Journal of Modern Research in English Language Studies
Print ISSN: 2676-5357
Online ISSN: 2676-5985

Analyzing the Development of a Longitudinal Error-Tagged English Learner Corpus: Exploring Accuracy Development in Iranian EFL Learners' Writing

Samaneh Yadollahi[1], Amirsaeid Moloodi[2*], Mohammad Rahimi[3]

[1]Ph.D. Candidate, Department of Foreign Languages & Linguistics, Shiraz University, Iran. *s.yadollahi@shirazu.ac.ir*

[2*](corresponding author) Assistant Professor, Department of Foreign Languages & Linguistics, Shiraz University, Iran. *amirsaeid.moloodi@shirazu.ac.ir*

[3] Professor, Department of Foreign Languages & Linguistics, Shiraz University, Iran. *rahimim@shirazu.ac.ir*

Article info

Article type: Research article

Received: 2024/2/18

Accepted: 2024/5/26

Abstract

Error-tagged learner corpora are helpful resources in language teaching, providing authentic samples of learners' errors. This longitudinal study aims to investigate accuracy development in three subsequent writing performances of Iranian EFL learners across beginner, intermediate, and advanced proficiency levels. This investigation involved developing and annotating the first error–tagged English written learner corpus for Iranian EFL learners, followed by analyzing the errors. The current corpus includes 35747 tokens from 219 written texts, manually transcribed and annotated based on the latest version of the Louvain Error Tagging Manual. A total of 6917 errors were identified. The developmental patterns of all error categories were detected using potential occasion analysis, specifically focusing on the most frequent error types (i.e., articles, noun numbers, and personal pronoun errors). The results indicated that grammar, lexical, and word redundant/missing/order errors increase significantly as proficiency levels increase. Conversely, form, lexico-grammatical, and punctuation errors exhibited a U-shaped trend, rising from beginner to intermediate levels and declining from intermediate to advanced levels.

Additionally, the accuracy of article and noun number usage improved from beginner to intermediate levels but showed little or no change from intermediate to advanced levels, suggesting that higher proficiency levels did not lead to much improvement in this area. However, there was a significant decrease in personal pronoun accuracy from beginner to intermediate levels, followed by a slight increase from intermediate to advanced levels. This study reveals error patterns across different proficiency levels, offering guidance for teachers to adapt their writing instruction methods and enhance learners' writing accuracy.

Keywords: error-tagged written learner corpora, error annotation, error analysis, accuracy development, Iranian EFL learners

Cite this article: Yadollahi, S., Moloodi, A., & Rahimi, M. (2025). Analyzing the development of a longitudinal error-tagged English learner corpus: Exploring accuracy development in Iranian EFL learners' writing. *Journal of Modern Research in English Language Studies*, *12*(1), 147-171.

DOI: 10.30479/jmrels.2024.19996.2333

Publisher: Imam Khomeini International University

1. Introduction

Learner corpora (LC) are collections of language data from learners that are prevalent in Second Language Acquisition (SLA) research (Granger, 2009). According to McEnery et al. (2006), these machine-readable authentic data from learners' spoken or written productions can be analyzed using corpus linguistic methods which contribute to new insights into the process of language learning and the reasons behind errors in writing or speaking performances (Lüdeling & Hirschmann, 2015).

One of the demanding issues in the SLA realm is investigating the degree of linguistic accuracy of learners' written productions (Larsen-Freeman, 2006). Housen et al. (2012) refer to accuracy or correctness as the learners' ability to produce linguistic productions similar to native language without deviations from its norms and conventions. These deviations, which include grammatical, lexical, orthographic, and pragmatic errors, are traditionally regarded as errors.

Due to the various functions of LC use in SLA research, numerous SLA research has shifted from general use towards more specific purposes. Error-tagged learner corpora are one of these powerful tools in investigating the correctness and well-formedness of the learners' writing. Such corpora serve as valuable resources for language teaching and learning as they reflect learners' authentic use of the target language and clarify the nature of learner errors for language teachers and researchers (Ruzaitė et al., 2020; Biber, et al., 1998). By analyzing the types and frequency of errors that learners commonly make, teachers can develop strategies to help learners avoid these errors in the future. Therefore, error-tagged learner corpora contribute significantly to language teaching and learning, updating effective instructional practices and enhancing language learning outcomes (Ortikov, 2023).

To conduct a well-defined computer-aided error analysis, it is necessary to develop an error annotation scheme and identify written errors that occur at a specific proficiency level (Lu, 2022). This enables researchers to develop a dynamic approach to tracing learners' errors across different acquisition stages over time (Lüdeling & Hirschmann, 2015).

Despite notable efforts made by Salmani Nodoushan (2018) and Khazaee et al. (2020a) and Khazaee et al. (2020b), there is an absence of error-tagged learner corpora developed specifically for Iranian EFL learners, highlighting a critical gap in identifying the linguistic problems faced by these language learners across different proficiency levels. Previous studies have not provided such a wide-ranging identification and classification of almost all types of errors Iranian EFL learners make in their written performances. Also, there were a few attempts to track the changes in error categories and error frequency in learners' written texts over a long time of an individual proficiency level, especially in EFL testing situations. Hence, by developing

and designing the first error-tagged written learner corpora and annotating complete categories of error types, this study can illuminate the nature of these errors and how linguistic accuracy develops over time across different proficiency levels. This study is a longitudinal study that compares the same participants as they progress through three different proficiency levels using similar types of writing tasks. According to Lüdeling and Hirschmann (2015), such longitudinal corpus study is rare. Therefore, by tracking learners' needs through their errors in their writing, teachers can use them as an authentic resource for making data-based decisions in instruction and providing constructive feedback.

This study's primary goal is to develop a longitudinal error-tagged learner corpus of English for Iranian EFL learners. It aims to analyze six out of seven main error categories based on the Louvain Error Tagging Manual (LETM) (Granger et al., 2022) including Formal errors (F), Grammatical errors (G), Lexico-grammatical errors (X), Lexical errors (L), Word Redundant/Missing/ Order errors (W), and Punctuation errors (Q) in written test tasks across beginner, intermediate, and advanced proficiency levels over three subsequent semesters.

More specifically, this study focuses on annotating and analyzing over 50 error types outlined in the latest version of the LETM (Granger et al., 2022). It particularly emphasized the three most common errors made by English language learners, regardless of their proficiency level (Thewissen, 2015). Thewissen (2015) found that these prevalent errors could impede communication and comprehension, negatively affecting the overall quality of written texts. Given their significance, this study investigates the developmental patterns of GA, GNN, and GPP across different proficiency levels. The second goal is to identify factors contributing to these errors and develop more effective teaching strategies to mitigate them. Ultimately, this study aims to create a valuable data source that can serve as a research tool and pedagogical aid for learning English as a foreign language in Iran.

2. Literature Review

2.1. Error-Tagged Learner Corpora

Learner corpora, a subfield of corpus linguistics, consist of electronically tagged collections of authentic language data from language learners, systematically gathered for language learning and teaching purposes (Granger, 2002). They enable researchers to gain a comprehensive understanding of learners' linguistic expressions, aiding in syllabus and materials design (Pravec, 2002).

Romer (2005) suggests that learner corpora provide real-life language examples, enhance teaching materials and allow teachers to identify common errors for targeted instruction. Additionally, learner corpora help learners

understand language elements and patterns, promoting autonomous learning. Learners can also use corpora to self-assess and improve, fostering responsibility for their learning (Ma & Mei, 2021). Thus, integrating learner corpora into teaching practices is crucial for effective language instruction (Flowerdew, 2001).

Based on the rich insights offered by learner corpora, error-tagged learner corpora deepen the analysis by studying the underuse, overuse, and misuse of various aspects of learner language, such as lexis, discourse, and syntax (Granger, 2003). Li (2022) emphasized that corpus-based error analysis contributes to researchers' and educators' better understanding of interlanguage errors caused by L1 transfer, learning strategies, and the overgeneralization of L1 rules. Furthermore, error-tagged learner corpora play a crucial role in both assessing learners' language progress and assisting the teachers and material designers to create instructional materials tailored to learners' proficiency levels by providing detailed annotations with specific tags or codes that identify and describe errors in learners' linguistic performances (Granger, 2002).

2.2. Common Written Errors by Iranian EFL Learners

Error analysis in Iranian EFL setting has been a well-researched area (Khansir & Pakdel, 2020), with studies highlighting the systematic nature of Iranian learners' errors. These investigations revealed that the errors made by Iranian EFL learners are not isolated incidents but rather part of a discernible pattern, underscoring the necessity for increased practice and an enhanced understanding of English grammar rules (Khansir & Shahhoseiny, 2013). Khansir and Pakdel's systematic review of research studies on written errors by Iranian learners from 2012 to 2018 uncovered a prevalence of grammatical errors. These errors included issues with prepositions, articles, verb tenses, and subject-verb agreement, alongside syntactic errors, spelling mistakes, and difficulties with English clauses. They added that these errors are often attributed to factors such as first language interference, inadequate mastery of English grammar rules, and insufficient practice in writing and language skills.

In more recent studies, various types of errors have been identified as prevalent among Iranian EFL learners in their English writings. Moazzeni Limoudehi and Mazandarani (2019) noted that grammatical errors were widespread across both lower intermediate and advanced proficiency levels. Gorjian (2022) specifically highlighted spelling, punctuation, and word order errors as the most frequently observed interlingual errors. Additionally, Zarabi et al. (2023) identified four primary areas of writing difficulty for Iranian EFL learners, including grammar, vocabulary, semantics, and mechanics.

Despite carried-out efforts to analyze written errors in Iranian EFL performances and acknowledging that the types of errors differ based on

proficiency level, there is a lack of investigation on the type and the frequency of a wide range of errors across different proficiency levels. Furthermore, as of our last update, the analysis of written errors among Iranian EFL learners has not been conducted using the error annotation approach across a broad spectrum of proficiency levels over an extended period. Therefore, this study represents the pioneering attempt to develop a learner corpus with annotated language errors.

For this study, the following questions were raised to be answered:

- Research Question 1: What are the changes in error categories and frequency in learners' written text in testing situations across three subsequent semesters of an individual proficiency level?
- Research Question 2: What are the error developmental patterns for three top error types, including GA, GNN, and GPP, across different proficiency levels?

3. Method

3.1. Participants

The research was conducted at the Navid Language Institute (NLI) in Shiraz, Iran. During the 2018-2019 academic year. The sample consisted of 73 female Iranian EFL learners who had been enrolled in NLI for at least three consecutive semesters and participated in all three written exams each semester. This sample was chosen to facilitate longitudinal research. To determine the sample size, Cochran's formula with a 95% confidence level and a 5% margin of error was used. The resulting sample size was 73, and participants were randomly selected from the population of 91 English learners using a simple random sampling method. Participants, ranging from ages 10 to 50, had diverse proficiency levels and were native speakers of Persian. Upon enrollment, learners underwent a CEFR-aligned placement test administered by NLI to determine their proficiency level. Subsequently, they were assigned to the corresponding instructional level based on their test results. Each semester spanned three months, and they attended English classes twice a week, each lasting 90 minutes. The language instruction followed the Top Notch series by Saslow and Ascher, aligning each proficiency level with the CEFR framework (Council of Europe, 2001). Three proficiency levels (A2, B2, and C1) were selected, and each of the 73 participants submitted three writing tasks at the end of three sequential semesters, resulting in a total of 219 texts.

3.2. Materials and Instruments

In total, 219 essays were collected as writing prompts. The written tasks ranged from 34 to 340 words, amounting to a corpus sample of 35747 tokens overall. This can be seen in more detail in Table 1. Moreover, learners provided

informed consent, ensuring anonymity, and the experimental procedure was approved by the Navid Institute's Ethics Committee.

Table 1

Corpus Sample

Proficiency level	Number of learners	Number of learners' written texts	Total tokens
A2	34	102	11567
B2	19	57	11089
C1	20	60	13091
Total	73	219	35747

3.2.1. Instruments 1

Firstly, the Google Docs Platform was used to convert the learners' PDF files of essays into editable text format. This streamlined the further processes of error detection, annotation, and analysis.

3.2.2. Instruments 2

Secondly, the most recent iteration of the Louvain error taxonomy (Granger et al., 2022), developed by the Centre for English Corpus Linguistics (CECL), was implemented to annotate the errors within the learners' essays. Acknowledged as a prominent framework within Computer-Assisted Error Analysis (CEA) research, the LETM holds a distinguished position (Thewissen, 2015). This hierarchical and descriptive manual provides the three precise levels of classifications for the error annotation process, including seven error categories (level 1), 23 error subcategories (level 2), and 54 error types (level 3). For instance, the sample sentence from A2 level involved an error of uncountable nouns.

Example:

Worldwide experts …. Study about the writings that were rich in *<XNUC corr="information">*informations*</XNUC>*.

The error was tagged as <XNUC>. X stands for lexico-grammatical error (first level), N stands for noun errors (second level), and UC stands for countable/uncountable (third level).

3.2.3. Instruments 3

The number of potential occurrences of the errors (POA) was used to identify the error frequency and developmental patterns of error categories in the data. POA quantifies errors based on an error-tagged and a part-of-speech-tagged (POS) version of the learner data. The LancsBox 5.1.2 software[1] (Brezina et al., 2020) was applied to count the POS across the learners' writing text using this instrument's KIWC module.

1. http://corpora.lancs.ac.uk/lancsbox/index.php.

3.3. Procedure

Over three subsequent semesters, a group of 73 students were instructed to write essays on topics that corresponded to their levels of proficiency in writing based on the Top Notch series. Each level of proficiency had a common writing topic, and the testing set included in-class (offline and paper and pencil version) tasks with time limitations. Moreover, the learners were not allowed to use references (e.g. textbooks, dictionaries, cellphones, etc.) during in-class essay writing. All in all, the written sets of data including three manuscripts of written exams from three subsequent semesters for each learner were gathered. After compiling the required data and metadata, an error-tagged learner corpus was developed based on the following procedure:

Firstly, design criteria were established based on the purpose and data availability (Biber et al., 1998), encompassing the major criteria for designing a learner corpus and the metadata specified for designing a learner corpus. A summary of the design criteria is outlined in Appendix 1.

Secondly, handwritten texts were digitized using Google Docs and saved in text format. To guarantee the accuracy of the digitized corpus, a two-step verification process was implemented. First, the first researcher meticulously compared each digitized text file with its corresponding handwritten original, identifying and rectifying any discrepancies. This ensured the digital versions faithfully reflected the handwritten content. Furthermore, to enhance reliability, the second researcher independently reviewed a subset of the digitized texts against the handwritten originals. This additional layer of verification minimized potential errors and maximized the accuracy of the corpus. Finally, the corpus dataset was developed.

Error analysis followed Gass and Selinker's (2008) process, involving data collection, error identification, classification, quantification, source analysis, and remediation. After teachers assessed prompts using NLI standard rubrics, the handwritten writing exams were collected as data. The rubrics focused on correcting grammar, spelling, and punctuation errors. It is worth mentioning that the teachers' error identification task was a part of their regular instructional duties and therefore it was distinct from the error tagging procedure conducted by researchers. However, it served as an initial screening tool to provide researchers with insights into potential errors in learners' writing. This initial input helped researchers focus their efforts on areas that might require closer examination. However, teachers' error identification did not play a direct role in the final error tagging process. Therefore, errors were identified through correcting the writing by the researchers. Then, the errors were annotated by the researchers using the LETM (Granger et al., 2022), featuring seven main error domains and subcategories. The Louvain error-tagging system, comprising seven main error domains and subcategories, is hierarchical. Six main categories used in this study are assigned error tags

(Table 2), forming a three-level error taxonomy from 53 out of 54 error types provided.

The Louvain system initially defined 7 categories at the first level, but this study included only 6 categories, omitting Style (Z) due to its infrequent occurrence in only two out of 73 texts. Additionally, metadata for each document were manually input into the database alongside the raw texts.

Table 2

The Seven General Error Categories in the Louvain Error Taxonomy

Error tags	Definition	Description
F	Formal errors	Errors in spelling or morphology lead to the formation of words that do not exist in English (including homophones).
G	Grammatical errors	Errors that deviate from the rules of English grammar
L	Lexical errors	Errors relating to the semantic properties of phrases or words (i.e. collocational, connotative or conceptual errors)
X	Lexico-grammatical errors	Errors deviate from word grammar, namely cases where the lexico-grammatical properties of a word have been violated including incorrect use of dependent prepositions, countable/uncountable noun confusion, or complementation patterns
Q	Punctuation errors	Errors concerning punctuation inaccuracies including missing, redundant, or misplaced punctuation markers
W	Word redundant/ missing/order errors	Redundant use of words, missing basic words, or misordered words
Z	Infelicities	Inquiries of political correctness, register and stylistic problems

To ensure annotation reliability and consistency, inter-annotator guidelines from the LETM (Granger et al., 2022) were followed. Two researchers independently coded error types in 25% of randomly selected writing tasks from each proficiency level. The Kappa statistic (Cohen, 1960) yielded a value of 0.821, signifying excellent agreement (Carletta, 1996). Any inter-annotator disagreements were resolved through discussion.

Thirdly, after double annotation, error tags were assigned to each error in three levels according to the error tag categorization of the LETM including the error tag, the suggested correction, and the error were inserted in front of the erroneous element. A sample of the coding convention is shown below.

We are best friend *<GNN corr="friends">*friend*</GNN>*.

In Iran, they should know each other to have *<GA corr="a">0</GA>* good life.

3.4. Data Analysis

To analyze changes in error patterns regarding question one, the frequency and percentage of errors in each category across proficiency levels were reported using descriptive statistics. Furthermore, to clarify the developmental trajectories of three prominent error types GA, GNN, and GPP, a one-way ANOVA was employed for initial group comparisons, followed by targeted analyses using Dunnett's T3 for pairwise comparisons and Levene's and LSD tests for assessing variance homogeneity. All analyses were conducted in SPSS 28.

4. Results and Discussion

4.1. Results

To answer the first question, the number of errors per category at each level of proficiency was determined as illustrated in Table 3. Then, the changes in error frequency per category in each proficiency level across the three writing performances were identified (see Table 4 to Table 6).

Table 3 shows that the most frequent errors were grammar errors (n=2503, 36.19%) and the least frequent ones were lexico-grammar errors (n=241, 3.48%). Moreover, there is a rise in the number of errors in categories of Grammar, Lexis, and Word redundant/missing/order as proficiency level increases. However, there was found a decrease in Form, Lexico-grammar, and Punctuation error frequency from A2 to B2 and this reduction continued as the learners reached C1 level.

Table 3

Distribution of Error Categories per Each Level of Proficiency in a Corpus Sample

Proficiency Level	A2		B2		C1		Total	
Error category	Total Errors	% of total errors	Total Errors	% of total errors	Total Errors	% of total errors	Total Errors	% of total Errors
F	459	27.29%	567	21.74%	561	21.36%	1587	22.94%
G	581	34.54%	947	36.31%	975	37.11%	2503	36.19%
X	41	2.44%	102	3.91%	98	3.73%	241	3.48%
L	137	8.15%	291	11.16%	336	12.79%	764	11.05%
W	73	4.34%	126	4.83%	164	6.24%	363	5.25%
Q	391	23.25%	575	22.05%	493	18.77%	1459	21.09%
Total errors per level	1682		2608		2627		6917	

Table 4 indicates a general increase in error categories across subsequent semesters, except for lexico-grammar errors. While lexico-grammar errors rose from the first to the second semester, they decreased from the second to the third semester.

Table 4

Distribution of Error Categories in All A2 Learner's Writings Across Their Three Writings

Error category	A2-		A2		A2+	
F	123	35.86%	150	28.90%	186	22.68%
G	111	32.36%	174	33.53%	296	36.10%
X	5	1.46%	20	3.85%	16	1.95%
L	20	5.83%	33	6.36%	84	10.24%
W	3	0.87%	23	4.43%	47	5.73%
Q	81	23.62%	119	22.93%	191	23.29%
Total	343		519		820	

Table 5 shows that lexis and word redundant/missing/order errors increased with higher semesters. Form, grammar, and punctuation errors followed an inverted U-shaped pattern, peaking in the first two semesters before decreasing.

Table 5

Distribution of Error Categories in All B2 Learner's Writings Across Their Three Writings

Error category	B2-		B2		B2+	
F	157	20.99%	249	24.51%	161	19.08%
G	243	32.49%	367	36.12%	337	39.93%
X	41	5.48%	27	2.66%	34	4.03%
L	86	11.50%	99	9.74%	106	12.56%
W	40	5.35%	41	4.04%	45	5.33%
Q	181	24.20%	233	22.93%	161	19.08%
Total	748		1016		844	

However, lexico-grammar errors exhibited a U-shaped pattern, declining from the first to the second semester and then increasing again in the third semester.

Table 6

Distribution of Error Categories in All C1 Learner's Writings Across Their Three Writings

Error category	C1-		C1		C1+	
F	206	24.73%	177	19.09%	178	20.53%
G	327	39.26%	340	36.68%	308	35.52%
X	20	2.40%	42	4.53%	36	4.15%
L	100	12.00%	109	11.76%	127	14.65%
W	40	4.80%	59	6.36%	65	7.50%
Q	140	16.81%	200	21.57%	153	17.65%
Total	833		927		867	

Table 6 illustrates an inverted U-shape pattern for grammar, lexico-grammar, and punctuation errors, increasing from the first to the second semester and decreasing in the third semester. Conversely, errors in lexis and word redundant/missing/order categories increased across subsequent semesters. The Form category saw a rise from the first to the second semester but remained stable between the second and third semesters.

Answering the second question involves identifying error developmental patterns using potential occasion analysis (POA). This method counts errors based on their potential occurrences rather than total tokens, utilizing both error-tagged and part-of-speech-tagged (POS) versions of learner data. Thewissen (2015) operationalized POS denominators for POA, annotating each word with its POS to create specific denominators. For example, a 'PP' POS denominator counts personal pronoun errors. Santorini (1990) provided 36 POS denominators, and in this study these were counted across each text using LancsBox software (Brezina et al., 2020).

The developmental patterns of error categories for each proficiency level were identified through two phases. Firstly, the total number of each error in each of the 219 texts was divided by the total number of that error type denominator per text, then multiplied by 100 to obtain the percentage of potential occasion analysis (POA) score. To count the POS denominator Keyword in Context (KIWC) module within LancsBox software was used (Figure 1). The KIWC module is a powerful tool for analyzing the frequency and distribution of words and parts of speech within a corpus. In this study, LancsBox and its KIWC module were used to calculate POA scores for various error categories within our learner corpus. For instance, personal pronoun POS denominator was created to count personal pronoun errors. In result, the total number of POS denominators 14 were considered to count the POS across as it depicted in KIWC module. Secondly, the sum of the resulting POA percentages per text was divided by the total number of texts across proficiency levels to yield the mean error percentage score for each proficiency level.

Due to word count limitations, detailed calculation for over 50 error types isn't feasible. Hence, grammatical errors, representing the highest number (n=2503, 36.19%), were selected. Among these, the top three error types (GA, GNN, GPP) were chosen for detailed investigation. For example, mean error percentage scores for GPP errors were 13.01%, 10.27%, and 10.64% for A2, B2, and C1 respectively, with a mean of 11.45% (Table 7). Similar calculations were conducted for GA and GNN errors to obtain POA scores.

Figure 1

Counting Personal Pronoun POS Denominator in a Sample Text Based on KIWC

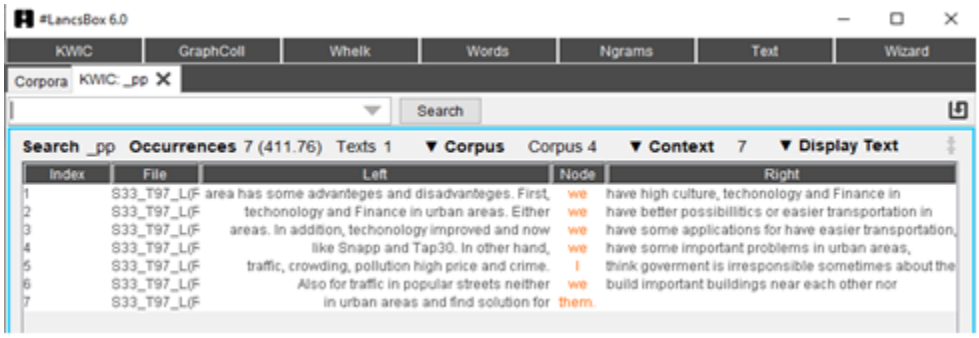


Table 7

The Potential Occasion Analysis Score per Text for GPP as a Sample Representation

Proficiency level	Text number	GPP error frequency	PP number denominator	Potential Occasion Analysis Score (%)
	Text 521	1	22	4.54 %
A2	Text 768	2	11	18.18 %
	Text 802	1	10	0.1%
Mean error percentage score			7.06%	
	Text 299	3	9	33.33%
B2	Text 243	1	13	7.69%
	Text 321	2	21	9.52%
Mean error percentage score				16.84%
	Text 5	2	34	5.88%
C1	Text 164	3	25	12%
	Text 36	4	26	15.38%
Mean error percentage score				11.08%
Mean of mean error percentage scores				11.66%

Thirdly, one-way ANOVA was applied to examine whether different error types did exist among proficiency levels (i.e. A2, B2, and C1) within various texts for GA, GNN, and GPP, shown in Table 8.

Table 8

One-way ANOVA Test Results for GA Errors Between the Three Proficiency Levels

Error type		Sum of Squares	Df	Mean of Squares	F	Significance: p-value*
	Inter-Groups	.989	2	.495	6.242	.002
GA	Intra-Group	13.946	176	.079		
	Total	14.935	178			
	Inter-Groups	.022	2	.011	3.089	.049
GNN	Intra-Group	.467	130	.004		
	Total	.489	132			
	Inter-Groups	413.553	2	206.776	4.275	.017
GPP	Intra-Group	4595.189	95	48.370		
	Total	5008.742	97			

**. The mean difference is significant at the 0.05 level.*

Table 8 shows that ANOVA results across proficiency levels were significant at $p<0.05$ for GA, GNN, and GPP error types. There's a significant difference in GPP and GA errors between proficiency levels, with $F(2, 97) = 4.275$; $p<0.05$ for GPP and $F(2, 178) = 6.242$; $p<0.05$ for GA. However, for GNN errors, the p-value across proficiency levels is 0.049 ($F(2, 130) = 3.089$), which is close to the alpha level (0.05). While statistically significant differences exist in GNN errors between proficiency levels, pinpointing exact mean differences between levels might be challenging.

To ensure variance homogeneity, Levene's test was conducted. Table 9 indicates significantly different variances for GA and GPP errors across proficiency levels ($p < 0.05$), implying unequal sample variances. Thus,

Dunnett's T3 multiple comparisons were employed as post-hoc analyses to pinpoint differences in error frequency among proficiency levels for GA and GPP (Table 10). Dunnett's T3 test, suitable for ANOVA with unequal variances and small samples, avoids bias from multiple comparisons.

Table 9

Test of Homogeneity of Variances

Error type	Levene Statistic	df1	df2	Sig.
GA	4.305	2	176	**.015**
GNN	1.881	2	130	.157
GPP	5.517	2	95	**.005**

Additionally, since Table 9 shows no significant difference in GNN variance, the LSD test was used to identify frequency differences between levels (Table 11).

Table 10

Dunnett T3 Post hoc Test Results for GPP and GA between the Three Proficiency Levels

Error type	Levels (I)	Levels (J)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	95% Confidence Interval
						Lower Bound	Upper Bound
GA	A2	B2	**.16647***	.05093	.004	.0432	.2898
		C1	**.13829***	.05255	.028	.0111	.2654
	B2	A2	**-.16647***	.05093	.004	-.2898	-.0432
		C1	-.02818	.04865	.916	-.1461	.0897
	C1	A2	**-.13829***	.05255	.028	-.2654	-.0111
		B2	.02818	.04865	.916	-.0897	.1461
GPP	A2	B2	**-5.15310***	1.37040	.001	-8.5228	-1.7834
		C1	**-4.40868***	1.50469	.015	-8.1169	-.7005
	B2	A2	**5.15310***	1.37040	.001	1.7834	8.5228
		C1	.74442	1.78643	.966	-3.6192	5.1080
	C1	A2	**4.40868***	1.50469	.015	.7005	8.1169
		B2	-.74442	1.78643	.966	-5.1080	3.6192

**. The mean difference is significant at the 0.05 level.*

Moreover, Dunnett's T3 test for multiple comparisons (significance was defined as $p < 0.05$) showed a significant difference within the proficiency levels for the number of GPP errors and that of GA errors, as shown in Table 10.

A significant difference existed between A2 and B2 proficiency levels in GA errors, with A2 exhibiting more article errors compared to B2 and C1. However, the mean difference in article error frequency between A2 and B2 (mean difference= 0.16647; $p < 0.05$) and A2 and C1 (mean

difference=0.13829; p < 0.05) was minor, approximately 0.152641. No significant difference was found between B2 and C1 regarding article errors.

Additionally, a significant difference was observed in GPP errors between A2 and C1 levels, with both B2 and C1 reporting higher GPP error frequencies than A2. Similarly, significant differences were found between A2 and both B2 and C1 levels regarding personal pronoun errors, indicating more errors in B2 and C1 than in A2. Notably, the mean difference in personal pronoun errors between A2 and B2 (mean difference=5.15310; p < 0.05) was more pronounced than between A2 and C1 (mean difference=4.40868; p < 0.05).

The LSD test revealed a significant difference in GNN error frequency between levels A2 and B2, with A2 learners committing more GNN errors than B2 learners. However, no significant change was found between the number of GNN errors between A2 and C1 levels, as well as between B2 and C1.

Dunnett's T3 test results aided in identifying error developmental trajectories using Thewissen's (2013) classification system. This system categorizes patterns as strong, weak, or non-progressive across a proficiency continuum from A2 to C1, with B2 as the midpoint. A strong pattern occurs when there's a significant difference in error presence between adjacent levels, while a weak pattern occurs between non-adjacent levels. If error types show no significant change across designated levels, it is considered non-progressive (Thewissen, 2013).

Table 11

LSD post hoc Test Results for GNN between the Three Proficiency Levels

Error type	Levels (I)	Levels (J)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
GNN	A2	B2	**.03181***	.01321	.017	.0057	.0579
		C1	.02538	.01332	.059	-.0010	.0517
	B2	A2	**-.03181***	.01321	.017	-.0579	-.0057
		C1	-.00643	.01211	.596	-.0304	.0175
	C1	A2	-.02538	.01332	.059	-.0517	.0010
		B2	.00643	.01211	.596	-.0175	.0304

This study classified 53 error types across A2, B2, and C1 proficiency levels into three developmental patterns. Strong patterns were expected in A2>B2>C1, A2>B2, or B2>C1 trajectories, while weak patterns might show progressive development between non-adjacent levels (A2>C1). Non-progressive patterns exhibited no significant developmental difference across A2, B2, and C1 (i.e., A2/B2/C1).

The analysis of ANOVA, Dunnett's T3, and LSD results indicates marked changes in error presence across personal pronoun, article, and noun number error types. According to Thewissen's taxonomy (2013), when

significant progress occurs between adjacent and non-adjacent levels, it constitutes a strong pattern 3. Specifically, GA errors (A2> B2 and A2> C1) and GNN errors (A2>B2) exhibit strong pattern 3, as displayed in Table 12. Figure 2 and Figure 3 also provide the graphical representation of these trends.

Table 12

The Developmental Patterns of GA, GNN, and GPP Errors

Developmental pattern	Error categories	F	p	Mean error percentages A2	B2	C1
Strong pattern 3: A2>B2 and (A2>C1)	GA	6.242	.002	51.84%	35.19%	38.01%
Strong pattern 3: A2>B2	GNN	3.089	.049	10.75%	7.57%	8.21%
Strong pattern 4: B2>A2	GPP	4.275	.017	9.25%	12.24%	11.49%

Figure 2

GA Errors: Developmental Pattern

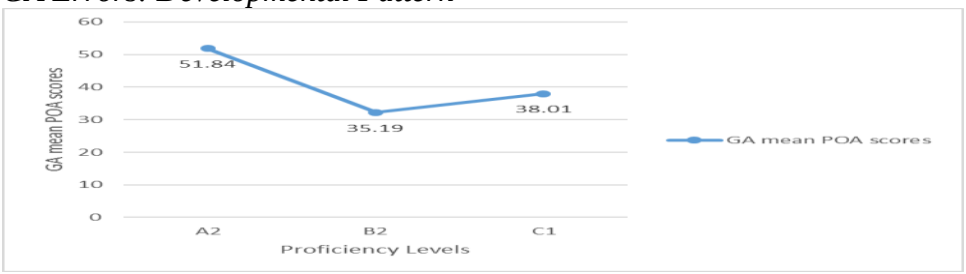
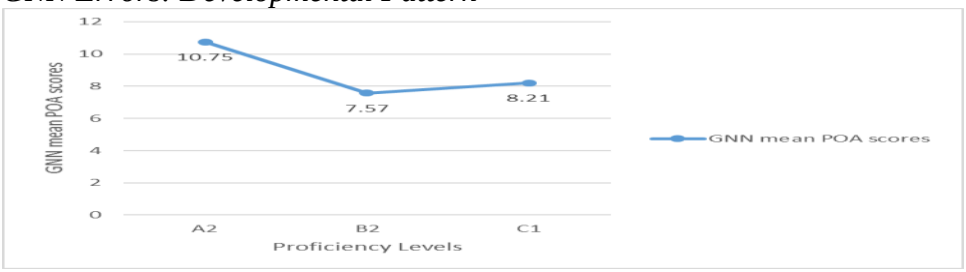


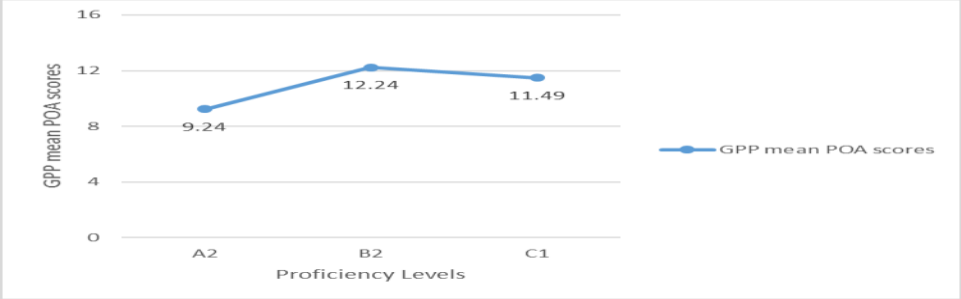
Figure 3

GNN Errors: Developmental Pattern



In contrast to previous error types, GPP errors displayed a regression in development, with fewer personal pronouns evident in A2 compared to B2 and C1. However, there was a minor significant change between B2 and C1, suggesting a slight improvement trend from B2 to C1, shown in Table 12. Thus, GPP errors corresponded to a B2>A2 pattern (see Figure 4). This reverse development suggests that B2 learners may employ longer sentences with more grammatical complexity, leading to potential confusion in using personal pronouns.

Figure 4
GPP Errors: Developmental Pattern



4.2. Discussion

The first research question investigated developmental changes in six error categories across three subsequent semesters among Iranian learners. Results showed a significant increase in errors in grammar, lexicon, and word redundant/missing/order with higher proficiency levels. Conversely, form, lexico-grammatical, and punctuation errors displayed a non-linear pattern, increasing from A2 to B2 and decreasing from B2 to C1.

Despite expectations, participants' increasing language proficiency did not lead to significant improvement in overall grammar, lexis, and word redundant/missing/order accuracy. This could stem from the heightened complexity of language attempted at higher proficiency levels. De Graaff (2019) notes that grammatical and lexical errors are inherent in second language acquisition. Complex tasks in higher proficiency levels demand more intricate language use, increasing the likelihood of errors (Yusuf et al., 2021; Wee et al., 2009). This finding aligns with findings of the research by Singh et al. (2017), Pek et al. (2019), Moazzeni Limoudehi and Mazandarani (2019), and Minh et al. (2022), suggesting persistent difficulties in mastering writing mechanics and syntax at intermediate and higher levels.

No significant progress was observed in the overall accuracy of redundancy, missing words, or word order, possibly due to inherent differences between Persian and English. Iranian language learners often deviate from English word order, influenced by Persian constructions, particularly at lower proficiency levels (Ridha, 2012). They may unintentionally apply Persian writing styles in English, leading to errors such as misordering and overuse of certain structures (Khaghaninejad & Mavadat, 2015; Mohseni & Samadian, 2019). Despite the Iranian language curriculum's emphasis on sentence structure and rules, Iranian EFL learners struggle with word

redundant/missing/order errors, possibly due to difficulty in understanding word use and function (Moazzeni Limoudehi & Mazandarani, 2019). Larsen-Freeman (2015) underscores that learners' grammar and lexis are continuously evolving, and stagnant frequency in lexical-related errors is not a reliable gauge of language progression (Abe & Tono, 2005). At higher levels, stabilization should not imply a lack of learning but rather should underscore the need for teachers to enhance their linguistic awareness for error detection and improve written corrective feedback. Iran's exclusive EFL context, coupled with distinct cultural and linguistic backgrounds, limits exposure to English, while institutional rules pose additional challenges for teachers and learners in developing writing skills.

A rise-and-fall pattern was observed in form, lexico-grammar, and punctuation categories across proficiency levels, with B2 learners exhibiting higher rates of spelling, punctuation, and lexico-grammatical errors compared to A2 and C1. This suggests that B2 learners may not have fully internalized spelling and punctuation rules, possibly influenced by Persian language mechanics or negative L1 transfer (Brown, 2014; Ngaiza, 2023; Ghafar Samar & Seyyed Rezaai, 2006). Additionally, B2 learners may experience cognitive load and lack automaticity in language construction, leading to more conscious sentence construction and increased errors (Thewissen, 2015). Mechanical skills improved at the C1 level, aligning with findings from Farooq et al. (2020), Nanda (2015), and Yusuf et al. (2021), who highlighted punctuation, capitalization, and spelling as common writing errors.

This study found no significant difference in error frequency between B2 and C1 levels across most error categories, consistent with Dagneaux et al. (1998), suggesting limited improvement from intermediate to advanced levels. Learners may reach a plateau in second language competency beyond intermediate stages (O'Keeffe & Mark, 2017: Richards, 2008; Thewissen, 2015).

Closer examination of writing performances per semester across proficiency levels revealed a significant increase in lexis and word redundant/missing/order errors as learners progressed through subsequent semesters. This suggests Iranian language learners may struggle with sentence complexity, and there may be no developmental endpoint for lexis due to the expansive nature of the English vocabulary domain (Bintz, 2011).

The results revealed an inverted U-shaped pattern for form in B2, grammar in C1 and B2, and punctuation in B2 and C1, indicating a turning point in development at the B2 level where errors begin to decrease towards C1. Therefore, instructors should focus on B2 as a crucial threshold for reconstructing language knowledge and reinforcing language development through appropriate strategies and real-time written corrective feedback on form, grammar, and punctuation.

The study found inconsistent trends in lexico-grammar errors, with B2 displaying a U-shaped pattern across three semesters, while C1 and A2 showed an inverted U-shape. Manual inspection revealed limited vocabulary and grammar familiarity, overgeneralization, overreliance on L1 translation, and attempts to use complex grammatical structures, particularly at the C1 level.

The second research question examined the error developmental patterns for articles, noun numbers, and personal pronouns. The analysis revealed improvement from A2 to B2 followed by relative stabilization or limited development from B2 to C1 for article errors and noun numbers. This finding aligns with previous research indicating that higher-level learners show less improvement compared to lower-level counterparts (Milton & Meara, 1995; Thewissen, 2013). However, stabilization should not be interpreted as a lack of progress in higher levels.

The study found a regression pattern for personal pronoun errors from A2 to C1. Common errors included confusion between grammatical functions, number, person, and gender. The Top Notch series aligned with the CEFR expects learners to acquire different kinds of personal pronouns before B2 levels. As Iranian learners progress, intralingual errors become more prominent, reflecting previous findings (Fati, 2013; Moazzeni Limoudehi & Mazandarani, 2019). Although these errors may not hinder intentional meaning, they can impact writing quality, particularly with frequent occurrence.

As Long (1990) states “language learning progress is not linear; backsliding is common, giving rise to so-called U-shaped behavior observed in first and second language acquisition” (p. 659). In this study, the most sudden and temporary drops off or upturns occurred in the B2 level. As Richards (2008) argues, the intermediate level is a point at which the learners seem to depart from their real language ability and progress unsteadily in their language proficiency. This may indicate that Iranian learners in the middle phases of language proficiency need to reorganize and reconstruct their current knowledge to assimilate new knowledge (Ellis, 1994).

This finding fairly challenges what Ellis (1994) proposed about the occurrence of U-shape learning behavior in the beginner stages of learning. In this regard, the A2 level showed a sharp rise in error frequency in all error categories except lexico-grammatical errors which need further investigation.

A notable finding was the consistent high frequency of certain error types. Iranian EFL teachers may lack insight into learners' linguistic backgrounds from previous levels, hindering the provision of effective remedial strategies to address recurring errors at higher levels and overcome the 'plateau effect' (Richards, 2008).

5. Conclusion and Implications

This study was an attempt to investigate the error types and the progress patterns observed in the written texts produced by a group of Iranian EFL learners from three different proficiency levels. This was done through the merits of error-tagged learner corpora and a longitudinal study. The findings revealed that learning language among Iranian learners is a non-linear process, as is often the case in reality. In general, except lexis and word redundant/missing/order errors with a rise across proficiency levels, all other categories represented non-linear patterns from one level to the next one with either U-shape or inverted U-shape trajectories.

This study holds significant implications for a wide range of individuals, including teachers, learners, curriculum designers, and researchers. In particular, the detailed analysis of error types about proficiency levels may empower teachers in identifying the potential errors in the course of each proficiency level and tailoring their teaching methods to the specific needs of the language learners (Rezaee et al., 2024). Addressing the areas of writing weaknesses avoids reinforcing incorrect usage of language in writing tasks and boosts their writing accuracy.

Moreover, exercises targeting common grammatical errors like personal pronouns, articles, and noun numbers should be incorporated into Iranian learners' practice routines. These exercises should utilize authentic passages, allowing learners to recognize, identify misuse or omission, and correct errors. Activities such as fill-in-the-blank exercises and reading authentic materials can facilitate this process effectively (Fati, 2013). Consequently, the learners become more aware of the proper usage of grammatical items and structures by decreasing the negative transfer of the Persian language which may be an obstacle to hinder the development of accuracy in writing skills in English (Derakhshan & Karimian Shirejini, 2020).

Overall, this study elucidates common error patterns across proficiency levels. One merit of this study is the contribution and collaboration between the teachers and the researcher in identifying the errors. However, this study is limited to female Iranian learners since it is recommended to conduct a similar study with male learners to compare error commitment between genders more accurately. Also, future research is expected to conduct further investigations involving more students and longer written texts. Conducting further research based on the more extensive error data may provide opportunities to gain a deeper understanding of more error patterns of other specific language items in learners' writing.

Acknowledgements

We would like to thank the participants who took part in this study.

References

Abe, m., & Tono, Y. (2005). Variations in L2 spoken and written English: investigating patterns of grammatical errors across proficiency levels. In P. Danielsson, & M. Wagenmakers, *Proceedings from the Corpus Linguistics Conference Series* (pp. 1-11). Birmingham.

Biber, D., Corad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and language use.* Cambridge University Press.

Bintz, W. P. (2011). Teaching vocabulary across the curriculum. *Middle School Journal, 42*(4), 44-53.

Brezina, V., Weill-Tessier, P., & McEnery, A. (2020). *LancsBox v. 5.x. [software]*. Retrieved from: http://corpora.lancs.ac.uk/lancsbox.

Brown, H. D. (2014). *Principles of language learning and teaching (6th ed.).* Pearson Education.

Carletta, J. (1996). Assessing agreement on classffication tasks The kappa statistic. *Computational Linguistics, 22*(2), 249-254. http://dl.acm.org/citation.cfm?id=230386.230390

Cohen, J. (1960). A coeffictent of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46. https://doi.org/10.1177/001316446002000104

Council of Europe. (2001). *Council of common European framework of reference for languages: Learning, teaching, assessment.* Press Syndicate of the University of Cambridge.

Dagneaux, E., Denness, S., & Granger, S. (1998). Computer-aided error analysis. *System, 26(2),* 163–74. https://doi.org/10.1016/S0346-251X(98)00001-3

De Graaff, R. (2019). You've got to crack a few eggs to make an omelette. Linguistic complexity and instruction in SLA: What's in it for language teaching? *Instructed Second Language Acquisitionn, 3*(2), 249–258. https://doi.org/10.1558/isla.39398

Derakhshan, A., & Karimian Shirejini, R. (2020). An investigation of the iranian efl learners' perceptions towards the most common writing problems. *SAGE Open, 10*(2), 1-10. https://doi.org/10.1177/2158244020919523

Ellis, R. (1994). *The study of second language acquisition.* Oxford University Press.

Farooq, M. S., Uzair-Ul-Hassan, M., & Wahid, S. (2020). Opinion of second language learners about writing difficulties in English language. *South Asian Studies, 27*(1), 183-194.

Fati, M. (2013). The effect of English writing proficiency in the type and amount of errors produced by Moroccan EFL students. *International Journal of Education and Research, 1*(9), 1-12. https://doi.org/10.30495/JAL.2019.671929

Flowerdew, L. (2001). The exploitation of small learner corpora in EAP materials design. In M. Ghadessy, A. Henry, & R. L. Roseberry, *Small corpus studies and ELT: Theory and practice* (pp. 363-379). Benjamins.

Gass, S. M., & Selinker, L. (2008). *Second language acquisition: An introductory course.* Routledge.

Ghafar Samar, R. G., & Seyyed Rezaai, S. H. (2006). Analysis of errors made by Iranian English learners in their written production: A new answer to an old question? *Iranian Foreign Language Learning Journal, 80*, 59-64.

Gorjian, B. (2022). Investigating the Interlingual errors in students' essay: The case of error classification. *Journal of Education, Teaching and Social Studies, 4*(3), 26-44. https://doi.org/10.22158/jetss.v4n3p26

Granger, S. (2002). A bird's-eye view of computer learner corpus research. In S. Granger, J. Hung, S. Petch-Tyson & J. Hulstijn (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 3-33). John Benjamins.

Granger, S. (2003). Error-tagged learner corpora and CALL: A promising synergy. *CALICO Journal, 20*(3), 465-480.

Granger, S., Swallow, H., & Thewi, J. (2022). The Louvain error tagging manual version 2.0. Centre for English Corpus Linguistics, Université catholique de Louvain.

Housen, A., Kuiken, F., & Vedder, I. (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. John Benjamins Publishing Company.

Khaghaninejad, M. S., & Mavadat, R. (2015). Using English discourse markers: a comparison of Persian and English dentistry authors. *Theory and Practice in Language Studies, 5*(1), 97-109. https://doi.org/10.17507/tpls.0501.13

Khansir, A. A., & Shahhoseiny, H. (2013). The study of written errors of EFLpre-university learners. *Journal of Language Teaching and Research, 4*(6), 1253-1258.

Khansir, A., & Pakdel, F. (2020). A study of written errors of Iranian learners: A systematic review. *Theory and Practice in Language Studies, 10*(8), 982-987. http://dx.doi.org/10.17507/tpls.1008.18

Khazaee, H., Maftoon, P., Birjandi, P., & Rezaie Golandouz, G. (2020a). Hedges in English for academic purposes: A corpus-based study of Iranian EFL learners. *International Journal of Foreign Language Teaching & Research, 8*(33), 115-129.

Khazaee, H., Maftoon, P., Birjandi, P., & Rezaie Golandouz, G. (2020b). Metadiscourse markers in a corpus of learner language: The case of

Iranian Efl learners. *The Journal of Language and Translation (JLT), 10*(3), 17-34. https://doi.org/10.30495/TTLT.2020.678752

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics, 27*(4), 590-619. https://doi.org/10.1093/applin/aml029

Larsen-Freeman, D. (2015). Saying what we mean: Making a case for 'language acquisition' to become 'language development'. *Language Teaching, 48*(4), 491-505. https://doi.org/10.1017/S0261444814000019

Li, Y. (2022). Corpus-based error analysis of Chinese learners' use of high-frequency verb take. *English Language Teaching, 15*(2), 21-30. https://doi.org/10.5539/elt.v15n2p21

Long, M. H. (1990). The least a second language acquisition theory needs to explain. *TESOL Quarterly, 24*(4), 649–666.

Lu, X. (2022). What can corpus software reveal about language development? In A. O'Keeffe, & M. J. McCarthy, *The Routledge handbook of corpus linguistics (2nd ed.)* (pp. 155-167). Routledge. https://doi.org/10.4324/9780367076399-12

Lüdeling, A., & Hirschmann, H. (2015). Error annotation systems. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 135-158). Cambridge University Press. https://doi.org/10.1017/CBO9781139649414

Ma, Q., & Mei, F. (2021). Review of corpus tools for vocabulary teaching and learning. *Journal of China Computer-Assisted Language Learning, 1*(1), 177-190. https://doi.org/10.1515/jccall-2021-2008

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book.* Routledge.

Milton, J., & Meara, P. M. (1995). How periods abroad affect vocabulary growth in a foreign language. *ITL Review of Applied Linguistics, 107/108*, 17-34. https://doi.org/10.1075/itl.107-108.02mil

Minh Anh, N. H., Yen, N. H., Thi Y Tho, N., & Minh Nhut, L. (2022). Grammatical errors in academic writing of English second-year students. *European Journal of English Language Teaching, 7*(6), 70-83. http://dx.doi.org/10.46827/ejel.v7i6.4547

Moazzeni Limoudehi, M., & Mazandarani, O. (2019). Iranian EFL learners' written grammatical errors: Different levels of language proficiency. *The Journal of English Language Pedagogy and Practice, 12*(24), 116-138. https://sanad.iau.ir/journal/jal/Article/671929?jid=671929

Mohseni, A., & Samadian, S. (2019). Analysis of cohesion and coherence in writing performance of Iranian intermediate EFL learners. *Issues in Language Teaching (ILT), 8*(2), 213-242. https://doi.org/10.22054/ilt.2019.21262.193

Nanda, R. P. (2015). *Teacher's efforts to cope with problems faced by students in writing skills (Bachelor's thesis).* Universitas Syiah Kuala.

Ngaiza, A. (2023). A linguistic error analysis of ESL written essays by Indian Tamil university students. *Journal of Modern Research in English Language Studies, 10*(4), 45-59. https://doi.org/10.30479/jmrels.2023.18208.2160

O'Keeffe, A., & Mark, G. (2017). The English grammar profile of learner competence: methodology and key findings. *International Journal of Corpus Linguistics, 22*(4), 457–489. https://doi.org/10.1075/ijcl.14086.oke

Ortikov, U. (2023). Practical uses of corpus analysis in designing language teaching materials. *Oriental Renaissance: Innovative, Educational, Natural and Social Sciences, 3*(7), 304-309. https://doi.org/10.1558/slte.25727

Pek, L. S., Mee, R. M., Shing, S. R., Theesmas, D., & Nadarajan, N. M. (2019). Strengthening tertiary students writing skills through TREND Model. *Asian Journal of Contemporary Education, 3*(1), 65-71. https://doi.org/10.18488/journal.137.2019.31.65.71

Pravec, N. A. (2002). Survey of learner corpora. *ICAME Journal, 26*(1), 8-14.

Rezaee, A. A., Norouzi, M. H., & Aryaeian, N. (2024). Iranian EFL teachers' cognition of corrective feedback as an element of classroom management: A qualitative study. *Journal of Modern Research in English Language Studies, 11*(2),2-24.https://doi.org/10.30479/jmrels.2023.18807.2216

Richards, J. C. (2008). *Moving beyond the plateau: From intermediate to advanced levels in language learning.* Cambridge University Press.

Ridha, N. (2012). The effect of EFL learners' mother tongue on their writings in English: Anerror analysis study. *Journal of the College of Arts*, 22-45.

Romer, U. (2005). *Progressives, patterns, pedagogy*: *A corpus-driven approach to English progressive forms, functions, contexts and didactics.* John Benjamins.

Ruzaitė, J., Dereškevičiūtė, S., Kavaliauskaitė-Vilkinienė, V., & Krivickaitė-Leišienė, E. (2020). Error tagging in the Lithuanian learner corpus. The Ninth conference of human language technologies - The Baltic perspective. In A. Utka, J. Vaičenonienė, J. Kovalevskaitė, D. Kalinauskaitė (Eds.), *Frontiers in Artificial Intelligence and Applications* 328 (pp. 253–260). IOS Press. https://doi.org/10.3233/FAIA200631

Salmani Nodoushan, M. A. (2018). Toward a taxonomy of errors in Iranian EFL learners' basic-level writing. *International Journal of Language Studies, 12*(1), 61-78. https://doi.org/10.5281/zenodo.7514650

Santorini, B. (1990). *Part-of-speech tagging guidelines for the Penn treebank project.* Technical Report No. MS-CIS-90-47, University of Pennsylvania, Department of Computer and Information Science.

Singh, C. S., Singh, A. J., Razak, N. A., & Ravinthar, T. (2017). Grammar errors made by ESL tertiary students in writing. *English Language Teaching, Canadian Center of Science and Education, 10*(5), 16-27. http://doi.org/10.5539/elt.v10n5p16

Thewissen, J. (2013). Capturing L2 accuracy developmental patterns: Insights from an error- tagged EFL learner corpus. *Modern Language Journal, 97*(1), 77-101. https://doi.org/10.1111/j.1540-4781.2012.01422.x

Thewissen, J. (2015). *Accuracy across proficiency levels: A learner corpus approach.* Presses universitaires de Louvain.

Wee, R., Sim, J., & Jusoff, K. (2009). Enhancing the quality of EAP writing through overt teaching. *English Language Teaching, 2*(3), 58–68.

Yusuf, Y. Q., Mustafa, F., & Iqbal, R. M. (2021). An inquiry into grammatical errors in writing committed by high achieving EFL students. *International Journal of Language Studies, 15*(2), 1-22.

Zarabi, H., Yamrali, N., & Gharani, N. (2023). An investigation of the most common difficulties Iranian pre-university EFL learners face in their paragraph writings. *International Journal of Language and Translation Research, 3*(1), 59-84. http://doi.org/10.22034/IJLTR.2023.169971

Appendix 1

A Summary of the Error-Tagged Learner Corpus Criteria

Major criteria			Metadata		
Corpus criteria	Purpose	To develop an error-tagged dataset for research and instructional purposes	Metadata for learner corpus data	Title of the text	Various
	Mode of the LC	Written		Year/month of the production	2018-2019
	Size	40009 words		Country of the production	Iran
	Availability	Limited access		Text genre	Argumentative, descriptive
	Users	Researchers, teachers, test developers, learners		Task setting	In-class, exam session
Data criteria	Text type	Written		Timing	Limited
	Task type	Composition, essay, controlled writing		Length of the text (min./max.)	34 to 340 words
	Genre	Argumentative, descriptive		References use	No references to be used
	First language	Persian			

Learner criteria	Target language	English		Text medium.	Handwritten
	Level of proficiency	A2, B2, C1		Age	10- 50
Type of error annotation	Form (F)	*	Metadata for learner	Gender	female
	Grammar (G)	*		Nationality	Iranian
	Lexico-grammar (X)	*		Mother tongue	Persian
	Lexis (L)	*			
	Word redundant/ missing/order (W)	*		Nativeness	Non-native speakers
	Punctuation (Q)	*		Educational institution	Private and non-governmental