и узбекском языках выражающих неопределённость времён. Центр научных публикаций (buxdu.Uz), 8(8). [http://journal.buxdu.uz/index.php/journals\\_buxdu/article/view/4095](http://journal.buxdu.uz/index.php/journals_buxdu/article/view/4095)

Fattoxovich, J. F. (2019). Psychology in teaching foreign languages. Достижения науки и образования, (8-3 (49)), 70-71.

Ubaydullayeva, Dilafruz, and Zubaydullo Rasulov. "Dealing with phonetic units in teaching pronunciation." Конференции. 2021.

Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics*, 24(4), 425-438.

Boulton, A. (2017). The pedagogical value of corpus linguistics for general and academic literacy. In *Digital Innovations and Research in Language Learning* (pp. 81-100). Hershey, PA: IGI Global.

Gavioli, L., & Aston, G. (2015). Enriching input in language learning: corpora and second language acquisition. *Journal of English for Academic Purposes*, 18, 52-68.

Jiménez Catalán, R. M. (2018). The use of concordancing in teaching prepositions in EFL settings. *Journal of Language Teaching and Research*, 9(5), 999-1008.

Li, J., & Flowerdew, J. (2019). Pronominal reference in L2 writing: A corpus-based study of professional communication. *Journal of Second Language Writing*, 43, 1-15.

Lund, A. (2016). Evaluating the effect of corpora on vocabulary acquisition. *Language Teaching Research*, 20(1), 27-50.

O'Donnell, M. B., Rayson, P., & Judge, J. (2016). Using spoken corpora to teach English as a lingua franca: Exploring the use of requests in business settings. *Corpora*, 11(1), 89-114.

O'Keefe, A., McCarthy, M., & Carter, R. (2019). *Routledge handbook of corpus linguistics*. London: Routledge.

Moon, J. (2018). Teaching academic vocabulary to ELLs: Using corpus linguistics-inspired, high-frequency word lists and activities. *TESOL Journal*, 9(4), e3892.

Römer, U. (2011). Corpus linguistics and language pedagogy: The state of the art and future prospects. In *Corpus linguistics and second language acquisition: Research and applications* (pp. 276-287). Amsterdam: Rodopi.

Sinclair, J. (2004). *Trust the text: language, corpus and discourse*. London: Routledge.

Wible, D., & Liang, D. (2017). Corpus linguistics and language education. In A. Barrón, E. Morillo, & M. del Pilar Agustín-Llach (Eds.), *The Routledge handbook of language learning and technology* (pp. 231-246). London: Routledge.

Xia, R. (2018). A comparative study of the effects of corpus-based vocabulary instruction and traditional vocabulary instruction on college EFL learners. *Modern Foreign Languages*, 41(2), 133-143.