



CORPUS-BASED ERROR ANALYSIS IN ENGLISH

Sattarova Umida Bolbekovna

*Faculty of Philology, Jizzakh State Pedagogical University, Jizzakh,
Uzbekistan*

umidabolbekovna0902@gmail.com

Abdullajonova Khakima Abduqadirovna

*Senior lecturer at Jizzakh State Pedagogical University, Jizzakh, Uzbekistan
hakimabotirova9@gmail.com*

Abstract: This article explores corpus-based error analysis as a valuable method for recognizing and comprehending typical learner mistakes in English. Through the examination of genuine learner corpora, the research emphasizes consistent grammatical, lexical, and syntactic patterns that indicate the progression of learners' interlanguage. Using examples from English as a Foreign Language (EFL) settings, the article illustrates how corpus data offers objective perspectives on challenges in language learning. The article explores the educational benefits of corpus-based research for enhancing English language instruction and evaluation.

Keywords: corpus-based analysis; error analysis; learner corpora; English a Foreign Language (EFL); interlanguage; grammatical errors; lexical errors; language teaching.

Since it enables researchers and educators to better understand how language learners pick up a new language, error analysis has long been a key component of applied linguistics and second language acquisition studies. Error analysis has historically depended on subjective assessments and tiny learner language samples. However, the creation of language corpora has revolutionized this subject by making it possible for researchers to methodically examine enormous sets of real learner data. The study of learner errors utilizing computerized corpora and quantitative



techniques is known as corpus-based error analysis, which produces more accurate and broadly applicable results. English learners are impacted by their first language, educational background, and exposure to English in multilingual and EFL environments like Uzbekistan. As a result, trainees frequently make the same mistakes. Analyzing these trends using corpus data is crucial for spotting recurring issues and comprehending the mechanisms behind language acquisition. Corpus, which is described as a sizable, organized collection of real texts; error analysis, which focuses on recognizing and categorizing learner errors; and interlanguage, the changing linguistic system created by language learners, are important terminology pertinent to this subject.

This article takes a corpus-based technique to evaluate typical English mistake patterns and discusses their causes and consequences for teaching the language.

Theories of interlanguage development and second linguistic acquisition serve as the foundation for corpus-based error analysis. Learner errors are systematic and significant, reflecting learners' assumptions about the target language, according to Corder's early study. As corpus linguistics gained popularity, researchers started using computational tools on learner language data, which resulted in more data-driven and empirical studies. Learner output may now be compared across skill levels and with native speaker norms thanks to learner corpora like the International Corpus of Learner English (ICLE).

The capacity of corpus-based analysis to identify frequency patterns is one of its main advantages. Frequent errors in a broad corpus are more likely to indicate real learning challenges than individual errors. Typically, corpus studies concentrate on lexical errors (such as improper word choice and collocations), syntactic faults (such as word order issues), and grammatical errors (such as tense misuse and article omission). Corpus-based error analysis offers a thorough knowledge of learner language by fusing quantitative frequency analysis with qualitative interpretation.



Typical Error Patterns in Learner Data

Grammatical errors are among the most common in English learner production, according to analyses of learner corpora. For instance, because English and their native tongue differ, learners frequently have trouble using articles, especially the definite and indefinite articles. Learners may completely ignore or employ articles inconsistently in EFL contexts when they do not exist in the native language. Additionally, corpus research shows that verb tenses and subject-verb agreement are frequently incorrect, particularly in lengthy written texts. Another significant category found by corpus analysis is lexical mistakes. Learners may use improper words that are pragmatically incorrect but conceptually equivalent, such as misusing prepositions in fixed sentences or mistaking do and make.

According to corpus-based research, overgeneralization of well-known vocabulary items or straight translation from the original language are frequently the causes of such errors. Furthermore, when comparing learner corpora to native speaker corpora, collocational errors—word combinations that are not natural—are very noticeable. Syntactic mistakes, such as improper word order and sentence construction, are also common. The learner's first language's syntactic patterns frequently have an impact on these mistakes. Researchers may follow these tendencies over numerous texts using corpus data, which shows systematic transfer rather than individual errors.

Reasons for Learner Mistakes

According to corpus-based research, learner errors come from a variety of sources. First language interference, in which students apply rules from their home language to English, is a significant contributing element. Overgeneralization, in which students apply a grammatical rule outside of its proper context, is another factor. For example, students may give irregular verbs normal past tense endings. Persistent mistakes are also a result of limited exposure to real English input, especially in situations where English is primarily acquired in the classroom.



Additionally, corpus evidence suggests that error patterns might be influenced by instructional approaches. Learners may build false or incomplete representations of the language if instructional materials oversimplify some rules or do not offer enough contextualized instances.

Thus, corpus-based error analysis aids in identifying both learner challenges and possible gaps in instructional strategies.

Pedagogical Consequences and Difficulties

One prevalent misperception is that making mistakes frequently is a sign of low intelligence or inefficiency. However, mistakes are a normal and essential aspect of language development from a linguistic standpoint. This perspective is supported by corpus-based error analysis, which demonstrates that errors follow predictable patterns and progressively decline as proficiency rises. Teachers can use corpus findings to create focused instruction that meets students' real needs instead than penalizing mistakes. Practically speaking, corpus-based insights can guide the creation of materials, curricula, and evaluations. To increase knowledge of proper usage, educators might concentrate on high-frequency error categories and give students real-world examples from corpora.

Nevertheless, there are still issues, such as the requirement for teacher training in corpus methodologies and the restricted availability of learner corpora. Corpus-based mistake analysis is a potent tool for enhancing English language instruction despite these difficulties.

As this paper has shown, corpus-based error analysis provides insightful information on the trends and reasons behind English learner errors. Researchers and educators can find systematic grammatical, lexical, and syntactic issues that might not be apparent using conventional techniques by examining extensive student data. Crucially, learner errors should be seen as markers of interlanguage development and learning advancement rather than as symptoms of linguistic deficiency.



For language instruction, especially in EFL contexts, comprehending mistake patterns using corpus evidence has important ramifications. It makes it possible for training to be more data-driven and promotes a more knowledgeable and accepting attitude toward student mistakes. Corpus-based mistake analysis in spoken learner data or digital communication may be investigated in future studies, increasing its applicability in contemporary language instruction.

1.

References:

1. Corder, S. P. (1967). The significance of learners' errors. *International review Applied Linguistics*, 5(4), 161–170.
2. Granger, S. (1998). *Learner English on computer*. London: Longman.
3. Romaine, S. (1995). *Bilingualism* (2nd ed.). Oxford: Blackwell.
4. Granger, S., Hung, J., & Petch-Tyson, S. (2002). *Computer learner corpora, second language acquisition and foreign language teaching*. Amsterdam: John Benjamins.
5. James, C. (1998). *Errors in language learning and use: Exploring error analysis*. London: Longman.
6. Larsen-Freeman, D., & Long, M. H. (1991). *An introduction to second language acquisition research*. London: Longman.
7. Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.
8. Romaine, S. (1995). *Bilingualism* (2nd ed.). Oxford: Blackwell.
9. Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
10. Xu, X., & Liu, X. (2018). A corpus-based study of learner errors in EFL writing. *Journal of Language Teaching and Research*, 9(3), 482–489.