A Computer-Assisted Applied Corpus Analysis of Writing Errors Made by Senior Students of The English Language Department Faculty of Arts at Tanta University

Moustafa Mohammad Shalabi[1*]

Alkhaleej Training and Education at King Saud University

mostafa.shalaby1970@gmail.com

Mohammed Said Negm, Professor of linguistics, Tanta University

ngm_msaid@hotmail.com

Hani Mohammed Helmy, Professor of English Literature

Tanta University, Faculty of Arts

hany.mohamed@art.tanta.edu.eg

Abstract

The dissertation investigates writing errors among Senior Students in the English Language Department using computer-assisted corpus analysis. It aims to explore various types of errors in Lexical, Syntactic, Semantic, and Mechanical aspects. A mixed approach was used, using virtualwritingtutor.com to identify different categories of errors. The study aims to uncover prevalent errors in English writing and compare and evaluate errors based on gender (male and female). The Ant Concordance Tool software was used to identify common sources and causes of errors. A quantitative analysis was conducted using SPSS with a sample size of 494 participants. Results showed that male students exhibited significantly fewer errors in syntactic, lexical, and semantic aspects, but not in mechanical aspects. The study highlights the importance of understanding and addressing writing errors among senior students in the English Language Department.

Key Words

Computer-Assisted Applied Corpus Analysis, Writing Errors, Interlanguage, Interalanguage, Language transfer.

Introduction

The dissertation investigates writing errors among Senior Students in the English Language Department using computer-assisted corpus analysis. It explores Lexical, Syntactic, Semantic, and Mechanical aspects. A mixed approach was used, using virtualwritingtutor.com to identify different categories of errors. The first chapter provides an overview of the research background, problem statement, research objectives, and connection between objectives and research questions. The dissertation holds significant theoretical, pedagogical, and practical implications for students and educators.

Writing is a challenging task, especially for second and foreign language learners, and is considered the most challenging language ability to master (Shawish, 2010). It is a progressing advanced skillfulness compared to other languages, making it a significant challenge for language learners at all stages (Erkan & Saban, 2011),(Abdel Latif, 2015). Writing skill is crucial for language learners as it reflects thoughts, ideas, and goals, reinforces syntax, diction, and idioms, develops personality, and helps readers understand

easily (Wright, 2011). It also refines oneself through feedback and explains intellectual flexibility and maturity through self-expression(Nadira, 2012).

Arab students studying English face difficulties in writing due to limited vocabulary, cultural influences, and limited practice (Huwari et al., 2011); Al-Sawalha & Chow, 2012). A study at Al-Azhar University found that most students felt confused, discomposed, and struggled with choosing topics, developing ideas, and ending articles. They also lacked procedural skills to write acceptable English compositions, with repeated ideas, errors, and irrelevant information Salem, (2007).

Language Use and Computer-Assisted Corpus Analysis (CACA)

Corpus linguistics is a growing methodology analyzing large collections of written or spoken data to investigate linguistic phenomena. It is based on real-life language use models and is collected, stored, and processed through computer systems. Sinclair, (2005) states that "Corpus refers to a collection of electronic text, selected for linguistic research, representing a language or language variety".

This study examines Senior Students' writing errors using computer-assisted corpus analysis at Tanta University. It aims to identify the most common errors and their causes as nonnative English speakers. The study utilizes corpus linguistics tools to save time and effort, focusing on real language process systems Leech, (2014).

Statement of the Problem

The research investigates writing errors made by target students using Computer-Assisted Applied Corpus Analysis. Writing is a challenging task, especially for second and foreign language learners, and is considered the most challenging language skill to master (Erkan et al., 2011). In Egypt, students may struggle to learn a second language because they are taught in their first language Richards & Schmidt (2002). This can lead to students borrowing forms and patterns from their first language, and using words and grammar that they already recognize. This can make it difficult for students to learn the new language's grammar and vocabulary.

The Objective of the Study

Arab and Egyptian students studying English encounter writing difficulties. It is crucial to analyze their writing errors using computer-assisted applied corpus analysis to understand the causes and find remedies. In Egypt, English is taught as a foreign language from an early grade. Utilizing technology enables a detailed exploration of linguistic behavior and facilitates effective communication.

Research Questions

The present study employs a mixed-methods approach, combining qualitative and quantitative research methodologies. Aligned with the predetermined objectives, the research inquiries are formulated.

1. Which types of errors, including lexical, syntactic, semantic, and mechanical errors, are frequently observed in the L2 writing of senior students studying English at the Faculty of Arts, Tanta University?

2. What are the prevailing errors that occur most frequently among these students?

3. Is there a correlation between the overall number of errors (lexical, syntactic, semantic, and mechanical) and the gender of the students?

4. What are the primary causes underlying these errors?

Significance of the Study

The answers to these questions aim to provide effective approaches for improving professional writers' expressions, stimulating their thinking, and serving as a reference for public writing. The findings will be valuable to the Egyptian Ministry of Higher Education and educational institutions in Egypt, enhancing understanding of writing errors and offering feedback on the challenges faced by Egyptian students in their writing.

Method of the Study

According to Kim, (1998), genre analysis in written texts reflects how people effectively use language in specific settings to accomplish tasks. In the case of written thesis procedures prepared by professors in universities, genre serves as a means of communication. Bhatia, 2008 emphasizes the need for analytical perspectives in analyzing professional genres to demonstrate knowledge transfer. Hyland, (2003) suggests that genre analysis can help L2 students understand the social relations that genres establish. Jones, (2006)explores ethnographic investigations involving interviews with individuals who read or write these genres, providing L2 students with access to authentic contexts and cultural practices. The establishment of a Corpus of written texts allows for systematic tagging of moves within the texts, visually representing the entire tagged documents and facilitating the identification of recurring moves.

Data Collection and Analysis

Written communication plays a vital role in daily interactions, fulfilling diverse language needs. Errors are common, particularly in lengthy English essays. This study analyzes 494 essays from 494 participants out of 700. Participants freely choose their topics and wrote essays of approximately 150 to 200 words within an hour, without the aid of dictionaries or smartphones. All participants have a non-English speaking background and limited English communication outside the university. Handwritten essays were collected, compiled into a MS Word file (with no "auto correct" enabled), and later processed in AntConc after conversion to a notepad file.

Literature Review

The literature review encompasses various aspects of corpus linguistics, including text collection, manual copra, corpus-based approach, representativeness of corpora,

computerized corpora, uses of corpora, corpus processing, error analysis, contrastive analysis, and a critical examination distinguishing between errors and mistakes.

The Arabic Language

Arabic primarily serves as the official written language, often used for communication among those who study and learn it, but it is predominantly utilized in written form. It serves as the standard language across the Arab world, spanning from Morocco to Saudi Arabia. The Arab population speaks various related dialects, including Egyptian-Sudanese, North African, Syro-Palestinian, Iraqi, Arabian, and Northwest African (Imaniar, 2018). The differences between these dialects are more pronounced compared to the variations between UK, USA, and Australian English. Additionally, Arabic holds significance as the language of the Holy Quran, making it familiar to Muslims worldwide who can even recite passages from it.

Linguistic Overview and Theoretical Background

"Thirty years ago, when research of corpus analysis started it was considered impossible to process texts of several million words in length. Twenty years ago, it was considered marginally possible but lunatic. Ten years ago, it was considered quite possible but still lunatic. Today it is very popular" (McEnery and Wilson 2001). Figure 1-2 demonstrates the rise of Corpora Studies.
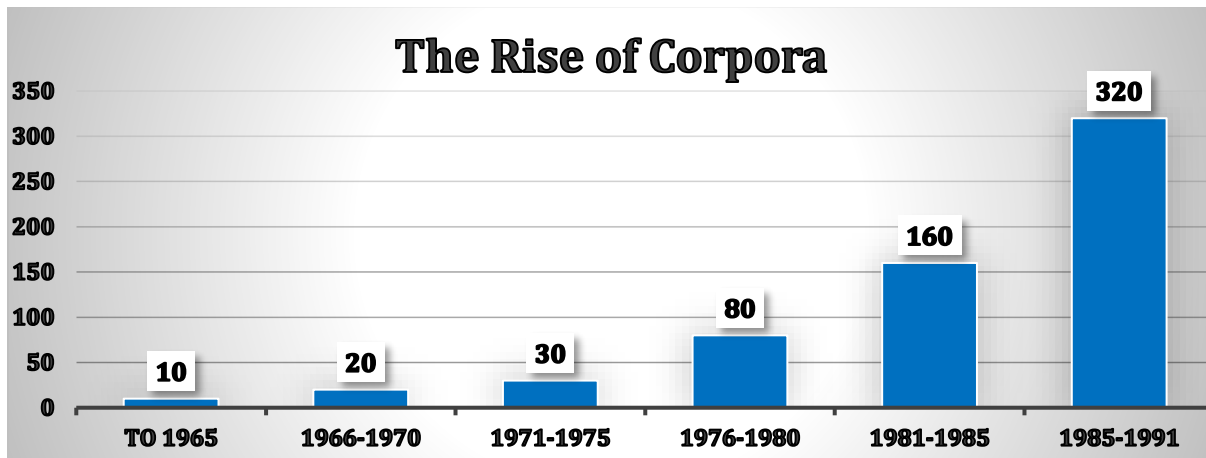

Figure 1-2 Demonstration of the rise of corpora Studies (McEnery and Wilson 2001)

Corpus Linguistics

The term "Corpus" comes from the Latin word for "body" and can refer to any context in spoken or written form. However, in modern Linguistics, it refers to large collections of texts that represent a particular variety or usage of language that can be accessed in machine-readable form (Hunston, 2006) . A Corpus is an electronic collection of examples of authentic language, usually consisting of full texts or large extracts from lengthy texts. The main objective of a corpus is to test hypotheses about language and to demonstrate how

to use a particular sound, term, expression, or grammar structure. Corpus linguistics deals with the principles and practice of using corpora in language study(Crystal, 1992).

Written Discourse

Writing is a common activity for learners of English as a foreign language (EFL) or English as a second language (ESL), whether they are native or non-native speakers. Written communication plays a crucial role in everyday conversations, as governments and societies rely on printed media and documents. According to Grabe and Kaplan, (1996)., written discourse is an important aspect of linguistic requirements in daily life. However, the ability to write is not necessarily as developed as spoken language skills. Kaplan and Grabe, (1996) emphasize that writing is a socially mediated activity and a set of skills that require practice and continuous improvement.

Text Collection

A corpus is a collection of texts, both spoken and written, that have been created and compiled based on specific criteria. The term "corpus" has been used since the 16th century to refer to a collection of real texts Francis (1992). However, from a linguistic perspective, simply any collection of texts cannot be considered a corpus. It must be a collection of texts that is representative of a particular language, dialect, or subset of a language, and is used for linguistic analysis. Similarly, Sinclair defined a corpus as a collection of naturally occurring language texts that are chosen to represent a specific state or variety of a language.

Manual Corpora

Arabs were the first to use corpus linguistics, first using it manually when Al-Khalil created a dictionary. Sibawayh later used a corpus to create rules and explain linguistic devices, (Eliwa, 2004). Although deeply rooted in Arab linguistics, most studies focus on English. Early corpus linguistics before Chomsky was exclusively corpus-like, as methodological pedagogy was based on observed language usage, (McEnery, 2016).

Corpus-Based Approach

The corpus-based approach has received widespread recognition and has been applied to nearly all areas of linguistics; for a general overview, see T. McEnery, Xiao, Tono, et al. (2006). However, corpora have also come under special criticism. For instance, Chomsky (2002) asserted that while language is "an infinite set of sentences," a corpus only contains a small number of sentences. For a critique of that concept of language, read Beaugrande and Williams (2002), and for the discussion of early generative linguistics, see article 2 in this book. Because a corpus does not include every possible sentence in the language, corpus data is by nature incomplete.

Corpora and Representativeness

Structuralists and generative linguists in the 20th century emphasized the importance of studying language usage (parole, performance) and language system (langue,

competence). Saussure and Chomsky believed that generalizations about language require studying language forms, regularities, and rules. Leech (1991) defines a corpus as typical of a given language variety if its conclusions can be applied to the whole language variety. Biber (1993) argued that sampling is necessary for the creation of any corpus, as a corpus is always a sample of a language or language variety.

Computerized Corpora

Before computers, analysts had to manually extract linguistic statistics, a time-consuming process. With the introduction of computers, software tools automate linguistic record extraction. The Brown corpus, introduced by Francis and Kucera in 1961, became the first machine-readable corpus of spoken English, establishing a public recognition and exploitation of corpora.

Uses of Corpora

Corpora and concordances are crucial tools for teaching and learning second languages. Working with corpus data, whether directly accessible or selected by instructors, can enhance learners' awareness of form and meaning, as well as word-context relationships. Johns' "learner-as-researcher" metaphor highlights the importance of this approach, as it allows learners to independently discover grammar and meaning, taking on increased responsibility for learning while the instructor serves as a studies coordinator (O'Sullivan, 2007).

Corpus Processing

A computer corpus is a huge quantity of natural sources computer-readable texts or text-extracts that can be used in research, particularly corpus studies, the current study was built on nearly about 494 essays of senior students' writings at Tanta University nearly about 93.324 words that would be used to serve the study and answer the research questions. AntConc's Concordancer Tool contains a variety of features that make it a useful tool not just for students, but also for teachers and researchers.

Error Analysis and Constructive Analysis

Error Analysis

Error analysis studies linguistic errors, categorizing them based on proficiency, linguistic levels, form, type, and cause. Learner errors are often caused by incorrect conclusions about new language principles. Interlanguage, a psychological structure triggered by learning a second language, can lead to communication strategy-based errors. Language transfer, also known as L1 interference or linguistic interference, refers to speakers applying knowledge from their native language to a second language. Positive transfer occurs when speakers and writers transmit items and structures different in both

languages, while negative transfer occurs when speakers and writers transfer items and patterns different in both languages. Negative transfer has been a common cause since the 1960s.

Contrastive Analysis

Larsen-Freeman (1991) highlights the progression of second language acquisition research through phases like contrastive analysis, error analysis, performance analysis, and discourse analysis. Error analysis became prevalent among researchers, and behaviorism, a prominent movement from the 1940s to the 1960s, is closely linked to contrastive analysis. Behaviorists believe learning is advanced by establishing stimulus-response conditions, new habits, and reinforcing existing links. Contrastive analysis theory appeals to researchers due to the potential for more effective pedagogy.

Errors Versus Mistakes

Errors are distinct from mistakes or slips in that they are not random guesses or slips. Mistakes are failures to use a recognized system, while errors are variations from the native speaker's adult grammar, demonstrating the learner's interlanguage ability. Beginners' errors differ from advanced learners', representing their proficiency level. James (1998) emphasizes the concept of intentionality, stating that errors are unintentionally aberrant language that can be self-corrected by the writer. Errors require additional learning before they can be self-corrected.

Edge (1989) suggests that errors are something that learners cannot self-correct, while slips can be self-corrected. Corder (1967) argued that errors are indicators of learners' efforts to understand an underlying rule-governed system, while mistakes are careless slips of the tongue. Chomsky's distinction between performance and competence exemplifies this point.

Significance of Errors

Errors in language learning are now considered markers of the learning process and require investigation. They inform teachers, researchers, and students about the learning process. Error analysis can help predict errors and develop teaching materials and methods. These findings can be applied to second language acquisition and language learning methodologies. Errors can be beneficial tools for both teachers and students, enabling them to recognize errors and improve their language skills. Teachers should be aware of the challenges of learning a foreign language and consider the probable causes of errors to create a beneficial approach.

Sources of Errors

Writing in a foreign language (EFL) is a challenging task, and researchers have sought to identify common errors made by EFL students. This awareness helps teachers recognize students' challenges and implement appropriate teaching methods. Error analysis (EA) is a crucial tool in language teaching, reorganizing perspectives and reevaluating methods to

fill gaps in students' knowledge. EA involves collecting samples of learner language, recognizing errors, describing them, classifying them, and assessing their significance. The objective is to determine what the learner knows and does not know, enabling teachers to provide the right information and data to help them form a sufficient understanding of the target language.

Participants

This research involved a group of 494 high school students, aged between 20 and 23, who are currently enrolled in the Faculty of Arts at Tanta University. These students specifically belong to the English department within the College of Arts, and comprise both male and female individuals. In Egypt, they have received a similar level of education, primarily through university studies. For the purpose of this study, the researchers collected a dataset consisting of 500 essays, each ranging from 150 to 200 words in length. These essays were written by the participants, both males and females, as part of a free writing activity. It is worth noting that all subjects, except for English, were taught in Arabic. The participants had been studying English for approximately 12 years, attending six English sessions per week, with each session lasting 45 minutes. Arabic is the first language for all the participants. During the study, each student was instructed to choose a topic they felt most comfortable writing about.

Analysis of Data Driven Learning (DDL)

Corpus is a collection of texts with specific connotations in modern linguistics, categorized into four classifications by McEnery and Wilson (1996): sampling and representativeness, finite size, machine-readable form, and a standard reference. It emphasizes that a corpus is essentially a computer-stored collection of written, spoken, or mixed texts, offering the advantage of compiling and analyzing large amounts of text with specialized software (O'Keeffe and McCarthy, 2010). This computer-stored data is accessible to researchers, educators, and textbook authors.

In the context of language teaching, corpus linguistics has shown limited impact, sparking a decade-long discussion on the role of corpora in English language teaching (ELT). Changes influenced by corpus linguistics have gradually integrated into ELT practices, in contrast to its revolutionary impact on EFL lexicography (Ranalli, 2003). The text concludes by noting that the present study aims to create a representative corpus tailored to specific research goals, consisting of texts produced by senior students in the English Department of the Faculty of Arts at Tanta University.

The findings of language corpora can be employed as a base for data driven learning (DDL). In other words, the data from the language corpus through utilizing the "AntConc" concordance as shown in (Figure 3.3) and the free website "virtual writing tutor", / virtualwritingtutor.com / as shown in (Figure 3.4) are used to provide insights on the university students, Faculty of Arts, department of English capability to avoid the errors in

their writings and how they can improve their writings, Consequently, the expected findings can be useful for both lecturers and textbook writers correspondingly.
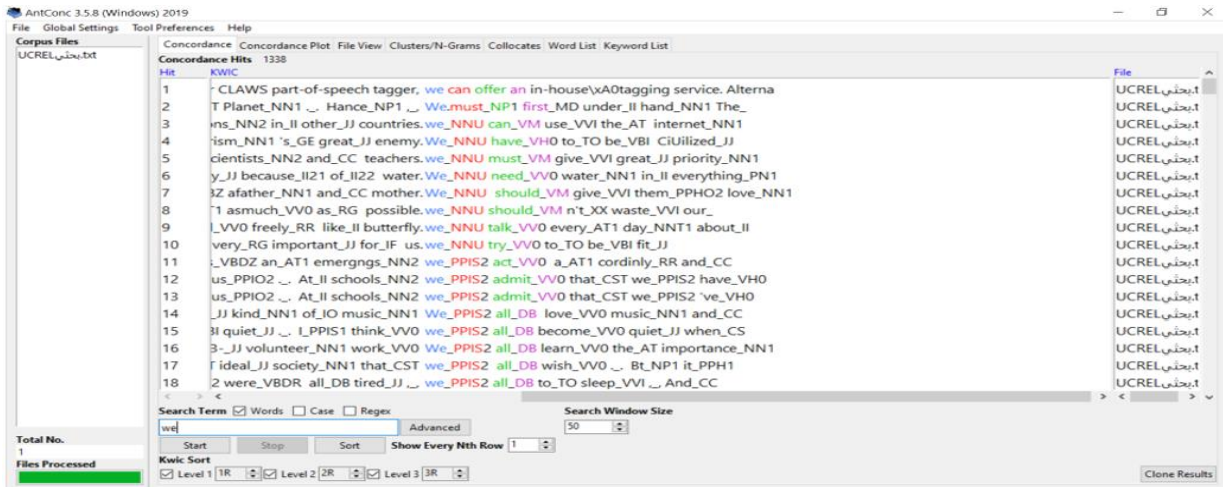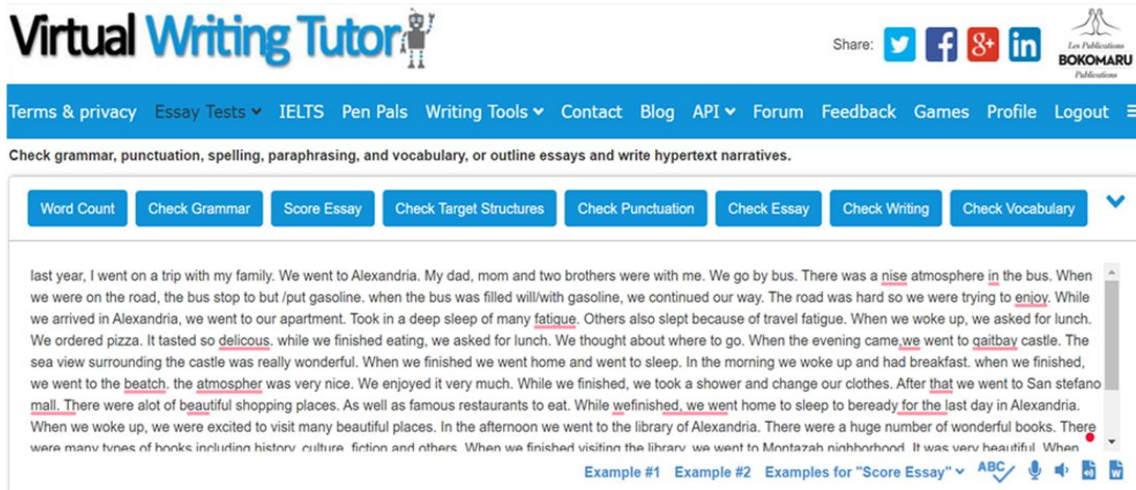

Figure 3-3. "AntConc" Concordancer Tool

The inclusion of a concordancer is a crucial feature found in most corpus analysis software, such as AntConc. Concordancers have proven to be valuable tools in second language acquisition, aiding in the learning of vocabulary, collocations, syntax, and writing styles, as noted by Sun and Wang (2003). To effectively learn new vocabulary, for instance, it is necessary to encounter words in various natural contexts and situations (Cobb, 1999). Relying solely on intuition makes it practically impossible to find a sufficient number of examples for a given word or phrase that meet these conditions. In contrast, a concordance program can swiftly and efficiently locate and display numerous examples from diverse settings and scenarios using a reasonably large corpus. Figure 3.1 provides a screenshot of AntConc, illustrating the use of the Concordancer Tool. This tool, along with others in the program, is designed to have the most commonly used operations readily accessible on the

main screen. This approach to software design, as highlighted by C. Lonfils and J. Vanparys (2001), is crucial for ensuring user-friendly functionality, eliminating confusion caused by excessive pull-down menus and additional windows.
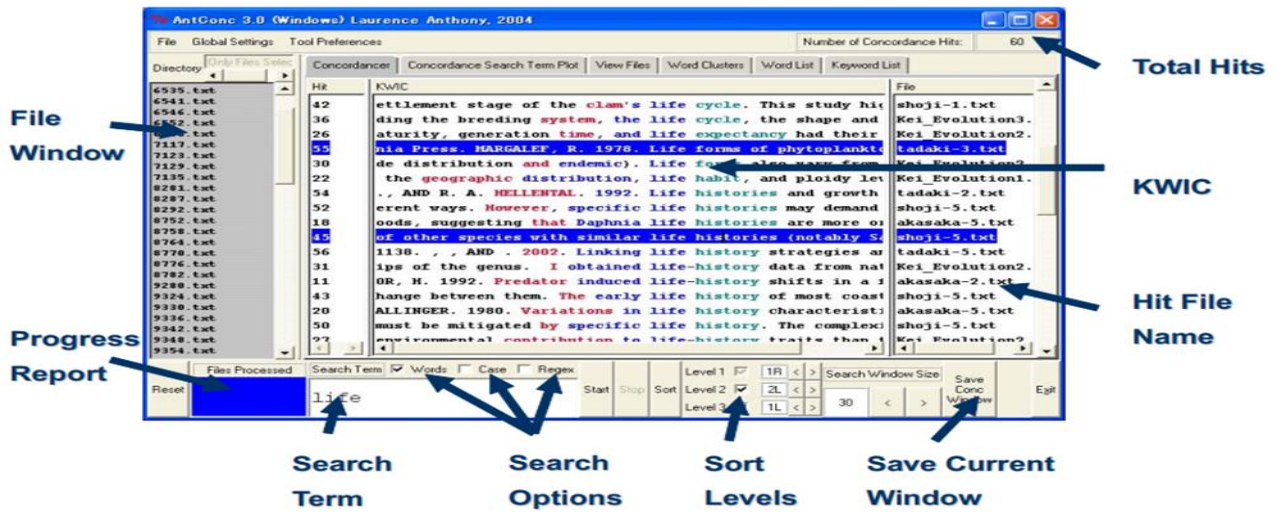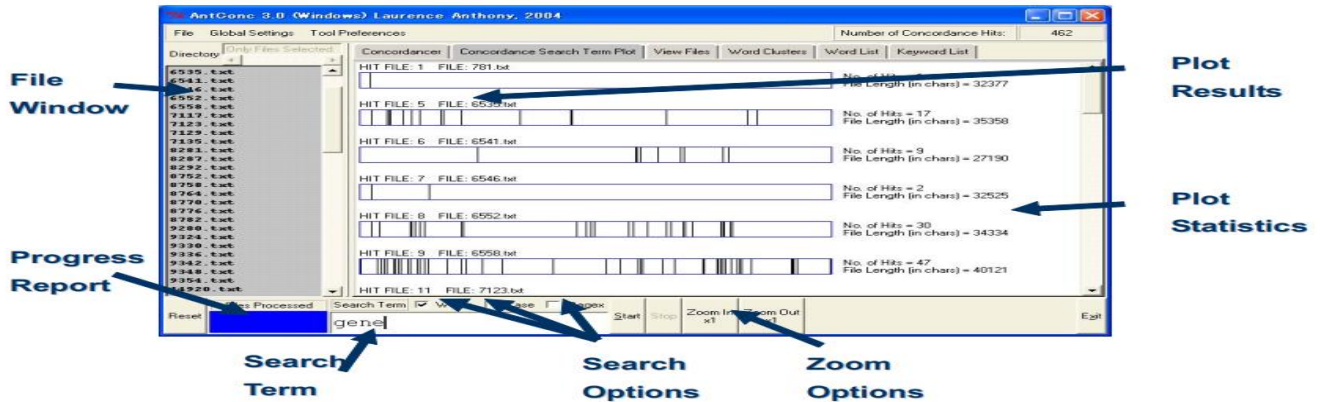

Figure 3-5 KWIC Concordancer Tool

Figure 3-6. Concordance Search Term Plot Tool

Statistical Analysis

Descriptive statistics play a pivotal role in this research as they provide a comprehensive overview of the data. These statistics offer concise summaries of the sample and its metrics, serving as the fundamental basis for most quantitative data analyses, alongside simple graphical analysis. In the context of this study, descriptive statistics are employed to quantitatively describe the lexical, syntactic, mechanical, and semantic errors made by senior students in the English department of the Faculty of Arts in their second language (L2) writing. Specifically, they are utilized to provide quantitative descriptions of the most prevalent errors. Additionally, correlation analysis is employed in this thesis to determine potential relationships between variables, examining the strength and direction (positive or

negative) of these relationships. This analysis helps identify possible connections between lexical, syntactic, mechanical, and semantic errors.

Results and Discussions

This part showed the stages of the analysis results utilizing SPSS software to extract the data necessary to answer the research questions and coin these findings with the literature and the theoretical framework.

Table 4-1. Demographic Data of the Current Study

Gender

	Frequency	
	Number	**Percent (%)**
Male	180	36.44
Female	314	63.56
Total	494	100.0

Table (4.1) shows Demographic Data of the Current Study (Gender). The results determines that the number of male students are (180) with percentage 36.44%, the female students are (314) with percentage 63.56%. Whereas the total number of male and female students are 494.
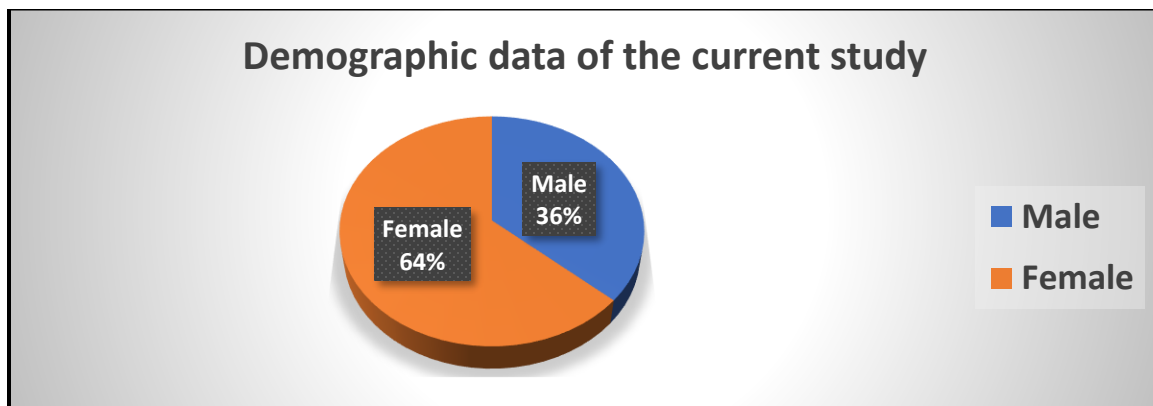

Figure 1- 4. Demographic Data of the Current Study

Descriptive statistics such as, range, means (M) and standard deviation (S.D) was conducted on the data collected for all the participants. Frequencies for the current sample can be found in Tables 4.2.

Table 4-2. Descriptive Statistics of Word, Sentence and Paragraph Count.

Descriptive Statistics

	Range	Min.	Max.	Sum	Mean	SD	N
Word count	1528	1448	2976	81949	2643.5	415.8	31
Sentence count	222	48	270	5467	176.4	52.7	31
Paragraph count	68	26	94	1711	55.2	13.8	31

Table (4.2) displays the data consisting of 31 samples. The total number of words examined in the study sample was "81.949" words at a rate of "2.643.5" words, the largest amount is "2.976 words", the smallest is "1.448" words with SD "415.8". The data in the sample included "5467" sentences per word at a rate of "176.4" with standard deviation "52.7'. The data in the study sample also included "1.711" paragraphs at a rate of "55.2" with standard deviation "١٣,٨".
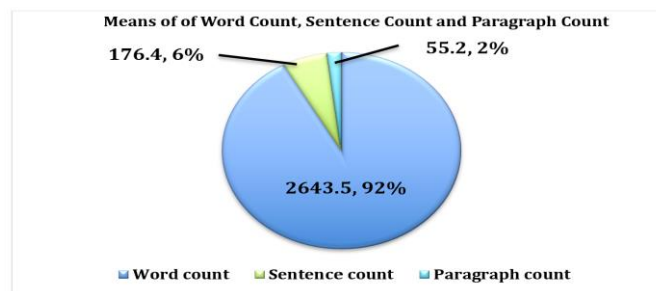

Figure 4-2. The Means of Word Count, Sentence Count and Paragraph Count.

The most frequent Errors done by Senior Students in their L2 writing?

The most frequent errors in the study sample is Average Semantic Errors "626.9" at a rate of "20.2" with a standard aberration "7.8". It was followed by "136.5" Average Mechanics Errors at rate "4.4" with standard deviation "1.3". The Syntactic   errors were "45.7" at a rate of "1.5" with a standard deviation "0.4". The last rank is Average Lexical Errors in the study sample of "19.9" at rate "0.6" with standard deviation " 0.3".
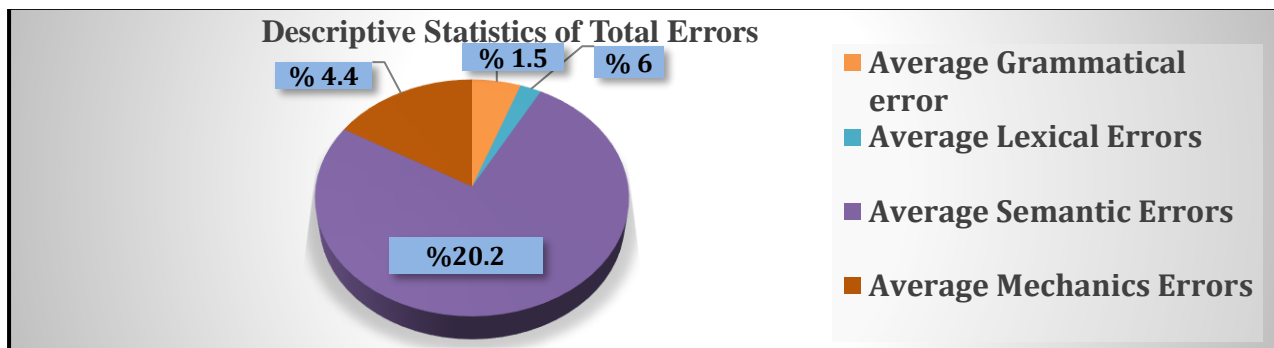

Figure 4.4. Descriptive Statistics of Total Errors

What are the Most Common Errors?

The following section gives a description of the most common errors committed by students' essays. Research question number two.

Analysis of Mechanical Errors

The mechanical errors in the participants' written essays are recognized and divided into 15 kinds in this section. The frequency and percentages of mechanical errors are statistically described in Figure 4.6. The total number of mechanical errors committed by the participants are "1862," as shown in this table. The frequency of mechanical errors ranged from 0.16 percent to 44.68 percent, with "832" capitalization errors accounting for 44.68 percent, "386" punctuation errors accounting for 20.73 percent, "252" capitalization errors accounting for 13.53 percent, and "3" quoted and reported speech errors accounting for 0.16 percent. The frequency and percentages of mechanical errors are shown in Figure 4.6.
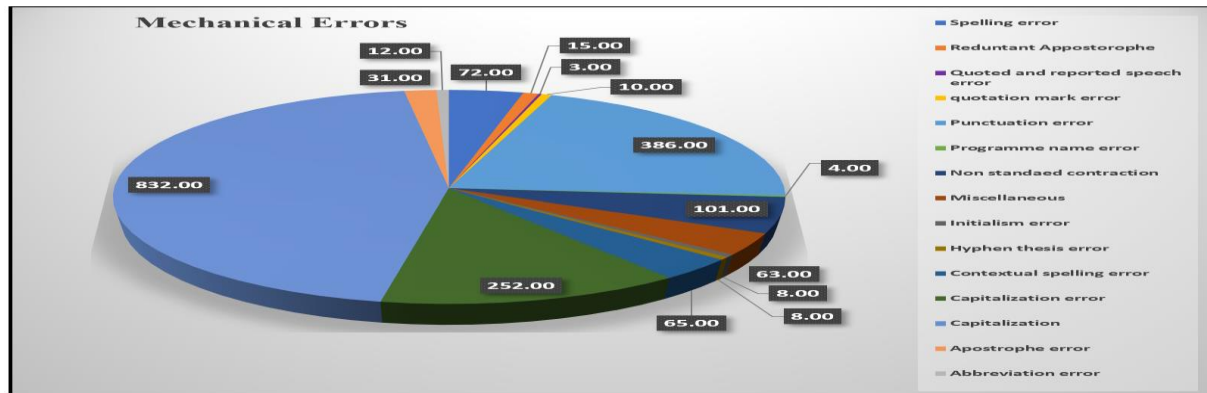

Figure 4-6. Mechanical Errors

Analysis of Semantic Errors

The semantic errors in the participants' written essays are discovered and classified into 16 kinds in this section. The frequency and percentages of semantic errors are statistically labelled in Figure 4. 7. The frequency of semantic errors ranged from 0.02 percent to 59.35 percent, with (5950) Possible Typo errors accounting for 59.35 percent of all errors, followed by (3673) Typography errors accounting for 36.64 percent of all errors, and only (2) semantic errors accounting for 0.02 percent of all errors. The frequency and percentages of semantic errors are shown in Figure 4.7.
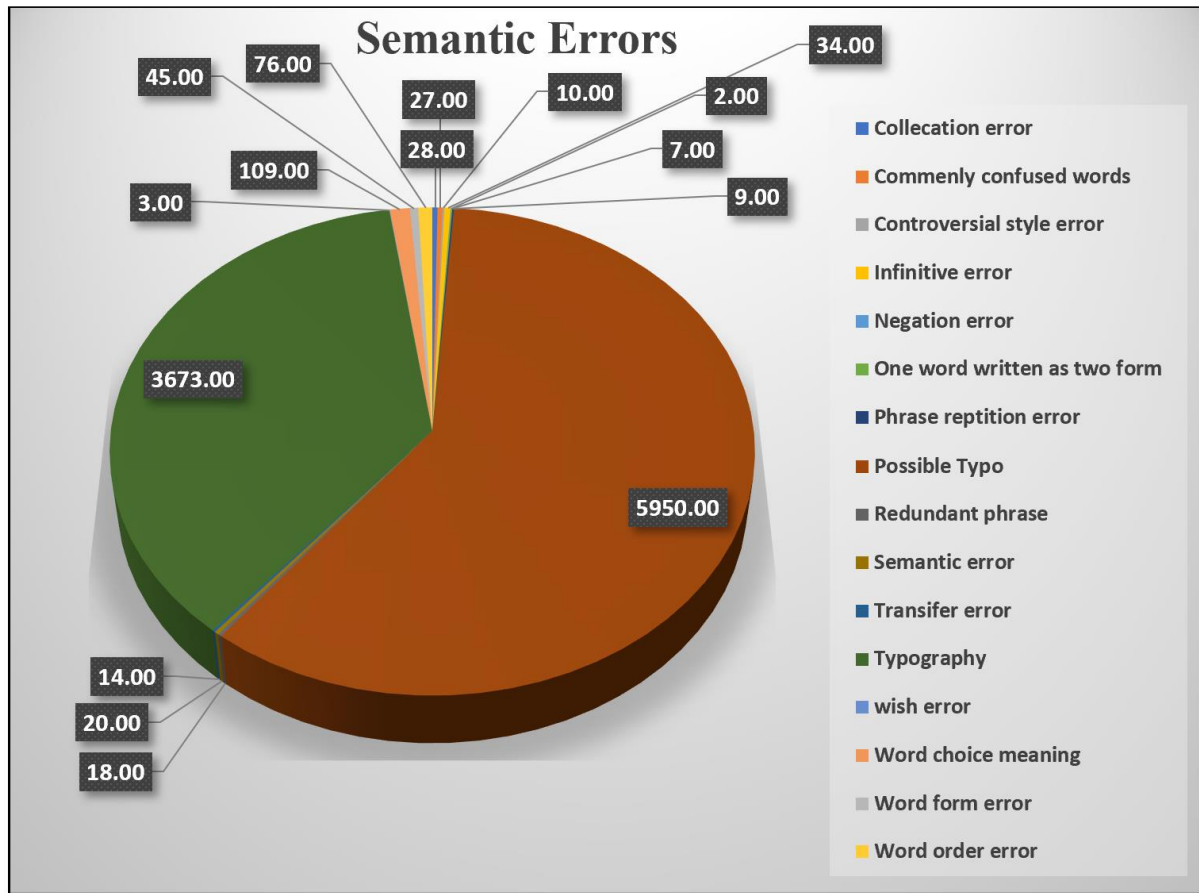
Semantic Errors

45.00 76.00 27.00 10.00 34.00
28.00 2.00
109.00 7.00
3.00 9.00

3673.00

5950.00

14.00
20.00
18.00

- Collecation error
- Commenly confused words
- Controversial style error
- Infinitive error
- Negation error
- One word written as two form
- Phrase reptition error
- Possible Typo
- Redundant phrase
- Semantic error
- Transifer error
- Typography
- wish error
- Word choice meaning
- Word form error
- Word order error

Figure 4-7. Frequencies and Percentages of Semantic Errors

Analysis of syntactical Errors

The syntactical errors in the participants' written essays are recognized and classified into 23 kinds in this section. The frequency and percentages of syntactical errors are statistically labeled in Figure 4-8. The total number of mechanical errors committed by the participants is shown in this table (1048). The frequency of syntactical errors varied from 0 to 23.09 percent, with (242) errors for Word order error accounting for 23.09 percent, (111) errors for Verb form error accounting for 10.59 percent, and (0) errors for Adjective Word error accounting for 0.0 percent. The frequency and percentages of syntactical errors are shown in Figure 4-8
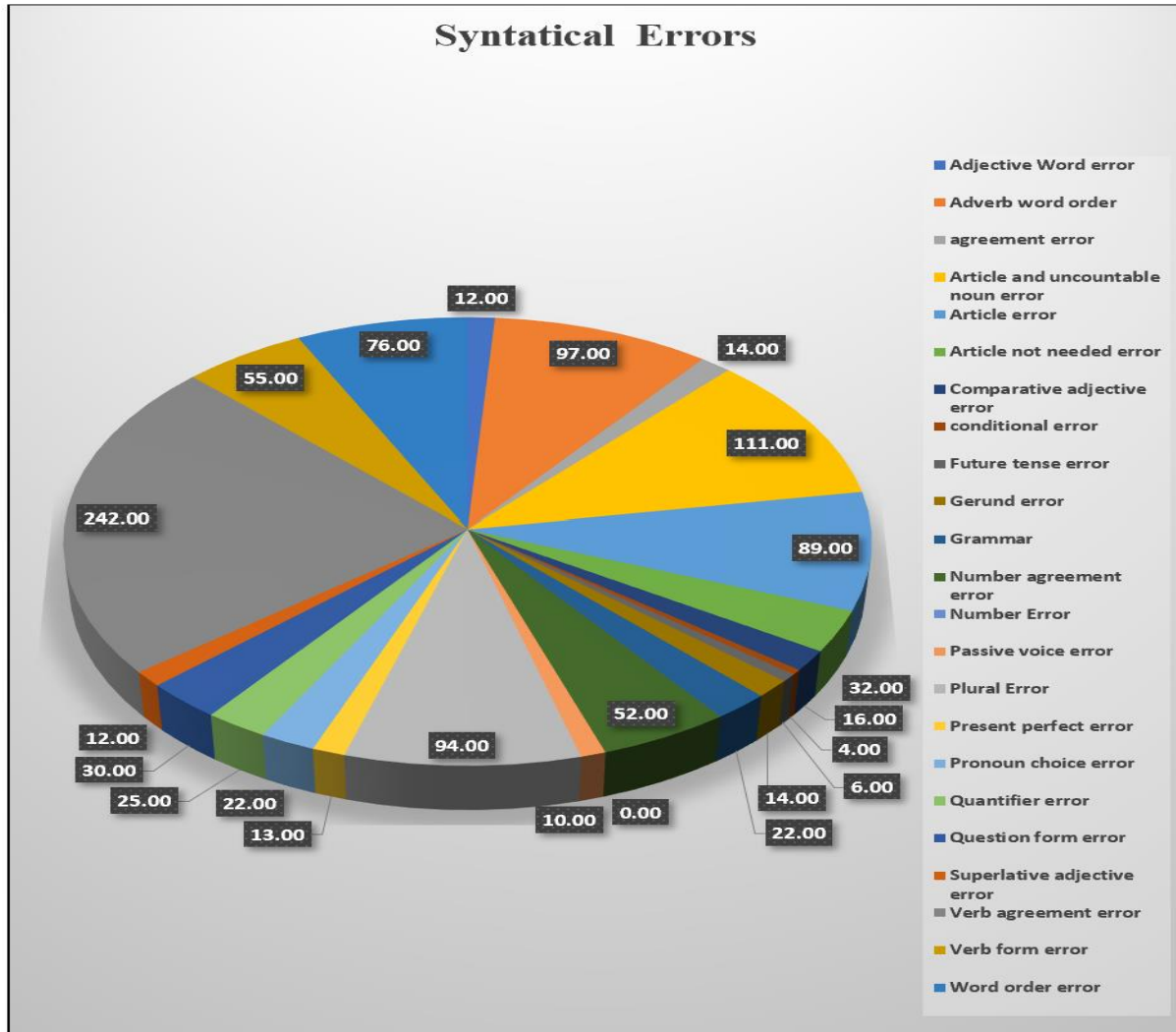
Figure 4-8. Frequencies and Percentages of syntactical Errors

Analysis of Lexical Errors

The lexical errors in the participants' written essays are discovered and classified into 19 kinds in this section. The frequency and percentages of lexical errors are statistically labeled in Figure 4.9. The total number of lexical errors committed by the participants is shown in this table (378). The frequency of lexical errors ranged from 0.53 percent to 20.37 percent, with (77) errors in total. With a percentage of 20.37 percent, word from error is followed by (55) errors. The is-are error had a rate of (13.23 percent), and the Fragment Problem error had a percentage of (0.53 percent) and (2) errors had the least lexical errors. The frequencies and percentages of lexical errors are shown in Figure 4.9
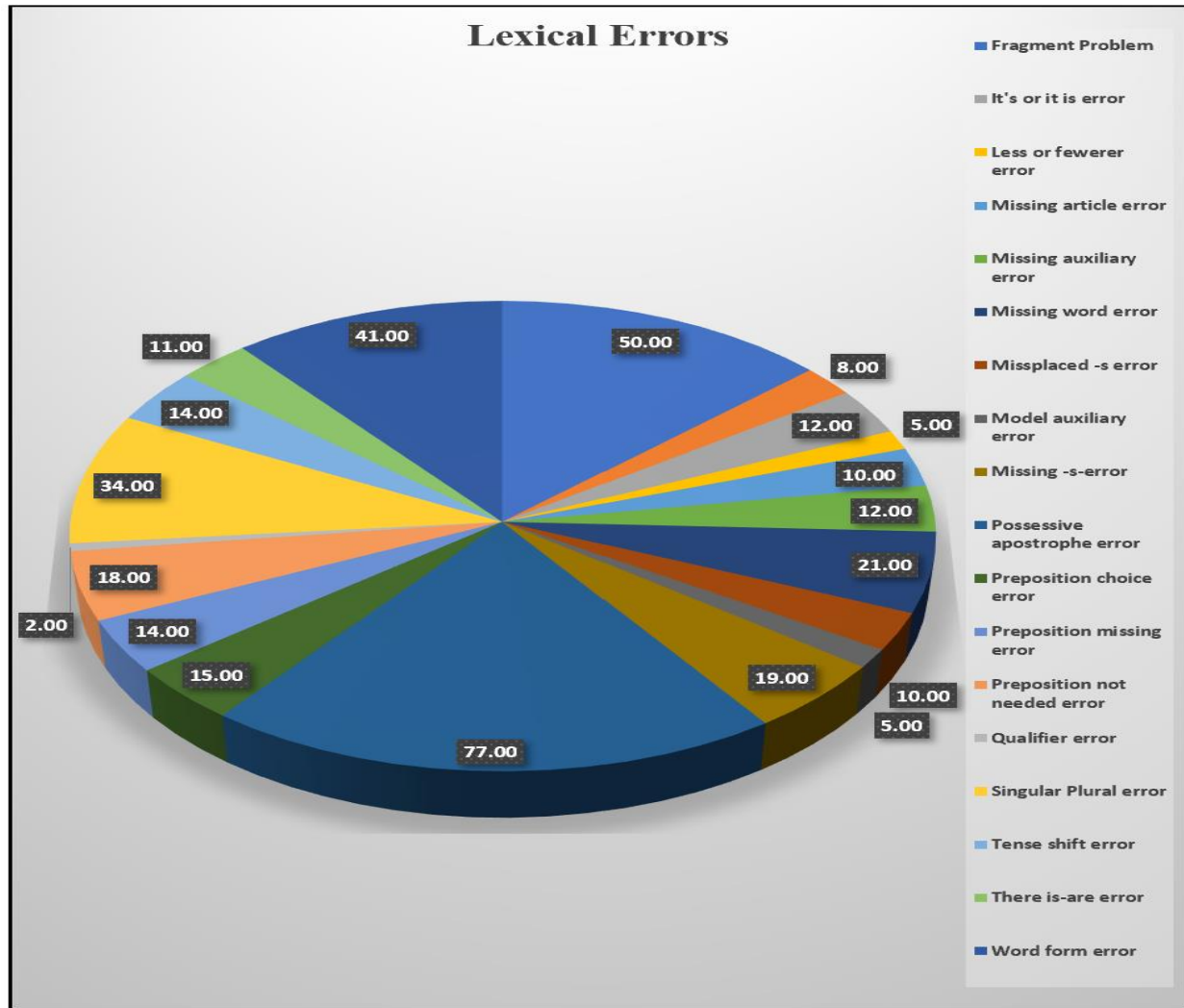
Figure 4-9. Frequencies and Percentages of Lexical Errors

What is the correlation between male and female students in the overall number of errors (Lexical, Syntactic, semantic, and mechanical errors?

The following analysis of the difference and correlation between male and female students in the overall number of errors in their descriptive essay writing gives an answer to research question number three.

Table 4.10 shows the overall errors that students make while writing descriptive essays, according to gender. When compared to male students (mean = 98, S.D = 13.482), female students committed more descriptive essay errors (mean = 187.4, S.D = 52.7). The independent samples T-test found that male students committed less errors than female students (P value =0.0000), which is a statistically significant difference.

In terms of total errors, the present study discovered that male students produced much less errors and scored better in descriptive essay writing than their female

counterparts. Male students, according to the English teacher, tend to make more mistakes in their writing than female pupils. These findings back up the findings of Boroomand & Rostami Abusaeedi ,(2013), who discovered that male students performed significantly better in ESL writing than female students. In descriptive essay writing, however, the independent samples T-test demonstrates that there is no significant difference in mechanical errors between male and female students (p-value = 0.12). Punctuation and capitalization errors are among the errors committed by students in this field. The Independent samples T-test findings, on the other hand, show that male students did considerably better than female students in grammar descriptive writing (P-value = 0.002).

Grammar errors can be found in word form, articles, prepositions, and conjunctions...etc. These findings support the findings by Muhammad & Nair, (2017), who found that female students committed more syntactic errors than male students. In terms of lexical errors, the independent samples T-test shows that male students committed considerably less lexical errors than female students (P-value = 0.015). The findings corroborate those of Zafari & Biria, (2014), who discovered that male learners outperform female learners in vocabulary tests. In addition, the data show that female students produced more semantic errors in descriptive essay writing than male students (p-value = 0.04). These results correspond those of Alhaisoni et al. (2015), who discovered that females made more semantic errors than males.

Table 4-10. Statistics of Errors Made by Male and Female Students

Gender		N	p - value
Average Syntactic error	Male	180	
	Female	314	0.0020
Average Lexical Errors	Male	180	
	Female	314	0.0150
Average Semantic Errors	Male	180	
	Female	314	0.0400
Average Mechanics Errors	Male	180	
	Female	314	0.1200

What are the most common reasons of these errors?

Interlingual errors:

It is now well known that errors can be caused by either intralingual or interlingual variables. Interlingual errors, according to Richards, (1971: P: 205), are errors caused by the native language's interference. These errors occur when students use elements of their native tongue in their target language verbal communication or written performance. When

confronted with a new language, people tend to correlate what they already know with what they don't know, whether consciously or unconsciously.

Interlingual is not regarded an operation in SLA under the effect of L1 or L2, unlike CA and EA. In other words, it is neither the L1 system, the L2 system, or the TL system, but rather an independent linguistic system that functions on its own. IL creates a space for L2 students to share what they already know about a new language they're learning. According to Selinker, (1972), IL is identifiable and discoverable in a learner's language as the transitional method between L1 and L2. He considers IL to be "a dialect whose rules reflect characteristics of two social dialects of languages, whether or not these languages share rules themselves" (Selinker, 1972 as cited in Corder, 1981:17). Figure 4.12 illustrates the concept of IL as it refers to him.
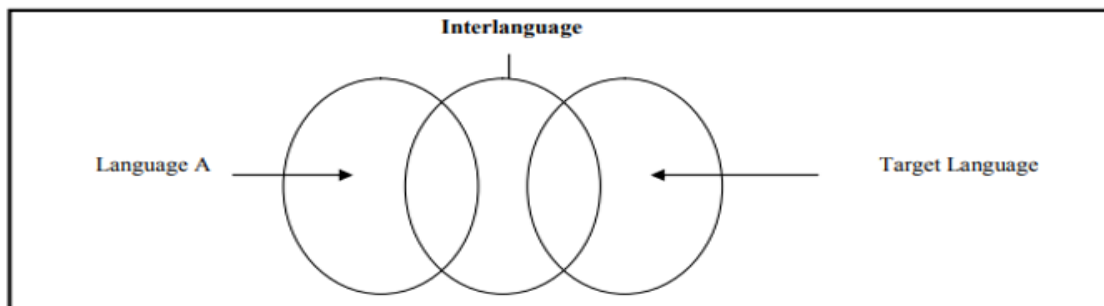

Figure 4-13. The notion of the IL. (Adopted from Corder, 1981:17). In this diagram, A represents the Learner's L

According to Corder, (1981), the learner's language could be considered a dialect in the linguistic sense. He proposes that two languages that share some grammatical principles form dialects. Based on this assumption, he claims that [A] and [B], as shown in Figure 4.13, are in a dialect relationship that leads to [IL].
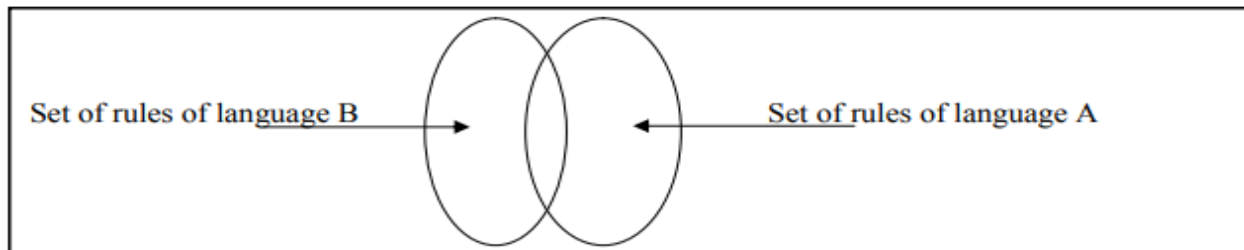

Figure 4-14. The Notion of Dialect Relation (Corder, 1981:14)

Learners apply what they already know about their native language to their target language performance (Ellis, 1997:28). In most circumstances, learning a foreign language without relying on some linguistic qualities of the language that has previously been acquired is unavoidable. In any case, interference can occur in phonology, morphology, grammar, syntax, lexis, and semantics, among other domains of linguistics (Ellis, 2008:350). Furthermore, the effect emerges in a form that is similar or different in both the native and target languages, as seen in Figure 4.14
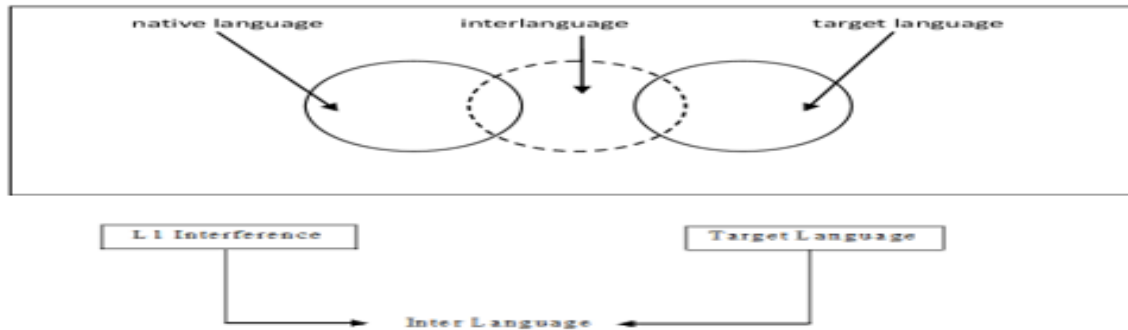
Figure 4-15. The Influence Native Language and Target Language

The effect is obvious wherever the differences appear (Odlin, 1989:7). Where the linguistic theories of a native language are significantly distinct from those of a target language, it is difficult for students to understand the rules and structures of a native language and to apply them in the learning process (Krashen, 1981:85). Interlingual errors are a consequence of the effect of the NL. Although the effects are considered minimal or non-existent when EA begins, in SLA it is now acceptable for learners to rely on their knowledge of L1 to try to learn a new language.

Intralingual Errors

These are errors that occur as a result of inadequate learning characteristics such as improper rule application and a lack of knowledge of rule limits (Richards, 1971). Intralingual errors are thus unaffected by native language interference and are instead driven by the target language. These errors typically arise throughout the language acquisition process when learners have insufficient knowledge (Kaweera, 2013). In addition, developmental errors arise when the learner trying to build up assumptions about the English language from his limited experience of it in the classroom or text-book," according to (Richards, 1971:209). It should be remembered that EFL learners' writing errors are the result of their lack of understanding of the target language. Learners' attempts to make sense of the material and construct rules are reflected in intralingual errors. They arise as a result of the TL structures' complexity, which the learner is unable to handle when initially confronted with them. They are the consequence of defective or incomplete TL learning, in which the TL components interact with one another. The error might be caused by (a) overgeneralization of a more productive rule to instances where it doesn't apply (They come with you.); or (b) developmental sequences - the learner has not reached the final stage; He not play tennis. She didn't saw the car; (c) Simplification: Yesterday I want go there.

Interlingual or Intralingual Errors

It's been a major task to analyze and describe the errors in writing produced by EFL students. Various studies have been performed to examine the errors made by learners from various nations, and it was determined that the predominant source of writing errors was the learners' native language. On the other hand, various studies focused on intralingual

sources of errors such overgeneralization, inadequate rule application, and improper analogies. As a result, prior research reviews are separated into two categories as follows.

Studies on Interlingual Sources

Ridha, (2012)began by researching at the errors made by EFL Iraqi college students when writing English essays. The errors were classified into the following categories: syntactic, lexical, semantic, mechanics, and word order errors. The most significant and common errors were syntactic and mechanical problems. The Arabic interference was to responsible for the majority of the students' errors (Ridha, 2012:44). Liu & Xu, (2013) investigated syntactic errors committed by Chinese University students in their English writings, using Chinese as the native language. According to the study's findings, tense, voice, and modality errors were the most common, and the main causes of errors were student carelessness and native language interference (Liu & Xu, 2013:188).

Watcharapunyawong & Usaha, (2013:67-78) investigated writing errors induced by Thai language interference in three writing genres: narration, description, and comparison/contrast in Thailand. Verb tense, word choice, sentence structure, article, preposition, modal/auxiliary, singular/plural form, fragment, verb form, pronoun, run-on sentence, infinitive/gerund, transition, subject-verb agreement, parallel structure, and comparison structure were all shown to be interlingual errors.

The findings of this study also revealed that the frequency of errors varied by writing type. In conclusion, the studies of written essays have revealed that the native language continues to have a detrimental impact on Thai EFL students' writing. Empirical research also shows that errors impacted by the learner's native language may be discovered at several levels of language, from lexicon to discourse, and that interlingual errors can occur in any form of writing.

Studies on Intralingual Sources

Kim, (2001) investigated over thirty composition samples written by Korean EFL University students. The overwhelming of errors was found to be caused by intralingual sources, according to the error analysis. However, there were a few errors caused by the native language's influence. In a similar manner, Sattayatham & Honsa, (2007) examined the errors made by first-year medical students at a Thai university. Order of adjectives, there is/are, subject-verb agreement, direct/indirect object, verbs of feeling, past tense, present perfect, reported speech, passive voice, and question tag were among the 10 categories of linguistic errors revealed in the study (Sattayatham & Honsa, 2007:184).

The study's findings also demonstrated that intralingual sources are mostly responsible for the following errors: overgeneralization, insufficient rule application, omission, and the creation of incorrect ideas. In a limited percentage of cases, native language interference was discovered (Sattayatham & Honsa, 2007:188). Teachers have challenges while teaching English writing to EFL students, despite the fact that many research studies have been undertaken to provide insights into the probable errors that arise

in EFL students' compositions. EFL students' writing challenges continue to persist. As a result, it is necessary to conduct continuous analyses of writing errors.

The causes of errors in students English essay writing

The fourth research question is addressed in the next section. The data show that while Egyptian students make errors for legitimate reasons related to L1 transfer, the majority of their errors are caused by interlingual and intralingual factors.

Language Transfer

When students write in TL, they use their knowledge of TL principles and structures incorrectly in new settings, resulting in unexpected or inadequately sentences. Selinker identified three sources of errors in 1974: 1. Language transfer. 2. Learning strategies for a second language 3. Over generalization of TL linguistic content However, in this study, the focus will be on two sorts of errors: First, if errors were made as a result of his or her native language, this procedure is referred to as "language transfer" or "interlingual transfer." These errors might be found objectively at the beginning of L2 learning.

Because the learner is a beginner, he can rely on a significant deal of transfer from his or her native tongue. Second, we are dealing with the over generalization of TL linguistic knowledge if they can be demonstrated to be the consequence of a clear over generalization of TL rules and semantic features. These are a case of negative intralingual transfer committed later in the learning development.

Language Transfer, Overgeneralization, and Teaching Methods or Materials were the three sources of errors identified by Corder in 1974. Richards and Simpson presented seven sources of errors in their study "The Study of Learner English," which was published in 1974.

1. Language transfer, which might account for one-third of the inaccurate sentences produced by second-language learners.

2. Intralingua interference: Richards 1970 identified four types of intralingua errors and their causes.

a. Over generalization: it is associated with redundancy reduction. It refers to situations in which a student builds a deviant model based on his prior knowledge of the target language's structures.

b. Ignorance of rule restrictions: using rules to contexts where they don't apply.

c. Rules that aren't consistently applied.

d. Semantic errors, such as the construction of incorrect concepts/systems: i.e., an incorrect understanding of differences in the Target language.

3. Sociolinguistic context: motivation and learning environments for second languages can have an impact.

4. Modality: the way the TL is exposed and the way it is produced.

5. Age: People's learning abilities changes as they get older.

6. Successions of approximation systems: the acquisition of new lexical, phonological, and syntactic components differs from person to person, as does the case of language learning.

7. Universal challenge hierarchal structure: It refers to the intrinsic difficulty of phonological, syntactic, or semantic items or structures.

Interlingual Transfer

In foreign language learning, interlingual transfer (i.e., transfer from the mother tongue or any other previously learned language) is a significant cognitive method that learners use when their linguistic means fall short of achieving their communication goals. The mother tongue, of course, provides an extra source of hypothesis formation that the first language learner lacks. The impact of the mother tongue and the widespread nature of interlingual transfer are undeniable, particularly in settings when students' exposure to the foreign language is limited to a few hours per week of formal classroom teaching (Mahmoud, 2000). As a result, interlingual transfer is a technique that learners might use to substitute for their weaknesses when communicating in a foreign language. At all language levels, deviations resulting from interlingual transfer have been reported (Douglas & Selinker, 1994)

Intralingual Transfer

Intralingual transfer is a second source of inaccuracy in learner interlanguage that is commonly observed. Those arising from the study of a foreign language (TL), regardless of the native tongue. They are items formed by the learner that indicate generalizations based on incomplete exposure to the target language rather than the structure of the mother tongue. In this situation, the learner is attempting to derive the principles behind the data to which he or she has been presented, and may form hypotheses that do not correlate to the mother tongue or the target language (Richards, 1971, p. 6). The effect of one target language items on another may result in such inaccuracies. H. Douglas Brown, (1994:225) reports studies demonstrating that interference (interlingual transfer) predominates in the early phases of language learning, but that after learners have begun to learn parts of a new system, intralingual transfer - generalization within the target language – is becoming increasingly evident.

Lack of Writing Practice

Due to a lack of writing exercises, students' writing skills may deteriorate, resulting in a high number of errors in their writing. Many students assume that their inability to write well in English is due to a lack of writing assignments and activities. Furthermore, the textbooks used in schools are responsible for students' writing errors. Students seldom

practice writing on a regular basis in writing exercises. Furthermore, it may be deduced that a lack of consistent writing activities may result in more writing errors.

Weakness in English Writing Skill

According to the findings and conclusions, the respondents make various errors for a variety of reasons that have previously been discussed. It's worth noting that there's a consistent lack of fundamental writing ability. In their compositions, students consistently make a variety of errors. On the other side, poor English writing skills might be attributed to a lack of vocabulary, motivation, writing tasks, and teaching approaches. Given the limited time allotted to the writing task, the researcher hypothesized that students would continue to produce numerous errors of various sorts, resulting in writing is below the assumed goal.

Recommendations & Conclusion

Summary of the Findings

This study attempted to answer four questions. The findings of the study showed that the most common frequent errors which up to 72 types of errors plus the sources of these errors in English writings which might be the impact of the first language. The students' lack of knowledge, inadequate application of structure, inexperience of certain rules, and negligence were amongst the most important probable causes of error occurrence. A complete description of the sources and causes have been discussed earlier.

Recommendations for Further Studies

This research has resulted in several recommendations and research directions for the future. Here are a few recommendations:

1. Since the sample size of student essays was limited, the findings could not be applied to all fourth-year university students. As a result, future research might be conducted with fourth-year students at multiple colleges to compare the various common errors received from different groups of students. If the frequently errors committed by those students are almost identical to those in the current study, it may shed light on common errors made by Egyptian University students, allowing language teachers to better address their students' problems.

2. To attain more precise and detailed descriptions of the fundamental reasons why students made the error. Furthermore, future studies should focus on errors detected in students' written work that indicate their linguistic competence and are reasonable. To accomplish so, the researcher can investigate employing self-correction as a method of eliminating non-systematic errors caused by students' inattention.

3. Further research of errors in speaking ability made up of different English language courses, particularly selective ones, will be fascinating. This is because conducting such study might aid teachers in learning about difficult categories of errors and determining whether they are similar to those seen in writing. If this is the case, teachers should alter their teaching techniques and create appropriate resources to help students get a fundamental understanding of syntactic structure. This may assist students in overcoming the basic difficulties they have in learning productive English skills.

4. It is significant to mention that the findings of this study are based on an analysis of errors committed by Egyptian students at Tanta University's Faculty of Arts. This might lead to a comparison of teaching methods and error correction strategies in experimental research. Based on the findings of such investigation, teachers may be able to demonstrate effective methods to dealing with students' errors during language learning.

References

Abbad, A. (1988). *An analysis of communicative competence features in English language texts in Yemen Arab republic*. https://www.ideals.illinois.edu/handle/2142/69184

Abdel Latif, M. M. (2015). Sources of L2 writing apprehension: A study of Egyptian university students. *Journal of Research in Reading*, *38*(2), 194–212. https://doi.org/10.1111/j.1467-9817.2012.01549.x

Al-Sawalha, A., & Chow, T. (2012). The effects of proficiency on the writing process of jordanian efl university students. *Academic Research*.

Alhaisoni, E. M., Al-Zuoud, K. M., & Gaudel, D. R. (2015). Analysis of spelling errors of beginner learners of English in the English foreign language context in Saudi Arabia. *English Language Teaching*, *8*(3), 185–192.

Badger, R., & White, G. (2000). *A process genre approach to teaching writing*.

Boroomand, F., & Rostami Abusaeedi, A. (2013). A gender-based analysis of Iranian EFL learners' types of written errors. *International Journal of Research Studies in Language Learning*, *2*(5), 79–92.

Brown. (1994). *Teaching by principles an interactive approach to language pedagogy*. Englewood Cliffs.

Corder. (1981). *Error analysis and interlanguage*. Oxford University Press, USA.

Crystal, D. (1992). *Profiling linguistic disability*.

Douglas, D., & Selinker, L. (1994). Research methodology in context-based second-language research. *Research Methodology in Second-Language Acquisition*, 119–131.

Eliwa, A. (2004). Synonymy and Collocation in Classical Arabic. *Unpublished PhD Thesis, University of Manchester, UK*.

Ellis. (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System*. https://doi.org/10.1016/j.system.2008.02.001

Ellis, R. (1997). *SLA research and language teaching.* ERIC.

Erkan, & Saban. (2011). Writing performance relative to writing apprehension, self-efficacy in writing, and attitudes towards writing: A correlational study in Turkish tertiary-level EFL. *The Asian EFL Journal Quarterly*, *13*(13), 164–192.

Francis. (1992). Language corpora BC. *Directions in Corpus Linguistics: Proceedings of Nobel*

Symposium, 82, 17–32.

Grabe, W., & Kaplan, R. (1996). Theory and practice of writing. *Longman. Pennington, MC (199Th). The Teacher ….*

Haq, F. A. (1982). An analysis of syntactic errors in the composition of Jordanian secondary students. *Unpublished MA Thesis. Jordan. Yarmouk University*.

Hunston, S. (2006). Corpus linguistics. *Linguistics*.

Huwari, M., Fathi, I., Aziz, A., Hashima, N., Huwari, I., Aziz, N., Huwari, M., Fathi, I., Aziz, A., & Hashima, N. (2011). Writing apprehension in English among Jordanian postgraduate students at Universiti Utara Malaysia (UUM). *Academic Research International*.

Hyland, K. (2003). Genre-based pedagogies: A social response to process. *Journal of Second Language Writing*.

Imaniar, F. (2018). Students' Writing Errors and What Lies Beyond Them. *Langkawi: Journal of The Association for Arabic and English*, 4(2), 71. https://doi.org/10.31332/lkw.v4i2.936

Jones. (2006). Uses of blogs in L2 instruction. *Languages for Today's World: Dimension*, 25–35.

Kaplan, R., & Grabe, W. (1996). Applied linguistics and the Annual Review of Applied Linguistics. *Annual Review of Applied Linguistics*, 1–289.

Kaweera, C. (2013). Writing error: A review of interlingual and intralingual interference in EFL context. *English Language Teaching*, 6(7), 9–18.

Kim. (2001). *A descriptive analysis of Korean and English apologies with implications for interlanguage pragmatics*. University of Florida.

Kim, M. (1998). *Genre-Based Approach to Teaching Writing*.

Krashen. (1981). Second Language Acquisition and Second Language Learning. In *Pidginization and Creolization as language acquisition*.

Leech. (2014). Teaching and language Corpora: A convergence. In *Teaching and Language Corpora* (pp. 1–24). https://doi.org/10.4324/9781315842677-1

Liu, M., & Xu, Y. (2013). An investigation of syntactic errors in Chinese undergraduate EFL learners' compositions: A cohort study. *International Journal of Applied Linguistics and English Literature*, 2(4), 182–191.

Mahmoud, A. (2000). Modern standard Arabic vs. non-standard Arabic: Where do Arab students of EFL transfer from? *Language Culture and Curriculum*.

McEnery. (2016). *Corpus Linguistics and 17th-century prostitution: computational linguistics and history*. Bloomsbury Academic.

Muhammad, A. S. idu, & Nair, S. M. (2017). Evaluating pragmatic competence in Nigerian undergraduates' language errors within descriptive ESL writing. *International Journal of Instruction*, 10(1), 255–272. https://doi.org/10.12973/iji.2017.10116a

Nadira. (2012). *Written English is as important as Spoken English*.

O'Sullivan, Í. (2007). Enhancing a process-oriented approach to literacy and language learning: The role of corpus consultation literacy. *The Journal of Eurocall*, *19*(3), 269.

Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*.

Richards. (1971). ERROR ANALYSIS AND SECOND LANGUAGE STRATEGIES. *Department of LinguisticsUniversity Laval, Quebec, CanadaThe Text of an Invited Lecture given AtIndiana University*.

Richards, & Schmidt. (2002). Longman dictionary of applied linguistics and language teaching. *Harlow, UK: Longman*.

Ridha, N. S. A. (2012). The effect of EFL learners' mother tongue on their writings in English: An error analysis study. In *iasj.net* (Issue 60). https://www.iasj.net/iasj?func=article&aId=58267

Salem, I. (2007). The lexico-grammatical continuum viewed through student error. *ELT Journal*, *61*(3), 211–219.

Sattayatham, A., & Honsa, S. (2007). Medical students' most frequent errors at Mahidol University, Thailand. *The Asian EFL Journal*, *9*(2), 170–194.

Selinker, L. (1972). Interlanguage. *IRAL-International Review of Applied Linguistics in Language Teaching*, *10*(1–4), 209–232. https://doi.org/10.1515/iral.1972.10.1-4.209

Shawish. (2010). *An Investigation of Palestinian EFL Majors' Writing Apprehension: Causes and Remedies*.

Sinclair, J. (2005). *Corpus and text-basic principles. Developing linguistic corpora: A guide to good practice, 1-16*.

Thomas, J., & Short, M. (1996). *Using corpora for language research: Studies in honour of Geoffrey Leech*. Longman Pub Group.

Watcharapunyawong, S., & Usaha, S. (2013). Thai EFL students' writing errors in different text types: The interference of the first language. *English Language Teaching*, *6*(1), 67–78.

Wright, S. (2011). *Why Writing Skills are More Important Than Ever*.

Zafari, M., & Biria, R. (2014). The relationship between emotional intelligence and language learning strategy use. *Procedia-Social and Behavioral Sciences*, *98*, 1966–1974.