



A Corpus-Based Study on the Development of Grammatical Accuracy in EFL Writing

Weilu Wang¹, Manfu Duan^{2(✉)}, and Jinyu Liu¹

¹ Foreign Languages College, Inner Mongolia University, Hohhot, People's Republic of China

² Division of International Cooperation and Exchange, Inner Mongolia University, Hohhot, People's Republic of China

duanmanfu@imu.edu.cn

Abstract. This paper takes a descriptive approach to the development of grammatical accuracy in EFL beginner writing. The types and frequency of errors in a longitudinal learner corpus are described and analyzed. The type of errors with the highest average frequency is misspelling, followed by typographical and grammatical errors. Besides, the average frequencies of misspelling and typographical errors are on the rise over six semesters. On the contrary, grammatical errors do not demonstrate any linear trend. It indicates that the development of grammatical accuracy in beginner language is a dynamic process. Further, we notice that the types of grammatical errors with a comparatively higher frequency are subject-verb disagreement, sentence fragments, misuse of noun plural, and errors concerning forms of verbs. A descriptive analysis of these types of grammatical errors indicates that the influence of the first language is strongly felt. The implication here is that the learner's first language is not a hindrance but a facilitator in learning a new language. Therefore, a comparative approach between the first language and English seems a potentially promising route for teaching beginners. Furthermore, teaching materials starting from the similarities and gradually evolving to the differences between the first language and a new one deserve a try.

Keywords: Longitudinal learner corpus · Grammatical accuracy · EFL writing

1 Introduction

Accuracy is one of the most reliable indicators for measuring the performance of a language user. As a system of symbols, conveying its meaning is undoubtedly based on the acceptable combination of the linguistic symbols on the part of the user. Learner language, especially that of beginners, as the rules are not set in, tends to be ungrammatical and unacceptable compared with that of native speakers. In light of this, errors of one kind or another are characteristic of learner language. Grammatical errors in specific are a necessary and inevitable part of learning (Corder 1967). A close and detailed description of the errors learners usually make in the longitudinal development of their learning facilitates our understanding of learners' actual range of difficulties. It provides precious guidelines for the compilation of teaching materials, the design of the syllabus, and the actual conduct of language teaching by an instructor.

© The Author(s) 2023

Y. Yau and F. Hage Chehade (Eds.): DESD 2022, ASSEHR 691, pp. 328–337, 2023.

https://doi.org/10.2991/978-2-494069-37-4_41

The errors language learners make in their learning process have been a widely concerned subject of research in language teaching and acquisition. Early studies on learner errors draw on the methodology of error analysis; they mainly focus on exploring the reasons behind these errors, highlighting the influence of learners' first language, the learning material used, individual learning style, and learner strategies (Xavier et al. 2020). It has been generally argued that a learner's first language may interfere with learning a foreign language. Individual factors such as efforts put into his studies and strategies employed may positively impact his learning. It has also been argued that a language learner is supposed to learn to get rid of the interference of the first language and be equipped with the proper way of thinking, and draw on the effects of favorable individual factors. However, studies of this kind were conducted with limited participants, even a single case in some instances. The results need to be further tested with a larger sample. In terms of the reasons for errors, in most cases, they are attributed to the negative transfer of the first language. Recent research suggests that learner language is working in its peculiar way; its comprehensive evolution and development can be and should be described (Tarone 2012).

Descriptive approaches have been employed to analyze the types and frequency of errors in learner language, especially in recent years. Fitrawati et al. (2021) find that verb-related errors are the most common in EFL writing, which account for 48% of all errors, and that verb forms and verb use in clauses are where learner difficulty lies. Sabtan (2019) asserts that errors with the highest frequency in EFL writing are misspelling and grammatical errors. The major types of grammatical errors are subdivided into wrong verb forms, verb-subject disagreement, misuse of single/plural forms of a noun, misuse of propositions, incorrect part of speech, inappropriate word order, misuse of articles, misuse of comparative degree/superlative degree of adjectives and adverbs. Ceterisi (2018) finds that the most common error types have something to do with the linguistic incompetence and pragmatic incompetence of the language user. Under linguistic incompetence, the high-frequency types are lexical, orthographical, and grammatical errors. Hamed (2018) asserts that in writing by EFL non-English majors, the error category with the highest frequency of occurrence goes to substance errors, including misspelling and punctuation misuse, incorrect verb tenses, misuse of articles, subject-verb disagreement, and misuse of propositions. Pouladian (2017) examines the English produced by EFL Iranian adult learners and notes that the most common error types are misuse of verb tense and attributive clauses. Previous research works on the micro-level of learner language and informs us of some of the static features of the actual use of EFL learners. Still, they fail to approach and describe the dynamic development of that language system.

A few studies also explore the differences in errors made between learners of diverse backgrounds or proficiency levels. Martínez(2017)studies the accuracy and complexity of advanced and intermediate learners, and the findings indicate that there are significant differences between the two groups of learners only in spelling and punctuation errors. Eng et al. (2020) investigate the grammatical errors in IELTS essays written by undergraduates from China and Malaysia. They find that the errors made by the participants mainly include: lack of words, form errors, misuse, incoherence, and wrong word order. They also find that the frequency of errors in essays by Chinese undergraduates

is significantly higher than those by undergraduates from Malaysia. Lahuerta (2018) compares the English compositions by 64 intermediate learners and 36 advanced learners and argues that the average frequencies of errors in spelling, grammar, diction, and punctuation of intermediate learners are higher than those of advanced learners. Still, significance can only be noticed in grammatical errors and punctuation errors. De Kleine (2018) draws a comparison between L1 learners and Generation 1.5 learners, and L1 and L2 learners respectively, and asserts that there is a significant difference between the two groups in grammatical errors, but quite unexpectedly, there is no significant difference between the L1 and L2 groups in terms of the frequencies of errors. Leroux et al. (2018) explore the acquisition of articles by Chinese college EFL students and Chinese professionals living and working in the United States. The results show a surprising level of conformity of accuracy rates across participants of different proficiency levels.

Previous studies mainly draw conclusions from data collected from synchronic corpora, and few studies focus on longitudinal data. Crosthwaite (2018) and Hamid et al. (2018) take a pre-test, and post-test design analyzing a longitudinal academic English corpus and finds that errors are significantly reduced in the post-test corpus while the types of errors made are not considerably cut down. One possible reason could be the short length of duration between the pre-test and post-test. And there are not so many studies focusing on describing the longitudinal changes of errors in learner language. This study, based on a self-built longitudinal learner corpus collected over three years, describes the development and changes in the writing of EFL beginners, which provides new data for longitudinal learner corpora research..

2 Method

The present study is descriptive in nature, which focuses on the errors in EFL writing of beginners. It takes a natural language procession tool as the source of error annotation and calculation to automatically extract the errors made in EFL beginner writing. Further, the data extracted is processed via Microsoft Excel to manually categorize the grammatical errors into a diversity of error types. The frequency and percentage of all types are calculated.

The extraction of learner error data is done by GAMET (GRAMMAR AND MECHANICS ERROR TOOL), which is designed by Christopher Kyle, an assistant professor of linguistics at the University of Oregon, and Scott Crossley, a professor of applied linguistics in the Department of Applied Linguistics and ESL, Georgia State University. It makes possible automatic calculation of grammatical errors, misspellings, and typographical errors. And practically, it extracts the sections of the text which contain the errors for the convenience of further assessment (Crossley et al., 2019).

3 Results and Discussion

Figure 1 shows the descriptive statistics of the average frequency (every one hundred words) of errors in compositions by EFL beginners over three years. It can be seen that the error with the highest frequency is misspelling with a frequency of 11.2 in every one hundred words. And the second-highest is typographical errors (small-letter words at

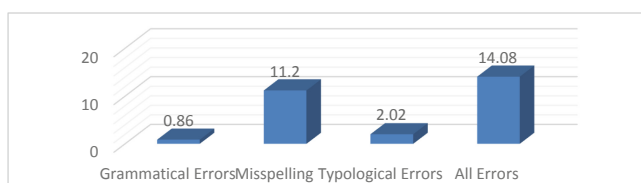


Fig. 1. Average Frequency of Errors

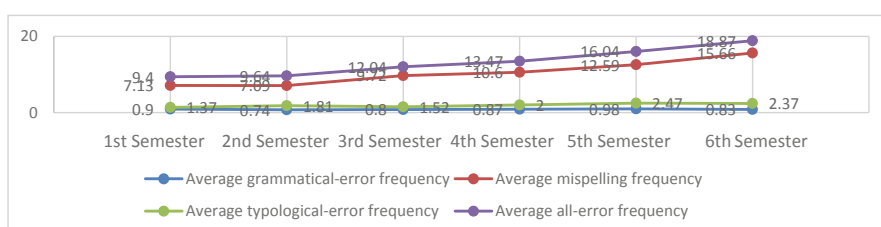


Fig. 2. Average Frequency of Errors (by semester)

the beginning of a sentence, etc.) with a frequency of 2.02 in every one hundred words, followed by grammatical errors with a frequency of 0.86. The total frequency of errors in every one hundred words comes to 14.08.

The findings are consistent with those of the previous studies. In most cases of writing, beginners are in a constant struggle to produce letter-correct language and are inevitably under the influence of their first language. For Chinese learners, the good news is that Chinese and English share a similar word order structure for most of the core elements of the syntactic types; the two languages share the five essential patterns of a sentence (subject + verb, subject + verb + object, subject + link verb + predicative, subject + verb + double objects, subject + verb + object + supplement).

Figure 2 shows the changes over time in the average frequency of types of errors, from which it can be noticed that the average frequency of misspelling is on a steady rise from 7.13 in the first semester all the way to 15.66 in the sixth. That is to say, the errors in spelling more than doubled over three years in the EFL learning program, which pushes the all-error frequency to be rising synchronously from 9.4 in the first semester to 18.87 in the sixth. One possible explanation for the rise could be that the students are learning more words with each passing day within the program, which puts them on the verge of a higher degree of anxiety in using the language and spelling the right word. As adults, they are constantly trying to put what's in their minds into the new language they are just beginning to learn; consequently, they are making more mistakes in writing as they are allowed a limited length of time in an exam. It may also have something to do with the length of their writing. As they progress over the years, they are required to register something longer, and the longer the writing, the greater chance to make more mistakes. A third reason may be that spelling is not a highly-demanded part of the learning program, which is quite common for courses intended for adult learners.

The other two types of error, namely grammatical and typographical, almost remain stable over time. It can be attributed to the fact that in a beginners' course, the primary learning task of the learners is to have a command of grammatical and syntactical rules, which denotes that along the way, they are learning new grammatical rules and

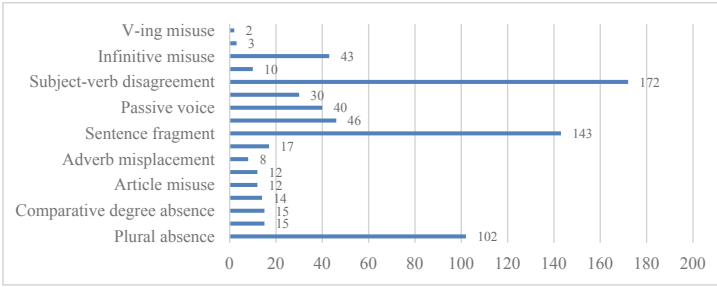


Fig. 3. Frequency of grammatical errors (by type)

morphemes and practicing the rules and making mistakes in the process. Errors are an indispensable part of the learning process, and grammatical errors prevail in the complete process of learning grammar.

Figure 3 shows the total frequency of grammatical errors in beginner language. Among the types of errors, subject-verb disagreement goes first (frequency = 172), followed by sentence fragments (143), plural absence (102), modal verb misuse (46), infinitive misuse (43), and tense-progressive (40). Therefore, a large part of the errors a beginner makes in his writing are related to the absence or misuse of English inflectional morphemes. The explanation lies in the tremendous differences between Chinese and English. As widely noticed, the English language partially depends on the binding of free morphemes with inflectional morphemes for users to construct the world around. In contrast, the Chinese language belongs to the non-inflectional type, which combines its free morphemes and draws on its context clues for interaction and negotiation of meaning on the part of both sides: the speaker and the listener, the writer and the reader. For a beginner EFL learner of Chinese origin, the first and most profound challenge is to make adjustments in their language use and build from the ground up the habit of adding grammatical and functional suffixes to the words he uses to express himself in English. In writing, they may naturally seek refuge in their first language, pour down what they are accustomed to, and take it for granted that what is acceptable in their first language is also transferable in any language. This common practice can be conceptualized as an overgeneralization of known rules in the first language.

On the syntactical level, two types of errors, subject-verb disagreement, and sentence fragments, are among the highest-frequency errors, which can also be attributed to the divergences in syntactic typology between Chinese and English. In Chinese, the predicate can be a noun or adjective in addition to a verb, and there is no such a rule as subject-verb agreement in number and gender. It seems beginners are just transferring the laws of their first language to a second language they are learning. Such grammatical or syntactical errors are just an embodiment of the negative transfer of the native language. Corpus data show that a learner's native language does not hinder. Instead, it furnishes the learner with a direct means to draw on for communication, without which the purpose of expressing could stand no chance of being achieved. Therefore, the destined approach for beginners is to seek references from their first language, test the transferred hypothesis in a new environment, and try to make amendments.

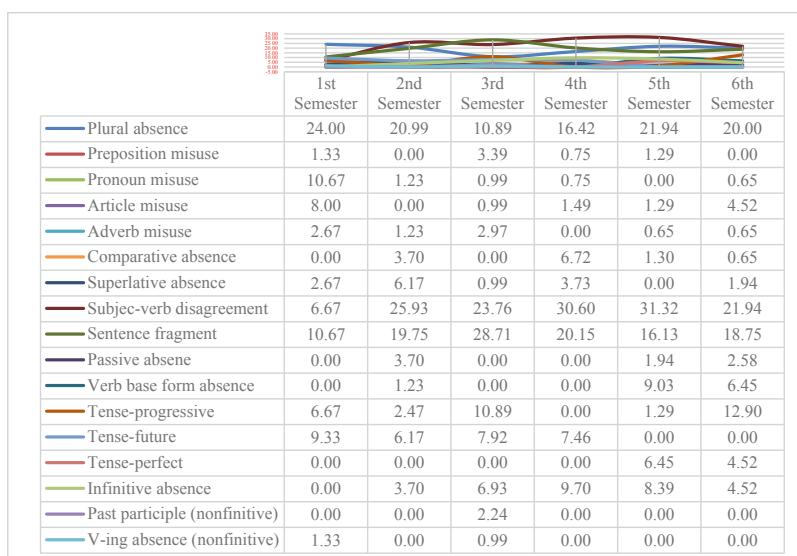


Fig. 4. Average frequency of grammatical errors (by semester, every one hundred words)

As shown in Fig. 4, none of the frequencies for the major types of grammatical errors is linear, which demonstrates the dynamics in the accuracy of beginner language development. Contrary to our expectations, types of grammatical errors do not dwindle significantly as the learner progresses over the years. But we notice that errors concerning complex structures such as non-finite forms of verbs and passive voice are increasing in the last semesters. In the case of infinitive absence, its average frequency rises from 0.00 in the first semester to 6.45 in the sixth semester, almost a sixfold increase. And for errors in the present progressive tense, the frequency grows from 6.67 in the first semester to 12.9 in the sixth, nearly doubling. While frequencies of errors in simple structures descend in the recent semesters. The frequency of plural absence lessens from 24 in the second semester to 20 in the sixth; the frequency of subject-verb disagreement drops from 25.93 in the second semester to 21.94 in the sixth.

A possible explanation is that learners are learning grammatical rules of a certain language by following a sequence, and complex structures are introduced at a later date compared to those simpler ones. Additionally, with the passage of time, learners are gaining more experience in handling grammar, which brings down the frequency of errors slightly but not significantly, as they are encountering new elements along the way and they are forming new hypotheses.

To be specific, errors regarding singular/plural forms of nouns in beginner language may mainly fall into three major types, of which the most common is the cluster “many/some/a lot of/lots of + singular form of a countable noun”, as exemplified in Table 1. Then it is followed by an absence of plural nouns in set expressions and phrases, such as “make friends with” can be “make friend with” in beginner language. The last type goes to the addition of the plural morpheme “-s” to abstract nouns. Errors of this kind

Table 1. Sub-types/Description and Examples of Prominent Types of Errors

Types of Grammatical Errors	Sub-types/Description	Examples of Beginner Language
Plural absence	many/some/a lot of/lots of + singular noun	*Because there are a lot of book in (the) library.
	absence of plural nouns in set expressions	*I want to make friend; I have to smile to them.
	abstract nouns in plural forms	*And in the free times I can read my favourite books.
Preposition misuse	preposition + adverb	*I hope you can have a wonderful school life in here.
	preposition absence	*So I read many books and listen English songs during the weekends
	wrong preposition	*My teachers and friends are all very kind of me.
Pronoun misuse	misuse of the nominal case, objective case and possessive pronoun	*We like he. He like we.
Article misuse	the + pronoun/cardinal number	*And I must tell you a one thing.
	misuse in set expressions	*We ate the lunch at 12:30.
Adverb misuse	an adverb used as an adjective	*Don't do strongly exercise.
	misuse of "hardly"	*And you will know. I can work hardly every day.
Comparative absence	absence of comparative	*Some people believe that technology make our life convenient than before.
Subject-verb disagreement	third-person singular + verb base	*As summer begin, our school become beautiful.
Sentence fragment	absence of predicate verb	*I hope you can healthy.
Tense-progressive	be + verb base	*I'm look forward to recovering quickly.

can be explained by the influence of Chinese since there is no plural morphine for nouns referring to things or concepts other than people.

For the error of preposition misuse, beginners may have an additional unnecessary preposition preceding an adverb ("in there", for example) or following a transitive verb or connective. First language influence is also obvious here, for example, in Chinese an adverbial of place is commonly preceded by a proposition, while there is no such

usage in English. On other occasions, a beginner tends to omit a preposition where it is a must, especially in the case of intransitive verbs. An error of this kind is most commonly noticed in situations where the verb is intransitive in English but is transitive in Chinese; for examples see Table 1. Besides, EFL beginners are aware that a preposition is needed in some cases where a preposition is also required in the Chinese equivalent. But as there is no direct one-to-one equivalent between Chinese and English, a language learner may choose the wrong preposition, for example, “in Sunday” in learner language.

As to errors in pronoun use, beginners seem to be confused about the correct use of nominal case, objective case, and possessive pronoun. Definitely, it’s another proof of the influence of Chinese, in which there is no distinction of cases, and in which possessives are two symbols: a pronoun followed by a possessive marker. Therefore, in beginner language, the nominal cases are prevalent for subjects, objects and possessives as well.

For errors on article misuse, beginners tend to add an unnecessary definite article preceding a pronoun or cardinal number (as in “...play games on the it” and “... tell you a one thing”). This may result from the overgeneralization of a rule of placing an article before a countable noun. Unnecessary definite article use can also be found in set expressions in beginner language (as in “...play the football.”). Furthermore, beginners are prone to miss an article preceding a countable noun, which is clearly another embodiment of the influence of Chinese. For a singular countable noun, there is no imperative need to place a preceding numerical quantifier in Chinese, unless the number is emphasized.

For errors concerning adverb use, in beginner language, they are mistakenly used as modifiers for nouns, which shows confusion about the distinction between an adjective and an adverb (as in “I have to get up early and have a quietly breakfast”). It seems that there is one adverb that causes most of the confusion: hardly, which is usually mistaken for “hard” (as in “I can work hardly every day”). And for comparatives and superlatives, the absence of such forms is a common occurrence in beginner language (as in “technology make our life convenient than ago”).

Syntactically, for errors of subject-verb disagreement, beginners may use a verb in the base form following a subject of third-person singular (as in “and he speak English very well”), a direct transfer from the first language, since in Chinese there is no such a rule as subject-verb agreement: verbs are always in their base form with all subjects. And for sentence fragments, beginners tend to make ungrammatical sentences without a predicate verb, or with more than one verb which are used as predicates but not in parallel construction. The most frequently forgotten predicate verb is “be”, since in Chinese the equivalent of “be” is mainly used for judgment, while predicates can be a noun or an adjective without any verb. For errors concerning verb tenses, the most prominent is the progressive tense, in which beginners may either omit “be” or have the wrong verb form rather than v-ing.

4 Conclusion

The present study takes a descriptive approach to errors in EFL beginner writing. We find that the type of error with the highest average frequency is misspelling, followed by typographical errors and grammatical errors. Besides, the average frequency of misspellings is on the rise over a span of six semesters. Similarly, the average frequency of

typographical errors is slightly increasing. On the contrary, the average frequencies of grammatical errors do not demonstrate any linear trend. It indicates that the development of grammatical accuracy in EFL beginner writing is a dynamic process. The growth in the average frequency of misspelling and typographical errors coincides with the growth of text length. As beginners are writing longer texts, their grammatical accuracy does not improve. With more new language items coming, time and effort are especially needed to enhance the letter-correctness of their writing.

Further descriptive analysis of the types of grammatical errors indicates that the influence of the first language is strongly felt. Beginners tend to draw on what is already known and transfer known rules to new situations. The types of errors with a comparatively higher frequency are subject-verb disagreement, sentence fragments, misuse of noun plurals, and errors concerning forms of verbs. The high frequency of these errors embodies the fact that EFL beginners are really struggling to produce correct English and the elementary grammatical, syntactical and lexical knowledge is also their real difficulty, which forms the real thresholds for beginners to cross.

The implication here is that the first language of the learner is not a hindrance but rather a facilitator. Therefore, a comparative approach between the first language and English seems a potentially promising route for beginners. Furthermore, teaching materials starting from the similarities and gradually evolving to the differences between the first language and a new one really deserve a try. Textbooks taking this approach and the empirical studies on it should be the focus of future research.

Acknowledgments. The present research is funded by the Innovative Practice Base for Developmental Integration of Information Technology with Foreign Languages Teaching and Research and the National Social Science Fund of China (17BY042, A Typological Study of the Function-order Interactions of the Modifiers of English and Chinese).

References

- Cetereisi Y, & Bostanci H B, Classification of written errors regarding the language competencies, *Modern Journal of Language Teaching Methods*, 2018, 8(10), pp.235-244.
- Corder S, The Significance of Learner's Errors, *International Review of Applied Linguistics in Language Teaching*, 1967, 5(1-4), pp. 161-170. <https://doi.org/10.1515/iral.1967.5.1-4.161>
- Crossley S A, Bradfield, F., & Bustamante, A, Using human judgments to examine the validity of automated grammar, syntax, and mechanical errors in writing, *Journal of Writing Research*, 2019, 11(2), pp. 251-270. <https://doi.org/10.17239/jowr-2019.11.02.01>
- Crosthwaite P, Evidence from a longitudinal corpus of L2 EAP essays and reports, *International Review of Applied Linguistics in Language Teaching*, 2018, 56(3), pp. 315-343. <https://doi.org/10.1515/iral-2016-0129>
- De Kleine C, & Lawton R, An analysis of grammatical patterns in generation 1.5, L1 and L2 students' writings: A replication study. *Journal of Second Language Writing*, 2018, 42, pp. 12-24. <https://doi.org/10.1016/j.jslw.2018.10.003>
- Eng L S, Luyue C & Lim C K, A Comparison of the English Grammatical Errors of Chinese Undergraduates from China and Malaysia, *International Journal of Instruction*, 2020, 13(1), pp. 931-950. <https://doi.org/10.29333/iji.2020.13160a>

- Fitrawati F & Safitri D, Students' Grammatical Errors in Essay Writing: A Pedagogical Grammar Reflection. *International Journal of Language Education*, 2021, 5(2), pp. 74–88. <https://doi.org/10.26858/ijole.v5i2.16128>
- Hamed M, Common linguistic errors among non-English major Libyan students writing. *Arab World English Journal (AWEJ)*, 2018, Volume, 9. <http://dx.doi.org/10.2139/ssrn.3258802>
- Hamid H A, Nasri N F, & Ghazali N, Colours as a Form of Corrective Feedback in EFL Learners' Writing, *GEMA Online Journal of Language Studies*, 2018, 18(4). <https://doi.org/10.17576/gema-2018-1804-08>
- Lahuerta A C, Study of accuracy and grammatical complexity in EFL writing, *International Journal of English Studies*, 2018, 18(1), pp. 71-89. <https://doi.org/10.6018/ijes/2018/1/258971>
- Leroux W & Kendall T, English article acquisition by Chinese learners of English: An analysis of two corpora, *System*, 2018, 76, pp. 13–24. <https://doi.org/10.1016/j.system.2018.04.011>
- Martínez A C L, Analysis of accuracy and grammatical complexity in the writing of upper intermediate and advanced learners of English, *RLA: Revista de lingüística teórica y aplicada*, 2017, (55), pp. 13–33. <https://doi.org/10.4067/S0718-48832017000100013>
- Pouladian N, Bagheri M S & Sadighi F, An analysis of errors in writing skill of adult Iranian EFL learners preparing for the IELTS. *International Journal of English Linguistics*, 2017, 7(3), pp. 85–96. <https://doi.org/10.5539/ijel.v7n3p85>
- Sabtan Y M N & Abdelkader Elsayed A M, Common writing errors among EFL students at Dhofar University in Oman: an analytical study, *International Journal of English Linguistics*, 2019, 9(2). <https://doi.org/10.5539/ijel.v9n2p402>
- Tarone E, Interlanguage. *The encyclopedia of applied linguistics*, 2012, pp. 1–7. <https://doi.org/10.1002/9781405198431.wbeal0561.pub2>
- Xavier C A, Hong H M & Renandya W A, Grammar in Writing: Teachers' Reflections, *PASAA: Journal of Language Teaching and Learning in Thailand*, 2020, 60, pp. 199-221

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

