

## ERROR ANALYSIS OF ENGLISH LANGUAGE IN THE COMPOSITION OF PAKISTANI ESL LEARNERS: A CORPUS BASED STUDY

**Izzah Rashid,**

Mphil, Department of English, Minhaj University Lahore

**Mariam Saleem**

Lecturer, Department of English, Minhaj University Lahore [mariamsaleem.eng@mul.edu.pk](mailto:mariamsaleem.eng@mul.edu.pk)

### ABSTRACT

*Errors are systematic deviation from the set norms of the language. Writing encourages students to explore their imagination, fostering creative thinking and self-expression (Fitria, 2024). Writing requires students to analyze, evaluate, and synthesize information by promoting critical thinking skills. Writing for digital platforms such as blogs and social media prepares students for the digital age. Writing skills reinforce language learning, consolidating students' understanding of syntax, semantics, and pragmatics.*

*The Data collected from forty (20 male and 20 female) ESL learners of middle class constituted the corpus of the study. Corder (1967)'s Theory of Error Analysis serves as the theoretical and analytical framework. The analysis of data not only revealed the most prevalent errors in the written composition of the learners but also revealed that the frequency of errors was prevalent more in females than males. The study also proposes valuable suggestions for future researchers. Developing writing skills enables students to effectively convey their thoughts, ideas, and opinions; hence fostering academic, professional, and personal growth. Writing skills facilitate clear and coherent expression of ideas, enhancing students' ability to communicate effectively.*

**Keywords:** Error Analysis, Second Language Acquisition, Linguistic Errors, Inter-language, Frequency of Errors

### Introduction

Error analysis (EA) is a technique for assessing correctness and evaluating the language of learners. Fundamentally, learner errors are identified, described, and possibly even explained by comparing concrete learner language samples to the target language norm, as Gráf (2015) succinctly states. In the 1960s, error analysis gained prominence in applied linguistics and according to Corder (1967); examining mistakes might provide insight into what was termed a learner's "built-in syllabus" which is the organic order in which they pick up a second language. It was intended for the systematic pattern of errors to serve as proof of the system and to assist educators and researchers in creating more student-centered. Error analysis is a crucial aspect of language learning and teaching, particularly in the English language. It involves identifying, analyzing, and correcting errors made by learners in their language production. Mistakes made by students are a natural part of the learning process and are not solely attributed to the teacher, materials or students themselves (Damayanti & Listyani, 2020).

Committing errors systematically is an essential part of language learning. Learners' errors play a role in comprehending the process of acquiring a foreign language (Mirzayev & Oripova, 2022). Acquiring the skill of writing is remarkably difficult, as it entails the intricate process of generating ideas, converting them into spoken language, and structuring them into a suitable text for a specific genre. After translating ideas into written text, it is highly important to revise the write up before it is ready for final submission (Zhai & Wibowo, 2023). Ever since the evolution

of the field of second and additional language learning, errors have been an area of study (Chong & Reinders, 2020).

The field of error analysis evolved in the 1960s and 70s when Corder (1967) and his other colleagues, in response to the limitations of behaviorism-based language acquisition theories, celebrated learners' errors as a testament of interlanguage development process rather than something to be avoided and criticized. Bates et al. (1993) emphasize that "errors are positive" and suggest that instructors should adopt this viewpoint with their students. Error analysis has been a common practice in examining learner language since the 1960s, as noted by Ellis and Barkhuizen (2005). This method is extensively utilized in applied linguistics and holds significant value in learner corpus research which has a prominent focus in current Second Language Acquisition research. Corder (1967) initially introduced the mistake as a performance error that may be due to physical stress, short of memory or the lack of automatization of language items (Šebesta & Škodová, 2012) and as problems that learners are able to correct on their own. Errors are defined as systemic defects in learners' competence that learners are unable to self-correct.

According to Corder (1967), the only way to uncover the "underlying knowledge of the language to date" is for error analysts to concentrate their efforts on errors in competence but even Corder (1967) acknowledges that it is frequently exceedingly difficult to distinguish mistakes from errors. One method that might be employed for this goal is demonstrated by Ellis and Barkhuizen (2005) that involves systematically analyzing the language that learners produce, with the analyst keeping an eye on whether the learners alternate between the accurate and incorrect forms of the target language. However, according to Richards and Sampson (1974), there is an issue with this process, since it is rare for a learner to use a replacement (error) or overuse a given structure 100% of the time. The next stage of error analysis, known as error description, involves two main aspects. Firstly, it involves categorizing errors based on error taxonomy. Secondly, it focuses on recording the frequencies of errors within each error category. According to Barrot (2023), the categorization of errors can be done using two types of taxonomies: linguistic taxonomy and surface structure taxonomy.

### **1.3 Objectives**

- 1 To investigate the writing errors committed by the Pakistani ESL learners
- 2 To determine the occurrence rate of writing errors committed by the Pakistani ESL learners in relation to gender
- 3 To identify the causes of writing errors made by Pakistani ESL learners

### **1.4 Rationale of the Study**

The current study attempts to highlight the emerging issue of errors found in the writing performance of Pakistani ESL learners, as writing skill is highly demanding in all the professional chores of daily life. Error analysis helps teacher's pinpoint specific areas where students struggle and are usually found deficient in the performance areas; hence, it provides teachers an ample room for targeted interventions and feedback. By understanding common errors, teachers can refine their instructional approaches and materials. Error analysis encourages students to reflect on their writing, developing their self-editing skills and writing awareness. Error analysis helps teachers in tailoring instruction to meet the diverse needs of students. Therefore, the current study in the Pakistani perspective is not only important for English language teachers but also for all the stake holders of academia, as the study has the practical potential of suggesting the ameliorative steps in Pakistani educational arena in order to bring educational reforms.

### 1.5 Research Questions

1. Which writing errors are committed by Pakistani ESL learners?
2. What is the frequency of writing errors made by Pakistani ESL learners in relation to gender?
3. What are the causes of writing errors made by Pakistani ESL learners?

### Literature Review

Error analysis is the act of locating, evaluating, and deciphering mistakes committed in a variety of settings, including experiments, measurements, and language learning. Accuracy and comprehension in the relevant subject are improved by using it to better understand the types and origins of errors. The goal of error analysis in science and engineering is to quantify the precision of experimental findings.

Language scientists and educators can better comprehend the precise reasons and trends behind language acquisition failures with the use of these kinds. The principal varieties are transfer errors occurring between languages. It is the learner's native language (L1) that causes interference with these errors. There are many mistakes made by learners when they transfer structures or principles from their first language (L1) to their second language (L2). For example, due to direct translation from Spanish, a learner of the language would say "He has 12 years rather than "He is 12 years. Mistakes Made in Other Languages Instead of interference from L1, these errors originate from the target language (L2). These frequently arise as a result of learners' overgeneralization of L2 rules or their poor comprehension of them.

Chomsky (1959) criticized contrastive analysis for language learning as habit formation process. Ellis (1994) claimed that children learn their native language as mental rules not as a set of habits. Chomsky's main focus was on universal grammar as every learner has a capacity to learn any language rules. Selinker and Gass (2008) also claimed language learning as a process of rule formation not as habit formation. This attitude of critics rejected contrastive analysis as its focus was on repetition and reinforcement. Dulay and Burt (1974) claimed in their study that only five percent of errors are due to first language interference. In addition, contrastive analysis was also rejected due to its ignorance of factors as learners' production and over-generalization (Keshavarz, 2003). Corder (1967) made a clear distinction between error and mistake. He describes errors occurrence as a result of learners' incomplete learning and lack of linguistic competence and self-correction of errors by learners is not possible. Learners commit mistakes due to poor performance. This poor performance of language learners may be due to many aspects like fatigue and negligence of learners etc. Learners have the awareness of the correct linguistic usage and mistakes can be self-corrected. Error leads to communication breakdown in spoken and written discourse and results in misunderstanding whereas mistakes do not cause any communication breakdown and learner can correct it himself in most cases.

Lee (2006) reconsidered the nature of error taxonomies in his research study on students of California. He discussed a learner-oriented taxonomy for errors committed by EFL learners and divided this taxonomy into three sub-categories; morphology, syntax and others. Morphological errors enjoy the concrete and transparent position. Otherwise, the errors are placed into syntactic errors. For example, I want to be a international lawyer. In it, 'a' instead of 'an' is categorized into

a morphological error.

According to Ellis and Barkhuizen (2005), four steps (data collection, error identification, error description, and error explanation) cover up the process of error analysis. Getting data is the easiest to understand. First, learner language samples are gathered. Nevertheless, natural language production would afford the best data for examining learner's competence according to SLA researchers. Clinically-elicited data are prioritized (Ellis & Barkhuizen, 2005).

### **Inter-Language Theory**

Selinker (1972) proposed the idea of Inter-Language which asserts a separate linguistic system independent of target language and native language. Tarone (1988) refers to Inter-Language as a system having their own significant internal consistency. It is a gradual developmental process which travels from mother tongue/native language to target language building an independent linguistic system called Inter-Language. It is neither L1 nor L2. Selinker (1972) suggests that in language learning process, it is observable and exportable. The emergence of Inter-Language theory also gave birth to psychological shift from behaviorism to mentalist. Inter-Language theory seems to borrow its assumptions from mentalist. (Tarone, 2001). During language learning process, learners formulate some hypotheses regarding the rules of target language. These rules are called mental grammars leading to Inter-Language system. This system is variable allowing language learners to add or delete rules for reconstructing an organized system, Inter-Language System. Khansir (2012) asserted that for second language competence, for making the teaching process easier, these three theories are inter-connected.

Abisamra (2003) found 214 errors in his research study that was conducted on secondary level ESL students. He categorized these errors into 29 grammatical, 120 spelling errors, 26 lexical, 35 syntactic and 3 semantic. Papp (2004) conducted a study on Chinese ESL students at tertiary level. The research tool was written composition. It was corpus-based study. It detected that Chinese faced a great difficulty in the use of articles. Tan (2005) conducted a research study on student of Kun Shan University at tertiary level. Amiri & Puteh (2017) conducted a study on 16 Iranian students at post-graduate level. They had to undertake an Intensive English Course for entering into their concerned academic disciplines. This study studied their term papers and Corder model of Error Analysis was applied. They found the errors description in four different categories: 20 % content based, 20 % language based, 6 % organization and format based, 4 % surrounded the bibliography list category. The most common errors were in the category of sentence structure, articles, punctuation and capitalization. This study categorized the errors into intra-lingual and developmental phases. Darus & Subramaniam (2009) traced errors committed by 72 Malay ESL students at secondary level through essay writing. Corder model of error analysis was utilized. The tool for analysis of data was Markin software. The most common errors found in this study were singular/plural level having 5.72 percentages, Verb Tense with 4.80 percentages, 4.51 percentages weigh word choice, 4 % in preposition, subject-verb agreement with 3.01 % and word-order carries 2.99 percentages. The effective teaching material can be planned through utilizing this study.

### **Methodology**

The current study employed Corpus-based descriptive and exploratory research design, as this design is helpful for understanding the phenomenon of the study under discussion.

### **Data Collection and Analysis Procedures**

The data was collected from the forty students selected through systematic



randomization sampling technique and the collection of corpus of written texts (essays) from Pakistani ESL students was handled by using a standardized writing prompt to ensure consistency. Systematic randomization sampling is a technique in research that involves selecting a sample from a population in a systematic and random manner. This technique is preferred, as a random number is generated to determine the starting point for selecting the sample and the researcher selects every  $n$ th element from the list, where  $n$  is a fixed interval (e.g., every 10th element). Systematic randomization sampling offers several benefits, including the sample is likely to represent the population's diversity. The systematic selection process minimizes bias and ensures equal probability of selection for each element. This technique is simple to apply, especially when working with large populations.

Systematic randomization sampling can be more efficient than other sampling techniques, such as simple random sampling. Then the collected data will be converted into corpus using Optical Character Recognition (OCR) technologies for cleaning and normalizing the data. OCR enables quick and accurate extraction of text data from images, scans, or PDFs, saving time and effort. This technology reduces errors associated with manual data entry, ensuring higher accuracy and reliability. OCR-processed texts become searchable, facilitating keyword searches and analysis and it helps digitize and preserve historical documents, making them more accessible for research. OCR enables automated data processing, reducing the need for manual data entry and increasing productivity and reduces the need for manual data entry, resulting in cost savings.

OCR facilitates quick analysis of large volumes of texts, accelerating literature reviews. It also enables researchers to analyze and mine large datasets, revealing patterns and insights and helps make texts more accessible for researchers with disabilities. By leveraging these benefits, OCR technology significantly enhances research efficiency, accuracy, and productivity across various disciplines. After compiling the corpus, Corder (1967)'s error analysis framework will be used to identify the types of errors e.g. grammatical errors (morphology, syntax, verb tense, etc.) and Lexical errors (vocabulary, word choice, collocations, etc.). For the compilation of the corpus, AntConc 3.2.2 was used. This software has several benefits in research, as it provides an intuitive interface that makes it easy to use for researchers, teachers, and learners. AntConc is available for free, making it an accessible tool for researchers. AntConc processes data quickly and accurately, even when handling large datasets. It can be used on various platforms, making it a versatile tool for researchers.

The Concordance Tool in AntConc offers a range of features that make it an effective tool for data analysis and visualization. It enables researchers to analyze and visualize corpus data, facilitating insights into language patterns and trends. AntConc provides data visualization capabilities, making it easier for researchers to present and communicate their findings. It is designed to be user-friendly, reducing the technical barriers for researchers who may not have extensive technical expertise. AntConc can be used for various research purposes including language learning, teaching, and corpus linguistics research. Descriptive statistics was applied using SPSS version 21 for computing the error types and frequencies, as it can find differences in the form of numbers. By using Descriptive statistics, social sciences researchers can gain a better understanding of complex phenomena, make informed decisions, and develop evidence-based policies.

### **Corder Model's Outline**

The complete framework is provided below:

### **4.3.2 Stages in Pit Corder's Error Analysis**

#### **4.3.2.1 Compilation of Learner Language sample**

Firstly, the collection or gathering of either written or spoken sample of a learner's language is needed.

#### **4.3.2.2 Recognition of errors**

Secondly, a comparison of the learner language sample with the target language is required. This comparison helps in evaluating the point where the deviations and divergences occur.

#### **4.3.2.3 Description of errors**

This involves the classification of errors into distinct subtypes like semantics, Syntax or morphology.

#### **4.3.2.4 Explanation/ justification of errors**

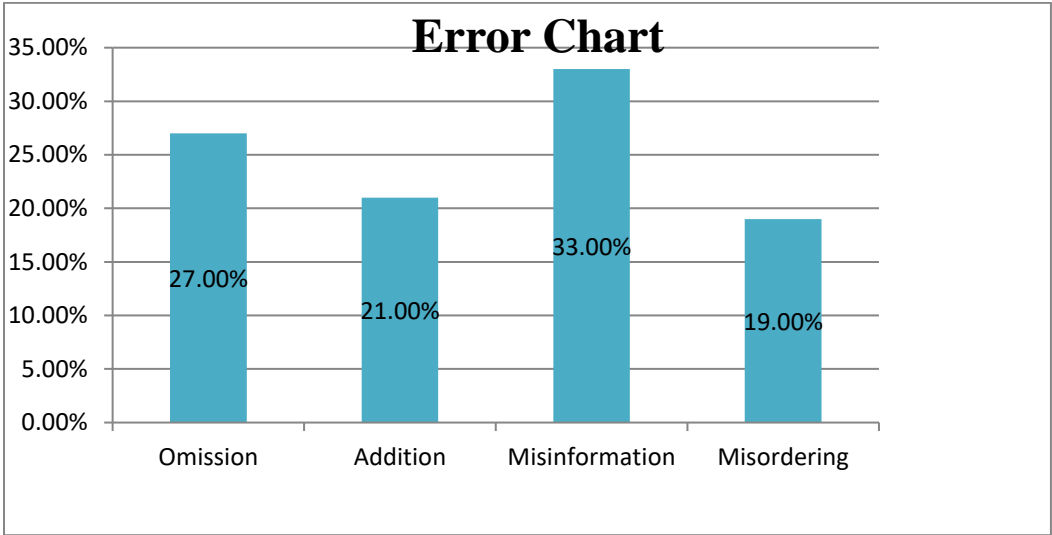
Explanation of the errors aims to see why and how the error has been occurred. The factors like Interlingua and intralingua are being checked in this stage.

#### **4.3.2.5 Evaluation of errors**

Last stage is characterized by examining the influence of the errors on the communication. The fact that which error must be corrected first is also evaluated in this stage.

<b>Type of Error</b>	<b>Error</b>	<b>Correction</b>
<b>Omission</b>	Rainy season starts July and lasts up to September.	Rainy season starts from July and lasts up to September.
	Rain is blessing God.	Rain is a blessing of God.
<b>Addition</b>	They do not goes to school.	They do not go to school.
	Rain may cause of flood.	Rain may cause flood.
<b>Misinformation</b>	She comed to see me.	She came to see me.
	I taken tea after the rain.	I took tea after the rain.
<b>Disordering</b>	Why he is not coming?	Why is he not coming?
	I take tea after rain usually.	Usually, I take tea after rain.

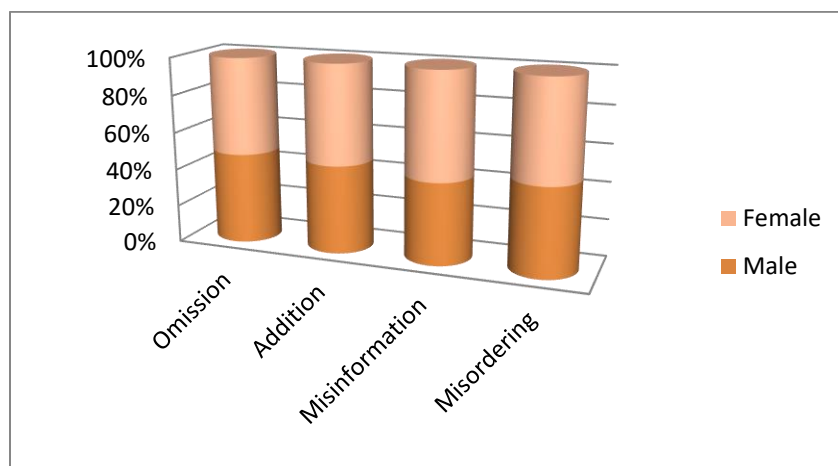
### **4.3.4 Frequency of Errors**



From the above table and chart, it’s clear that error of misinformation outweighs all the types of errors. However, gender wise analysis of the types of errors is presented below.

**Errors frequency in given samples with respect to males and females:**

Types of Errors	Male	Female
Omission	62	66
Addition	48	54
Misinformation	68	88
Disordering	40	50



#### 4.4 Causes of Errors

Explaining the nature of errors is a central issue in second language acquisition process. Ellis and Barkhuizen (2005) state that "explaining errors involves determining their sources in order to account for why they were made". According to Ellis (1994), explanation of errors involves an attempt to establish the processes responsible for L2 acquisition and also traced out the psycholinguistic sources of errors by categorizing them into errors of competence and errors of performance. Second language learners commit errors of competence when they apply the rules of target language incorrectly whilst learners commit errors of performance when they commit mistakes repetitively in language use. The researcher has collected primary data from its samples of the study. Therefore, after reviewing considerable amount of literature and describing research participants' responses, he found the following sources of errors in his study.

#### Conclusion

The research participants committed overt errors more frequently than covert errors. The number of overt errors and covert errors were 622 and 162 respectively. The overt errors found in this research study were about subject-verb disagreement, errors while handling transitive and intransitive verb, errors whilst handling modal verbs, miss election of appropriate lexicon, spellings, capitalization, punctuation, errors in syntactic structures. Their percentage were 25%, 30%, 20%, 10% and 15% respectively. While handling overt errors, EFL learners committed error more frequently in the use of transitive and intransitive verb and subject-verb agreement simultaneously. On the other hand, the research participants committed 20% covert errors in their written productions. On the whole, the research participants committed 50% errors on linguistic level. It was found that 65% research participants handled statements related to linguistic taxonomy correctly. It was so as they have been learning English for ten years. As far as the surface structure taxonomy is concerned, ESL learners committed 16% errors. This taxonomy analyzed the errors on morphological, syntactic and lexical level. The ESL learners committed more errors on linguistic level than surface structure taxonomy errors, as learners have to face more problems



while handling linguistic categories. The main sources of errors found in this research study are interlingua factors (interference of native language), intralingual factors (overgeneralization, ignorance of rule restriction, incomplete application of rules and false concept hypothesis).

## References

- Al-Khasawneh, F. (2014). Error analysis of written English paragraphs by Jordanian undergraduate students: a case study. *International Journal of English Language, Literature, and Humanities*. 112-132.
- Almusharraf, N., & Alotaibi, H. (2020). Gender-based EFL writing error analysis using human and computer-aided approaches. *Educational Measurement: Issues and Practice*, 40(2), 60–71. doi:10.1111/emip.12413
- Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, 100-107.
- Burt, C. (2015). *Communicative methodology in language teaching. The roles of fluency and accuracy*. Cambridge: Cambridge University Press.
- Chong, S. W., & Reinders, H. (2020). Technology-mediated task-based language teaching: A qualitative research synthesis. *Language Learning and Technology*, 24(3), 70-86.
- Choudhary, A., & Arora, A. (2021). Linguistic feature based learning model for fake news detection and classification. *Expert Systems with Applications*, 169, 114-128.
- Corder, P. (1975). The language of second-language learners: The broader issues. *The Modern Language Journal*, 59(8), 165–183.
- Damayanti, M. E., & Listyani, L. (2020). An analysis of students' speaking anxiety in academic speaking class. *Eltr Journal*, 4(2), 152-170.
- Dörnyei, Z. (2016). *Research methods in applied linguistics: quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press.
- Dušková, L. (1969). On Sources Of Errors In Foreign Language Learning. *International Review of Applied Linguistics in Language Teaching*, 7(1).
- Ellis, R., & Barkhuizen, G. P. (2005). Analysing learner language. *International Journal of Applied Linguistics and English Literature*, 8(3), 93-103.
- Gráf, T. (2015). Accuracy and fluency in the speech of the advanced learner. *Journal of Language Research*, 18(3).
- Granger, S. (2003). Error-tagged Learner Corpora and CALL: A Promising Synergy. *CALICO Journal*, 20(3), 465-480.
- Granger, S. (2013). Learner Corpora and CALL. *CALICO Journal*, 30(1), 187-198.

- Fitria, T. N. (2024). Creative Writing Skills in English: Developing Student's Potential and Creativity. *EBONY: Journal of English Language Teaching, Linguistics, and Literature*, 4(1), 1-17.
- Jain, S. (1974). *A Comprehensive Grammar of the English Language*. New York
- Khoso, A.R., Pathan, H., & Shah, S.W. (2018). Identifying the Grammatical Errors Committed by EFL Learners: A study of Public Sector Universities Sindh, Pakistan. *Journal of Literature, Languages and Linguistics*, 47, 66-71.
- Lee, S. M. (2023). The effectiveness of machine translation in foreign language education: a systematic review and meta-analysis. *Computer Assisted Language Learning*, 36(1-2), 103-125.
- Magsi, S. A., Khaskheli, P. N., Soomro, A. R., & Lashari, A. A. (2023). Error Analysis In Academic Writing Of Post Graduate Engineering Students Of Sindh. *Journal of Positive School Psychology*, 1185- 1191.
- Maitlo, S. K., Soomro, A. R., & Lashari, A. A. (2023). The impact of picture series learning on the creative writing skills of ESL learners. *Global Digital & Print Media Review*, VI, 211-223.
- Mirzayev, A., & Oripova, S. (2022). Communicative method—a new approach in the practice of teaching foreign language. *Science and innovation*, 1, 778-783.
- Phuket, P.R., & Othman, N.B. (2015). Understanding EFL Students' Errors in Writing. *Journal of Education and Practice*, 6, 99-106.
- Poehner, M. E., & Wang, Z. (2021). Dynamic assessment and second language development. *Language Teaching*, 54(4), 472-490.
- Qizi, E. Y. A. (2023). Learning foreign languages is vital in the development of the professional competence of future engineers. *Строительство и образование*, 4(6), 147-153.
- Ramzan, M. et al.(2023). Subject-verb agreement errors in ESL students' academic writing: a surface taxonomy approach. *Linguistic Forum*, 5, 16-21.
- Richards, J.C., & Sampson, G.P. (1974). The Study of Learner English In J.C. Richards (Ed.), *Error Analysis: Perspectives on Second Language Acquisition*. Great Britain: Longman.
- Šebesta, K., & Škodová, S. (Eds.). (2012). *Čeština - cílový jazyk a korpusy*. Liberec: Technická univerzita v Liberci. *Harf-o-Sukhan*, 7(3), 101-113.
- Sermsook, K., Liamnimitr, J., & Pochakorn, R. (2017). An Analysis of Errors in Written English Sentences: A Case Study of Thai EFL Students. *English Language Teaching*, 10(3), 10-31.

- Schachter, J. & Murcia, M. (1977). Some Reservations Concerning Error Analysis. *TESOL Quarterly*, 2, 441-451.
- Shadiev, R., & Feng, Y. (2024). Using automated corrective feedback tools in language learning: a review study. *Interactive learning environments*, 32(6), 2538-2566.
- Siddiqui, K. A., Abbasi, R. H., & Soomro, A. (2023). Paragraph organization errors in the writing of Pakistani college-going students: An error analysis study. *Academy of Education and Social Sciences Review*, 3(2), 131-139.
- Stojanov, A. (2023). Learning with ChatGPT 3.5 as a more knowledgeable other: An autoethnographic study. *International Journal of Educational Technology in Higher Education*, 20(1), 35-49.
- Thomas, J. (2017). *Discovering english with sketch engine*. versatile.
- Zhai, C., & Wibowo, S. (2023). A systematic review on artificial intelligence dialogue systems for enhancing English as foreign language students' interactional competence in the university. *Computers and Education. Artificial Intelligence*, 4, 100-134.
- Zheng, G., & Park, T. (2013). An analysis of errors in English writing made by Chinese and Korean university students. *Theory and Practice in Language Studies*, 3, 1342-1351,